

Siddhartha Bhattacharyya  
Sankar K. Pal  
Indrajit Pan  
Abhijit Das *Editors*

# Recent Trends in Signal and Image Processing

Proceedings of ISSIP 2018

# **Advances in Intelligent Systems and Computing**

Volume 922

## **Series Editor**

Janusz Kacprzyk, Systems Research Institute, Polish Academy of Sciences, Warsaw, Poland

## **Advisory Editors**

Nikhil R. Pal, Indian Statistical Institute, Kolkata, India

Rafael Bello Perez, Faculty of Mathematics, Physics and Computing, Universidad Central de Las Villas, Santa Clara, Cuba

Emilio S. Corchado, University of Salamanca, Salamanca, Spain

Hani Hagrass, Electronic Engineering, University of Essex, Colchester, UK

László T. Kóczy, Department of Automation, Széchenyi István University, Győr, Hungary

Vladik Kreinovich, Department of Computer Science, University of Texas at El Paso, EL PASO, TX, USA

Chin-Teng Lin, Department of Electrical Engineering, National Chiao Tung University, Hsinchu, Taiwan

Jie Lu, Faculty of Engineering and Information Technology, University of Technology Sydney, Sydney, NSW, Australia

Patricia Melin, Graduate Program of Computer Science, Tijuana Institute of Technology, Tijuana, Mexico

Nadia Nedjah, Department of Electronics Engineering, University of Rio de Janeiro, Rio de Janeiro, Brazil

Ngoc Thanh Nguyen, Faculty of Computer Science and Management, Wrocław University of Technology, Wrocław, Poland

Jun Wang, Department of Mechanical and Automation Engineering, The Chinese University of Hong Kong, Shatin, Hong Kong

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 such as: computational intelligence, soft computing including neural networks, fuzzy systems, evolutionary computing and the fusion of these paradigms, social intelligence, ambient intelligence, computational neuroscience, artificial life, virtual worlds and society, cognitive science and systems, Perception and Vision, DNA and immune based systems, self-organizing and adaptive systems, e-Learning and teaching, human-centered and human-centric computing, recommender systems, intelligent control, robotics and mechatronics including human-machine teaming, knowledge-based paradigms, learning paradigms, machine ethics, intelligent data analysis, knowledge management, intelligent agents, intelligent decision making and support, intelligent network security, trust management, interactive entertainment, Web intelligence and multimedia.

The publications within “Advances in Intelligent Systems and Computing” are primarily 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.

**\*\* Indexing: The books of this series are submitted to ISI Proceedings, EI-Compendex, DBLP, SCOPUS, Google Scholar and Springerlink \*\***

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

Siddhartha Bhattacharyya · Sankar K. Pal ·  
Indrajit Pan · Abhijit Das  
Editors

# Recent Trends in Signal and Image Processing

Proceedings of ISSIP 2018

 Springer

*Editors*

Siddhartha Bhattacharyya  
Department of Information Technology  
RCC Institute of Information Technology  
Kolkata, West Bengal, India

Sankar K. Pal  
Indian Statistical Institute  
Kolkata, India

Indrajit Pan  
RCC Institute of Information Technology  
Kolkata, West Bengal, India

Abhijit Das  
RCC Institute of Information Technology  
Kolkata, West Bengal, India

ISSN 2194-5357                      ISSN 2194-5365 (electronic)  
Advances in Intelligent Systems and Computing  
ISBN 978-981-13-6782-3              ISBN 978-981-13-6783-0 (eBook)  
<https://doi.org/10.1007/978-981-13-6783-0>

Library of Congress Control Number: 2019932685

© Springer Nature Singapore Pte Ltd. 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

*Dr. Siddhartha Bhattacharyya would like to dedicate this book to his beloved wife Rashni, his colleagues Paramartha, Kousik, Koushik, Krishnendu, Atanu, Mridul, Tapas, Sandip, Syamasis, Kinsuk, and Sanjeeb.*

*Prof. Sankar K. Pal would like to dedicate this book to Prof. C. A. Murthy, his sincere, humble colleague and co-author, who passed away in a street accident in front of our Institute (ISI) on March 14, 2018.*

*Dr. Indrajit Pan would like to dedicate this book to his beloved parents Dr. Sitansu Kumar Pan and Smt. Ratna Pan.*

*Dr. Abhijit Das would like to dedicate this book to his beloved wife Amrita and his daughter Abhipsa.*

# **ISSIP—2018 Committee**

## **Chief Patron**

Prof. (Dr.) Ajoy Kumar Ray, Chairman, RCC Institute of Information Technology, India

Dr. Sparsamani Chatterjee, Ex-Chairman, RCC Institute of Information Technology, India

## **Honorary Chairs**

Prof. (Dr.) Witold Pedrycz, FIEEE, University of Alberta, Edmonton, Canada

Prof. (Dr.) Janusz Kacprzyk, FIEEE, Polish Academy of Sciences, Poland

Prof. (Dr.) Sankar K. Pal, ISI, Kolkata, India

## **General Chairs**

Prof. (Dr.) Siddhartha Bhattacharyya, RCC Institute of Information Technology, India

Prof. (Dr.) Ujjwal Maulik, Jadavpur University, India

Dr. Abhijit Das, RCC Institute of Information Technology, India

## **Program Chairs**

Dr. Indrajit Pan, RCC Institute of Information Technology, India

Ms. Satabdwi Sarkar, RCC Institute of Information Technology, India

Ms. Alokanda Dey, RCC Institute of Information Technology, India  
Ms. Abantika Choudhury, RCC Institute of Information Technology, India

## **Organizing Secretaries**

Mr. Ranjan Jana, RCC Institute of Information Technology, India  
Dr. Arindam Mondal, RCC Institute of Information Technology, India  
Mr. Hiranmoy Roy, RCC Institute of Information Technology, India  
Mr. Hrishikesh Bhaumik, RCC Institute of Information Technology, India

## **International Advisory Committee**

Dr. Swagatm Das, Indian Statistical Institute, Kolkata, India  
Prof. (Dr.) Sushmita Mitra, Indian Statistical Institute, Kolkata, India  
Prof. (Dr.) Aboul Ella Hassanien, Cairo University, Egypt  
Prof. (Dr.) Debotosh Bhattacharjee, Jadavpur University, India  
Prof. (Dr.) Kazumi Nakamatsu, University of Hyogo, Japan  
Prof. (Dr.) Valentina Emilia Balas, Aurel Vlaicu University of Arad, Romania  
Prof. (Dr.) Eduard Babulak, National Science Foundation, USA  
Prof. (Dr.) B. K. Tripathy, VIT University, India  
Prof. (Dr.) Elizabeth Behrman, Wichita State University, Kansas, USA  
Prof. (Dr.) Susanta Chakraborty, IEST, India  
Prof. (Dr.) Anupam Basu, IIT Kharagpur, India  
Dr. Ashish Mani, Amity University Noida, India  
Dr. Tandra Pal, NIT Durgapur, India  
Dr. Kaori Yoshida, Kyushu Institute of Technology, Japan  
Prof. (Dr.) Swapan Bhattacharya, Jadavpur University, India  
Dr. Md. Hasanuzzaman, University of Dhaka, Bangladesh  
Prof. (Dr.) Malay Kundu, Indian Statistical Institute, Kolkata, India  
Prof. (Dr.) Leonard Barolli, Fukuoka Institute of Technology (FIT), Japan  
Prof. (Dr.) Cesare Alippi, Politecnico di Milano, Italy  
Prof. (Dr.) Nik Bessis, Edge Hill University, UK  
Prof. (Dr.) P. N. Basu, Jadavpur University, India  
Prof. (Dr.) Prasanta Narayan Dutta, RCC Institute of Information Technology, India  
Prof. (Dr.) Zbigniew Michalewicz, The University of Adelaide, Australia  
Prof. (Dr.) Hisao Ishibuchi, Osaka Prefecture University, Japan  
Prof. (Dr.) Subhansu Bandyopadhyay, Brainware University, India  
Dr. Abhishek Das, Aliah University, Kolkata



## Technical Program Committee

Dr. Goran Klepac, University College Algebra, Croatia  
 Mr. Arup Bhattacharjee, RCC Institute of Information Technology, India  
 Dr. Indrajit Pan, RCC Institute of Information Technology, India  
 Dr. Abhijit Das, RCC Institute of Information Technology, India  
 Mr. Hiranmoy Roy, RCC Institute of Information Technology, India  
 Mr. Soumen Mukherjee, RCC Institute of Information Technology, India  
 Dr. Sandip Dey, Om Dayal Group of Institutions, India  
 Dr. Koushik Ghosh, University Institute of Technology, The University of Burdwan, India  
 Mr. Biswanath Chakraborty, RCC Institute of Information Technology, India  
 Dr. Kousik Dasgupta, Kalyani Government Engineering College, India  
 Dr. Anasua Sarkar, Jadavpur University, India  
 Mr. Debanjan Konar, Sikkim Manipal Institute of Technology, India  
 Dr. Jay Lofstead, Sandia National Laboratories, USA  
 Dr. Sourav De, Cooch Behar Government Engineering College, India  
 Mr. Subhadip Chandra, B. P. Poddar Institute of Management and Technology, India  
 Dr. Debashis De, West Bengal University of Technology, India  
 Dr. Elizabeth Behrman, Wichita State University, USA  
 Dr. Srijan Bhattacharya, RCC Institute of Information Technology, India  
 Dr. Amlan Chatterjee, California State University, USA  
 Dr. Pijush Barthakur, KLS Gogte Institute of Technology, India  
 Dr. David Meyer, University of Oregon, USA  
 Dr. Khaled El-Naggar, Benha University, Egypt  
 Dr. Rajat Mehrotra, Desert Research Institute, USA

## Publicity and Sponsorship Chairs

Mr. Soumen Mukherjee, RCC Institute of Information Technology, India  
 Mr. Biswanath Chakraborty, RCC Institute of Information Technology, India  
 Mr. Hrishikesh Bhaumik, RCC Institute of Information Technology, India  
 Mr. Pankaj Pal, RCC Institute of Information Technology, India

## Local Hospitality Chairs

Mr. Jayanta Datta, RCC Institute of Information Technology, India  
 Mr. Amit Khan, RCC Institute of Information Technology, India  
 Mr. Falguni Adhikary, RCC Institute of Information Technology, India  
 Mr. Kundan Kumar Singh, RCC Institute of Information Technology, India  
 Mr. Deepam Ganguly, RCC Institute of Information Technology, India

## Finance Chairs

Mr. Chinmoy Ghosal, RCC Institute of Information Technology, India  
Mr. Arup Kumar Bhattacharya, RCC Institute of Information Technology, India  
Mr. Manas Ghosh, RCC Institute of Information Technology, India  
Mr. Hrishikesh Bhaumik, RCC Institute of Information Technology, India

## Organizing Committee

Mr. Soumen Mukherjee, RCC Institute of Information Technology, India  
Mr. Jayanta Datta, RCC Institute of Information Technology, India  
Mr. Ranjan Jana, RCC Institute of Information Technology, India  
Mr. Arup Kumar Bhattacharya, RCC Institute of Information Technology, India  
Ms. Satabdwi Sarkar, RCC Institute of Information Technology, India  
Mr. Biswanath Chakraborty, RCC Institute of Information Technology, India  
Ms. Alokanda Dey, RCC Institute of Information Technology, India  
Dr. Arindam Mondal, RCC Institute of Information Technology, India  
Mr. Manas Ghosh, RCC Institute of Information Technology, India  
Mr. Hrishikesh Bhaumik, RCC Institute of Information Technology, India  
Dr. Dipankar Majumdar, RCC Institute of Information Technology, India  
Mr. Pankaj Pal, RCC Institute of Information Technology, India  
Mr. Hiranmoy Roy, RCC Institute of Information Technology, India  
Mr. Soumyadip Dhar, RCC Institute of Information Technology, India  
Ms. Abantika Choudhury, RCC Institute of Information Technology, India  
Ms. Moumita Deb, RCC Institute of Information Technology, India  
Dr. Abhijit Das, RCC Institute of Information Technology, India  
Mr. Amit Khan, RCC Institute of Information Technology, India  
Mr. Sudarsan Biswas, RCC Institute of Information Technology, India  
Dr. Srijan Bhattacharya, RCC Institute of Information Technology, India  
Ms. Sudipta Bose (Nandi), RCC Institute of Information Technology, India  
Mr. Ripan Mandal, RCC Institute of Information Technology, India  
Ms. Jayanti Das, RCC Institute of Information Technology, India  
Mr. Kundan Kumar Singh, RCC Institute of Information Technology, India  
Mr. Deepam Ganguly, RCC Institute of Information Technology, India  
Ms. Kamalika Banerjee, RCC Institute of Information Technology, India  
Ms. Moumita Chatterjee, RCC Institute of Information Technology, India

# Preface

Processing and analysis of signals have been at the helm of affairs in recent times, thanks to the revolution of data and signals. The signals might be speech, audio, images, video, sensor data, telemetry, electrocardiograms, or seismic data among others. The possible application areas include transmission, display, storage, interpretation, classification, segmentation, or diagnosis. Scientists and researchers across disciplines are venting out their efforts in designing efficient algorithms for processing and faithful analysis of different types of signals encountered in daily life.

A plethora of classical techniques has been in existence over the years to efficiently deal with signals and systems to foster the field of signal and image processing encompassing the theory and practice of algorithms and hardware that convert signals produced by artificial or natural means into a form useful for a specific purpose. The filtering techniques have played a recommendable role in this regard. However, the classical techniques often fall short when the signals become corrupted by noise artifacts.

In order to do away with the limitations of the classical techniques, scientists have evolved several intelligent methods entailing soft computing paradigm and machine intelligence. Moreover, the present era sees a great thrust in this direction with the advent of voluminous data and signal contents.

This volume, comprising 12 well-versed chapters, aims to open up a discourse of new realm of research in signal and image processing using intelligent computing paradigms.

Facial expression recognition is the process of identifying the expression that is displayed by a person, and it has several applications in the fields of medicine, human–computer interaction, and others, where recognition of expressions displayed on a face is of vital. Chapter “[Face Detection and Expression Recognition Using Haar Cascade Classifier and Fisherface Algorithm](#)” introduces a process for recognition of facial expressions using Haar classifier and Fisherface based on Fisher’s linear discriminant analysis (FLDA) for dimensionality reduction.

Chapter “[Pyramid-Based Multi-scale Enhancement Method for Iris Images](#)” proposes a pyramid-based image enhancement through multiscale image processing. It can increase the contrast of the image to compensate for the illumination

problem, compared to the existing methods. A comparative analysis with CLAHE and divide-and-conquer method is also presented in this chapter. Simulation results show that the proposed method gives promising results.

Character recognition from handwritten images is of great interest in the pattern recognition research community for their good application in many areas. To implement the system, it requires two steps, viz. feature extraction followed by character recognition based on any classification algorithm. Convolutional neural network (CNN) is an excellent feature extractor and classifier. The performance of a CNN for a particular application depends on the parameters used in the network. A CNN is implemented for the MNIST dataset with appropriate parameters for training and testing the system in Chapter “[Character Recognition from Handwritten Image Using Convolutional Neural Networks](#)”. The system provides accuracy up to 98.85%, which is better with respect to others. It also takes a very low amount of time for training the system.

Deep learning has been shown to efficiently identify patterns from images and signals from various application domains. The use of deep learning in medical image analysis is, however, limited till date. Two well-known malignant lesion image datasets, namely Dermofit and MED-NODE, are both separately and together used to analyze the performance of a proposed deep convolutional neural network (CNN) named as CNN malignant lesion detection (CMLD) architecture in Chapter “[Malignant Melanoma Classification Using Cross-Platform Dataset with Deep Learning CNN Architecture](#)”.

Soft biometrics such as gender significantly works to enhance the performance of biometric systems and is also having applications in human–computer interaction, content-based indexing and retrieval, demographic studies, security, and surveillance. Gender identification among adults is easy as compared to that in children, due to the similarity of facial skin texture and appearance of faces. The authors have introduced a technique to identify/classify the gender from facial images of children in Chapter “[Automatic Gender Identification from Children Facial Images Using Texture Descriptors](#)”. In this method, the authors have applied three basic steps, namely preprocessing, feature extraction, and classification. In the preprocessing stage, face detection and normalization are performed. The authors have computed different texture descriptors for extraction of powerful features.

At the physical layer, underwater acoustic communication (UWAC) channels present many challenges for efficient communications, which features both long delay spreads and serious Doppler effects. The waves and motions at the surface of the sea are the main causes for these effects. The complexity at the receiver increases along with the symbol rate, thereby increasing the burden of the channel. Orthogonal frequency-division multiplexing (OFDM) which is a form of multi-carrier modulation has been effectively used for UWAC recently because of its ability to handle long dispersive channels reducing the complexity of the channel equalization. However, intercarrier interference (ICI) is introduced by the Doppler effects that destroy the orthogonality of the sub-carriers. To overcome the effects of OFDM, an emerging technique known as filter bank multicarrier communication (FBMC) system is used in UWAC. The high capacity of the transmission

bandwidth in FBMC systems can be accomplished by applying offset quadrature amplitude modulation (OQAM). Because of cyclic prefix deletion, FBMC/OQAM can arrange higher spectral efficiency when compared to OFDM. Chapter “[Assessment of UWAC System Performance Using FBMC Technique](#)” uses both OFDM and FBMC/OQAM by considering the losses and noise effects in UWAC environment and illustrates their performances with respect to bit error rate (BER).

Simple exponential smoothing (SES) is a popular form of smoothing which can be regarded as a recursive system with the infinite impulse response (IIR). As a consequence, it suffers heavily from the threat of instability and phase nonlinearity. An effort has been made to get a finite impulse response (FIR) estimation of SES to acquire the benefits of an optimal speed and computational efficiency in addition to the usual advantages like stability and phase linearity in Chapter “[Simple Exponential Smoothing and Its Control Parameter: A Reassessment](#)”.

The domain of image processing has been very popular among researchers since a very long time. It is a well-known fact that there are myriad objectives that can be met with this subject. Chapter “[A Novel Cross-Dimensional Image Processing Technique: Toward a 3D View of a 2D Image](#)” proposes a novel technique of training by processing images on a two-dimensional plane, thereby subsequently producing a resultant three-dimensional view of an image. This transformation is attained through cross-dimensional image processing by studying, understanding, and analyzing the various attributes and properties of an image. This cross-dimensional technique explores the existing digital image processing domain where the inputted 2D image is mapped to produce a 3D view of the image in 3D space.

Chapter “[Quantum-Inspired Bat Optimization Algorithm for Automatic Clustering of Grayscale Images](#)” introduces a quantum-inspired framework with the bat optimization algorithm (QIBOA) for automatic clustering of image datasets. The aim of this chapter is to find out the optimal number of clusters from an image dataset on the run. A comparison has been made between the quantum-inspired bat optimization algorithm and its classical counterpart. As a result, it is seen that the quantum-inspired version outperforms its classical counterpart. Computational experiments have been conducted on four Berkeley image datasets.

With the advancement of technology and ease of the access of digital information and its creation and transference, the protection of copyrights for information security is of the prime concern. Digital watermarking gives copyright protection of digital images by impregnating additional information in the original cover image. Thus, watermarking prevents illegal copying by providing a method of enacting the ownership of a redistributed copy. The authors propose a wavelet-based dynamic LSB replacement method, which serves the purpose by providing excellent robustness and imperceptibility. Moreover, the watermark is encrypted by a symmetric key cryptographic algorithm to make it secure from further eavesdropping in Chapter “[A Wavelet-Based Blind Digital Image Watermarking using Dynamic LSB Replacement \(DLSBR\) and Symmetric Key Cryptography](#)”.

Chapter “[Fuzzy-Based Classification for Fusion of Palmprint and Iris Biometric Traits](#)” presents a technique of the fusion of palmprint and iris biometric traits for the purpose of classification. The region of interest (ROI) of the palm is detected and segmented based on the valley detection algorithm, and ROI of the iris is extracted based on the Neighbor-Pixels Value Algorithm (NPVA). The statistical local binary pattern (SLBP) is used for extracting the local features from the iris and palm. For enhancing the palm features, the combination of discrete cosine transform (DCT) and the histogram of oriented gradient (HOG) is applied. The Gabor–Zernike moment is used to extract the iris features. This experimentation was carried out in the identification system. The fuzzy Gaussian membership function was used as classification in the matching stage for the fusion system of palm and iris. The CASIA datasets of palm and iris were used as test datasets.

In the most recent years, numerous approaches have been accomplished in the discipline of human voice recognition for building speaker identification systems. Frequency- and time-domain techniques are widely used in extracting human voice features. Chapter “[Employing FPGA Accelerator in Real-Time Speaker Identification Systems](#)” presents the most robust and most popular Mel-frequency cepstral coefficient (MFCC) technique to parameterize voices to be used later in the voiced/unvoiced different feature extraction process methods. In addition, the direct classical techniques for human voice feature extraction purposes are used. A field-programming gate array (FPGA) is used to speed up the system performance for use in real-time applications.

The primary objective of the book is to bring a broad spectrum of the recent intelligent methods for signal and image processing so that it is able to trigger further inspiration among various research communities and contribute to their respective fields of applications.

The editors would like to take this opportunity in expressing their heartfelt regards to Mr. Aninda Bose, Senior Editor (Hard Sciences), Springer, India, for his constant support and guidance during the book project tenure.

Kolkata, India  
November 2018

Siddhartha Bhattacharyya  
Sankar K. Pal  
Indrajit Pan  
Abhijit Das

# About This Book

The book is aimed to report the recent trends in signal and image processing given the fact that almost every modern tools and gadgets resort to signal processing in one way or the other in this advanced era of human civilization. The relevant algorithms govern different application areas such as mobile communications, medical imaging, gaming and host of other technologies. As a matter of fact, scientists and researchers are continually engaged in developing time-efficient, robust, and failsafe signal processing algorithms for the benefit of mankind. The main reason behind these intensive efforts is to foster the field of signal and image processing encompassing the theory and practice of algorithms and hardware that convert signals produced by artificial or natural means into a form useful for a specific purpose.

From time immemorial, scientists have developed several efficient algorithms to process signals and images faithfully and extract meaningful information from them. However, the classical methods often fall short of offering robust solutions since the real-life signals are often uncertain and imprecise, which remains insurmountable to process in practice using the conventional classical methods.

Hence, advanced computational techniques have come up for addressing the complex real-life problems like handling the uncertainty in signal processing. The objective of such computational paradigms is to give rise to failsafe and robust solutions to the emerging problems faced by mankind. Several intelligent techniques have been in vogue over the year in this direction. Among these techniques, the latest hybrid soft computing techniques stand in good stead.

The proposed book aims to introduce to the prospective readers the latest trends in signal and image processing technologies with reference to both the classical and hybrid computational paradigms.

# Contents

<b>Face Detection and Expression Recognition Using Haar Cascade Classifier and Fisherface Algorithm</b> . . . . .	1
Indrasom Gangopadhyay, Anulekha Chatterjee and Indrajit Das	
<b>Pyramid-Based Multi-scale Enhancement Method for Iris Images</b> . . . . .	13
J. Jenkin Winston and D. Jude Hemanth	
<b>Character Recognition from Handwritten Image Using Convolutional Neural Networks</b> . . . . .	23
Ranjan Jana and Siddhartha Bhattacharyya	
<b>Malignant Melanoma Classification Using Cross-Platform Dataset with Deep Learning CNN Architecture</b> . . . . .	31
Soumen Mukherjee, Arunabha Adhikari and Madhusudan Roy	
<b>Automatic Gender Identification from Children Facial Images Using Texture Descriptors</b> . . . . .	43
Ayesha Iftikhar, Rehan Ashraf, Asim Saeed, Srinivas Alva and Rajmohan Pardeshi	
<b>Assessment of UWAC System Performance Using FBMC Technique</b> . . . . .	53
B. Pranitha and L. Anjaneyulu	
<b>Simple Exponential Smoothing and Its Control Parameter: A Reassessment</b> . . . . .	63
Dipta Chaudhuri, Moloy Mukherjee, Mofazzal H. Khondekar and Koushik Ghosh	
<b>A Novel Cross-Dimensional Image Processing Technique: Toward a 3D View of a 2D Image</b> . . . . .	79
Dipnarayan Das and Sumit Gupta	



**Quantum-Inspired Bat Optimization Algorithm for Automatic Clustering of Grayscale Images** ..... 89  
Alokananda Dey, Siddhartha Bhattacharyya, Sandip Dey, Jan Platos and Vaclav Snasel

**A Wavelet-Based Blind Digital Image Watermarking using Dynamic LSB Replacement (DLSBR) and Symmetric Key Cryptography** ..... 103  
Ananya Ghosh, Subhadeep Koley and Sagnik Acharyya

**Fuzzy-Based Classification for Fusion of Palmprint and Iris Biometric Traits** ..... 113  
Akram Alsubari, Preeti Lonkhande and R. J. Ramteke

**Employing FPGA Accelerator in Real-Time Speaker Identification Systems** ..... 125  
Omran Al-Shamma, Mohammed A. Fadhel and Haitham S. Hasan

**Author Index** ..... 135

# About the Editors

**Siddhartha Bhattacharyya** did his Bachelors in Physics, Bachelors in Optics and Optoelectronics and Masters in Optics and Optoelectronics from University of Calcutta, India in 1995, 1998 and 2000 respectively. He completed Ph.D. in Computer Science and Engineering from Jadavpur University, India in 2008.

He is currently a Senior Research Scientist in VSB Technical University of Ostrava, Czech Republic. He is also the Principal of RCC Institute of Information Technology, Kolkata, India. He is a co-author of four books and the co-editor of 20 books and has more than 200 research publications in international journals and conference proceedings to his credit. He has also published two PCTs. He is the founding Editor-in-Chief of International Journal of Hybrid Intelligence, Inderscience. He is also on the editorial board of reputed international journals.

His research interests include soft computing, pattern recognition, multimedia data processing, hybrid intelligence and quantum computing. Dr. Bhattacharyya is a life fellow of OSI. He is a fellow of IETE and IEI. He is also a senior member of IEEE, IETI and ACM. He is a life member of CRSI, CSI, ISTE, IUPRAI and CEGR. He is a member of IET (UK), IRSS, IAENG, CSTA, IAASSE, IDES, ISSIP and SDIWC.

**Sankar K. Pal** a Distinguished Scientist and former Director of Indian Statistical Institute, is currently an INSA Distinguished Professor Chair. He is an alumnus of Calcutta University; Indian Statistical Institute; and Imperial College, London.

He worked at NASA-Johnson Space Center, Houston; University of California, Berkeley; University of Maryland, College Park; and US-Naval Research Lab, Washington DC. He is an IEEE-CS Distinguished Visitor for Asia-pacific region, and visited forty five countries as a Keynote/Invited speaker or an academic visitor.

Prof. Pal is a Fellow of TWAS (The World Academy of Sciences), IEEE, IAPR, IFSA, IRSS, and all the four Indian academies in science/engineering; decorated with most coveted S.S. Bhatnagar Prize (Science award) and Padma Shri (Civilian award) in India, and Khwarizmi International Award in Iran, among others; and co-author of 20 books, and 400+ publications in pattern recognition, machine intelligence, soft computing, granular mining and Big data. He serves(d) in editorial boards of 20+ international journals, including IEEE Transactions.

**Indrajit Pan** has done his Bachelor of Engineering (B.E.) in Computer Science and Engineering with Honors from The University of Burdwan in 2005 and completed Master of Technology (M.Tech.) in Information Technology from Bengal Engineering and Science University, Shibpur in 2009. He was the recipient of University Medal for his academic performance during Masters. He obtained Ph.D. in Engineering from Department of Information Technology, Indian Institute of Engineering Science and Technology, Shibpur in 2015. His present research interest includes social network analysis and cloud computing. Indrajit is now Associate Professor of Information Technology Department at RCC Institute of Information Technology, Kolkata. He has around thirty research publications in different International journals, Edited books and conference proceedings. He also co-authored couple of edited research volumes with prestigious publishers like Wiley, Taylor and Francis, Springer Nature. Indrajit is the Senior member of IEEE, USA and Member of ACM, USA and CSI.

**Abhijit Das** received his B.Tech. in IT from University of Kalyani in the year 2003, M.Tech. in IT from University of Calcutta in the year 2005 and Ph.D. (Engineering) from Department of CSE Jadavpur University in the year 2016. He has over 15 years of Teaching and Research experience and more than 30 publications of international repute. Presently he is serving as an Associate Professor in the department of IT, RCC Institute of Information Technology, Kolkata, India. He has organised and chaired various international conferences and seminars. Four research scholars are currently working with him in different topics like IoT, Precision Agriculture, e-Waste Management, Object Oriented Categorization etc. Prof. Das has acted as reviewer for many reputed journals and is professional member of IEEE and ACM.

Abhijit is also a professional singer and is a regular artist of All India Radio Kolkata and various Television channels and has more than 10 music albums on his credit.

# Face Detection and Expression Recognition Using Haar Cascade Classifier and Fisherface Algorithm



Indrasom Gangopadhyay, Anulekha Chatterjee and Indrajit Das

**Abstract** Facial expression recognition is the process of identifying the expression that is displayed by a person, and it has several applications in the fields of medicine, human–computer interaction others; where recognition of expressions displayed on a face is of vital. The process mainly comprises face detection and expression recognition using Haar classifier and using Fisherface based on Fisher’s linear discriminant analysis (FLDA) for dimensionality reduction, respectively. The dataset from which the faces were presented to the classifiers yielded a precision of 96.3% with a recognition speed of 8.2 s. An improvement in recognition accuracy of 3.4% is observed by this algorithm from other algorithms, viz. eigenfaces, LBPH recognizer, and artificial neural network; although with a drawback of incorrect recognition in cases of uneven illumination or low-light conditions. This result may be considered as efficient both with respect to accuracy and speed of recognition of the facial expressions.

**Keywords** Haar cascade classifier · Fisherface · Cohn-Kanade database · Facial expression recognition · Face detection

## 1 Introduction

Facial expression analysis is a difficult challenge which has become an interesting topic for the research. The expression on one’s face provides an insight about their mood or feelings and can play a vital role in social communication. Facial expressions are also useful in forensic, biometric, and neuroscience studies [1]. It also has applications in image comprehension, psychological evaluation, animation of face,

---

I. Gangopadhyay (✉) · A. Chatterjee · I. Das  
Meghnad Saha Institute of Technology, Kolkata, India  
e-mail: [indrasomgangopadhyay@yahoo.com](mailto:indrasomgangopadhyay@yahoo.com)

A. Chatterjee  
e-mail: [anulekha1312@gmail.com](mailto:anulekha1312@gmail.com)

I. Das  
e-mail: [indrajitdas1979@hotmail.com](mailto:indrajitdas1979@hotmail.com)

© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_1](https://doi.org/10.1007/978-981-13-6783-0_1)

human–computer interaction (HCI) and in fields like medicine to determine state of a patient, or in criminology for determining the truthfulness of a criminal, etc. A human face has salient and absolute features due to which the detection of expressions of human faces is possible. Once a face is identified in an image, most suitable face characteristics are extracted for expression detection [2].

The critical procedure of the facial expression recognition (FER) technology includes acquiring the image, detecting the face, extracting features and lastly classification of facial expressions. Firstly, the face needs to be detected, which is a primary step in expression recognition. The feature extraction methods are described based on whether they emphasize on motion of faces and its features or the deformation of faces. Motion-based approaches can be implemented by using dense optical flow, pattern tracking, feature point extracting, etc. Deformation-based approaches have to bank on neutral face in order to retrieve facial features and not get affected by changes such as aging signs like wrinkles. Here, the relevant features of a face like mouth, eyebrow, and nose are utilized for expression analysis [3]. Finally the classification of the facial expression based on the features extracted. Feature extraction is the most vital step to get the region of interest (ROI) [4].

In this paper, faces are classified into eight expressions: happy, angry, surprise, contempt, disgust, fear, surprise, and neutral. The Cohn-Kanade database is used for the process of training and testing the classifier. As for the detection of the face, the Haar cascade classifier is used here. The classifier is trained with positive images (images containing faces) and negative images (images lacking face). After training of the classifier is done, it is applied to an ROI on an image and it scans for the specific object. Fisherface algorithm is used for feature extraction, recognition, and classification of facial expressions. It is very important for expression recognition because if features used are not adequate, then the system would fail to generate the correct expression.

The remaining part of the paper is presented in this way: In Sect. 2, a thorough literature survey has been conducted to this domain. Section 3 presents the method that has proposed in this paper. Section 3.4 put forth the experimental results, and Sect. 4 describes the conclusion.

## 2 Literature Survey

Over the time, various researchers worldwide have already worked on the subject of facial expression recognition and they have utilized and conceived a heap calculation that assesses the execution of their proposed approaches and plans. In this segment, a few surveys of such existing research along with their corresponding accuracy of computational executions is discussed below.

Researchers Ranade et al. [5] proposed a method by implementing local binary pattern (LBP) for analysis of facial expression. First is the splitting up of the input image of faces into a group of smaller regions, and from that, LBP histograms are drawn which are fabricated into one feature histogram. Use of uniform LBP is the

primary reason for high performance. The highest and average recognition accuracy rates are 94.1 and 82.08%.

Researchers Dhavalikar et al. [6], in their paper, put forward a method where skin color detection and segmentation are used for face detection. The process implemented to recognize facial features is termed as active appearance model (AAM). Then, simple Euclidean distance and artificial neuro-fuzzy inference system (ANFIS) are used for categorized the expressions as sad, happy, fear, anger, surprised, and disgust. Based on Euclidean distance, the true recognition rate is around 90–95%. Modifying the technique using ANFIS, an accuracy of around 100% is achieved.

Researchers Sarode et al. [7] implemented a method using 2D appearance-based local method for extracting facial features and recognizing four facial expressions. The algorithm devises radial symmetry transform and additionally makes use of edge projection analysis for feature extraction. This subsequently leads to the classification into one of the four expression classes, viz. happy, surprise, sad, and anger. The accuracy has been measured around 81.0% for facial recognition from grayscale images.

Researchers Zhao et al. [8] devised a method in their paper by implementing a new kernel-based multifarious learning method called kernel discriminant isometric mapping (KDIsoMap). It is used for performing nonlinear dimensionality depletion on the extracted facial features, and it generates low-dimensional discriminant embedded data representations with noticeable advancement in the performance on functions recognizing facial expressions. For facial expression classification, the nearest neighbor classifier along with the Euclidean metric is applied. Two databases are used for facial expression recognition experiments, i.e., the Cohn-Kanade database (accuracy is 94.88%) and the JAFFE database (accuracy of 81.59%).

Researchers Chitra et al. [9] in their paper proposed a method by implementing Gabor filtering (also known as Gabor wavelengths or kernels) for facial recognition. In this paper, preprocessing is done using contrast and illumination equalization, fuzzy filtering, and histogram equalization. After that, detection of faces and extraction of facial features are done with the assistance of local binary pattern (LBP). The expressions are then classified into happy, surprise, anger, sad, fear, disgust, neutral and are categorized. SVM algorithm is adopted here for face recognition and facial expression recognition and produces accuracy of approximately 90%.

Researchers Anil. et al. [10] have applied some tailor-made facial expression recognition algorithms like regional registration, gradient feature matching, patched geodesic texture transform, FARO, bag-of-words method, and curvelet feature extraction. They have considered several databases for demonstration purposes which include Cohn-Kanade, Yale, JAFFE, CMU AMP, AR, BU-3DFE, and databases. The main steps performed here are facial image preprocessing, face detection, detecting facial components, extracting features, and finally classifying into the respective expression category. The expressions that are contemplated for identification are happy, disgust, anger, fear, sadness, and surprise. FARO and bag-of-words method produce best results (recognition rate of 94.74%) even with occluded images.

Researchers Kumar et al. [11] describe different stages used for recognizing facial expressions, viz. face acquisition, face normalization, face segmentation, facial fea-

ture extraction, facial feature representation, and expression classification. For face normalization scale, illumination and pose are enhanced. In segmentation, the face is differentiated from the image background and then features are separated. For facial feature extraction, image-based and model-based techniques are employed. To represent facial features, muscle-based or component projection methods are used. Finally, during facial expression recognition, expression interpretation is done. An accuracy of 96.83 and 92.22% was found on JAFFE and Yale databases, respectively.

### **3 Proposed Methodology**

The process for recognizing facial expressions mainly consists of three steps as follows:

Image acquisition  
Face detection  
Facial expression recognition.

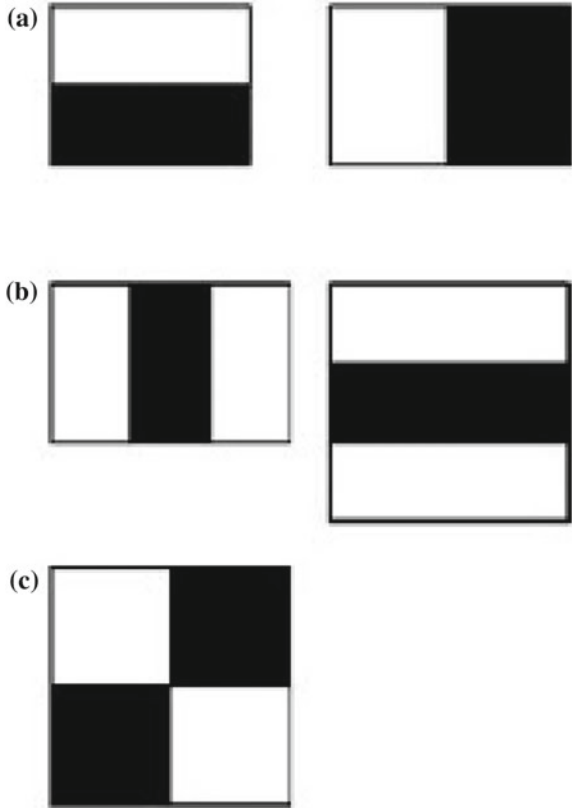
#### ***3.1 Image Acquisition***

The Cohn-Kanade database is used for training and testing of images in this paper. The expressions are divided into eight folders according to the expression contained in the image under investigation. Out of all the database images, most of the images contain archetypical expressions such as neutral, sadness, surprise, happiness, fear, anger, contempt, and disgust. The Cohn-Kanade database is a posed expression database with spontaneous smiles. The images of this database are mostly gray and stored in the following manner: The first face image in every folder displays no expression, whereas the last image consists of an expression, and the images in between contain the transformation from a neutral face to the expression face.

#### ***3.2 Face Detection***

Several techniques have been proposed for face detection as can be seen from the literature survey. While there are many algorithms which have been proven to be effective under small variations of lighting, facial pose and expression, camera focus, or shadowing yet for large variations of such factors, the results of the algorithms are less impressive. However, keeping in mind the above-mentioned conditions, two algorithms have been used which have been proven to be highly compelling in the past.

**Fig. 1** **a** Edge features, **b** line features, and **c** rectangular features

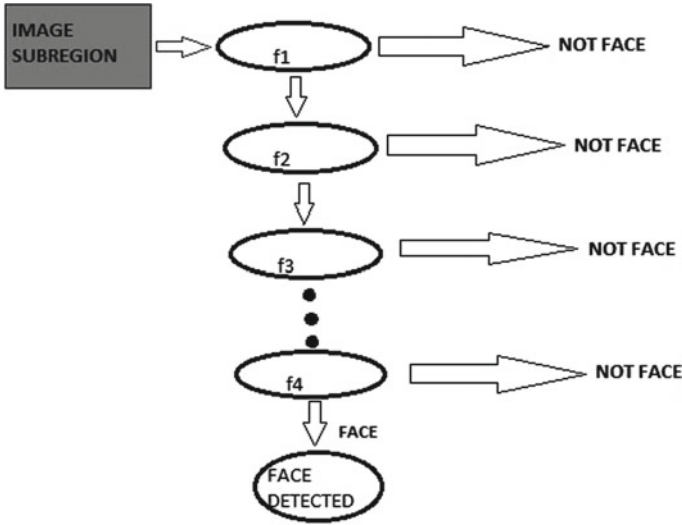


Haar feature-based classifier is first trained with images from Cohn-Kanade dataset where few images are stated as positive examples; i.e., the images with faces and the remaining images are called negative examples, i.e., images without faces. After the classifier is trained, it can be applied to an ROI in a given image. The classifier produces the output as 1 if the region shows the face of a person or 0 otherwise. A Haar object detector uses edges, lines, and combination of four rectangles for detection of face. Countless rectangular features are detected by the Haar classifier, and it detects them into regions to create the definition of an object shown in Fig. 1.

Haar cascade classifier makes use of the Haar-like features to detect a face in an image. This improved feature set is known as two-rectangle feature. The areas surrounding the eyes are commonly found to be comparatively darker than the cheeks. A case of a Haar-like component for detecting the face is an arrangement of two neighboring rectangular zones over the eye and cheek regions.

Figure 2 shows the cascade of classifier containing a list of stages of weak learners. The classifier detects an object by placing a window over the given image. Each sub-window labels the region it moves upon as positive, meaning an object or a face is





**Fig. 2** Stages of classification

detected, or negative, meaning the specified object, or the face, is not present. If a negative result (face not found) is produced by the labeling, the classification of that region can be termed as complete and the window is shifted to the next location. If the labeling yields positive result (face found), then region moves on to the next stage. The classifier finally gives a result when each and every stage yields a positive result; that is, the face in the image is detected. After detecting a face using Haar cascade classifier, the facial regions from the images are cropped and are separated into the respective expression folder for training.

### 3.3 Facial Expression Recognition

Facial expression recognition—Fisherface [12] method is a derivative of Fisher's Linear Discriminant (FLD). Based on the feature set, it can be implemented in a lot of ways for classification. This algorithm contemplates the disparity between two persons. Variance may also arise in database images due to contortions such as illumination, facial expression, and pose variation.

The mathematical representation of the Fisherface algorithm is as follows. Let  $x_{ij}$  be the  $i$ th sample of class  $j$ ,  $\mu_j$  be the mean of class  $j$ , and  $n_j$  be the number of samples in class  $j$ . The computation for more than two classes is done by minimizing intraclass differences and maximizing interclass distances. Intraclass differences can be estimated using the intraclass scatter matrix, given by  $S_w$ , and the between class, differences are computed using the interclass scatter matrix,  $S_b$ . To minimize the

determinant of intraclass scatter matrix and maximize that of interclass scatter matrix, the  $S_w^{-1}S_b$  matrix is constructed to get Fisherfaces.  $S_w$  and  $S_b$  are defined as follows:

$$S_w = \sum_{i=1}^{n_j} \sum_{j=1}^C (x_{ij} - \mu_j)(x_{ij} - \mu_j)^T \quad (1)$$

$$S_b = \sum_{j=1}^C (\mu_j - \mu)(\mu_j - \mu)^T \quad (2)$$

where  $\mu$  denotes the mean of all classes. Now, the task is to find those basis vectors  $V$ , where  $S_w$  is minimized and  $S_b$  is maximized, where  $V$  is a matrix whose columns  $v_i$  are the basis vectors defining the subspace.

$$|V^T S_b V| / |V^T S_w V| \quad (3)$$

It is solved by eigenvalue decomposition,  $S_b V = S_w V \Lambda$ , where  $\Lambda$  denotes a diagonal matrix of corresponding eigenvalues and  $V$  denotes a matrix of eigenvectors. The nonzero eigenvalues of eigenvectors  $V$  are the Fisherfaces. There is a maximum of  $C - 1$  Fisherfaces. Here, Fisherface is used for training a certain percentage of dataset for training the system to detect expressions, and the rest serves as test data set. In the upcoming section, the experimental results from the experiment performed using the proposed algorithm are explained.

### 3.4 Experimental Results

The experiment was done in Python IDLE 3.6.4 on Windows 8.1 Operating System with Intel Core i3 running at 2.6 GHz using Cohn-Kanade dataset. The data was organized into the following eight categories: happy, anger, neutral, contempt, sad, fear, surprise, and disgust. The training set consists of 456 images with 57 images present in each expression category, i.e., happy, anger, neutral, contempt, sad, fear, surprise, and disgust. From the whole dataset, about 69% was utilized for training the classifier, while the remaining 31% was used for testing it. This process was repeated 10 times, calculating the accuracy every time, and the average accuracy was measured.

In the experiment, firstly, the detection of the face using Haar classifier is done which is shown in Fig. 3. After that, in Fig. 4, cropping the face region in the image and resizing are done. Finally, the classification algorithm is used to detect the expressions of the faces which are shown in Fig. 5.

The average accuracy of the proposed expression recognition algorithm was found to be 96.3% as depicted in Fig. 6. The classifier, if had not been trained, would randomly be assigned classes, and the probability of recognizing a correct expression would have been 12.5% ( $(1/8) * 100$ ). Hence, it can be said that the proposed algorithm



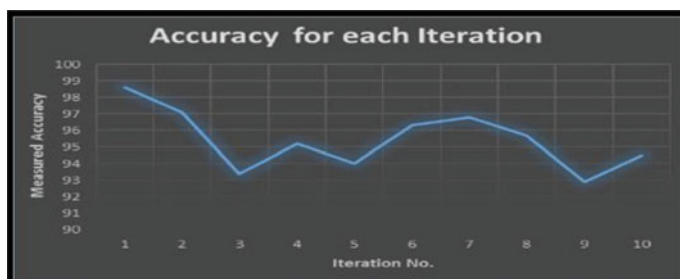
Fig. 3 Detection of faces



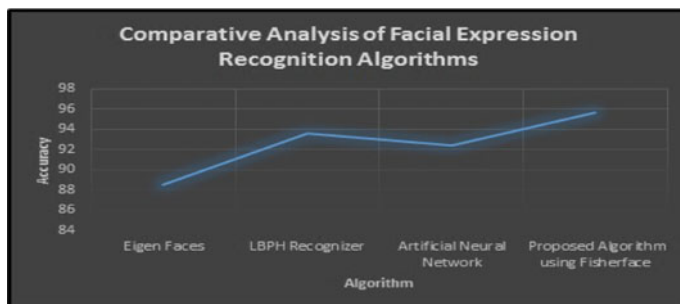
Fig. 4 Original faces and cropped faces



**Fig. 5** Expressions of faces detected

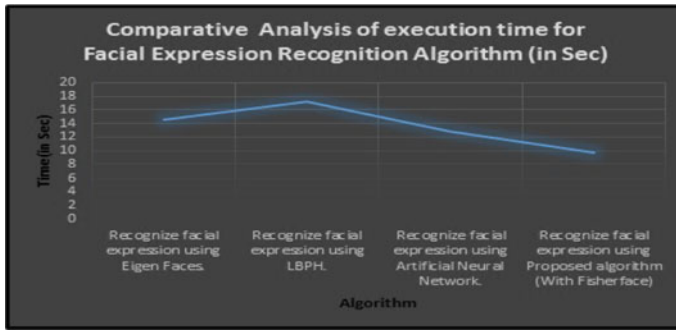


**Fig. 6** Accuracy of the proposed algorithm using Fisherface



**Fig. 7** Comparative analysis of facial expression recognition algorithms

shows a performance improvement of 83.8% over a classifier that has not been trained. Figures 7 and 8 show a comparative analysis of the proposed method with other algorithms in terms of accuracy and speed. The results show that the proposed algorithm gives the best result against other algorithms. The following section puts forth the conclusion.



**Fig. 8** Comparative analysis of execution time for facial expression recognition algorithm

## 4 Conclusion

A process of face detection and expression recognition is proposed and implemented in this paper. Eight expressions, viz. happy, angry, surprise, contempt, disgust, fear, surprise, and neutral, are recognized from the faces. The face detection process is accomplished using Haar classifier. The expression recognition process is performed using Fisherfaces which produces 96.3% accuracy which is better by at least 3.4% in comparison to other algorithms. The proposed algorithm has an expression recognition execution time of 8.2 s which is faster than other algorithms by at least 3.2 s. These results may be used to conclude that the algorithm proposed in this paper is efficient both in terms of accuracy and recognition speed.

## References

1. Sarode N, Bhatia S (2010) Facial expression recognition. *Int J Comput Sci Eng* 2(5):1552–1557. ISSN: 0975-3397
2. Zeng W, Liu C (2016) Facial expression recognition based on texture and shape. In: 25th wireless and optical communication conference (WOCC), pp 1–5. ISBN: 978-1-4673-9958-6
3. Kumar Y, Sharma S (2017) A systematic survey of facial expression recognition techniques. In 2017 international conference on computing methodologies and communication (ICCMC), pp 1074–1079. ISBN: 978-1-5090-4890-8
4. Happy SL, Routray A (2015) Automatic facial expression recognition using features of salient facial patches. *IEEE Trans Affect Comput* 6(1):1–12. ISSN: 1949-3045
5. Ranade SK, Mangat AM (2016) Facial expression recognition based on local binary pattern (LBP). *Int J Adv Res Comput Sci* 85–94. ISSN: 0976-5697
6. Dhavalikar AS, Kulkarni RK (2014) Face detection and facial expression recognition system. In: 2014 International Conference on Electronics and Communication System, pp. 1–7. ISBN: 978-1-4799-2320-5
7. Sarode N, Bhatia S (2010) Facial expression recognition. *Int J Comput Sci Eng* 2(5): 1552–1557. ISSN: 0975-3397
8. Zhao X, Zhang S (2011) Facial expression recognition based on local binary patterns and kernel discrimination isomap. ISSN: 1424-8220

9. Chitra N, Nijhawan G (2016) Facial expression recognition using local binary pattern and support vector machine. *Int J Innovative Res Adv Eng* 103–108. ISSN: 2349-2763
10. Anil J, Suresh LP (2016) Literature survey on face and face expression recognition. In: 2016 international conference on circuit, power and computing technologies [ICCPCT]
11. Kumar Y, Sharma S (2017) A systematic survey of facial expression recognition. In: 2017 international conference on computing methodologies and communication, pp 1074–1079. ISBN: 978-1-5090-4890-8
12. Zhang S, Zhao X, Lei B (2011) Facial expression recognition using local fisher discriminant analysis. In: *Communications in computer and information science*, pp 443–448

# Pyramid-Based Multi-scale Enhancement Method for Iris Images



J. Jenkin Winston and D. Jude Hemanth

**Abstract** The uniqueness of the iris makes it an effective physiological biometric trait which helps in personal identification. In an iris-based personal identification system, image enhancement plays a vital role. An apt enhancement will help in the proper localization of iris in the image. This paper proposes a pyramid-based image enhancement through multi-scale image processing. It can increase the contrast of the image to compensate for the illumination problem, compared to existing methods. A comparative analysis with CLAHE and divide-and-conquer method is also performed in this paper. Simulation results show that the proposed method gives promising results.

**Keywords** Pyramid · Iris image enhancement · Iris recognition · Biometrics

## 1 Introduction

Security has become a very crucial issue to be solved. The breaches in security have caused huge loss in almost every sector of life. Many manual and automated standards have emerged to overcome this menace. One such system introduced with the advancement of digital technology is biometric system. Various biometric features like signature, voice, typing, walking gait, fingerprint, iris, and face can be thought for easy access. However, out of these features, iris is more unique and noninvasive. Hence, it becomes more effective for personal identification system [1].

Iris is a meshwork of connective ligaments and ciliary muscles forming crypts, furrows, and freckles. This texture-like pattern is highly unique for each individual and does not change over the life span. It is captured using digital cameras and further processed to identify the person. During the capturing process, because of

---

J. J. Winston (✉) · D. J. Hemanth  
Department of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India  
e-mail: [jenkinwinston@karunya.edu](mailto:jenkinwinston@karunya.edu)

D. J. Hemanth  
e-mail: [judehemanth@karunya.edu](mailto:judehemanth@karunya.edu)

© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_2](https://doi.org/10.1007/978-981-13-6783-0_2)

the non-ideal environment, the captured image is prone to be affected by artifacts. These artifacts reduce the visibility of textural features in the iris. Artifact caused due to low illumination is one such problem faced by the iris-based biometric system. Elimination of this artifact is done by a process called image enhancement.

Image enhancement is a method of improving the brightness, sharpness, and contrast of the image. This helps in proper localization of the iris in the image.

Raffei et al. in [2] have used an iso-CLAHE approach on iris images with low lighting and low contrast. The clip parameter is selected based on the entropy of the sub-image in luminous plane for histogram equalization. However, the sharpness of the image needs to be further improved for better results. Sanpachai et al. in [3] have done image enhancement through different variants of histogram equalization and studied its performance. Among them, multi-scale adaptive histogram equalization and contrast limited adaptive histogram equalization perform better. Yet, the enhancement is nonlinear. Sajjad et al. in [4] have enlarged the iris region using bilinear interpolation and then applied CLAHE in order to get a well-distributed textural pattern. Despite enlarging, the features are not sharp. Hemalatha et al. in [5] have tried a series of filtering followed by histogram equalization to improve the quality of the image. Median filter is used to remove the noise. Later, Gabor filter is used to sharpen the edges, and then, histogram equalization is done to enhance the contrast. In this method, filtering has not contributed much to the performance of histogram equalization.

Tammam et al. in [6] have used a 2D complex matched filter to eliminate the illumination spots in the pupil region. The region is replaced with black pixel. This may lead to false localization of pupil boundary. Kumar et al. in [7] have experimented with a different series of steps. The image is converted to gray scale. Later, histogram equalization is done. Then, median filtering is done. And then gamma intensity correction is done to improve the brightness. And finally, top-hat and bottom-hat filtering is done to further improve the contrast. However, this method fails to trace the vertical boundary. Kiruthiga et al. in [8] have demonstrated smoothing through fractional derivative mask in both  $x$  and  $y$  direction. This is done to highlight the important features. Kumar et al. in [9] have tried high-boost median filtering to improve the PSNR of the iris image. Amplification is done on high-frequency and low-frequency components to achieve this. Although not any constraints on amplification factor is discussed. A similar kind of work is also reported in [10–12].

From the literature, it is clear that very few works are done on low-illuminated iris images. In this paper, the performance of our proposed multi-resolution enhancement with CLAHE and divide-and-conquer method on low-illuminated iris images is evaluated. The experimental results show improvement in the luminance, and the texture of the image is preserved.

New technologies continue to evolve in biometric recognition systems. In systems, the first input that it receives should be free from artifacts. In non-ideal situations, it is practical to consider images with low illumination. To improve the visual quality of the image, a multi-resolution fusion-based enhancement method is proposed in this paper.



The rest of the paper is arranged as follow. In Sect. 2, the proposed multi-resolution algorithm for iris image enhancement in sequence and short notes on the divide-and-conquer method and CLAHE method is given. Later, in Sect. 3, a discussion on the results and comparison with other methods is made. At last, conclusions are discussed in Sect. 4.

## 2 Proposed Methodology

The outline of the proposed work is presented in Fig. 1. The proposed methodology along with CLAHE and divide-and-conquer method is used to enhance the image, and then, their quality is measured.

### 2.1 Multi-resolution Enhancement

According to retinex theory [13], each digitally captured image can be considered as a product of brightness and radiance. Several enhancement algorithms based on this theory are there. However, these methods for enhancing iris images captured in low illumination are not explored.

The basic idea of this method is to combine the several derived inputs from the brightness and radiance image generated from a single iris image. The global brightness is improved, local contrast is bettered, and the texture of the iris image is preserved to enhance the low-illuminated iris image. Initially, the image is separated as brightness and radiance images. The radiance image. From each derived image, the Laplacian and weights Gaussian image pyramid are fused. The reconstructed

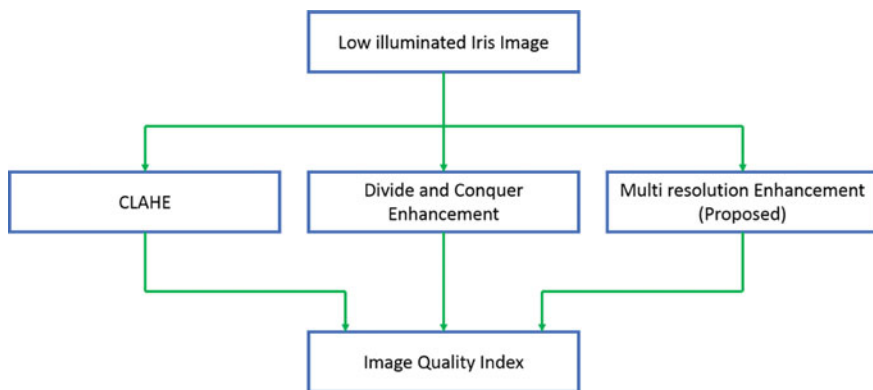
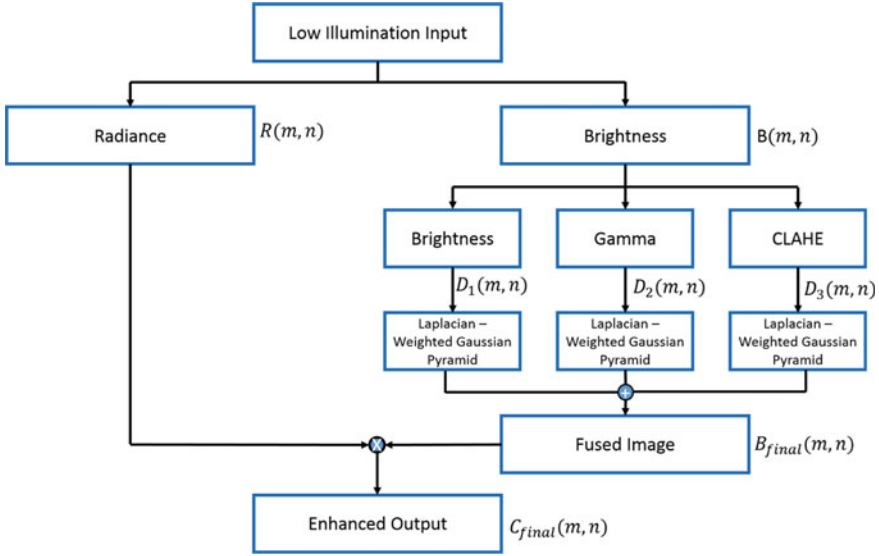


Fig. 1 Outline of the proposed work



**Fig. 2** Flowchart of the multi-resolution enhancement method

image is then combined with the brightness image to get the enhanced image. This is shown in Fig. 2.

From the retinex theory, the image from the database is taken as the product of brightness and radiance.

$$C^p(m, n) = B^p(m, n) * R(m, n) \quad (1)$$

where  $C$  is the captured image,  $B$  is the brightness,  $R$  is the radiance,  $p$  is the index for RGB color channels, and  $(m, n)$  is the index for a pixel location. The brightness image contains information about the illumination variation and can be assumed to be locally smooth [14]. The illumination profile of the image is generated by finding the maximum intensity value from the three color planes. This is given by

$$L(m, n) = \max C^p(m, n) \quad (2)$$

The brightness plane is estimated by performing morphological operation on the illumination profile. This smoothens the local features. The brightness image is given by

$$B = \frac{L \bullet S}{255} \quad (3)$$

where  $S$  is the structuring element, and  $\bullet$  denotes the closing morphological operation. Equation 2 maps the illumination profile to the range  $[0, 1]$ . As the pupil and the iris

region are circular, choosing disk as the structuring element would be the best for performing the closing operation. This operation smoothens the contours by filling the gaps and fusing the discontinuity.

This estimated brightness image is further enhanced by using guide filter. The guide filter is a derived linear model to preserve the edges. The filter kernel  $G$  is chosen from the  $V$  channel of HSV color space of the input image. The filter output is the weighted average at pixel  $i$ , given by

$$B_i = \sum_j G_{i,j}(V) B_j \quad (4)$$

where  $G$  is the filter kernel that is a function of guiding image  $V$ . The kernel is given by

$$G_{i,j}(V) = \frac{1}{N^2} \sum_{w:i,j} 1 + \frac{(V_i - \mu_w)(V_j - \mu_w)}{\sigma_w^2 + \varepsilon} \quad (5)$$

where  $N$  is the number of pixels in the filter kernel window,  $\mu_w$  and  $\sigma_w^2$  are the mean and variance of  $V$  within the window  $w$ , and  $\varepsilon$  is the normalization parameter. The radiance image can now be computed from Eq. 1, after the brightness image is estimated.

In the next stage, the derived images D1, D2, and D3 are computed. The image D1 has the same brightness image estimated. The image D2 is the gamma-corrected image. This is done to address the global illumination problem. To enhance the local contrast feature, the third derived image D3 is computed using contrast limited adaptive histogram equalization (CLAHE).

The next stage is the creation of normalized weight map for each derived image. The weight for the  $k$ th derived image is given by

$$W_k(m, n) = W_{b,k}(m, n) * W_{c,k}(m, n) \quad (6)$$

where  $W_{b,k}(m, n)$  is the brightness weight, and  $W_{c,k}(m, n)$  is the contrast weight for the  $k$ th derived image. It is given by

$$W_{b,k}(m, n) = \exp \left\{ -\frac{D_k(m, n) - 0.5^2}{2 * 0.25^2} \right\} \quad (7)$$

$$W_{c,k}(m, n) = D_k(m, n)(1 + \cos(\alpha H(m, n) + \phi)S(m, n)) \quad (8)$$

where  $H$  and  $S$  are the hue and saturation channel in HSV color space of the low-illuminated input iris image. For this image set, the parameters are chosen as  $\alpha = 3$  and  $\phi = 240$  to preserve the color features. This highlights the region of high contrast. The normalized weight image for the  $k$ th derived image is given by

$$W_{\text{nor},k}(m, n) = \frac{W_k(m, n)}{\sum_k W_k(m, n)} \quad (9)$$

Next is the fusion of images in the pyramid level. The derived images are low-pass filtered and decomposed into Laplacian pyramid. It extracts the edge features. The normalized weight image is decomposed into the Gaussian pyramid. It extracts the smooth image features. The fusion of pyramid at each level is given by

$$F_l(m, n) = \sum_k \text{Ga}_l\{W_{\text{nor},k}(m, n)\} \text{La}_l\{D_k(m, n)\} \quad (10)$$

where  $l$  denotes the level of decomposition,  $\text{Ga}\{W_{\text{nor},k}\}$  is the Gaussian pyramid of  $k$ th normalized weight image, and  $\text{La}\{D_k\}$  is the Laplacian pyramid of  $k$ th derived image. Then, the enhanced brightness adjusted image is obtained by summing fused pyramid at each level.

$$B_{\text{final}}(m, n) = \sum_l \text{Up}\{F_l(m, n)\} \quad (11)$$

where Up is the up-sampling operator. The final enhanced image is obtained by

$$C_{\text{final}}(m, n) = B_{\text{final}}(m, n)R(m, n) \quad (12)$$

### 3 Divide-and-Conquer Method

As high frequency components and low frequency components of an image are totally different parts. So, in this method they operate on image subspaces individually to enhance the image. A detailed study on this method is presented in [15].

### 4 Contrast Limited Adaptive Histogram Equalization (CLAHE)

The images are split into smaller contextual regions. On each region, the pixel values that have maximum occurrence are redistributed by limiting the histogram. This improves the contrast. A detailed study on this method is available in [16].

## 5 Experimental Results and Discussions

The experimental results of this proposed method are discussed here. The experiments were conducted on 100 low-illuminated iris images from UBIRIS. v2 run on PC with 2.7 GHz Intel Core i5 processor and 4 GB RAM using Matlab R2014a.

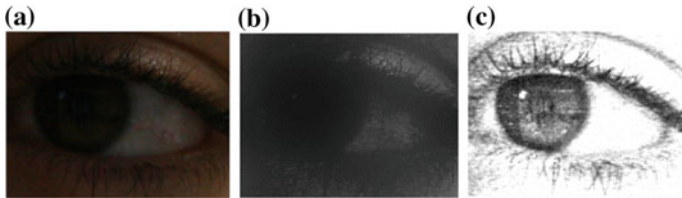
### 5.1 Qualitative Analysis

The input image is separated as radiance image and brightness image. The brightness image captures the texture and intensity information of the input image. The naturalness of the color in the input image is captured in the radiance image. An example for this is shown in Fig. 3.

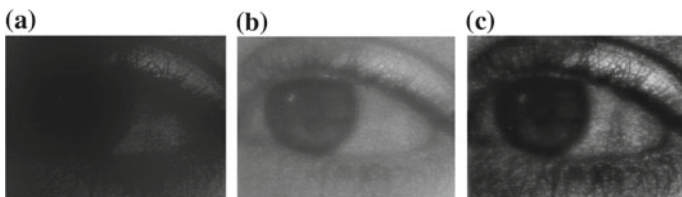
From the brightness image, the images D1, D2, and D3 are derived. An example for this is shown in Fig. 4.

Combining these derived images by fusing it at various levels of pyramidal resolution, the enhanced image is computed.

An example for the low-illuminated input image and image enhanced through multi-resolution-based fusion technique is shown in Fig. 5. From Fig. 5b, it can be seen that the proposed method enhances the low-illuminated regions of the iris image.

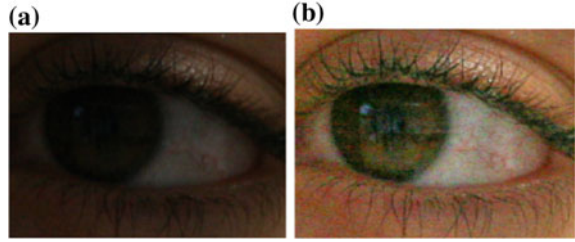


**Fig. 3** a Low-illuminated input image, b Estimated brightness image, and c Estimated radiance image



**Fig. 4** Derived images D1, D2, and D3 a Brightness image, b Gamma-corrected c CLAHE

**Fig. 5** **a** Input image, **b** Multi-resolution enhanced image



**Table 1** Comparative analysis on the performance of the proposed method

Method	Average MSE	Average PSNR
Divide-and-conquer	22.91	20.96
CLAHE	0.8737	49.33
Proposed	0.0041	95.87

## 5.2 Quantitative Analysis

The average mean square error and the average peak signal-to-noise ratio of the methods discussed are tabulated in Table 1.

In terms of MSE, CLAHE is 96% better, and the proposed method is 99% better compared to the divide-and-conquer method. The proposed method achieves an increase in average PSNR of 74.91 compared to the divide-and-conquer method and 46.54 compared to CLAHE method. Overall the proposed multi-resolution method performs better for low-illuminated iris images.

## 6 Conclusions

A dynamic multi-scale image enhancement for low illumination iris images is proposed and tested on UBIRIS. v2 iris database. CLAHE and divide-and-conquer methods were not able to enhance the images with low illumination. Compared to these techniques, the proposed method enhances the visibility of the iris image details much better. This method improves the contrast under exposed iris images. And it directly operates on the pixel and more computation is required. However, it is efficient compared to other preprocessing techniques in the frequency domain. In the future, the recognition rate for this preprocessing approach on iris recognition will be evaluated. Other multi-resolution methods for image enhancement will be analyzed. Impact of this method on iris images from other databases will also be analyzed in the future work as well.

## References

1. Daugman JG (1993) High confidence visual recognition of persons by a test of statistical independence. *IEEE Trans Pattern Anal Mach Intell* 15(11):1148–1161
2. Farihan A, Raffei M, Asmuni H, Hassan R, Othman RM (2015) A low lighting or contrast ratio visible iris recognition using iso-contrast limited adaptive histogram equalization. *Knowl-Based Syst* 74: 40–48
3. Sanpachai H, Malisuwan, S (2015) A study of image enhancement for iris recognition. *J Ind Intell Inf* 3(1)
4. Sajjad M, Ahn C-W, Jung J-W (2016) Iris image enhancement for the recognition of non-ideal iris images. *Trans Internet Inf Syst* 10:1904–1926
5. Hemalatha G, Sumathi CP (2016) Preprocessing techniques of facial image with Median and Gabor filters. In: *International conference on information communication and embedded systems (ICICES)*, Chennai, pp 1–6
6. Tammam AA, Khalil AH, Kader, NSA (2017) Image enhancement and iris localization based on 2D complex matched filter for noisy images. In: *28th international conference on micro-electronics (ICM)*, Giza, pp 161–164
7. Kumar D, Sastry M, Manikantan, K (2016) Iris recognition using contrast enhancement and spectrum-based feature extraction. In: *International conference on emerging trends in engineering, technology and science (ICETETS)*, Pudukkotta, pp 1–7
8. Kiruthiga AR, Arumuganathan, R (2017) Smoothing of iris images and pupil segmentation using fractional derivative and wavelet transform. In: *Fourth international conference on signal processing, communication and networking (ICSCN)*, Chennai, pp 1–6
9. Kumar SA, Pannirselvam S (2017) Preprocessing of IRIS image using high boost median (HBM) for human personal identification. *Int J Comput Sci Mob Comput* 6:142–151
10. Yan F, Tian Y, Zhou, C, Cao, L, Zhou, Y, Wu, H (2015) Non-ideal iris image enhancement algorithm based on local standard deviation. In: *The 27th chinese control and decision conference (2015 CCDC)*, Qingdao, pp 4755–4759
11. Ismail AI, Ali HS, Farag FA (2015) Efficient enhancement and matching for iris recognition using SURF. In: *5th national symposium on information technology: Towards new smart world (NSITNSW)*, Riyadh, pp 1–5
12. Cui C, Wang X, Shen H (2016) Improving the face recognition system by hybrid image preprocessing. In: *IEEE international conference on cyber technology in automation, control, and intelligent systems (CYBER)*, Chengdu, pp 442–447
13. Land EH, McCann JJ (1971) Lightness and retinex theory. *J Opt Soc Am* 61(1):1–11
14. Wang S, Zheng J, Hu HM (2013) Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE Trans Image Process* 22(9):3538–3548
15. Zhuang P, Fu X, Huang Y, Ding X (2017) Image enhancement through divide and conquer strategy. *J Vis Commun Image R* 45:137–146
16. Zuiderveld K (1994) Contrast limited adaptive histogram equalization. In: *Graphic gems IV*. Academic Press Professional, San Diego, pp 474–485

# Character Recognition from Handwritten Image Using Convolutional Neural Networks



Ranjan Jana and Siddhartha Bhattacharyya

**Abstract** Character recognition from handwritten images is of great interest in the pattern recognition research community for their good application in many areas. To implement the system, it requires two steps, viz., feature extraction followed by character recognition based on any classification algorithm. Convolutional neural network (CNN) is an excellent feature extractor and classifier. The performance of a CNN for a particular application depends on the parameters used in the network. In this article, a CNN is implemented for the MNIST dataset with appropriate parameters for training and testing the system. The system provides accuracy up to 98.85%, which is better with respect to others. It also takes very low amount of time for training the system.

**Keywords** Handwritten character recognition · Optical character recognition · Convolutional neural networks · Deep neural networks · MNIST dataset

## 1 Introduction

Character recognition from printed text images or handwritten image documents is very important in the domain of optical character recognition (OCR). Some of the research areas include automatic number plate recognition, automatic postal address checking from envelopes, processing of bank cheques to name a few, as illustrated in

---

R. Jana (✉)

Department of Information Technology, RCC Institute of Information Technology, Kolkata, West Bengal, India

e-mail: [ranjan.rcciit@gmail.com](mailto:ranjan.rcciit@gmail.com)

S. Bhattacharyya

Department of Computer Application, RCC Institute of Information Technology, Kolkata, West Bengal, India

Faculty of Electrical Engineering and Computer Science, VSB Technical University of Ostrava, Ostrava, Czechia

e-mail: [siddhartha.bhattacharyya@vsb.cz](mailto:siddhartha.bhattacharyya@vsb.cz)

© Springer Nature Singapore Pte Ltd. 2019

S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,

Advances in Intelligent Systems and Computing 922,

[https://doi.org/10.1007/978-981-13-6783-0\\_3](https://doi.org/10.1007/978-981-13-6783-0_3)



[1]. Extraction of text from real images is always a challenging proposition in many applications [2]. Recognizing text from real images becomes very complicated due to a wide range of variation in textures, backgrounds, font size and shading. The three basic steps of character recognition are segmentation, feature extraction, and feature classification. In the domain of computer vision, the Multi-layer perceptron (MLP) has been a revolution. However, the performance of MLP fully depends on the performance of feature selection algorithms [3]. After the invention of Deep Neural Network (DNN), it is proved that, DNN is an excellent feature extractor as well as a good classifier [4]. However, it takes huge amount of time for training the network due to large numbers of nonlinear hidden layers and connections. Convolutional neural network (CNN) has been discovered to solve various problems in computer vision by using lesser amount of hidden layers compared to DNN [5]. So, CNN is used to extract the position-invariant features in a reasonable amount of time for its simple structure. CNN takes relatively few parameters and it is very easy to train the system. CNN is able to map between input dataset to output dataset with temporal sub-sampling to offer a degree of rotation, distortion, and shift invariance [6]. So, CNN is used in this article to implement the system to recognize character from handwritten image.

## 2 Related Work

This section surveys related works on character recognition systems. Hanmandlu and Murthy have implemented a character recognition system that used different priorities for different features based on the accuracies of individual features [7]. The recognition system has been implemented using a fuzzy model. The average recognition accuracy was 98.4% for the numeric digits in English and 95% for the numeric digits in Hindi. A recurrent neural network (RNN) using Hidden Markov Model has been used to discover patterns in a script to find the sequence of characters [8]. The model has been implemented by Graves and Schmidhuber and it classified the IFN/ENIT database of handwritten Arabic words with 91% accuracy. For recognition of handwritten English characters, Pal and Singh have implemented the Multi-layer Perceptron [9]. The features have been extracted and analyzed by comparing its features for character recognition using boundary tracing and its fourier descriptors. It has been also analyzed to determine the number of hidden layers required to attain high accuracy. It has been reported with 94% accuracy of Handwritten English characters with very less amount of training time. Neves et al. have implemented a Support Vector Machine (SVM) based offline handwritten character recognition, which gave better accuracy compared to the Multi-layer perceptron classifier for the NIST SD19 standard dataset [10]. Although, MLP is suitable for segmentation of nonlinear separable classes, but it can easily be trapped into a local minima. An implementation of deep neural network models has been established by Younis and Alkhateeb so that it can extract the important features without unnecessary pre-processing [11]. Such a deep neural network has demonstrated an accuracy of 98.46% on MNIST dataset for

the handwritten OCR problem. A multilayer CNN using Keras with Theano has been implemented by Dutt and Dutt in [12] with accuracy of 98.70% on MNIST dataset. It provided better accuracy compared to SVM, K-Nearest Neighbor and Random Forest Classifier. A comparative study of three neural network approaches has been provided by Ghosh and Maghari [13]. The results showed that Deep Neural Network was the most efficient algorithm with accuracy of 98.08%. However, it was noted that each neural network has an error rate because of their similarity in digit shape for the digit tuples (1 and 7), (3 and 8), (6 and 9), (8 and 9).

### 3 Overview of CNN Architecture

A CNN is a special type of artificial neural network, which consists of one input and one output layer with multiple hidden layers. The hidden layers are convolutional layers, pooling layers, and fully connected layers [6]. The network consists of repetitive convolutional and pooling layers. Finally, it ends with one or more fully connected layers.

#### 3.1 Convolutional Layer

A convolutional layer applies sliding filters vertically and horizontally through the input image. This layer learns the features of all the regions of input image while scanning. It computes a scalar product of values of the filter with the values of image regions and adds a bias for each region. A Rectified Linear Unit applies element wise activation function, viz.,  $\max(0, x)$ , Tanh, Sigmoid:  $1/(1 + e^{-x})$  to the output of this layer for thresholding.

#### 3.2 Pooling Layer

Pooling layer is generally used after one or more convolutional layers to shrink the volume of the data to make the network computationally faster. It restricts the networks from overfitting as well as provides the network into translation invariance. Max pooling and average pooling are generally used to implement pooling. It applies sliding filters vertically and horizontally through the input image to get max value or average value for each region of the input data.

### 3.3 Fully Connected Layer

After all the convolutional and pooling steps, the networks generally use fully connected layers with separate neurons for each pixels like a standard neural network. The last fully connected layer contains  $n$  numbers of neurons, where  $n$  is the number of predicted classes. For digit classification problem, it should be 10 neurons for 10 classes (digit 0–9) and for English character classification problem, it should be 26 neurons for 26 classes (character a to z).

## 4 Proposed Methodology

Character recognition from handwritten images has been applied for extraction of text. To implement a character recognition system, one initially needs features extraction technique supported by a classification algorithm for recognition of characters based on the features. Several feature extraction and classification algorithms have been used for this purpose before the advent of deep-learning. Deep-learning has proved that, no separate algorithm is required for feature extraction and feature classification. Deep-learning is an excellent performer in the field of computer vision for both feature extraction and feature classification. DNN architecture consists of many non-linear hidden layers with enormous number of connections and parameters. So, it is very difficult to train the network with a small set of training samples. CNN is the solution, which takes relatively few set of parameters for training the system. So, CNN is capable to map the input dataset to output dataset correctly by changing the trainable parameters as well as the number of hidden layers. Therefore, CNN architecture is proposed for character recognition from the images of handwritten digits. For performance verification of the system, the standard normalized MNIST dataset has been used.

### 4.1 Database Used for Training and Testing

The MNIST dataset is a subset of NIST database [14]. The MNIST dataset is a collection of 70,000 images of handwritten digits. The dataset is divided into 60,000 images for training and 10,000 images for testing [14]. All images have resolution of  $28 \times 28$  and the pixel values are in the range of 0–255 (gray-scale). The background of digit is represented by 0 gray value (black) and a digit is represented by 255 gray value (white) as shown in Fig. 1.

The MNIST database is a collection of training and test set image files as well as training and test set label files. The pixel values are organized in row-wise for training and test set image files. So, the training set image file consists of 60,000 rows and 784 columns ( $28 \times 28$ ) and the testing set image file consists of 10,000 rows and

**Fig. 1** Examples of MNIST data set



784 columns. In the training and test label files, the labels values are 0–9. So, the training label file consists of 60,000 rows and 10 columns (0–9) and the testing file consists of 10,000 rows and 10 columns.

## 4.2 Design of CNN Architecture for MNIST Dataset

The performance of a CNN for a particular application depends on the parameters used in the network. So, the CNN architecture with convolutional layers is implemented for MNIST digit recognition as shown in Fig. 2.

At first, 32 filters of window size  $5 \times 5$  with a ReLU activation function for nonlinearity are used in first convolutional layer. After that, a pooling layer is used to perform a down-sampling operation using a pool size  $2 \times 2$  with stride by 2. As a result, the image volume is reduced. After that, 64 filters with window size  $7 \times 7$  with a ReLU activation function for nonlinearity is used in another convolutional layer. Then, another pooling layer is used to perform a down-sampling operation using a pool size  $2 \times 2$  with stride by 2. After that, a fully connected layer is used with 1024 output nodes. Finally, another fully connected layer with 10 output nodes is used to get network results for ten digits (0–9).

## 5 Experimented Result

Experiments have been conducted in Intel Xeon 2.2 GHZ processor with 128 GB RAM using python programming. The experimented result of the proposed method is detailed in this section. A comparative study of proposed CNN with other state-of-the-art works is shown in Table 1.

The proposed CNN gives 98.85% accuracy, which is better with respect to others. The accuracy with required training time for different training steps of the proposed CNN and CNN in KERAS are presented in Table 2. The results shows that proposed CNN architecture takes less amount of time for better accuracy with respect to CNN in KERAS as shown in Fig. 3.

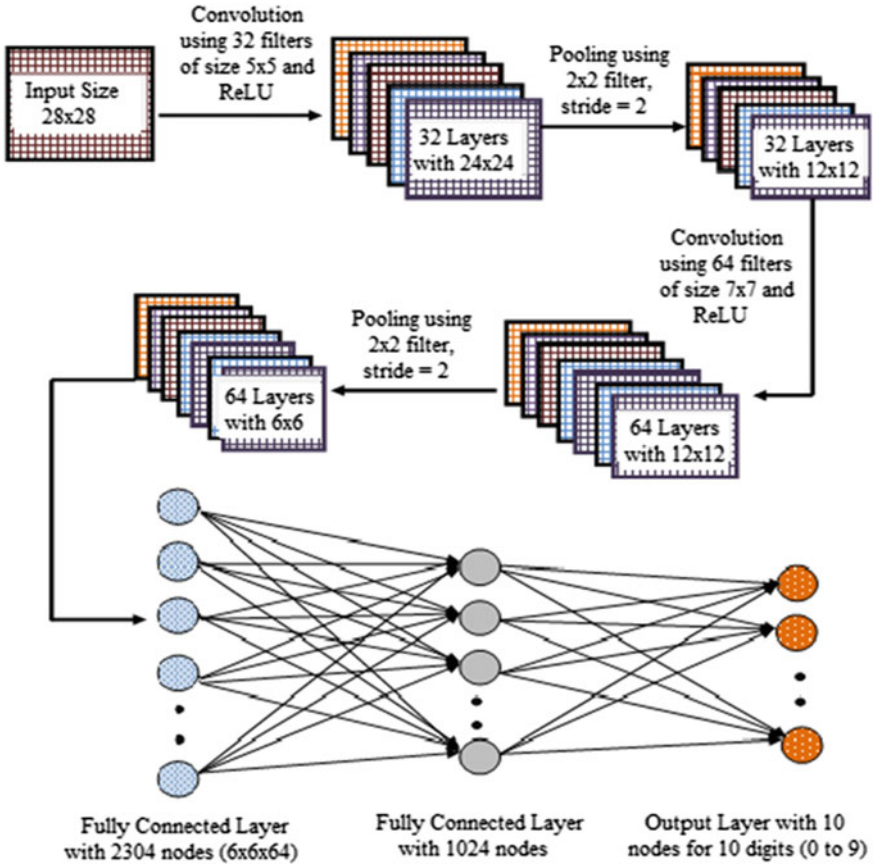


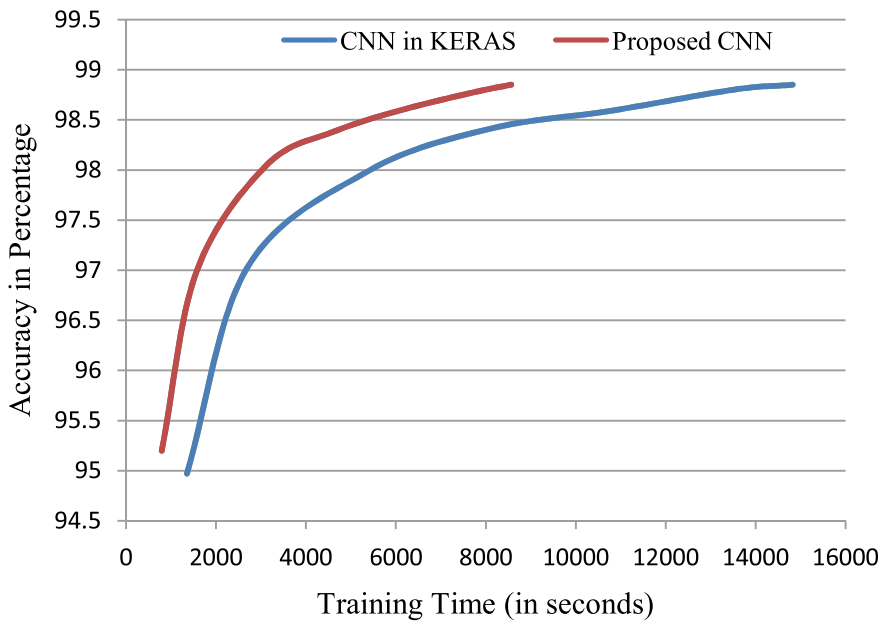
Fig. 2 The proposed CNN architecture

Table 1 Accuracy of different CNN architectures

Author	Method	Accuracy (%)
Younis and Alkhateeb [11]	CNN	98.46
Dutt and Dutt [12]	CNN (Keras + Theano)	98.70
Ghosh and Maghari [13]	DNN	98.08
Proposed method	CNN	98.85

**Table 2** Performance of proposed CNN and CNN in KERAS

Training steps	CNN in KERAS		Proposed CNN	
	Accuracy (%)	Training time (s)	Accuracy (%)	Training time (s)
10,000	94.97	1353	95.20	797
20,000	97.03	2701	96.99	1573
40,000	97.99	5394	98.05	3136
60,000	98.41	8087	98.39	4674
80,000	98.59	10,785	98.61	6232
100,000	98.80	13,508	98.78	7793
110,000	98.85	14,827	98.85	8569



**Fig. 3** Accuracy versus training time

## 6 Conclusion

In this paper, an implementation of handwritten digit recognition using CNN is implemented. The proposed CNN architecture is designed with appropriate parameters for good accuracy of the MNIST digit classification. The time required to train the system is also considered. The CNN architecture is designed with 32 filters with window size  $5 \times 5$  for the first convolutional layer and 64 filters with window size  $7 \times 7$  for the second convolutional layer. The experimented results proved that, the proposed CNN architecture is the best in term of accuracy and time for the MNIST

data set as compared to others. It is worth mentioning here that more filters can be used for better accuracy at the cost of higher training time. For further improvement of the accuracy, the system needs to do more training, which requires a huge amount of time. The parallelism technique of GPU machines can be used for better accuracy with extensive training in a less amount of training time.

**Acknowledgements** The authors are grateful to IT department, RCC Institute of Information Technology (RCCIIT), Kolkata for providing the requisite infrastructure for the research work. The authors are also grateful to Prof. (Dr.) Ajoy Kumar Roy, Chairman, RCCIIT for his constant inspiration.

## References

1. Arica N, Vural FTY (2001) An overview of character recognition focused on offline handwriting. *IEEE Trans Syst Man Cybern—Part C: Appl Rev* 31(2):216–233
2. Epshtein B, Oyek E, Wexler Y (2010) Detecting text in natural scenes with strokes width transform. In: *IEEE conference on computer vision and pattern recognition*, pp 1–8
3. Yang JB, Shen KQ, Ong CJ, Li XP (2009) Feature selection for MLP neural network: the use of random permutation of probabilistic outputs. *IEEE Trans Neural Netw* 20(12):1911–1922
4. Lee H, Grosse R, Ranganath R, Ng AY (2009) Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *26th Annual international conference on machine learning*. ACM, pp 609–616
5. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
6. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
7. Hanmandlu M, Murthy OVR (2007) Fuzzy model based recognition of handwritten numerals. *Pattern Recogn* 40(6):1840–1854
8. Graves A, Schmidhuber J (2009) Offline handwriting recognition with multidimensional recurrent neural networks. In: *Advances in neural information processing systems NIPS'22*, vol 22. MIT Press, Vancouver, pp 545–552
9. Pal A, Singh D (2010) Handwritten english character recognition using neural network. *Int J Comput Sci Commun* 1(2):141–144
10. Neves RFP, Filho, ANGL, Mello CAB, Zanchettin C (2011) A SVM based off-line handwritten digit recognizer. In: *International conference on systems, man and cybernetics, IEEE Xplore, Brazil, 9–12, October 2011*. pp 510–515
11. Younis KS, Alkhateeb, AA (2017) A new implementation of deep neural networks for optical character recognition and face recognition. In: *Proceedings of the new trends in information technology, Jordan, April 2017*. pp. 157–162
12. Dutt A, Dutt A (2017) Handwritten digit recognition using deep learning. *Int J Adv Res Comput Eng Technol* 6(7):990–997
13. Ghosh MMA, Maghari AY (2017) A comparative study on handwriting digit recognition using neural networks. In: *International conference on promising electronic technologies*, pp 77–81
14. <http://yann.lecun.com/exdb/mnist/>

# Malignant Melanoma Classification Using Cross-Platform Dataset with Deep Learning CNN Architecture



Soumen Mukherjee, Arunabha Adhikari and Madhusudan Roy

**Abstract** Melanoma, a malignant skin lesion, is the deadliest of all types of skin cancer. Deep learning has been shown to efficiently identify patterns from images and signals from various application domains. Use of deep learning in medical image analysis is, however, limited till date. In the present paper, two well-known malignant lesion image datasets, namely Dermofit and MEDNODE, are both separately and together used to analyze the performance of a proposed deep convolutional neural network (CNN) named as CNN malignant lesion detection (CMLD) architecture. When Dermofit and MEDNODE datasets are used separately with tenfold data augmentation, the CNN gives 90.58 and 90.14% classification accuracy. When the datasets are mixed together the CMLD gives only 83.07% accuracy. The classification accuracy of the MEDNODE dataset using deep CNN is considerably high in comparison with the results found in the related literature. The classification accuracy is also high in case of Dermofit dataset in comparison with the traditional feature-based classification.

**Keywords** ABCD rule · Image augmentation · MNIST · Convolutional layer · CMLD · MEDNODE · Dermofit · Stochastic gradient descent with momentum

---

S. Mukherjee

Department of Computer Application, RCC Institute of Information Technology, Kolkata, India  
e-mail: [soumou601@gmail.com](mailto:soumou601@gmail.com)

A. Adhikari (✉)

Department of Physics, West Bengal State University, Barasat, West Bengal, India  
e-mail: [arunabha.adhikari@gmail.com](mailto:arunabha.adhikari@gmail.com)

M. Roy

Surface Physics and Material Science Division, Saha Institute of Nuclear Physics, Kolkata, India  
e-mail: [roy.madhusudan57@gmail.com](mailto:roy.madhusudan57@gmail.com)

© Springer Nature Singapore Pte Ltd. 2019

S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_4](https://doi.org/10.1007/978-981-13-6783-0_4)



## 1 Introduction

Melanoma is the most [1] deadly type of skin cancer which requires surgery to cure in most of the cases [1]. In nearly 20% cases, surgery also fails to cure from this fatal disease. Out of all the skin cancer death, 75% happens for melanoma itself. It is the only cause for more than 50,000 cancer deaths around the world [1, 2]. Ultraviolet radiation exposure in bare skin is the leading cause of melanoma. In most of the European countries like Switzerland and Denmark, patients are diagnosed with melanoma at a much high rate. It is found from the literature that early diagnosis of melanoma plays a significant role for the successful treatment [1]. Melanoma is a skin lesion type which can only be identified by expert dermatologists because it is a close similarity with benign lesions. Dermatologists use several quantitative marking techniques like ABCD rule, 7-point checklist, 3-point checklist, etc. [3] other than the gold-standard skin biopsy to identify melanoma in an early stage. In recent year, computer vision, machine learning, and deep learning play a vital role in the detection of melanoma. In the present work, use of deep learning in malignant melanoma detection is carried out with a proposed CNN architecture with benchmark datasets. The paper is organized in the following way. In the related work section, two different approaches, i.e., identification of melanoma using the traditional feature and using deep learning CNN, are discussed. In the next section, a small introduction to deep learning is given. The proposed work description is given after that with dataset description and discussion on proposed CNN malignant lesion detection (CMLD) architecture. Finally, in the conclusion part, the importance of the work and the future scope is discussed.

## 2 Related Work

Melanoma and other malignant skin cancer detection and identification can be done using hand-crafted traditional features and also with deep learning features. In this section, both aspects are discussed in detail.

### *2.1 Melanoma Classification Using Traditional Feature*

Malignant skin cancer detection using image processing and machine learning technique is primarily started with hand-crafted features, where a considerably large amount of traditional features like shape, size, color, texture, etc. are extracted from each lesion image. Then, a more computationally efficient subset of the extracted feature set is found with different feature ranking algorithms and feature optimization algorithms, and finally, different supervised and unsupervised classification algorithms are used on the feature subset to identify the proper class of the skin lesion

image. Quite a large number of researches are carried out using the hand-crafted features with well-known Dermofit and MEDNODE datasets which are used in this work. Using Dermofit dataset, Tan et al. [4] got 88% classification accuracy with optimized 1472 traditional features by a genetic algorithm with SVM classifier. Mukherjee et al. [5] used four features ranking algorithms and found selected 163 features which ranked very high in all four ranking. These features with multilayer perceptron (MLP) classifier yielded 86.2% classification accuracy. Ballerini et al. [6] achieved 93% with only non-melanoma skin cancer and benign lesion images and 74% accuracy using 960 images of Dermofit dataset with 2 class and 5 class, respectively, with several color features and 15,552 texture features using K-NN classifier with threefold cross-validation. Laskaris et al. [7] found 80.64% classification accuracy with 31 images of Dermofit dataset. MEDNODE dataset is also used by different researchers in this domain. Giotis et al. [8] got 81% accuracy with 675 features using the 170 images of MEDNODE dataset. Mukherjee et al. [9] got 83.33% classification accuracy with only ten highest ranked traditional features using MLP. The ranking of the features is done by ReliefF algorithm. In the same work, they have shown that with only the best 25 PCA components, the classification accuracy level can be increased by up to 87%. Mukherjee et al. [10] got 91.02% accuracy when they used 1886 features but optimized the number of neurons in the two hidden layers of MLP by particle swarm optimization (PSO) using the MEDNODE dataset. In their work, they have also worked with Dermofit dataset where they have found 88% accuracy with 1886 features. In conclusion, the highest accuracy achieved in full Dermofit (1300 images) and MEDNODE (170 images) dataset is 88 and 91.02% so far. Other than Dermofit and MEDNODE dataset, DermIS, DermQuest, PH2, etc. are the other popular datasets in this domain.

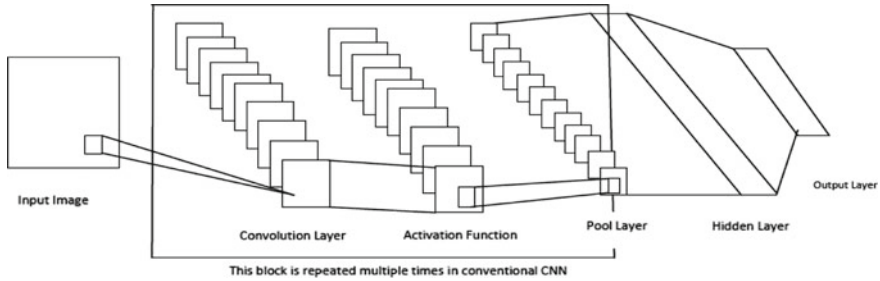
## ***2.2 Melanoma Classification Using Deep CNN Architecture***

In recent times, several researchers have worked on malignant lesion identification using deep learning architecture. Nasr-Esfahani et al. [11] in their work used MEDNODE dataset with 170 color images. They have augmented images to make it total 6120 including the original ones. Different image augmentation techniques were used like image cropping (5 and 10% cropping) and image rotation (with an angle  $0^\circ$ ,  $90^\circ$ ,  $180^\circ$ , and  $270^\circ$ ) generating 35 augmented images from a single original image. Then, each image is rescaled to  $188 \times 188$  pixels. The authors have used deep CNN to classify the images into two classes, malignant and benign. In their work, they have used two convolution layers with 20 and 50 feature maps of  $5 \times 5$  kernel each in the first and second layer, respectively. One max-pooling layer is present between each convolution layer. Finally, fully connected two-class layer is given for the classification task. The system is implemented in Intel i7 processor with 32 GB RAM and with two linked GPU of Titan and GeForce series. They have maintained 80/20% training and testing ratio. They have achieved 81% classification accuracy. Pomponiu et al. [12] have proposed a method of deep CNN, where they used the images from

DermIS and DermQuest image library. Their original datasets are divided into 217 benign and 182 malignant images which are augmented to make it around 10,000 images. They introduced Gaussian noise, histogram equalization, and compression and motion blur to augment the images. The images are rescaled to  $224 \times 224$  pixels to feed into the CNN. The CNN they have used has eight layers, which has five convolution layers at the beginning and three fully connected layers at the end. They have got a two-class classification accuracy of 93.64% in their work. Ayan and Unver [13] in their work have shown how data augmentation helps to improve the classification rate of melanoma using deep learning. In their work, they have used different sized 1000 original images from popular ISIC dataset which are divided into two equal set of 500 images of malignant and benign class. All the images are resized to  $224 \times 224$  pixels before feeding to the CNN. The CNN architecture they have used has total 11 layers with four convolution layers, two max-pool layers, three dropout layers, and two fully connected dense layers. They have achieved 78% classification accuracy without data augmentation, and when data augmented with five times of the original images to make the total image count up to 5000 with 75/25% training and testing ratio, the two-class classification accuracy increased to 81% with NVIDIA K4200 GPU card and 64 GB RAM. The entire program of their proposed system is implemented in Python. Kwasigroch et al. [14] made a comparison of the performance of three popular CNN architectures in melanoma classification including VGG19, ResNet50, and VGG19-SVM. In their work, they have used ISIC dataset with different types of image augmentation scheme like transformation (i.e., rotation, flip, shift, etc.). They have used an imbalanced dataset of 9300 benign and 670 malignant images. With VGG19, ResNet50, and VGG19-SVM, they have achieved 81.2, 75.5, and 80.7% classification accuracy. They have used GeForce GPU with 6 GB RAM in their work, and the program implemented in Python. Maia et al. [15] got 92.5% accuracy with PH2 dataset consisting of 200 images (40 malignant and 160 benign) using popular VGG19 CNN architecture. Lopez et al. [16] also used ISBI 2016 challenge dataset with training and testing images of 900 and 379 images and got 81.33% classification accuracy. They have used VGG16 architecture for their deep CNN network. Ali and Al-Marzouqi [17] have used ISBI 2016 challenge dataset with 727 benign and 173 malignant images to achieve 81.6% two-class classification accuracy with modified LightNet architecture, consisting of 17 layers including five blocks. Every block has convolution, ReLU, max-pooling, and dropout layer. Last block has fully connected and soft-max classification layer in this deep CNN.

### 3 Deep Learning

Deep learning is the most popular technique for image recognition in recent times for its efficiency and accuracy. It is not only used in the domain of image processing but also in natural language processing, bioinformatics, audio processing, etc. Deep learning application in image processing acts very similar to the visual cortex of the brain and finds out deep features from the images. Convolutional neural network



**Fig. 1** Deep CNN architecture

(CNN) is one of the mostly used deep learning techniques [18–21]. It is first developed in 1980s, but around 2012 the use of CNN in deep learning application comes into reality for the development of graphics processing unit (GPU) as CNN requires high computational time. The main area of concern is the training time required by CNN, which is considerably very high in complex CNN rather than traditional techniques. CNN has two major components, convolutional layer and pooling layer. Each convolutional layer produces a new image from the original image known as feature map. Convolutional layer contains small square shape convolutional filters which are applied to the input image. The size of the filter used varies and depends on different types of application. After convolutional layer, a batch normalization layer is present which normalizes the output of the previous convolutional layer. Activation function like ReLU (rectified linear unit) or sigmoid function is used at the end of the convolutional layer. The pooling layer is used to reduce the size of the feature map generated from the convolutional layer. There are different types of pooling like max-pooling, mean-pooling, and average pooling which are also application-specific. Filter shift step is known as stride. In the last stage of CNN, one or few fully connected layers, usually soft-max layer of classification, are present. Figure 1 shows small sample deep CNN architecture. More complex architecture has multiple numbers of convolutional and pooling layers. Different convolutional neural networks (CNNs) are used for the image classification task, e.g., AlexNet, VGGNet, GoogLeNet, ResNet, CIFAR10/100, MNIST, SVHN, etc. The CNN architecture varies for different types of application [18, 19].

## 4 Dataset

In the present work, two different popular malignant lesion image datasets are used. The first one is Dermofit [6] image set, which contains 1300 malignant (both melanoma and non-melanoma) and benign skin lesion optical images. Every image of this dataset is tagged by dermatopathologists, and it is identified as a gold-standard dataset. Images in this dataset are in “.png” color format with varying image pixel

**Table 1** Details of the two datasets used in the work

Dataset	Total images	Malignant	Benign	Format	Size
Dermofit	1300	526	774	.png color	Varying size
MEDNODE	170	70	100	.jpg color	Varying size

**Fig. 2** a, b Malignant and benign skin lesion images from Dermofit dataset

(a) Malignant

(b) Benign

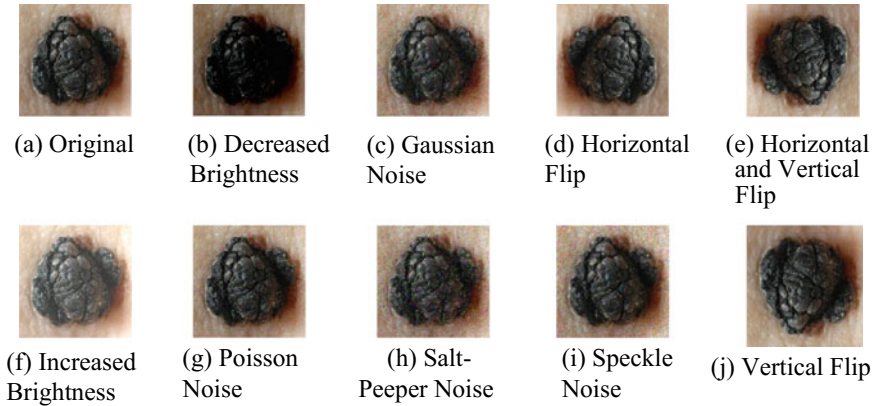
size. The second dataset used in the work is MEDNODE [8] dataset, which has total 170 number of “.jpg” color malignant (melanoma) and benign optical images. MEDNODE is collected by University Groningen. Nikon D3 or D1x cameras are used to take these images. Every image of this dataset is validated by a dermatologist and identified as a gold-standard dataset. Details of the two datasets are given in Table 1. In Fig. 2, the malignant melanoma and benign skin lesions are shown from Dermofit dataset.

## 5 Proposed Work

In this proposed work, MATLAB 2018a is used for writing code for image augmentation and for creating proposed deep learning CNN architecture. The codes are executed in Intel i5 processor with 4 GB RAM without GPU. The original images are augmented with replication of images by making three types of transformations, namely horizontal flip, vertical flip, and both horizontal and vertical flip together and introducing four types of noise, i.e., Gaussian, Poisson, Salt-and-Pepper, and speckle in the original image. Two other ways of augmentation done are increment and decrement of brightness by a value of 50 in the original image. This image augmentation makes malignant images to 5260 ( $526 \times 10$ ) images and benign image to 7740 ( $774 \times 10$ ) images for Dermofit dataset. Similarly, MEDNODE dataset [8] original images (170) are also extended by ten times (1700) of the original dataset after data augmentation. Figure 3 shows the results of different transformations on one of the single malignant melanoma images from MEDNODE dataset.

## 6 CNN Malignant Lesion Detection (CMLD) Architecture

In the present work, a new CNN malignant lesion detection (CMLD) architecture is proposed which is lightweight and similar to the MNIST deep CNN architecture [22]. Several deep CNN architectures are available at present like AlexNet, VGGNet,



**Fig. 3** a–j Original and different types of augmented images of MEDNODE melanoma dataset

ResNet, GoogleNet, etc. Each of them requires a considerable amount of time and space overhead and hence may be called heavyweight. Some CNN architectures require less computational load and branded as lightweight. In comparison with the lightweight CNNs like CIFAR10/100, MNIST, SVHN, etc., the above-mentioned heavyweight CNNs give very small enhancement in classification accuracy. In this proposed CNN malignant lesion detection (CMLD) architecture after the input image layer in the top layer, three 2D convolution layers are present and separated by two max-pool layers in between, and finally, fully connected layers and soft-max layer are present for classification. The detail of the architecture is given in Table 2. Architecture of CMLD is similar to MNIST deep CNN was originally developed to classify a 10-class (digits of 0–9) problem of handwritten digit identification.

In this proposed work, the authors have used input image layer of size  $128 \times 128$  to minimize data loss unlike  $28 \times 28$  input image size used in MNIST deep CNN architecture. Increase in image size will increase the training time of the deep CNN. Yu et al. [23] have shown that increasing the input image size will increase the complexity of the network but may not increase the classification accuracy.

For the present work, 75% of the total images are taken for training, and 25% of images are taken for testing. The training options used for the whole experiment are given in Table 3.

## 7 Result and Discussion

Two datasets, Dermofit and MEDNODE, are separately classified with the deep learning CNN architecture. Using augmented MEDNODE (1700 images) and Dermofit (13,000 images) datasets, an accuracy of 90.14 and 90.58% are found. When these two datasets are merged together, i.e., then a total image set of 14,700 gives

**Table 2** Details of the layers present in CNN malignant lesion detection (CMLD) architecture from top to bottom

[128 × 128] Input image layer
Two-dimensional convolve layer with eight number of [3 × 3] filters with padding so that output size is same like input
Batch norm layer
ReLU layer
Two-dimensional max-pool layer with [2 × 2] pool and [2 × 2] stride (step in CNN)
Two-dimensional convolve layer with 16 number of [3 × 3] filters with padding so that output size is same like input
Batch norm layer
ReLU layer
Two-dimensional max-pool layer with [2 × 2] pool size and [2 × 2] stride (step in CNN)
Two-dimensional convolve layer with 32 number of [3 × 3] filters with padding so that output size is same like input
Batch norm layer
ReLU layer
Fully connected layer (2)
Soft-max layer
Classify layer

**Table 3** Showing different values of the options chosen for training CMLD

Method	Learning rate	Max. epochs	Validation frequency
Stochastic gradient descent with momentum	0.01	500	100

**Table 4** Results found from the CMLD architecture

Proposed scheme	MEDNODE dataset (1700 images) (%)	Dermofit dataset (13,000 images) (%)	MEDNODE and Dermofit dataset both (14,700 images) (%)
Accuracy	90.14	90.58	83.07

an accuracy of 83.07%, which is much lesser than the accuracy of the individual dataset. The results are given in Table 4. In Table 5, comparison of different results found from the present work and related work are shown.

**Table 5** Comparison of results with other related works using two datasets

Results found with hand-crafted features			
Dermofit dataset		MEDNODE dataset	
Related work	Accuracy (%)	Related work	Accuracy (%)
Tan et al. [4] with 1472 features	88	Giotis et al. [8] with 675 features	81
Mukherjee et al. [5] with 163 features	86.2	Mukherjee et al. [10] with 1886 features	91.02
Mukherjee et al. [10] with 1886 features	88	Mukherjee et al. [9] with 10 features	83.33
		Mukherjee et al. [9] with 25 PCA components	87
Results found with deep learning architecture			
Proposed work	90.58	Nasr-Esfahani et al. [11]	81
		Proposed work	90.14 (without GPU)

## 8 Conclusion

It is given in Table 5 that deep learning gives 2.5% better accuracy result in case of Dermofit and 1% lesser accuracy in MEDNODE in comparison with the best results obtained from hand-crafted features. The classification accuracy of the present CMLD fares much better on MEDNODE than CNN used by Nasr-Esfahani et al. [11] although no GPU is used in CMLD and two linked GPUs are used by Nasr-Esfahani et al. When the mixed platform datasets are used, the classification accuracy drops through around 7% as it is commonly seen with hand-crafted features. It is a common belief that data augmentation improves classification accuracy in deep CNN [13], but the present work shows that when instead of synthetic data replication data is augmented from another dataset deep CNN does not always perform better and accuracy may decrease. It also shows deep CNN is not always superior to hand-crafted features in terms of classification accuracy but definitely is much more computationally expensive. Opportunity for the traditional machine learning approaches using the hand-crafted features and image preprocessing to find out an algorithm for higher accuracy and search for a feature set which may work across the datasets is still open. On the other hand, CNNs may have better classification accuracy if more complex network similar to ImageNet, GoogleNet, AlexNet is applied and more augmented data being introduced in the network. In conclusion, research for both the approach is still worth perusing.



## References

1. Corrie P, Hategan M, Fife K, Parkinson C (2014) Management of melanoma. *Br Med Bull* 111(1):149–162. <https://doi.org/10.1093/bmb/ldu019>
2. Riker AI, Zea N, Trinh T (2010) The epidemiology, prevention, and detection of melanoma. *Ochsner J* 10(2):56–65
3. Marks R (2000) Epidemiology of melanoma. *Clin Exp Dermatol* 25:459–463. <https://doi.org/10.1046/j.1365-2230.2000.00693.x>
4. Tan TY, Zhang L, Jiang M (2016) An intelligent decision support system for skin cancer detection from dermoscopic images. In: 12th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD), pp 2194–2199
5. Mukherjee S, Adhikari A, Roy M (2018) Malignant melanoma identification using best visually imperceptible features from Dermofit dataset. Accepted and presented in 1st international conference on emerging trends in engineering and science (ETES-2018)
6. Ballerini L, Fisher RB, Aldridge RB, Rees J (2013) A color and texture based hierarchical K-NN approach to the classification of non-melanoma skin lesions, color medical image analysis. In: Celebi ME, Schaefer G (eds) *Lecture notes in computational vision and biomechanics*, vol 6, pp 63–86
7. Laskaris N, Ballerini L, Fisher RB, Aldridge B, Rees J (2010) Fuzzy description of skin lesions. In: *SPIE conference proceedings*, vol 7627. <https://doi.org/10.1117/12.845294>
8. Giotis I, Molders N, Land S, Biehl M, Jonkman MF, Petkov N (2015) MED-NODE: a computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Syst Appl* 42:6578–6585
9. Mukherjee S, Adhikari A, Roy M Malignant melanoma detection using multi layer perceptron with visually imperceptible features and PCA components from MED-NODE dataset. *Int J Med Eng Inf Inderscience (UK) Enterprise Ltd. Publishers (in press)*
10. Mukherjee S, Adhikari A, Roy M (2018) Melanoma identification using MLP with parameter selected by metaheuristic algorithms, intelligent innovations in multimedia data engineering and management. Published from IGI Global, pp 241–268. ISBN 9781522571070
11. Nasr-Esfahani E, Samavi S, Karimi N, Soroushmehr SMR, Jafari MH, Ward K, Najarian K (2016) Melanoma detection by analysis of clinical images using convolutional neural network. In: 38th Annual international conference of the IEEE engineering in medicine and biology society (EMBC), pp 1373–1376
12. Pomponiu V, Nejati H, Cheung NM (2016) DEEPMOLE: deep neural networks for skin mole lesion classification. *IEEE ICIP*, pp 2623–2627
13. Ayan E, Unver HM (2018) Data augmentation importance for classification of skin lesions via deep learning. In: *Electric electronics, computer science, biomedical engineering's meeting (EBBT)*, Istanbul, pp 1–4
14. Kwasigroch A, Mikolajczyk A, Grochowski M (2017) Deep neural networks approach to skin lesions classification—a comparative analysis. In: 2017 IEEE conference, pp 1069–1074
15. Maia LB, Lima A, Pereira RMP (2017) Evaluation of melanoma diagnosis using deep features. In: 2017 IEEE conference, pp 1069–1074
16. Lopez AR, Giro-i-Nieto X, Burdick J, Marques O (2017) Skin lesion classification from dermoscopic images using deep learning techniques. In: *Proceedings of the IASTED international conference biomedical engineering (BioMed 2017)*, Innsbruck, Austria, pp 49–54
17. Ali AA, Al-Marzouqi H (2017) Melanoma detection using regular convolutional neural networks. In: 2017 International conference on electrical and computing technologies and applications (ICECTA). IEEE, pp 1–5
18. Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E (2018) Deep learning for computer vision: a brief review. *Hindawi Comput Intell Neurosci* 2018(Article ID 7068349):13. <https://doi.org/10.1155/2018/7068349>
19. Pak M, Kim S (2017) A review of deep learning in image recognition. In: 4th International conference on computer applications and information processing technology (CAIPT), Kuta Bali, pp 1–3

20. Patterson J, Gibson A (2017) Deep learning: a practitioner's approach, 1st edn. O'Reilly, p 507
21. Buduma N (2017) Fundamentals of deep learning, 1st edn. O'Reilly, p 283
22. Hasanpour SH, Rouhani M, Fayyaz M, Sabokrou M (2016) Let's keep it simple, using simple architectures to outperform deeper and more complex architectures. [arXiv:1608.06037](https://arxiv.org/abs/1608.06037)
23. Yu Z, Jiang X, Zhou F, Qin J, Ni D (2018) Melanoma recognition in dermoscopy image via aggregated deep convolutional features. IEEE Trans Biomed Eng. <https://doi.org/10.1109/tbme.2018.2866166>

# Automatic Gender Identification from Children Facial Images Using Texture Descriptors



Ayesha Iftikhar, Rehan Ashraf, Asim Saeed, Srinivas Alva  
and Rajmohan Pardeshi

**Abstract** Soft biometric such as gender significantly works to enhance the performance of biometric systems and also having applications in human–computer interaction, content-based indexing and retrieval, demographic studies, security, and surveillance. Gender identification among adults is easy as compared to in children, due to similarity of facial skin texture and appearance of faces. In this paper, we have introduced a technique to identify/classify the gender from facial images of children. In our method, we have applied three basic steps namely preprocessing, feature extraction, and classification. In preprocessing stage, face detection and normalization are performed. To extract the powerful features, we have computed different texture descriptors and after feature extraction process and SVM is applied for classification purpose. We have achieved the encouraging results in our experiments.

**Keywords** Gender identification in children · Facial images · Texture descriptors · Support vector machines · Soft biometric

## 1 Introduction

Biometric technology facilitates unique identification and verification based on physiological and behavioral attributes of a person. Physiological attributes include fingerprint, retina, palm, and face. Body movement, voice, handwriting, gait, etc., are considered as behavioral biometrics. Whereas in large biometric systems, soft bio-

---

A. Iftikhar · R. Ashraf

Department of Computer Science, NTU, Faisalabad, Pakistan

A. Saeed

Department of Computer Science, Beijing Jiaotong University, Beijing, China

S. Alva

Department of Computer Science, Gujarat University, Ahmedabad, India

R. Pardeshi (✉)

Department of Computer Science, Karnatak College, Bidar, India

e-mail: [madhurrajmohan1@gmail.com](mailto:madhurrajmohan1@gmail.com)

© Springer Nature Singapore Pte Ltd. 2019

S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,

Advances in Intelligent Systems and Computing 922,

[https://doi.org/10.1007/978-981-13-6783-0\\_5](https://doi.org/10.1007/978-981-13-6783-0_5)

metrics such as age, gender, eye color, hair color, gives additional information about the person and helps to enhance the performance. With rapid development in technology, biometric became popular in different applications such as national citizen identification framework, access control, law enforcement, security and surveillance [1], and automatic monitoring. But success of these applications is restricted only up to adult population and not scaled for children.

As per the report of The National Database and Registration Authority (NADRA) of Pakistan [2], being one the largest citizens biometric data agency, it is lagging for the children biometric services. In the future, major required applications of biometric technology for children are vaccination monitoring, flood relief, offering of financial benefits, reunification during disaster management, etc. In [3, 4], authors studied and they explained the issues faced by children such as child sexual abuse, exploitation, child pornography, and prostitution. To overcome these kinds of problems, biometric will play a vital role. Now, while developing such large biometric system for children of any country, soft biometric such as gender will play a major role in enhancement of accuracy.

Biometric data acquisition from children is a major challenge, in such case, the current study is focused on facial information due to its easy way of recording and it also does not need any special technical requirements. The remainder of the paper is structured as follows: In Sect. 2, we have presented related work. In Sect. 3, we have briefed our method and Sect. 4 dedicated for experiments and discussion. We concluded in Sect. 5.

## 2 Related Work

Gender classification (GC) from face image is getting attention of researcher from past few years. In comparison with humans, machines find difficult to classify the gender automatically. GC from face is very useful research because face has many discriminant features which can be used to identify human gender. Many methods have been proposed for GC from face images. In [5], author proposed novel approach from face images for GC by using local texture pattern as feature descriptor. Support vector machine (SVM) is used for classification purpose. FERET database is used and compared with other comparable works. Author uses LZP (local zigzag pattern) as a feature descriptor. Gender classification from face is becoming very popular. In [6], author uses hybrid approach by fusion of face appearance features and geometry feature. Researcher picks Haar wavelet for representation of appearance features and Adaboost algorithm for the selection of strong features. In [7], complex value neural network serves as a classifier and LBP with PCA feature extraction techniques for gender categorization by face. This paper also compares the complex value with real value neural network. Author in [8], proposed gender recognition by top half face based on local appearance feature. This research is suitable in that condition where the lower part of face is hidden. In [9], author classifies gender by face from appearance base method. Compare the nonlinear SVM classifier performance with

traditional classifier like linear, quadratic, nearest neighbor as well as with modern RBF (radial base function) classifier. Study focuses on low-resolution thumbnail ( $21 \times 21$ ) images of front face images. In [10], authors presented a novel approach for gender classification from frontal facial images with SVM classifier and ICA Features. In paper [11], author presented a scheme for identification of children based on multimodal biometric. In paper [12], authors presented algorithms for newborn face recognition using principal component analysis (PCA), linear discriminant analysis (LDA), integrated principal analysis (IPA), speeded up robust features (SURF), local binary pattern (LBP), integrated component analysis (ICA). In [13], author proposes a novel approach by integrating face and soft biometric trait to improve the result about 5.6% on face biometric by use of 210 newborn data.

In [14], researcher investing the GC by using face image of per person. For this purpose, two types of features have been used. First appearance base and other is geometric base for appearance base (LBP and DCT) used and for geometry base (GDF) is used. GDF is based on the physical changes in male and female faces. In [15], author said that gender classification is challenging task on real-world face images due to nonsupport environment. Authors used (LFW) database. LBP with Adaboost is employed for the selection of discriminant features and obtained 94.81% result. Viola–Jones algorithm is used for face detection and image dimension is  $250 \times 250$ . In [16], author claims that LBP histogram is good and different feature for the gender classification task from face images. Author compares his technique with more sophisticated Gabor filter technique and result was outperformed with raw pixel values as a feature are computationally less expensive. GC used in many other problems like it can reduce the search space for face recognition software. In [17], author projected 2DPCA (2D principal component analysis) for face recognition and test on ORL, AR, and Yale well-known databases. The ORL database assesses the performance under varied limitations of light and movement. The AR database computes the system working under the limitation of change of time in face expressions in light variation. When light and face expressions both vary, then Yale database is used to evaluate the system performance. 2DPCA is more simple than PCA. In accuracy term, it is most suitable than PCA. In [18] paper, author proposed a novel technique for GC by designing multi-resolution feature extraction method on the idea of local phase quantization (LPQ) by monogenic and intensity image. Features with LPQ are obtained from input image. SVM is used for classification. Author compares experiment result with two unconstrained LWF and groups public database. When the proposed result compares with other state-of-the-art approaches, the competed accuracy rate was 97.0% with LWF and 91.58% on group dataset. In [19], authors shown that, gender identification accuracy greatly affects with different age groups. Author claims that by using state-of-the-art method experimental study on big dataset of 8000 images from 0 to 93 age range display accuracy of GC on adult faces is 10% higher than the young and senior faces. Author studies that aging effect on human motivates to find which features on shape and texture differences on face can combine to tell about gender of human. LBP and HOG used to estimate the gender characterization difference with age. In [20] paper author introduce child robotic method for

interaction to estimate age and gender by 3D body matrix in real-world multi-party circumstances.

From our quick review presented in above paragraphs, it is clear that soft biometric plays a major role in human–computer interaction and enhancement of accuracy of biometric-based personal identification. Most of the work is dedicated toward adult biometric, whereas very limited focus is given toward children. In this paper, we have presented a method for gender identification from facial images of children.

### 3 Proposed Method

Our aim is to identify the gender from facial images of the children. To achieve this, we employed the three steps, preprocessing, feature extraction, and classification. In preprocessing, we have detected the face and normalized to a standard size. Histogram of oriented gradients (HOG) [21], uniform local binary patterns (ULBP), and Gabor wavelet transform (GWT) [22] based texture descriptors are used for feature extraction. These global approaches are found suitable for facial feature representation. SVM is considered for the classification. Block diagram of the proposed method is shown in Fig. 1.

#### 3.1 Preprocessing

Images in our dataset are unconstrained in nature and having huge variation in illumination, background, size, pose and expressions; therefore, FEI tool [23, 24] is used to detect and normalize the face region. In preprocessing, we first detected the face region. Input image of size  $240 \times 260$  is provided as input and FEI tool returns cropped normalized facial region. Output of face detection is shown in Fig. 2.



**Fig. 1** Flow diagram of proposed method



Fig. 2 Face detection and normalization using FEI tool

### 3.2 Feature Extraction

Selection of suitable features extraction technique is the significant step in designing of any image classification or retrieval system. Our main focus is on efficient feature extraction from face images to obtain discriminating features for gender classification. In this regard, we used uniform local binary patterns (ULBP) variant of LBP [22, 25], histogram of oriented gradients (HOG) [23], Gabor wavelet transforms (GWT).

**Uniform Local Binary Pattern (ULBP):** Local binary pattern (LBP) is one of the very efficient texture descriptors, and it labels the pixel of an image based on thresholding operation, later post-thresholding sequence of 0 and 1 considered as binary number further which is represented by its equivalent decimal. The same process is repeated for whole image. The histogram of these labels  $2^8 = 256$  is used as texture descriptor. Further, it can be optimized by considering only limited intervals as given in (Ojala et al.), which gives only 59 descriptors. Once the labeled image  $IL(x, y)$  is obtained, we can define the histogram of LBP as:

$$H_i \sum_{xy} IL \{ (x, y) = i \}, i = 0, 1, 2, \dots, n - 1 \tag{1}$$

where  $n$  is number of labels and  $IL \{A\}$  is 1 if true or 0 if false. In Fig. 3, we have shown working of local binary patterns.

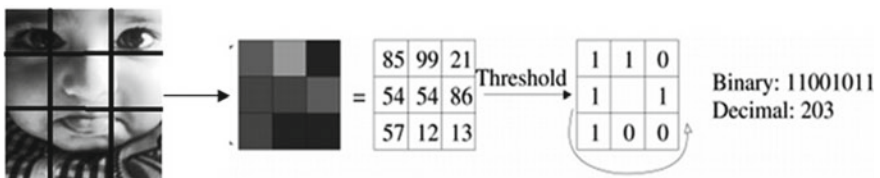
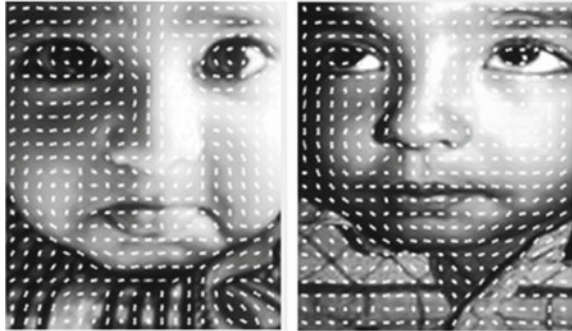


Fig. 3 Computation of local binary pattern

**Fig. 4** Visualization of HOG descriptors applied to facial image of children



**Histogram of Oriented Gradients (HOG):** In image processing, HOG [20, 24, 26, 27] is very popular descriptor used for object detection. In local image portion, it counts the occurrence of gradient. The basic idea behind the HOG descriptor is: (1) It divides the image into small linked section and each section is called cell, and for each cell, it calculates the HOG magnitude and direction for each pixel within the cell and then, (2) According to the gradient, orientation discretizes each cell into angular bins. (3) Each cell pixel has gradient weight according to angular bin. (4) Groups of connected cell are called blocks and grouping of cells into block is base for grouping and normalization of histogram. Following Fig. 4 shows the HOG visualization on our dataset. We use different HOG cell size in our work like  $16 \times 16$ ,  $32 \times 32$  and  $64 \times 64$ .

**Gabor Wavelet Transform (GWT):** Gabor wavelets [28, 29] are powerful tool to highlight the texture regions in image. In the literature, Gabor wavelets are widely used to extract the texture information for various image processing and computer vision applications. Gabor wavelet is the group of Wavelets; [30] mathematically represented as:

$$\phi(x, y, w_0, \Theta) = \frac{1}{2\pi\sigma^2} e^{-(x_0^2+y_0^2)/2\sigma^2} [e^{jw_0x_0} - e^{-w_0\sigma^2/2}] \quad (2)$$

To compute the texture descriptors based on Gabor wavelets, we have applied GWT at six different orientations and five scales which gives us total 30 transformed images. Then, from each transformed image, we have computed mean and standard deviation to form fixed size compact feature vector of size 60.

### 3.3 Classification

**Support Vector Machine:** SVM is supervised classifier based on statistical learning theory [31], introduced by Vapnik. Due to its strong learning ability and discriminating power, we have chosen SVM for difficult task of gender classification in



childrens' facial images. The basic principle of SVM is that it tries to construct hyperplane separating the two linearly separable classes. Suppose we have  $n$  data vectors  $X_i$ . Separation of each data item in two classes is carried out by a discriminant function  $g(X) = W^T.X - b$ .

$$g(X) = W^T.X_i - b \geq 1 \quad (3)$$

where  $y_i$  is the class either +1 (boy) or -1 (girl). In case on nonlinearly separable data, then the SVM can be extended with kernel function, for instance some of them are Cubic, Gaussian, and Polynomial.

## 4 Experiments

### 4.1 Dataset and Evaluation Protocol

Dataset: Due to the unavailability of the standard dataset for gender classification for child, we created our own dataset. To introduce more complexities, we have captured the face images with different smartphones having different configuration to get more variation in images. Facial images in our dataset having different poses, expressions, background, and resolutions. The informed consent is recorded from parents of volunteers during the process of data collection. We have considered total 752 images for our experiment purpose out of which 441 are boys and are 311 are girls. Samples from our dataset are shown in Fig. 5.

We have used tenfold cross-validation technique instead of traditional classification, to evaluate the performance of our method. The whole dataset is divided into 10 sub-parts. When any of one sub-part is used for testing, the other nine sub-parts used for training. This method is repeated 10 times, in such way that each sub-part will serve for both training and for testing.

### 4.2 Results and Discussion

The exhaustive experimental tests are carried out on our dataset to measure the performance of the proposed method. We have tested the performance of ULBP, HOG, and GWT with SVM Classifier. Linear SVM further extended with kernel trick to enhance the classifier performance. Detailed results of our experimental tests are given in Table 1. We have observed that quadratic SVM with histogram of oriented gradients of cell size  $16 \times 16$  given high accuracy in gender identification in child with an average accuracy of 89%, for boy 92% and 86% for girl. ULPB and GWT



**Fig. 5** Samples from our dataset **a** boys, **b** girls

**Table 1** Gender identification accuracy using SVM with ULBP, GWT and HOG

Descriptors	Linear SVM		Avg. (%)	Quadratic SVM		Avg. (%)	Cubic SVM		Avg. (%)
	Boy (%)	Girl (%)		Boy (%)	Girl (%)		Boy (%)	Girl (%)	
ULBP	85	53	69	87	82	85	84	78	82
GWT	83	34	58	84	67	75	85	80	82
HOG (16 × 16)	88	74	81	91	86	88	<b>92</b>	<b>86</b>	<b>89</b>
HOG (32 × 32)	89	72	81	91	84	87	90	87	88
HOG (64 × 64)	78	62	70	85	77	81	86	84	85

with Cubic SVM given weak performance with the classification accuracy of 82 and 82%, respectively. Further we have also witnessed that, the Cubic SVM performed better as compared to Linear SVM and Quadratic SVM.

## 5 Conclusion

In this paper, we have presented a method for gender identification among children. This is the first approach when children gender identification considered for smartphone captured complex dataset having huge variations in pose, lighting, and expressions of children. We have evaluated the performance of different texture descriptors such as ULBP, HOG and GWT; it is observed that HOG with cell size  $16 \times 16$  has good discriminating capacity as compared to ULBP and GWT. Support vector machine has performed better when extended with Cubic Kernel and given average accuracy of 89% for gender identification in child.

In the future, we will consider the combination of classifiers approach to enhance the performance of our system.

**Acknowledgements** We thank to the parents of children who given the consent to capture the facial data of their children, with known fact that data will be used in publication and facial images will reveal the identity of their child. We also thank to the children who took part in the process of database collection.



## References

1. Nixon MS, Correia PL, Nasrollahi K, Moeslund TB, Hadid A, Tistarelli M (2015) On soft biometrics. *Pattern Recognit Lett* 68:218–230
2. Malik T (2014) Technology in the service of development: the NADRA story. Essay, Center for Global Development, Washington, DC
3. Sadruddin MM (2011) Study on the important issues of child rights in Pakistan. *Dialogue* 6(1):14
4. Satta R, Galbally J, Satta R, Galbally J, Beslay L (2014) Children gender recognition under unconstrained conditions based on contextual information, July 2016
5. Li Y-J, Lai C, Wu H, Pan S-T, Lee S-J (2015) Gender classification from face images with local. *Int J Ind Electron Electr Eng* 3(11):15–17
6. Xu Z, Lu L, Shi P (2008) A hybrid approach to gender classification from face images. In: 19th international conference on pattern recognition, p 14
7. Amilia S, Sulistiyo MD, Dayawati RN (2015) Face image-based gender recognition using complex-valued neural network. In: 2015 3rd international conference on information and communication technology ICoICT, pp 201–206
8. Andreu Y, Mollineda RA, Garcia-Sevilla P (2009) Gender recognition from a partial view of the face using local feature vectors. *Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics)*, LNCS 5524, 481–488
9. Moghaddam B, Yang MH (2002) Learning gender with support faces. *IEEE Trans Pattern Anal Mach Intell* 24(5):707–711
10. Jain A, Huang J, Fang S (2005) Gender identification using frontal facial images. In: IEEE international conference on multimedia expo, ICME 2005, pp 1082–1085
11. Basak P, De S, Agarwal M, Malhotra A, Vatsa M, Singh R (2017) Multimodal biometric recognition for toddlers and pre-school children, pp 627–633
12. Tiwari S, Singh A, Singh S (2012) Integrating faces and soft-biometrics for newborn recognition. *Journalshub.com* 2(2):201–209
13. Tiwari S, Singh SK (2012) Face recognition for newborns. *IET Biometrics* 1(4):200–208
14. Mozaffari S, Behravan H, Akbari R (2010) Gender classification using single frontal image per person: combination of appearance and geometric based features. In: Proceedings of the international conference on pattern recognition, pp 1192–1195
15. Shan C (2012) Learning local binary patterns for gender classification on real-world face images. *Pattern Recognit Lett* 33(4):431–437
16. Datta S. Gender identification from facial images using local texture based features. <https://samyak-268.github.io/pdfs/gender-report.pdf>. Retrieved on 11 Oct 2018
17. Yang J, Zhang D, Frangi AF, Yang JY (2004) Two-dimensional PCA: a new approach to appearance-based face representation and recognition. *IEEE Trans Pattern Anal Mach Intell* 26(1):131–137
18. Nguyen H, Thi T, Huong N (2017) Gender classification by LPQ features from intensity and monogenic images, pp 96–100

19. Guo G, Dyer CR, Fu Y, Huang TS (2009) Is gender recognition affected by age? 2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops, Kyoto, 2009, pp 2032–2039. <https://doi.org/10.1109/ICCVW.2009.5457531>
20. Sandygulova A, Dragone M, OHare GMP (2014) Real-time adaptive child-robot interaction: age and gender determination of children based on 3D body metrics. In: 23rd IEEE international symposium on robot and human interactive communication human-robot co-existence adaptive interfaces system dly life, ther assist social engagement interact, pp 826–831
21. Dniz O, Bueno G, Salido J, De La Torre F (2011) Face recognition using histograms of oriented gradients. *Pattern Recognit Lett* 32(12):1598–1603
22. Balamurugan V, Srinivasan M, Vijayanarayanan A (2012) A new face recognition technique using Gabor wavelet transform with back propagation neural network. *Int J Comput Appl* 50(3):41–46; Thomaz CE, Giraldi GA (2010) A new ranking method for principal components analysis and its application to face image analysis. *Image Vis Comput* 28(6):902–913
23. Tenorio EZ, Thomaz CE (2011) Analise multilinear discriminante de formas frontais de imagens 2D de face. In: Proceedings of the X Simposio Brasileiro de Automacao Inteligente SBAI, pp 266–271
24. Universidade Federal de Sao Joao del Rei (2011) Sao Joao del Rei, Minas Gerais, Brazil, 18–21 Sept 2011
25. Hadid A, Pietikainen M, Ahonen T (2004) A discriminative feature space for detecting and recognizing faces. In: Proceedings on 2004 IEEE computer society conference on computer vision pattern recognition, CVPR 2004, vol 2(i), pp 797–804
26. Dalal N, Triggs W (2004) Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision pattern recognition, CVPR05, vol 1(3), pp 886–893
27. Zhao Y, Zhang Y, Cheng R, Wei D, Li G (2015) An enhanced histogram of oriented gradients for pedestrian detection. *IEEE Intell Transp Syst Mag* 7(3):29–38
28. Mustapha S, Jalab HA (2013) Compact composite descriptors for content based image retrieval. In: Proceedings-2012 international conference on advanced computer science application and technologies, ACSAT 2012, pp 37–42
29. Jalab HA (2011) Image retrieval system based on color layout descriptor and Gabor filters. In: 2011 IEEE conference on open systems, ICOS 2011, pp 32–36
30. Zhang H, Jiang X (2015) A method using texture and color feature for content-based image retrieval. In: 2015 IEEE international conference on computer and communications, pp 122–127
31. Vapnik V (2013) The nature of statistical learning theory. Springer Science & Business Media, Berlin

# Assessment of UWAC System Performance Using FBMC Technique



B. Pranitha  and L. Anjaneyulu 

**Abstract** At the physical layer, Underwater Acoustic Communication (UWAC) channels present many challenges for efficient communications, which feature both long delay spreads and serious Doppler effects. The waves and motions at the surface of the sea are the main causes for these effects. The complexity at the receiver increases along with the symbol rate, thereby increasing the burden of the channel. Orthogonal Frequency Division Multiplexing (OFDM) which is a form of multicarrier modulation has been effectively used for UWAC recently because of its ability to handle long dispersive channels reducing the complexity of the channel equalization. However, Inter Carrier Interference (ICI) is introduced by the Doppler effects that destroy the orthogonality of the subcarriers. To overcome the effects of OFDM, an emerging technique known as Filter Bank Multicarrier Communication system (FBMC) is used in UWAC. The high capacity of the transmission bandwidth in FBMC systems can be accomplished by applying Offset Quadrature Amplitude Modulation (OQAM). Because of cyclic prefix deletion, FBMC/OQAM can arrange higher spectral efficiency when compared to OFDM. This paper worked on both OFDM and FBMC/OQAM by considering the losses and noise effects in UWAC environment and explained their performances with respect to Bit Error Rate (BER).

**Keywords** UWAC · OFDM · FBMC · OQAM · BER

## 1 Introduction

Underwater Acoustic Communication (UWAC) is becoming more challenging to researchers these days because of its high applications [1]. The establishment of

---

B. Pranitha (✉) · L. Anjaneyulu  
Department of ECE, National Institute of Technology Warangal, Hanamkonda 506004,  
Telangana, India  
e-mail: [pranitha.boddu@nitw.ac.in](mailto:pranitha.boddu@nitw.ac.in)

L. Anjaneyulu  
e-mail: [anjan@nitw.ac.in](mailto:anjan@nitw.ac.in)

© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_6](https://doi.org/10.1007/978-981-13-6783-0_6)

things such as Underwater Sensor Network (UWSN), Remotely Operated Vehicle (ROVs), and submarine communications initiated a fresh range of research and development. The different applications of UWAC include oceanography, marine archaeology, offshore oil survey, rescue operations, underwater sports, military, and navigation motives [2]. In contrast to the wireless radio channel, UWAC channel is very rough because of its uncommon characteristics. They are the large delay caused by the time-varying multipath propagation which in turn results in Inter Symbol Interference (ISI), limited bandwidth that is available because of frequency-selective attenuation, and high spatial correlation [3]. The important method for transmitting wireless signals underwater is by acoustic waves. However, radio waves can propagate to any distance through underwater but at extra low frequencies (30–300 Hz), and they need large antennae and high transmitter power. Optical waves generally do not have large attenuation; rather, they are affected by scattering. Not only this, optical signals' transmission needs high precision toward the fine laser beams. Therefore, acoustic waves are the appropriate solution for transmitting underwater in applications where tethering is improper [4]. To lessen the bandwidth limitation, multicarrier modulation is introduced in UWAC, which stops the long-time delays in UWAC channels [5]. Orthogonal Frequency Division Multiplexing (OFDM) in the usage of multicarrier modulation technique finds its utilization in the vastly dispersive channel and is advisable for attaining high data rate in frequency-selective UWAC channels. Recent researches affirm the application of OFDM for UWAC for high data rate with improved performance [6]. But because of the use of long symbols in OFDM receiver, the performance is degraded and the channel also introduces division by the frequency axis which results in Inter Carrier Interference (ICI). Moreover, the cyclic prefix (CP) that is used reduces the efficiency of bandwidth as it does not purely use terms of information [7]. So to overcome this, Filter Bank Multicarrier Communication (FBMC) technique is introduced. This technique is introduced to reduce both ISI and ICI [8]. Le Floch et al. [9] illustrated the implication of FBMC techniques in working with the channels displaying time spread and frequency spread. Here, time spread is caused due to the multipath effects, while the time variation in the channel causes frequency spreading. These two are the major aspects in UWAC. To overcome OFDM limitations and meet 5G requirements, several European projects have been launched recently. In the 5GNow project, for instance, various modulations have been suggested, namely FBMC, Generalised Frequency Division Multiplexing (GFDM), and universal filtered multicarrier (UFMC). Among these multiple access techniques, FBMC seems to be a good candidate for 5G systems. The transmission bandwidth's high capacity in FBMC systems can be attained by applying Offset Quadrature Amplitude Modulation (OQAM). In the literature, many types of research shed light on the advantages of FBMC with OQAM (FBMC/OQAM) systems compared to OFDM systems. Among these advantages, one can cite their robustness to channel time and frequency spreading since their waveforms are restricted in time and frequency [10]. In this paper, these techniques are applied to multipath UWAC considering various factors that are involved in it and a comparison study between these two techniques in UWAC is made.

The paper is organized as follows: Sect. 2 describes the two different techniques used in the paper, namely OFDM and FBMC, while the underwater channel is explained in Sect. 3. Simulated results are exhibited in Sect. 4 followed by conclusions and future scope in Sect. 5.

## 2 Techniques Used

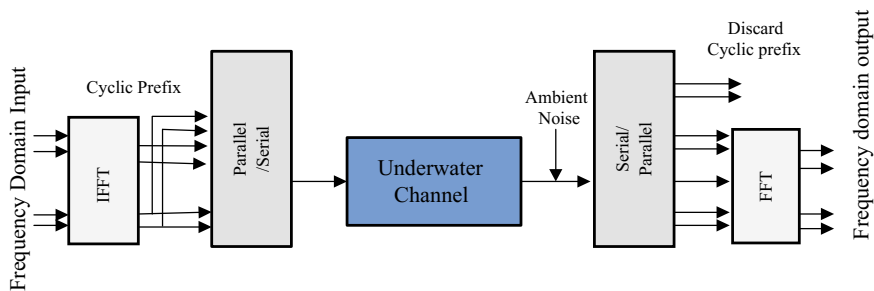
### 2.1 Orthogonal Frequency-Division Multiplexing (OFDM)

The design of extraordinary data rate in UWAC is quite a challenging task because of the Underwater Acoustic (UWA) channel being both time and frequency selective. The existing techniques that are used in RF communications probably are not applicable to the adverse UWA channel. Nevertheless, OFDM as multicarrier modulation in an underwater environment is appropriate to use. This is because of the receiver's low complexity design that deals with vastly scattering channels. Here, the bandwidth available is branched into many subcarriers. The space of the frequency between them is arranged in such a manner that they are orthogonal to each other by which interference between the signals can be reduced. Channel equalization becomes much simpler in OFDM than by using adaptive equalization techniques with single-carrier systems. The multipath spread and the ISI that is caused by this spread in UWA channel are reduced by using OFDM. OFDM eradicates the demand of complicated time-domain equalizers. Therefore, it is appropriate for real-time communications in the acoustic medium. The block diagram of OFDM in UWAC is as depicted in Fig. 1. To convert from frequency to time series, the input stream is passed through the block having Inverse Fast Fourier Transform (IFFT). These time sequences can be extended by adding CP. The duplicate of the symbol's last part being added to the start of the sequence is known as CP. This should be generally bigger than the delay spread of the network to relieve the ISI caused by the advent of distinct OFDM symbols that has a distinct delay. This parallel data is altered into serial form by using Parallel to Serial (P/S) converter. The final digital form of the signal is transformed into analog and transferred through the underwater medium. The ambient noise is also considered here which is explained in Sect. 3. All the reverse operations are performed at the receiver to collect the transmitted data.

OFDM is apt to enlarge spectral efficiency without constructing channel interference. Among many decisive reasons, the most important one for choosing OFDM in UWAC is its ability to tackle with multipath propagation delay. This can be used for outrageous bit-rate underwater communication. After the FFT demodulation, the received signal is modeled as [11]:

$$y_m(n) = H_m(n)d_m(n)e^{j\theta_m(n)} + z_m(n) \quad m = 0, 1, \dots, M - 1. \quad (1)$$

Here,  $H_m(n)$  is the transfer function of the channel that is calculated at a frequency:



**Fig. 1** OFDM for UWAC

$$f_m = f_0 + M \Delta f \quad (2)$$

During the  $n$ th block,  $d_m(n)$  is the data symbol that is communicated on the  $M$ th subcarrier along the  $n$ th OFDM symbol and  $z_m(n)$  represents ambient noise.

The spacing between the subcarrier is given by:

$$\Delta f = B/M \quad (3)$$

The period of single OFDM symbol is given by:

$$T = 1/\Delta f \quad (4)$$

Also, the span of a single OFDM block is given by:

$$T' = T + T_g \quad (5)$$

Here,  $T_g$  is the guard interval.

The distortion of the phase  $\theta_m(n)$  in acoustic channel is modeled as:

$$\theta_m(n) = \theta_m(n-1) + a(n) \cdot 2\pi f_m \cdot T' \quad (6)$$

Here,  $a(n)$  is termed as the Doppler rate that is generally expected to be constant in single OFDM block, however it may differ from one block to another. The channel transfer function in OFDM is given by the following equation:

$$H_m(n) = \sum_{l=0}^{M-1} h_l(n) \cdot e^{-j2\pi ml/M} \quad (7)$$

As in wireless communication, the OFDM block diagram almost remains the same except that the channel considered here is the multipath underwater channel.



Because of its wideband nature, OFDM-based UWAC is very delicate in handling the frequency offset. The slight alteration may result in overlapping of the subcarriers wherein the entire signal can get completely distracted. Subcarrier overlapping is of another form of Doppler shifting. This is because of the ocean dynamics that is caused by the slow speed of the sound when compared to that of RF. This sort of overlapping is known as ICI which will damage the orthogonality between the subcarriers [12]. It also requires a more linear power amplifier [13].

To overcome all these disadvantages, FBMC technique is used which is clearly explained in Sect. 2.2.

### 2.2 Filter Bank Multicarrier Communications (FBMC)

To overcome the time scattering of UWA channels, OFDM technique preferably with a long CP is recommended. The disadvantages of OFDM technique are conquered by a multicarrier communication system called FBMC. This was initially proposed by Saltzberg that delivers an improved spectral shaping of subcarriers than OFDM systems [14]. Due to small sideband power leakage, FBMC technique is considered as interesting alternatives for general OFDM. The block diagram of FBMC is shown in Fig. 2.

FBMC is an advancement of OFDM. Some researchers also say that it is an evolved version of OFDM that can be improved for better presentation in the doubly dispersive UWA channels. In FBMC, each of the individual subchannels is filtered by its own. It uses very narrow band filter with sufficiently long time length. This gives us good control over each filter bank. Each of the subcarriers passes through a filter called ‘pulse-shaping filter.’ The disadvantages of OFDM can be conquered by adding generalized pulse-shaping filters in FBMC. This yields a fine-bounded subchannel in both time and frequency domain. Non-rectangular pulse shapes (prototype filters) are used by FBMC waveforms wherein the time duration of the symbol is undoubtedly larger than OFDM. This is done so that neighboring FBMC symbols overlay in time. These arrangements in this prototype filters are made such that interference between the symbols is low [15]. Correspondingly, FBMC technique has high spectral

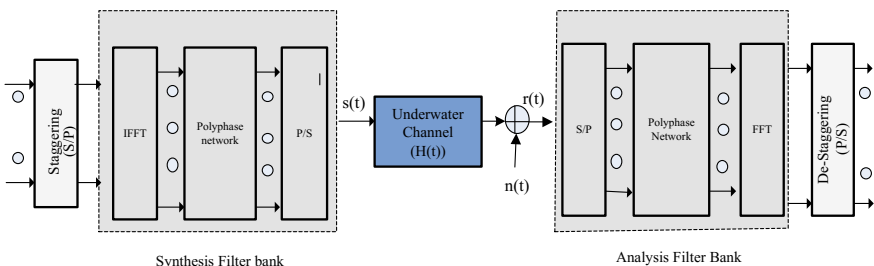


Fig. 2 FBMC for UWAC

constraint signals, and they offer further efficient use of the means in the absence of CP. Orthogonality is must for all the carriers in OFDM, while in FBMC, it is used only for neighboring subchannels. OFDM divides the available frequency bandwidth into some carriers, in the case of FBMC; the transmission channel is divided into some subchannels. OQAM is used to divide the channel bandwidth. Also, for the subchannel modulation to get accustomed to the adjacent orthogonality, OQAM is used. Maximum bit rate without the requirement of guard time can be achieved by the combination of filter banks and OQAM. The key variance among OFDM and FBMC is that the OFDM is replaced with a multicarrier system that is built on filter banks, where the IFFT and input cyclic prefix are substituted by the Synthesis Filter Bank (SFB) at the transmitter side, whereas FFT and output cyclic prefix are substituted by the Analysis Filter Bank (AFB) by the receiver side. As depicted in Fig. 2, the input signal is first transformed from serial to parallel form and then passed through SFB, and then, it is converted back to serial form after coming out of synthesis bank. After this, it can be seen that in the receiver side after the signal passes through the channel, it is converted to parallel form by S/P converter and passed through AFB. Finally, when the output signal is obtained, it is again converted to serial form by P/S converter. A number of digital filters are laid out parallel in a filter bank. These filters have digital up-samplers at the transmitter and down-samplers at the receiver. Time indexes  $n$  and  $m$  stand for low and high sampling rate, respectively, which are used to emphasize the multirate nature of the system [16]. The prototype filters that are in FBMC are sketched such that the interference among overlapping symbols is low.

Because of cyclic prefix deletion, FBMC/OQAM can provide high spectral efficiency when compared to that of OFDM. The transmitted signal in FBMC/OQAM is expressed as follows [17]:

$$s(t) = \sum_{m=0}^{M-1} \sum_{n=-\infty}^{\infty} a_{m,n} i^{m+n} g(t - n\tau_0) e^{j2\pi f_0 t} \quad (8)$$

where  $a_{m,n}$  stands for the data that is sent on the  $m$ th subcarrier at the  $n$ th OQAM symbol,  $M$  denotes the subcarriers,  $g$  stands for prototype function, and  $f_0$  represents the spacing of the subcarrier which fulfills:

$$f_0 = 1/T_0 = 1/(2\tau_0) \quad (9)$$

$T_0$  is the symbol duration of QAM,  $\tau_0$  is the OQAM symbol duration, and  $g_{m,n}$  represents the prototype function's shifted version. This can be expressed as:

$$g_{m,n} = g(t - n\tau_0) e^{j2\pi f_0 t} i^{m+n} \quad (10)$$

$g_{m,n}$  that is between the subcarriers are entirely orthogonal in OFDM, but this is not the situation in FBMC/OQAM due to interference.

The received signal is expressed as:

$$r(t) = \sum_{m=0}^{M-1} \sum_{n=-\infty}^{\infty} a_{m,n} H_{m,n} g_{m,n}(t) + n(t) \quad (11)$$

where  $n(t)$  denotes the ambient noise and  $H_{m,n}$  are complex-valued channel frequency coefficients.

### 3 Underwater Channel

The three main factors that characterize the underwater acoustic propagation are attenuation which increases with signal frequency, low speed of sound, and time-varying multipath propagation. The noise which affects the underwater channel is the ambient noise or the background noise. The waves caused by the sea surface, interior turbulence, fluctuations in the speed of the sound, etc., contribute to random signal variations. Turbulence, waves, shipping, and thermal noise are generally used in modeling ambient noise. Generally, ambient noise sources are illustrated by continuous Power Spectral Density (PSD) and Gaussian statistics. The PSD of the noise as a function of frequency in kHz is given by the following practical formula. Overall PSD is given by [18],

$$N(f) = N_t(f) + N_s(f) + N_w(f) + N_{th}(f) \quad (12)$$

where  $N_t(f)$  is the turbulence noise,  $N_s(f)$  stands for shipping noise,  $N_w(f)$  is the noise caused by waves, and  $N_{th}(f)$  denotes the thermal noise.

The main quality of UWA channel is the dependency of path loss on the frequency of a signal which is caused by absorption. Apart from the absorption loss, spreading loss is also experienced by the signal, which is proportional to the distance among them. The final path loss is expressed as:

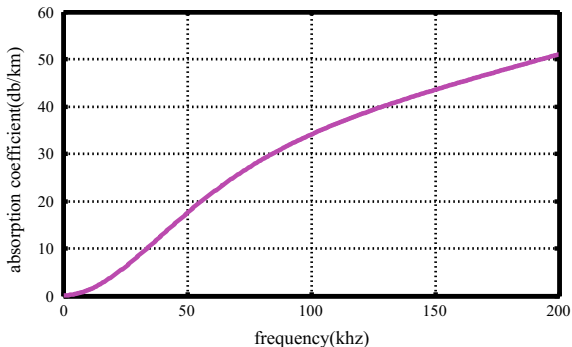
$$A(l, f) = \left(\frac{l}{l_r}\right)^k a(f)^{l-l_r} \quad (13)$$

Here,  $f$  denotes the frequency of the signal and  $l$  denotes the transmitted distance which is taken about reference transmission distance  $l_r$ . The spreading loss is described by the path loss exponent ( $k$ ). Its value is between 1 and 2. The absorption coefficient  $a(f)$  is obtained by using a practical formula [19]:

$$10 \log a(f) = 0.11 \frac{f^2}{1 + f^2} + 44 \frac{f^2}{4100 + f^2} + 2.75 \times 10^{-4} f^2 + 0.003 \quad (14)$$

Figure 3 shows the rapid increase of the absorption coefficient with frequency. It is clearly shown in the graph that the absorption coefficient increases exponentially concerning the frequency.

**Fig. 3** Frequency versus absorption coefficient in UWAC



Signal-to-noise ratio (SNR) is affected by attenuation and the noise that changes with the signal bandwidth. If we consider narrow frequency band of width  $\Delta f$  with some frequency  $f$ , then SNR is expressed as:

$$\text{SNR}(l, f) = S_l(f) / A(l, f)N(l, f) \quad (15)$$

where  $S_l(f)$  is power spectral density of the transmitted signal,  $A(l, f)$  is the overall path loss, and  $N(f)$  is the noise figure.

The received signals from both OFDM and FBMC techniques that are given in Eqs. (1) and (11) consist of the noise that is mentioned in this section. These noise elements are one of the main reasons for the error in the propagation of UWAC. So the performance evaluation factor BER is calculated as:

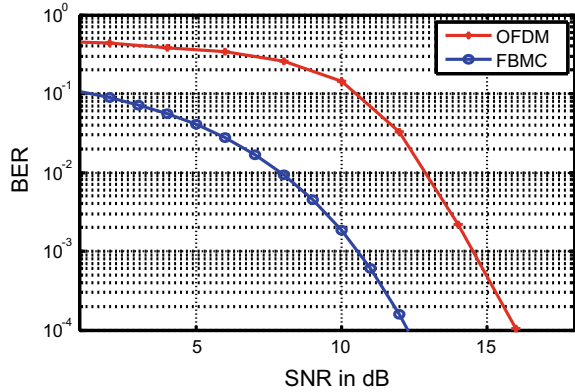
$$\text{BER} = \frac{\text{No. of errors}}{\text{Total no.of bits transmitted}} \quad (16)$$

## 4 Simulation Results

MATLAB simulations have been carried out for OFDM and FBMC for the underwater channel characteristics that are mentioned in Sect. 3. These characteristics mainly consider the losses and noise coefficients considered in UWAC.

From Fig. 4, it can be seen that FBMC outperforms the OFDM modulation technique regarding bit error rate and is found to be a better technique to reduce ICI. There is almost a 3 dB difference between OFDM and FBMC regarding SNR. The much known limitations of OFDM such as reduced spectral efficiency and harsh synchronization requirements can be surpassed by using FBMC technique. It is less sensitive to Doppler effect and therefore performs naturally with the development of the user mobility. The FBMC technique subdues the limitations in OFDM by involving simplified pulse-shaping filters which delivers a well-contained subchannel together in time and frequency domain. Because of the usage of prototype filters in FBMC, ICI

**Fig. 4** BER to SNR for UWAC



and ISI are greatly reduced when compared to OFDM which are considered to be the most important factors which affect the UWAC.

## 5 Conclusions and Future Scope

FBMC outperforms OFDM in the matter of spectral efficiency, robustness, and spectral protection at the cost of a slight increase in the complexity of UWAC. The shortcomings of the OFDM technique have been addressed and removed by the FBMC. The future scope is the optimization of FBMC for MIMO UWA channels. The implementation of further coding and estimation techniques at the receiver are to be explained with FBMC to compare with the existing OFDM UWA transceivers that exist.

## References

1. Sharma SK, Sharma U (2017) UWA communication using MIMO OFDM. *Int J Eng Manag Res (IJEMR)* 7(3):346–452
2. Pranitha B, Anjaneyulu L (2016) Review of research trends in underwater communications—A technical survey. In: 2016 international conference on communication and signal processing (ICCSP), pp 1443–1447, April 2016
3. Zhang L, Xu X, Feng W, Chen Y (2016) Multi-array iterative receiver for underwater acoustic OFDM communications with EXIT chart evaluation. *Appl Acoust* 114:307–316
4. Stojanovic M (1999) Underwater acoustic communication. In: *Wiley encyclopedia of electrical and electronics engineering*, pp 1–12
5. Le Floch B, Alard M, Berrou C (1995) Coded orthogonal frequency division multiplex. *Proc IEEE* 83:982
6. Chithra K, Sireesha N, Thangavel C, Gowthaman V, Narayanan SS, Sudhakar T, Atmanand MA (2015) Underwater Communication implementation with OFDM. *Indian J Geo-Mar Sci* 44(2): 259–266, Feb 2015

7. Tensubam BD, Chanu NL, Singh S (2014) Comparative analysis of FBMC and OFDM multi-carrier techniques for wireless communication networks. *Int J Comput Appl* 100(19)
8. Amini P, Chen RR, Farhang-Boroujeny B (2015) Filterbank multicarrier communications for underwater acoustic channels. *IEEE J Oceanic Eng* 40(1):115–130
9. Esmail H, Jiang D (2013) Multicarrier communication for underwater acoustic channel. *Int J Commun Netw Syst Sci* 6:361–376
10. Khrouf W, Siala M, Abdelkefi F (2018) How much FBMC/OQAM is better than FBMC/QAM? A tentative response using the POPS paradigm. *Wirel Commun Mob Comput*
11. Stojanovic M (2008) OFDM for underwater acoustic communications: adaptive synchronization and sparse channel estimation. In: *IEEE international conference on acoustics, speech and signal processing, ICASSP 2008*, pp 5288–5291, Mar 2008
12. Khan AIR, Gang Q, Mehboob K (2013) Investigation of channel modeling and simulation of OFDM based communication near northern regions of arabian sea. *Res J Appl Sci Eng Technol* 5(4):1169–1182
13. Bhardwaj M, Gangwar A, Soni D (2012) A review on OFDM: concept, scope & its applications. *IOSR J Mech Civ Eng (IOSRJMCE)* 1(1):07–11
14. Kansal P, Shankhwar AK (2017) FBMC vs OFDM waveform contenders for 5G wireless communication system. *Wirel Eng Technol* 8(04):59
15. Moles-Cases V, Zaidi AA, Chen X, Oechtering TJ, Baldemair R (2017) A comparison of OFDM, QAM-FBMC, and OQAM-FBMC waveforms subject to phase noise. In: *2017 IEEE international conference on communications (ICC)*, pp 1–6. IEEE, May 2017
16. Tensubam BD, Singh S (2014) A review on FBMC: an efficient multicarrier modulation system. *Int J Comput Appl* (0975–8887), pp 17273
17. He X, Zhao Z, Zhang H (2012) A pilot-aided channel estimation method for FBMC/OQAM communications system. In: *2012 international symposium on communications and information technologies (ISCIT)*, pp 175–180. IEEE, Oct 2012
18. Stojanovic M, Preisig J (2009) Underwater acoustic communication channels: propagation models and statistical characterization. *IEEE Commun Mag* 47(1):84–89
19. Joshy S, Babu A (2010) Capacity of underwater wireless communication channel with different acoustic propagation loss models. *Int J Comput Networks Commun (IJCNC)* 2

# Simple Exponential Smoothing and Its Control Parameter: A Reassessment



Dipta Chaudhuri, Moloy Mukherjee, Mofazzal H. Khondekar  
and Koushik Ghosh

**Abstract** Simple exponential smoothing (SES) is a popular form of smoothing which can be regarded as a recursive system with infinite impulse response (IIR). As a consequence, it suffers heavily from the threat of instability and phase nonlinearity. Here an effort has been made to get a finite impulse response (FIR) estimation of the SES to acquire the benefits of an optimal speed and computational efficiency in addition to the usual advantages like stability and phase linearity. The optimal order of the filter, its corresponding transfer function, has been worked out, and the frequency response has been estimated for this FIR form of the SES. The frequency response has been compared with the actual IIR form of the SES. In addition to this an effort has been made to get a suitable estimation of the control parameter also called smoothing constant ( $\alpha$ ) in order to have an effective smoothing taking the cut-off frequency, computational limitation of the transfer function, minimum MSE, SNR improvement and suitable window function realisation of the FIR form of SES into consideration. A magnitude of 0.5615 is found to be the most suitable value of the control parameter.

**Keywords** Infinite impulse response (IIR) · Finite impulse response (FIR) · Simple exponential smoothing (SES) · Smoothing constant

---

D. Chaudhuri · M. Mukherjee

Department of Electronics and Communication Engineering, Dr. B.C. Roy Engineering College, Durgapur, India

e-mail: [dipta.chaudhuri@bcrec.ac.in](mailto:dipta.chaudhuri@bcrec.ac.in)

M. Mukherjee

e-mail: [moloy.mukherjee@bcrec.ac.in](mailto:moloy.mukherjee@bcrec.ac.in)

M. H. Khondekar (✉)

Department of Applied Electronics and Instrumentation Engineering, Dr. B. C. Roy Engineering College, Durgapur 713206, India

e-mail: [hossainkm\\_1976@yahoo.co.in](mailto:hossainkm_1976@yahoo.co.in); [km.hossain@bcrec.ac.in](mailto:km.hossain@bcrec.ac.in)

K. Ghosh

Department of Mathematics, University Institute of Technology, University of Burdwan, Burdwan 713104, India

e-mail: [koushikg123@yahoo.co.uk](mailto:koushikg123@yahoo.co.uk)

© Springer Nature Singapore Pte Ltd. 2019

S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,

Advances in Intelligent Systems and Computing 922,

[https://doi.org/10.1007/978-981-13-6783-0\\_7](https://doi.org/10.1007/978-981-13-6783-0_7)

## 1 Introduction

Simple exponential smoothing (SES) is an effective time series technique, which is though primitive, yet not lost its implicational significance particularly in the domain of forecasting and filtering owing to its simplicity and accuracy [1]. Detailed literature review on SES has been made by some authors like Gardner [2, 3], Hyndman et al. [4]. The formulation of exponential smoothing forecasting methods arose in the 1950s from the work of Brown [5] and Holt [6] who were working on creating forecasting models for inventory control systems. SES is appropriate for short-term forecasting [7] of a time series having no trend and seasonality. Though it has the limitation of handling the nonlinear time series [8], still it is one of the most popular and successful techniques used in various applications of not only forecasting [9, 10] but also in data preprocessing to eliminate the irregularities in the raw data [11–13].

Exponential smoothing is basically exponential weighted moving average (EWMA) process where different weights are given to the recent and previous observations. The weights depend on the magnitude of the control parameter (smoothing constant)  $\alpha$ , causing the forecast values or the filter output to change accordingly. The two commonly used forms of SES are expressed in Eqs. 1 and 2.

$$y(n) = \alpha x(n) + (1 - \alpha)y(n - 1) \quad (1)$$

[14, 15]

and

$$y(n) = \alpha x(n - 1) + (1 - \alpha)y(n - 1) \quad (2)$$

[16, 17]

where  $y(n)$  is the output, and  $x(n)$  is the input at the  $n$ th time instant. Both the equations above are ARIMA (0, 1, 1) [1]. Equation (1) represents a filter whereas Eq. (2) not only corresponds as a filter but also as a forecaster [18–20], and hence, the last one is taken for analysis in this paper.

In this paper, the primary objective is to explore the performance of this simple yet very effective SES. As evident from Eqs. (1) and (2), the system is with feedback, and hence, it may be treated as recursive system with infinite impulse response, i.e. IIR system or filter. The analysis of the SES has been made from two perspectives.

Firstly, though IIR filter is advantageous in speed, computational efficiency and excellent magnitude response, it suffers massively from instability and phase nonlinearity [21]. Finite impulse response (FIR) filters are free from these problems. Here initiative has been taken to get an FIR estimation of the above IIR filter.

Secondly, depending on the value of  $\alpha$  performance of the SES changes. It is evident from Eqs. (1) and (2) that  $\alpha$  determines the degree of weight given to the recent and past values of the signal. For  $\alpha > 0.5$ , recent inputs get more significance while for  $\alpha < 0.5$  the current value depends more on past outputs [22]. And depending on the contribution of the recent inputs or past outputs, the signal is smoothed accord-



ingly. The choice of the proper value of  $\alpha$  is of paramount need to use the SES as a filter or predictor [23]. Using the error functions like MSE, MAD, MAPE, etc., the optimised value of  $\alpha$  has been estimated by many authors [18, 24–27]. In all these works, the estimation of the smoothing constant has been made by taking the SES as a forecasting tool. However, to the best of the knowledge of the authors, major works have not been endeavoured by the research communities to find a suitable value of the  $\alpha$  while SES acts as a filter. In this paper along with one of the error function MSE, the other parameters associated with a filter like cut-off frequency, transfer function limitation, SNR improvement and appropriate choice of window function are considered to estimate the suitable value of  $\alpha$ .

## 2 FIR Realisation of the IIR Form of SES Filter

### 2.1 Estimation of Order of the SES in FIR Form

The speed and computational efficiency of a filter are high or low for less or more number of coefficients used. On unfolding Eq. (2), it can be obtained that

$$y(n) = \alpha x(n-1) + \alpha(1-\alpha)x(n-2) + \dots = \alpha \sum_{k=1}^{\infty} (1-\alpha)^{k-1} x(n-k) \quad (3)$$

So Eq. (3) is the non-recursive feed-forward representation of the recursive filter (IIR) of Eq. (2).

The FIR form can be realised from Eq. (3) by truncating the infinite non-recursive filter coefficients.

Let  $\varepsilon$  be the threshold value of the exponentially decreasing coefficients  $\alpha(1-\alpha)^{k-1}$  beyond which they can be assumed to be zero. If  $N$ th coefficient attains the value of  $\varepsilon$ , then  $N$  will be the order of the finite non-recursive filter.

$$\therefore \varepsilon = \alpha(1-\alpha)^{N-1} \quad (4)$$

And the truncated form of Eq. (3) will be the

$$y(n) = \alpha \sum_{k=1}^N (1-\alpha)^{k-1} x(n-k) \quad (5)$$

From Eq. (4), it can be written as

$$N = \left\{ \log(\varepsilon/\alpha) / \log(1-\alpha) \right\} + 1 \quad (6)$$

The value of  $\varepsilon$  gives the order of the system  $N$ . Here it is to be noted that to have a finite value of  $N$ , it must satisfy that  $0 < \alpha < 1$ . To avoid the high order of the filter that will reduce the speed and efficiency, a low value  $\varepsilon$  is recommended.

From Eq. (6), the sensitivity ( $S$ ) of the variation of  $N$  with  $\varepsilon$  is

$$S = |dN/d\varepsilon| = |1/\varepsilon \log\{\alpha(1 - \alpha)\}| \quad (7)$$

It is seen now that for very low values of  $\varepsilon$ ,  $S$  becomes too much sensitive to  $\alpha$ . So it is recommended to use not very low value of  $\varepsilon$  and some intermediate value of  $\alpha$ .

## 2.2 Transfer Function of SES in FIR Form

Now FIR filter's output,  $y(n)$ , is the convolution of the input sequence  $x(n)$  and the filter's impulse response (coefficients)  $h(n)$

$$y(n) = h(n) * x(n) = \sum_{k=1}^N h(k)x(n - k) \quad (8)$$

Comparing Eq. (8) with Eq. (5),

$$h(k) = \alpha(1 - \alpha)^{k-1} \quad \text{for } k = 1, 2, \dots, N \quad (9)$$

The TF,  $H(z)$ , of the non-recursive form of the SES is the z-transform of  $h(k)$

$$H(z) = \sum_{k=1}^N h(k)z^{-k} = \sum_{k=1}^N \alpha(1 - \alpha)^{N-k} z^{-k} \quad (10)$$

Again from Eq. (2), the TF of the recursive form of the SES can be deduced to

$$H(z) = \alpha/\{z - (1 - \alpha)\} \quad (11)$$

The TF of recursive form (Eq. 11) is found to be simpler and hence computationally more efficient than that of FIR form (Eq. 10). But due to the pole  $(1 - \alpha)$  in the TF of the IIR form, it may get unstable when the pole lies outside the unit circle of the  $z$ -plane. On the contrary, as the poles are at the origin of the  $z$ -plane, FIR filter is free from this menace. Therefore, it is confirmed that the condition of  $0 < \alpha < 1$ , is the absolute need to be satisfied both by the FIR and IIR form of the SES.

### 2.3 Frequency Response of FIR Form of the SES

The frequency response of the FIR filter can be obtained by simply substituting  $z$  with  $e^{j\omega}$  in Eq. (10)

$$H(e^{j\omega}) = \sum_{k=1}^N \alpha(1 - \alpha)^{N-k} e^{-j\omega k} \tag{12}$$

Let  $A(\alpha) = \sum_{k=1}^N \alpha(1 - \alpha)^{N-k} \cos \omega k$  and  $B(\alpha) = \sum_{k=1}^N \alpha(1 - \alpha)^{N-k} \sin \omega k$ . Therefore, magnitude response is  $|H(e^{j\omega})| = \sqrt{A^2 + B^2}$  and phase response:  $\angle H(e^{j\omega}) = \tan^{-1} \frac{B}{A}$ .

Figures 1 and 2 show the magnitude response for different values of  $\alpha$  and phase response for  $\alpha = 0.5$ , respectively, of a non-recursive SES filter with  $\varepsilon = 0.001$  in Eq. (6).

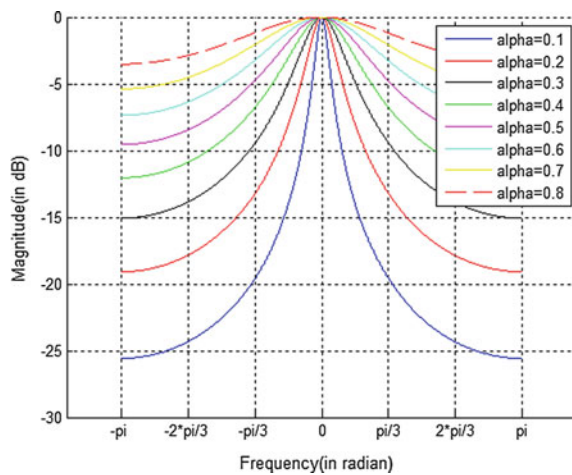
From Eq. (11), the magnitude and phase response of recursive SES are obtained as in Eqs. (13) and (14)

$$|H(e^{j\omega})| = \alpha / \{1 - 2(1 - \alpha) \cos \omega + (1 - \alpha)^2\}^{0.5} \tag{13}$$

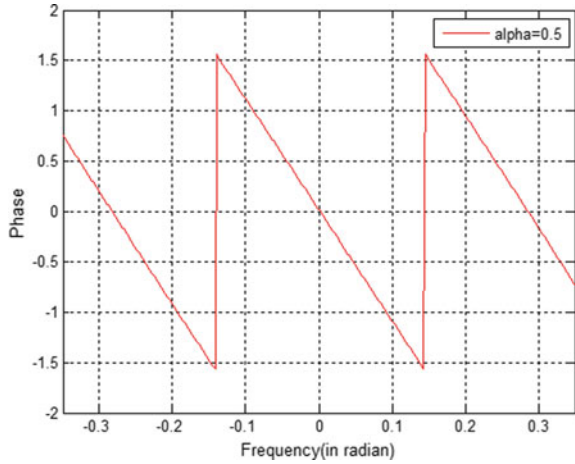
$$\angle H(e^{j\omega}) = \tan^{-1} [(-\sin \omega) / \{\cos \omega - (1 - \alpha)\}] \tag{14}$$

Figures 3 and 4 show the magnitude response for different values of  $\alpha$  and phase response for  $\alpha = 0.5$ , respectively, of recursive SES filter. Comparing Fig. 1 with Fig. 3, it is found that the magnitude response of the non-recursive SES nearly resembles that of the recursive SES, but the phase response is linear in case of the former compared to the later as shown in Figs. 2 and 4. However, this nonlinearity

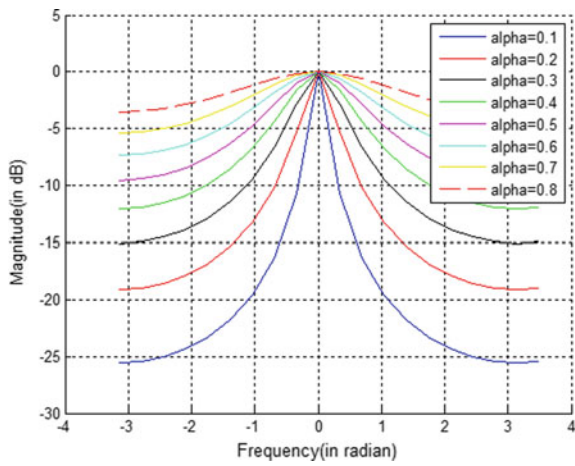
**Fig. 1** Magnitude response of non-recursive SES filter



**Fig. 2** Phase response of FIR SES filter for  $\alpha = 0.5$



**Fig. 3** Magnitude response of recursive SES filter



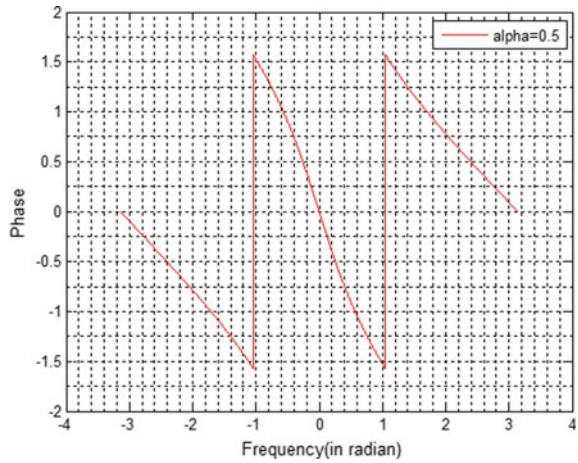
in the phase response of recursive SES reduces for lower values of  $\alpha$  but at the cost of narrow bandwidth as shown in Fig. 5.

So, to have a low-pass filter with an optimised bandwidth and linear phase response, the values of  $\alpha$  may safely be chosen in the mediocre range.

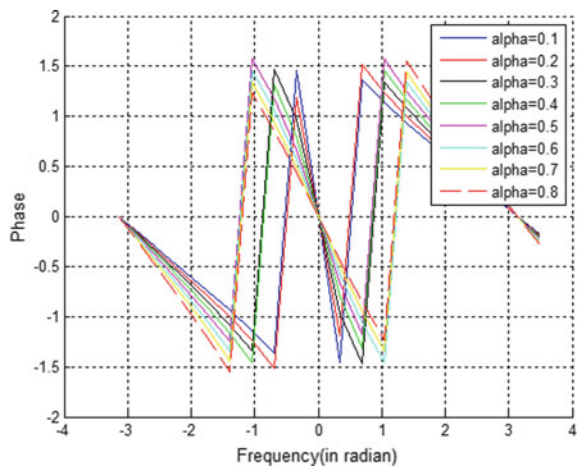
### 3 Estimation of Optimised Control Parameter

Since the smoothing constant ( $\alpha$ ) decides the performance of the SES, an optimised value of  $\alpha$  is highly desirable. Here the factors like cut-off frequency, MSE, algebraic limitation of TF, SNR improvement and identifying suitable window function of the

**Fig. 4** Phase response of recursive SES for  $\alpha = 0.5$



**Fig. 5** Phase response of recursive SES for different values of  $\alpha$

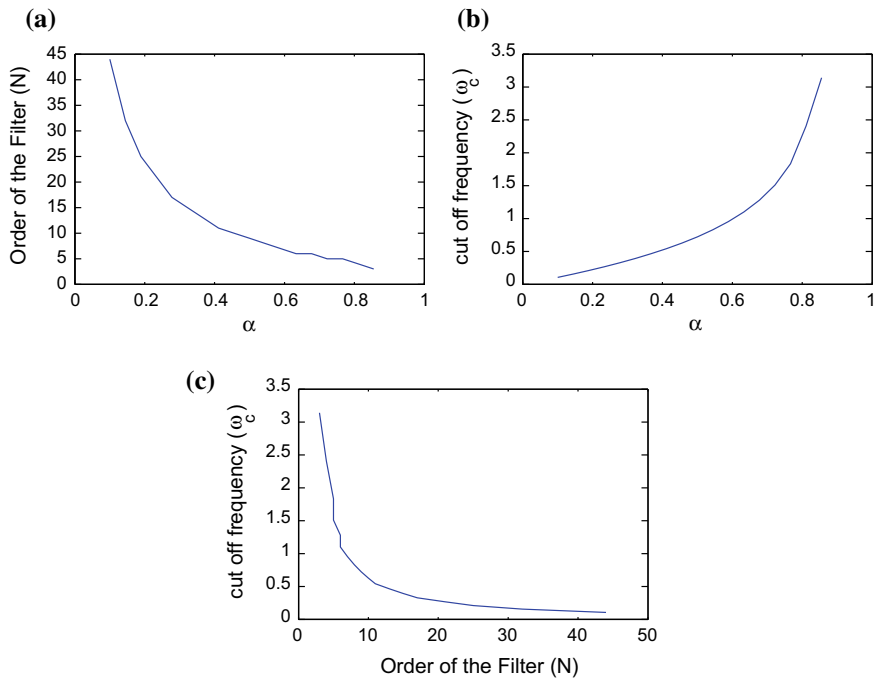


FIR form of the SES have been considered to assess the most favourable value of the smoothing constant or control parameter.

### 3.1 Cut-off Frequency in the Frequency Spectrum

From the definition of cut-off frequency or corner frequency or break frequency ( $\omega_c$ ) [28], it is to be said that at 3 dB point the power of the filter gets halved to that of the passband, i.e.

$$|H(z)^2| \approx 0.5 \tag{15}$$



**Fig. 6** Relationship between cut-off frequency ( $\omega_c$ ), smoothing constant ( $\alpha$ ) and order of the filter ( $N$ ) for non-recursive SES filter obtained from Eqs. 6 and 16

From Eqs. (13) and (15) it can be written that

$$\omega_c = \cos^{-1}[1 - \{\alpha^2/2(1 - \alpha)\}] \quad (16)$$

So, to design the SES with higher  $\omega_c$  the value of the  $\alpha$  needs to be increased as shown in Fig. 6b. It means for the FIR form of the SES, the signals with high frequency need to be filtered with SES having a large value of  $\alpha$ , and for low-frequency signal smaller value of  $\alpha$  is sufficient. The change of the order of the filter ( $N$ ) with respect to  $\alpha$  and the cut-off frequency ( $\omega_c$ ) is shown in Fig. 6a, c.

### 3.2 Computational Bounds of the Transfer Function

From Eq. (16), as the magnitude of  $\cos(\omega_c)$  lies within  $-1$  to  $+1$ , two cases may arise:

#### Case 1

$$1 - \{\alpha^2/2(1 - \alpha)\} \geq -1$$

$$\Rightarrow \alpha \leq \pm 0.82843 \quad (17)$$

Now  $\alpha$  cannot be negative as this will force the pole of the transfer function of Eq. (11) to move out of the unit circle making the system unstable. So,  $\alpha \leq +0.82843$ .

### Case 2

$$1 - \{\alpha^2/2(1 - \alpha)\} \leq 1 \quad (18)$$

This is possible when

$$\left. \begin{array}{l} (a) \alpha^2 \geq 0 \\ (b) 1 - \alpha \geq 0 \end{array} \right\} \Rightarrow 0 \leq \alpha \leq 1 \quad \text{or} \quad \left. \begin{array}{l} (a) \alpha^2 \leq 0 \\ (b) 1 - \alpha \leq 0 \end{array} \right\} \\ \Rightarrow \text{This combination is not possible as } \alpha \text{ becomes imaginary.}$$

Therefore from the above argument, it can be said that the value of  $\alpha$  lies between 0 and 0.82843, and this is evident from Fig. 6a, b.

### 3.3 Conventional Mean Square Error (MSE) Approach

Mean square error (MSE) [18] is a signal fidelity measure. The goal of a signal fidelity measure is to compare two signals by providing a quantitative score that describes the degree of similarity between them.

If  $X = \{x_i | i = 1, 2, 3, \dots, N\}$  and  $Y = \{y_i | i = 1, 2, \dots, N\}$  are the two finite length discrete signal of length  $N$  and  $x_i$  and  $y_i$  are the value of the  $i$ th samples in  $X$  and  $Y$ , respectively, then the MSE between the signals is

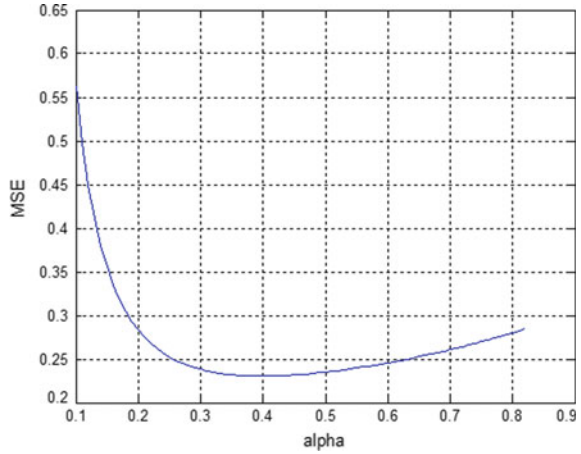
$$\text{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 \quad (19)$$

Let a non-stationary signal,  $X = A \sin 2\pi f_m t^2 = A \sin(\omega t)$  where  $\omega = 2\pi f_m t$ , is sampled with a sampling frequency of  $f_s = 2.5 \times f_m$  and contaminated with a noise. This noisy signal is filtered by the non-recursive SES filter with different smoothing constant  $\alpha$  to yield different smoothed output signals denoted by vector  $Y$ . From Eq. (19), the MSE is being calculated for different values of  $\alpha$  and shown in Fig. 7. The lowest MSE is obtained for  $\alpha = 0.4$ , and this value of  $\alpha$  may be considered as suitable for smoothing.

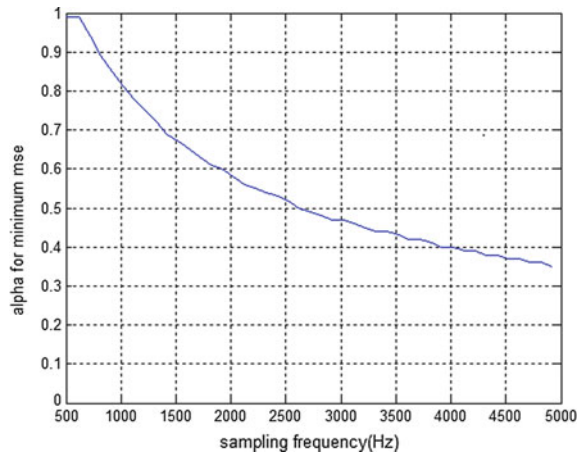
Again with variation in  $f_s$ , the  $\alpha$  for which MSE is minimum also varies as shown in Fig. 8. Since  $f_m$  the value of  $\alpha$  for which MSE will be minimum must be a function of the frequency of the signal.

From the above plots, an empirical relation may be framed between the smoothing constant  $\alpha$  and the sampling frequency ( $f_s$ ) as

**Fig. 7** Variation of MSE with smoothing constant  $\alpha$



**Fig. 8** Variation of  $\alpha$  for minimum MSE with sampling frequency ( $f_s$ )



$$\alpha = 0.9431e^{-0.00022f_s} \tag{20}$$

In Fig. 9, it can be seen Eq. (20) nearly fits (in red) the actual curve between the  $\alpha$  and sampling frequency ( $f_s$ ).

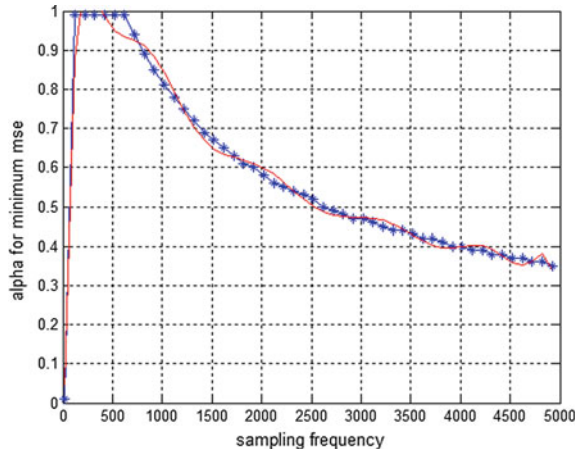
Another observation from all the plots above in this section is that within the span  $0.25 \leq \alpha \leq 0.6$  the MSE remains low and not varies sharply.

### 3.4 Signal-to-Noise Ratio (SNR) Improvement

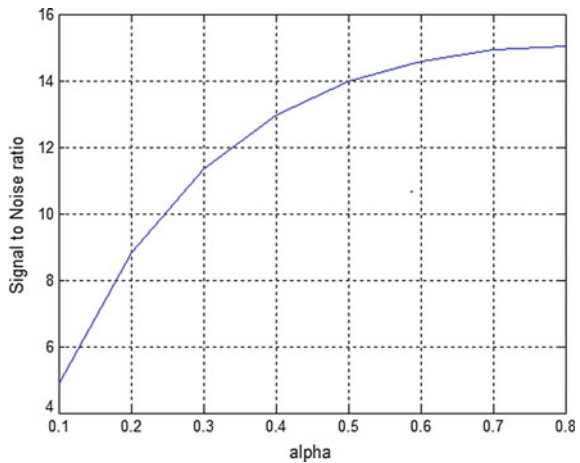
SNR is a measure to compare the level of the desired signal to the level of background noise [29]. A noisy signal is made to pass through the non-recursive SES filter for



**Fig. 9** Comparison between the variation of the actual  $\alpha$  and the empirical  $\alpha$  with  $f_s$



**Fig. 10** Variation of SNR with smoothing constant  $\alpha$



different values of  $\alpha$ , and the SNR of the output of the filter is calculated which is plotted in Fig. 10. It is noticed that with an increase of  $\alpha$ , SNR improves and gets nearly stagnant after  $\alpha = 0.5$ . So, to have an appreciable SNR at the output, it is safe to select  $\alpha \geq 0.5$ .

### 3.5 Identification of the Window Function of FIR Form of SES

The impulse response of any ideal low-pass filter (LPF),  $h_d(n)$  is basically a sinc function: [30]

$$\begin{aligned}
 h_d(n) &= \frac{\omega_c}{\pi} \frac{\sin(\omega_c n)}{\omega_c n} \quad n \neq 0 \\
 &= \frac{\omega_c}{\pi} \quad n = 0
 \end{aligned} \tag{21}$$

Considering the LPF as a causal system, i.e. for  $0 \leq n \leq \infty$ ,  $h_d(n)$  is an infinite impulse response. To have a finite impulse response,  $h_d(n)$  needs to be truncated by some window function  $W(k)$  with a finite length ( $N$ ) which when multiplied with  $h_d(n)$  will generate an impulse response  $h(k)$  with finite length  $N$ , i.e.

$$h(k) = h_d(n)W(k) \tag{22}$$

where  $k = 1, 2, \dots, N$ , and  $N$  is the order of the FIR.

Depending upon the window function,  $W(k)$ , the new LPF with impulse response  $h(k)$  will behave closely like an ideal LPF. The impulse response for SES in its FIR form can be obtained from Eq. (10). To have linear phase, impulse response must be symmetric as below

$$\begin{aligned}
 h(k) &= h(N - k) \\
 &= \alpha(1 - \alpha)^{N-k} \quad \text{for } k = 1, 2, 3, \dots, N
 \end{aligned} \tag{23}$$

Therefore from Eqs. (21), (22) and (23),  $W(k)$  can be derived as

$$\begin{aligned}
 W(k) &= h(k)/h_d(n) = \pi n \alpha(1 - \alpha)^{N-k} / \sin(\omega_c n) \quad \text{for } n \neq 0 \\
 &= \alpha(1 - \alpha)^{N-k} / (\omega_c / \pi) \quad \text{for } n = 0
 \end{aligned} \tag{24}$$

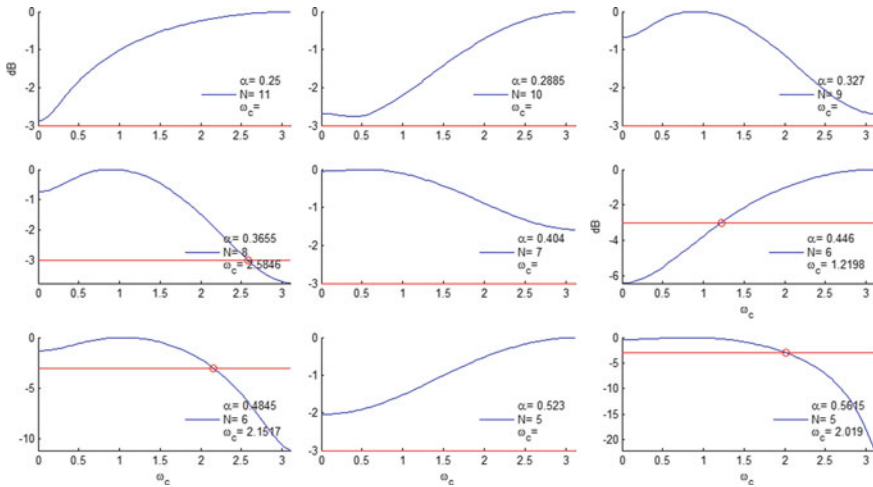
Now since for  $k > N$ ,  $W(k) = 0$  and as  $n$  varies synchronously with  $k$ , it can be argued that though  $n$  ranges from 0 to  $\infty$ , it will be appropriate to take the range of  $n$  to be from 1 to  $N$  like that of  $k$ . So replacing the  $n$  by  $k$  in the above equation,

$$\begin{aligned}
 W(k) &= \pi k \alpha(1 - \alpha)^{N-k} / \sin(\omega_c k) \quad \text{for } k = 1, 2, \dots, N \\
 &= \alpha(1 - \alpha)^N / (\omega_c / \pi) \quad \text{for } k = 0
 \end{aligned} \tag{25}$$

From Eq. (25), it is evident that  $W(k)$  depends on the cut-off frequency ( $\omega_c$ ) and the order of the filter ( $N$ ) for different values of the smoothing constant. Now by Eqs. (6) and (16) the smoothing constant  $\alpha$  depends on both the ( $N$ ) and ( $\omega_c$ ), respectively, as shown in Fig. 6. So, effectively the window function is a function of the smoothing constant  $\alpha$ .

Taking the z-transform of  $W(k)$ , the frequency response  $W(\omega)$  is obtained as follows:

$$W(e^{j\omega}) = \sum_{m=1}^N \frac{\pi m \alpha(1 - \alpha)^{N-m} e^{-j\omega m}}{\sin(\omega_c m)} \tag{26}$$



**Fig. 11** Magnitude response of the window function of the SES for different values of  $\alpha(0.25 \leq \alpha \leq 0.6)$  which yields low MSE. Red curve represents the  $(-3 \text{ dB})$  line

Now the main lobe width of the frequency response of the window function, i.e.  $W(\omega)$  of an ideal LPF, ought to be narrow, and its side lobes should be small which result in a decrease in the transition width between pass-band and stop-band [31]. Though the main lobe width is inversely related to  $N$ , side-lobe amplitude is essentially independent of  $N$ .

The magnitude response of the SES window, i.e.  $|W(\omega)|$ , is plotted in Fig. 11 with respect to the frequency  $\omega$  for different values of  $\alpha$  within the range between 0.25 and 0.6 (considering low MSE as discussed earlier).

From the above figure, it is noted that for  $\alpha = 0.3655, 0.4845$  &  $0.5615$  the window function of the SES filter nearly exhibit the characteristic of the ideal LPF window function. Comparing these three plots, it can be noticed that except for  $\alpha = 0.5615$ , the magnitude responses of other two show a considerable amount of ripple in the pass-band which is not desirable. Moreover, for  $\alpha = 0.5615$ , the transition band of the magnitude response is relatively sharp compared to those of the other two magnitude responses for  $\alpha = 0.3655$  &  $0.4845$ .

## 4 Conclusion

In order to get optimal speed, computational efficiency and magnitude response, stability and phase linearity, the simple exponential smoothing used for forecasting and filtering can be optimally unwrapped to non-recursive form. From Eq. (6), it is confirmed that the value of  $\alpha$  should not be greater than 1 and less than 0, i.e.  $0 < \alpha < 1$ . Again in order to secure the stability of the SES both in recursive and

non-recursive form, the same condition of  $0 < \alpha < 1$  has to be satisfied in Eq. (11). To have an appreciable sensitivity of the SES, Eq. (7) suggests that the threshold value of  $\varepsilon$  should not be very low, and the value of  $\alpha$  should be some intermediate values between 0 and 1. On inspecting the frequency responses of the transfer function of the FIR and IIR form of the SES, the advocacy in favour of this middle range values of  $\alpha$  has also been observed to get an optimised bandwidth and linear phase response. Equations (17) and (18) along with Fig. 6a, b suggest that the value of  $\alpha$  will lie between 0 and 0.82843. It seems from Fig. 6b and Eq. (16) that the choice of higher values of  $\alpha$  for high-frequency signals and lower values of  $\alpha$  for low-frequency signal in case of a non-recursive form of the SES filter is appropriate. But taking the criteria for minimum MSE into consideration, the smoothing constant  $\alpha$  found to decay exponentially with the sampling frequency ( $f_s$ ) and the empirical relation may be framed as in Eq. (20),  $\alpha = 0.9431e^{-0.00022f_s}$ . The span  $0.25 \leq \alpha \leq 0.6$  is suitable to have consistently low MSE over a large frequency range or bandwidth. Within this span, the values of  $\alpha \geq 0.5$  are suitable to have steady and nearly constant high SNR at the SES output as shown in Fig. 10. Though the range of  $0.3 \leq \alpha \leq 0.9$  will make the window function (in time domain) of the SES in FIR form effective to work as a low-pass filter (LPF), the values of  $\alpha = 0.3655, 0.4845$  &  $0.5615$  are appropriate for which the window function (in frequency domain) of the SES behaves as that of a LPF. From an another aspect, it is also clear from Eq. (2) that if we consider the positional importance of  $\alpha$ , i.e. more priority, is given to the most recent input than the previous output, i.e.  $\alpha > 1 - \alpha$ , then also  $\alpha$  must be greater than 0.5. So from these values of  $\alpha$ , the magnitude of 0.5615 is found to satisfy the criteria of the window function of a well-designed LPF. For this particular value of  $\alpha$ , Eq. (6) suggests that the threshold value ( $\varepsilon$ ) may be chosen as 0.02076 and the order ( $N$ ) of the filter will be 5. The value of  $\varepsilon$  is found to be moderate as per the requirement to have good sensitivity ( $S = 79.14$ ), and the order ( $N$ ) is found to be low causing less computational complexity with good speed.

## References

1. Chatfield C, Koehler A, Ord K (2001) A new look at models for exponential smoothing. *Stat* 50 (Part 2):147–159
2. Gardner E (1985) Exponential smoothing: the state of the art. *J Forecast* 1–28
3. Gardner E (2006) Exponential smoothing: the state of the art—Part II. *Int J Forecast* 22:637–666
4. Hyndman R, Koehler A, Ord J, Snyder R (2008) *Forecasting with exponential smoothing: the state space approach*. Springer, Berlin Heidelberg
5. Brown RG (1959) *Statistical forecasting for inventory*. McGraw-Hill, New York
6. Holt CC (1957) *Forecasting trends and seasonals by exponentially*. O.N.R. Memorandum, p 52
7. Ge SY, Zheng CJ, Hou MM (2013) Forecast of bus passenger traffic based on exponential smoothing and trend moving average method. *Appl Mech Mater* 1374:433–435
8. Lai KK, Yu L, Wang S, Huang W (2006) Hybridizing exponential smoothing and neural network for financial time series prediction. In *Lecture notes in computer science* 3994, pp 493–500

9. Chusyairi A, Ramadar NSP, Bagio (2017) The use of exponential smoothing method to predict missing service e-report. In: 2nd international conferences on information technology, information systems and electrical engineering (ICITISEE), Yogyakarta, Indonesia
10. Rahamneh A (2017) Using single and double exponential smoothing for estimating the number of injuries and fatalities resulted from traffic accidents in Jordan (1981–2016). *Middle-East J Sci Res* 25(7):1544–1552
11. Ray S, Ray R, Khondekar M, Ghosh K (2018) Scaling analysis and model estimation of solar Corona index. *Adv Space Res* 61(8):2214–2226
12. Sarkar T, Ray R, Khondekar M, Banerjee S (2015) Periodicity and chaos of solar wind speed: cycle 23. *Astrophys Space Sci* 357(2):128
13. Ray R, Khondekar M, Ghosh K, Bhattacharjee AK (2016) Memory persistency and nonlinearity in daily mean dew point across India. *Theoret Appl Climatol* 124(1–2):119–128
14. Kalekar PS (2004) Time series forecasting using holt-winters
15. Roberts S (1959) Control chart tests based on geometric moving average. *Technometrics* 42(1):97–101
16. Romer M (2016) Smoothing time series: PennState Eberly college of science (Online). Available: <https://onlinecourses.science.psu.edu/stat510/node/70>. Accessed 20 Dec 2016
17. Hunter J (1986) The exponentially weighted moving average. *J Qual Technol* 18(4):203–210
18. Paul SK (2011) Determination of exponential smoothing constant to minimize mean square error and mean absolute deviation. *Global J Res Eng* 11(3)
19. Dalrymple D (1987) Sales forecasting practices: results from a United States Survey. *Int J Forecast* 3(3–4):379–391
20. Winklhofer H, Diamantopoulos A, Witt S (1996) Forecasting practice: a review of the empirical literature and an agent for future research. *Int J Forecast* 12(2):193–221
21. Dwivedi S, Shrivastav P (2015) Survey paper on digital IIR filter using evolutionary algorithms. *Int J Innov Res Comput Commun Eng* 5106–5111
22. Yapar G, Yavuz I, Selamlar HT (2017) Why and how does exponential smoothing fail? An in depth comparison of ATA-simple and simple exponential smoothing. *Turk J Forecast* 01(1):30–39
23. Sahu P, Kumar R (2013) Survey of demand fluctuation of dairy products at Raipur Dugdh Sangh, Chhattisgarh, India for selection of appropriate forecasting. *Uncertain Supply Chain Management. Uncertain Supply Chain Manag* 1(3):133–144
24. Kumar R, Sahu PK (2013) Demand forecasting for sales of milk product (Paneer) in Chhattisgarh. *Int J Inventive Eng Sci (IJIES)* 1(9):10–13
25. Ravinder HV (2013) Determining the optimal values of exponential smoothing constants—does solver really work? *Am J Bus Educ* 1(3):347–360
26. Engineering Statistics Handbook (Online). Available: <http://www.itl.nist.gov/div898/handbook/pmc/section4/pmc43.htm>
27. Booranawong T, Booranawong A (2017) Simple and double exponential smoothing methods with designed input data for forecasting a seasonal time series: in an application for lime prices in Thailand. *Suranaree J Sci Technol* 24(3):301–310
28. Valkenburg MV (2014) Network analysis, 3rd edn. PHI Learning, pp 383–384
29. Das N, Chakraborty M (2017) Performance analysis of FIR and IIR filters for ECG signal denoising based on SNR. In: 2017 Third international conference on research in computational intelligence and communication networks (ICRCICN), Kolkata
30. Huang X, Jing S, Wang Z, Xu Y, Zheng Y (2016) Closed-form FIR filter design based on convolution window spectrum interpolation. *IEEE Trans Sig Process* 64(5):1173–1186
31. Loganathan M, Deepa D (2018) Performance analysis of FIR filter design for secure applications—a review. *Int J Adv Inf Eng Technol* 5(3):11–23

# A Novel Cross-Dimensional Image Processing Technique: Toward a 3D View of a 2D Image



Dipnarayan Das and Sumit Gupta

**Abstract** The domain of image processing has been very popular among researchers since a very long time. It is a well-known fact that there are myriad objectives that can be met with this subject. Through this paper, we are proposing a novel technique whereby we will train our system by processing images on a two-dimensional plane and subsequently produce a resultant three-dimensional view of an image. We have attained this transformation through cross-dimensional image processing by studying, understanding and analyzing the various attributes and properties of an image. This cross-dimensional image processing technique explores the existing digital image processing domain where the inputted 2D image is mapped to produce a 3D view of the image in 3D space.

**Keywords** Image processing · Z-Buffer · Depth of field · LED cube · Surface detection

## 1 Introduction

In the field of computer science, digital image processing is a well-known term, both in the research community as well in the academic circle. It is one of the most crucial categories of digital signal processing, capable enough in dealing with distortions and noise that occur when images are processed. Since images are defined over two dimensions (or perhaps more), digital image processing is generally modeled as multidimensional systems. But post-processing, what we observe is the change in the image attributes, keeping the dimension of the image unaffected.

---

D. Das (✉) · S. Gupta  
Department of Computer Science & Engineering, University Institute of Technology,  
The University of Burdwan, Golapbag (North), Burdwan 713104, West Bengal, India  
e-mail: [dipnarayan.das35@gmail.com](mailto:dipnarayan.das35@gmail.com)

S. Gupta  
e-mail: [sumitsayshi@gmail.com](mailto:sumitsayshi@gmail.com)

© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_8](https://doi.org/10.1007/978-981-13-6783-0_8)

The primary difference between 2D and 3D objects is that 3D objects have three dimensions, viz. height, width, and depth, whereas 2D objects have only two dimensions, viz. height and width. The real-world objects are three-dimensional in nature because they have depth as an extra dimension along with height and width. A drawing on a paper is often 2D, but on analyzing an image from a linear perspective, a 2D image can be made to appear as having three dimensions. So, what is crucial and worth mentioning over here is that even if the linear perspective helps to create a 3D view of a 2D image, yet the resultant image lacks the  $z$ -value or the depth value.

Through this paper, we are proposing a novel cross-dimensional image processing technique for obtaining a 3D view of a 2D image. Our system is based on a hardware layer where the main microcontroller is Arduino Mega. As we know, open source microcontrollers do not have an inbuilt library for image processing, so we have developed a novel indigenous algorithm that detects the color spaces for converting the 2D images into a 3D physical (e.g., 3D hologram) view.

The rest of the paper is organized as follows: Sect. 2 discusses the previous related works. In Sect. 3, we have proposed our methodology in terms of architecture, system components required, working principle, and theories related to the concept of cross-dimensional image processing. Section 4 discusses the implementation part and the results so obtained. Section 5 analyzes our work through a discussion on the merits and demerits of our technique. Section 6 concludes the paper by pointing out some future scope of improvements that can be taken up.

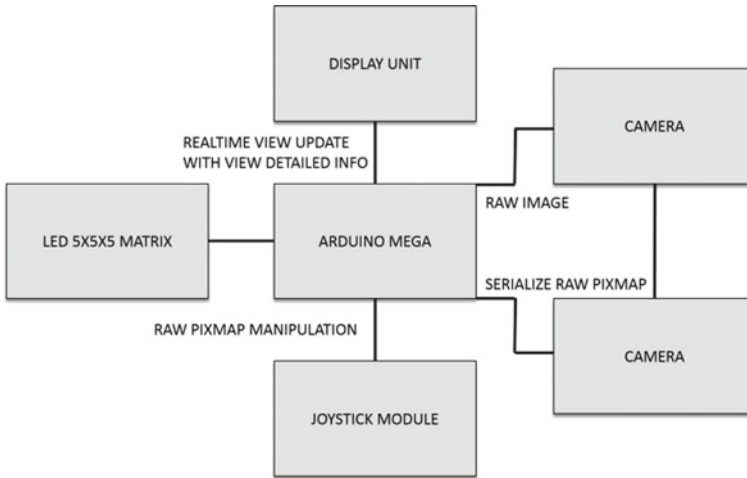
## 2 Previous Related Works

We have studied that in most of the existing systems [1–3], the stereoscopic vision technique is generally used to capture the depth of real-time view, and/or depth value is used for constructing the 3D shape. In papers [4, 5], the authors have proposed 2D-to-3D conversion algorithms that fairly deal with the depth quality and computational complexity but suffer due to the limited size of the training dataset. Some existing systems [6, 7] are in fact manual, cumbersome and lack automation. Moreover, all the current systems utilize the system-defined or inbuilt library functions, but we have developed a new image processing library compatible with low-performance microcontrollers like Arduino.

## 3 Our Proposed Methodology

### 3.1 System Architecture

See Fig. 1.



**Fig. 1** Architecture of our proposed system

### 3.2 System Components

Our proposed system (see Fig. 2) comprises the following components and/or modules:

1. **Arduino Mega:** The Arduino Mega 2560 is a microcontroller board based on the ATmega2560 [8].
2. **TFT Color LCD:** As we know that interfacing a TFT LCD with an Arduino is pretty easy, so we require an Arduino Board and a 2.4 in. TFT Shield to manage the hardware section and an Arduino IDE and the TFT Library for managing the software section. Several libraries exist for operating upon the TFT shield, but we need to identify that particular TFT driver out of multiple available TFT LCD drivers for which we will install our library to serve our purpose [9].
3. **Joystick Module:** The joystick module has 5 pins, viz. VCC, Ground, X, Y, and Key. The thumb stick is analogous in nature and provides more precise readings compared to the other joysticks with buttons and/or switches. Further, the joystick can be pressed down for activating a ‘press to select’ push button [10].
4. **5 × 5 × 5 LED Matrix:** The light emitting diodes (LEDs) have been arranged as 5 × 5 matrices in five different planes. The number of LEDs can be extended as per need.



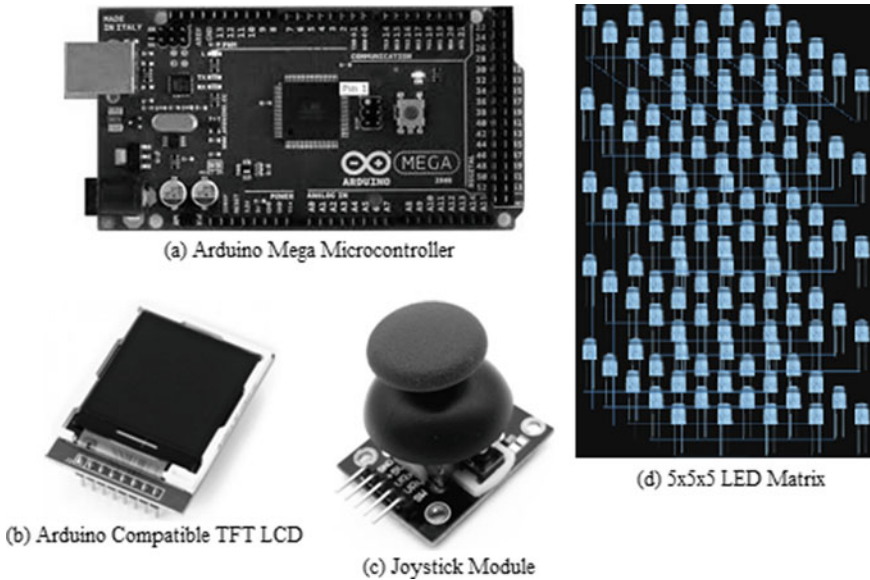


Fig. 2 Components of our proposed system

### 3.3 Working Principle

Our proposed system works as follows: Firstly, the raw image of any object will be entered into the Arduino Mega with raw (or unchanged) format after acquiring the image directly from any standard camera. As an alternative, we can also send serialized data which will be further formatted by the Arduino microcontroller. Further, as we know that Arduino does not have any inbuilt image processing library, so we have created a library termed as color space library, which is compact because microcontrollers have very less amount of built-in storage space. The color space library is able to detect a color region out of the three regions for 16,777,216 types of colors. We have used a specialized technique to detect the colors (that works in the same fashion as the boundary-fill algorithm in graphics designing for detecting the sudden change of values) and the variable rate of three zone colors. Our proposed library has the following features:

1. Two-dimensional image/space to 3D space conversion.
2. Texture detection and object boundary evaluation.
3. Depth recognition.
4. Real-world live streaming from 2D to 3D.
5. Object symmetry detection and rest production.
6. Color space detection.

In our work, in lieu of using a camera module, we are using a mobile phone to capture a real-world image. The pixmap is converted into serialized data and then sent

to the Arduino. The serialized data is also in raw format, and the actual processing will be done on the raw image. The Arduino Mega is attached to a TFT LCD for providing all the information about the raw image and its transmission, viz. transmission rate, image resolution, FPS (frames per second), PPI (pixels per inch), etc. The joystick module is used to simulate the image in TFT LCD so that operations such as cropping, resizing, texturing, and zooming can be performed while converting an image from 2D to 3D mode. The 3D object will be shown from the 2D plane onto the  $5 \times 5 \times 5$  LED cube. The LED cube is in the monochrome mode, and so, all the pictures or video that are captured will be threshold by the microcontroller during the time of processing. For depth analysis, the LED intensity is variable as well as the level of the cube is changed.

### 3.4 Related Theories

1. **Surface Detection:** In graphics designing, we have heard about the visible surface detection methods, namely the object-space method (implemented in the physical coordinate system) and the image-space method (implemented in the screen coordinate system). We are not concerned about the object-space method because in the input panel itself, the 3D coordinate system is already used. But in case of the image-space method, there is one method (depth-buffer method) which can recognize the depth and by that detection, it removes the back faces.
2. **Depth-Buffer Method (or Z-Buffer Method):** This method compares surface depths at each pixel position on the projection plane. An object's depth is generally measured from the view plane along the Z-axis of a viewing system. It requires two buffers, viz. the image buffer and the Z-buffer (or the depth buffer), each having the same resolution as the capturing image [11]. During the process, the image buffer stores the color values of each pixel position, and the Z-buffer stores the depth values for each  $(x, y)$  position. But here the system using a 2D image has all the  $z$ -values for each pixel which helps us to determine the actual back and front faces.
3. **Image-specific attributes:** We observe that for a 2D image, the Z-buffer method will not work to display in 3D space. Now to resolve this 2D image to 3D view objective, our system focuses on few image-specific attributes, viz. image pixel values, rate of change of color values, color region detection, boundary detection, and object symmetry detection.
4. **Depth of Field:** Depth of field is basically the distance between the nearest and the farthest objects in a photograph that appears to be acceptably sharp. It is known that a camera can focus sharply on only one point. Scientifically, it is based on the circle of confusion. Using these properties, every object is considered individually, and also the whole image is converted into a 3D object whose properties are structured into a table and depending upon the user activity, different types of 3D image-cum-night hologram are made in the LED cube. Based on the LED cube configuration (see Fig. 3a), the structured values made from the image are

again re-formatted to fit into the display by image processing operations like cropping, resizing, etc., and according to that values, the microcontroller sends the monochromatic color lookup table to the display controller to show the 3D space information as well as to activate the digital pins of the LED cube.

This method is very effective in case when there are not enough pins in the microcontroller. But this method also suffers from a problem which is that if we connect the common lines vertically and horizontally, the level selector will be connected, and then there will be no possible way to create an object angle in the vertical direction. Only horizontal axis can be used for any inclination then. So to solve this problem, we have to use a matrix connection due to which although the number of pins will increase, but still it will be quite less when compared to a normal setup. In the matrix connection, there will be 25 pins spread horizontally as well as vertically. So, in total 50 pins will be needed to enable cent percent working of the system. In this case, we observe that there still lies the problem that there are not sufficient pins to make a connection. So here we have used 74HC595 shift registers IC (see Fig. 3b) to convert three pins to (8\* number of shift registers) pins [12]. So now we can connect the whole circuit. The shift register has three control pins, namely the latch clock, the shift clock, and the data line. Here, the latch clock is used to enable the rest of the control pins. The shift clock will change the pin sequentially, and on using the data line, we can pass the data or a command to that pin. Hence, when connected, we can extend the pins from 3 to 8. And when there will be two shift registers then there will be multiple 74HC595 connection with 3–16 pins orientation allowed. It is worth noting that the LED cube can be extended into any dimension by adding an extra IC but the dimension should have an odd order for achieving an accurate range approximation.

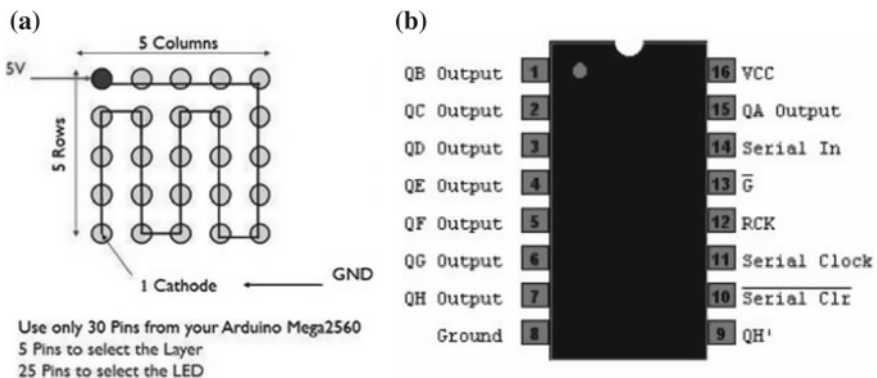
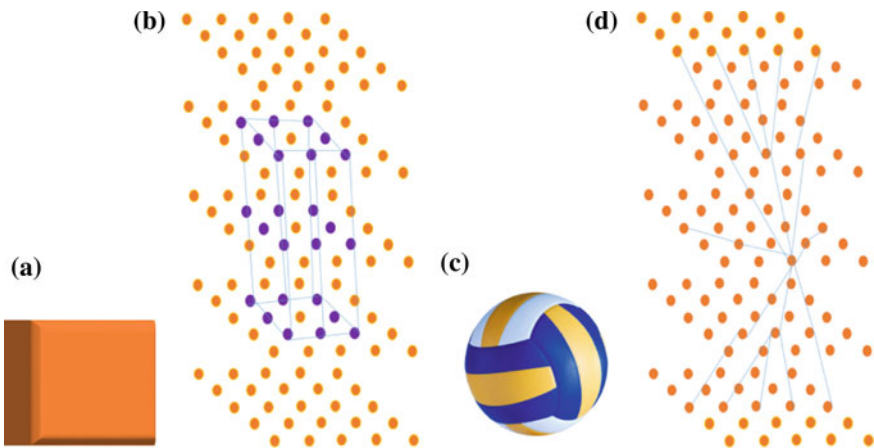


Fig. 3 a LED cube configuration, b 74HC595 IC

### 4 Implementation and Result

We have implemented our system to observe, test, and analyze the performance of our proposed approach. To understand the working of our system, we have initially taken the 2D view of a 3D Rectangle object (see Fig. 4a). Here, our algorithm first checks where the color deviation works and how it occurs. When the system finds that the same color space is becoming darker, then it considers it to be the depth of the object. Many other factors and conditions are also significant for mapping the object into a 3D display and to visualize it like a 3D hologram which we can see in the following image (see Fig. 4b) where the system approximates based on the fact that the object is symmetric in nature.

We have also considered one real-world image of volleyball (see Fig. 4c) downloaded from the Web [13]. This picture consisted of  $201 * 251 = 50,451$  pixels in the 2D layer. Now our physical device consists of only 25 pixels in the 2D layer and there are a total of 125 pixels in the three layers. So when we have fed this image into our device, then the occupied cross-sectional area has a resolution of  $5 * 5 = 25$  pixels which can be moved across the whole image using the joystick. We also have another option to scale the image to 25 pixels in order to get the  $5 * 5$  pixels resolution. Depending upon the shades of the pixels, firstly our system checks for the changes in the color space. Our system analyzes on the percentage of same value pixels, new color pixel, direction of change, and how color is varying in a specific direction. Depending on the result, the data is mapped onto our physical system. Here, the spread or the percentage of the white color is maximum, so the white color is considered to be the background of the image, and the resultant view is depicted in the adjoining figure (see Fig. 4d).



**Fig. 4** a A 2D view of a 3D rectangle object, b corresponding 3D rectangle view in the LED cube, c a real-world volleyball's image, and d corresponding 3D inclined view of volleyball in the LED cube

## 5 Analysis and Discussion

There are various situations where our proposed approach can prove to be useful. For instance, NASA, ISRO, and other such organizations send different types of robots and intelligent devices into space for observing the nature, activity, land texture, etc. of the planets. But the device on reaching the spot transmits a continuous frame of the picture to the space control room, which is further processed to detect if there is any unknown surface contact or not. Our device can show all the land curves and offer different types of other processing in real-time 3D systems.

To watch a 3D movie we use 3D glasses. But there will be no need for any special glass or modification of the 3D movie by using our proposed device. This device can be used in defense for 3D mapping with hidden spy because during that moment, the monochrome LEDs are replaced with the infrared LEDs. Our radar system provides 2D plotting to a good extent for different objects whose position details (altitude, longitude, latitude) are shown in the display. If we split every level of this display with a specific height, then we can easily convert this device into an advanced 3D monitoring system. In ultrasonography, the imaging is in monochromatic mode situated in a 2D display. So, an observer (unlike an expert) cannot understand where the actual problem lies. On using our device, the depth analysis can be done efficiently. In endoscopy, the camera goes into our body channel for taking images of our internal system. But if we have to target any internal organ and operate on it, then we have to perform the approximation of its depth relative to the camera, resulting in the chances of getting hurt but our device can be pretty helpful.

Our proposed system provides a low cost, portable solution for performing 3D imaging for any 2D view. We can easily extend the depth of the display by adding an extra layer and by changing the value of the mapping function. So, our system is very flexible. The limitation of our system is that the LED cube is not very strong in tolerating any physical hazard or heavy jitter. Our system is currently not so effective in performing complicated computer vision processing on dynamically changing images.

## 6 Future Work and Conclusion

In the future, we aim to incorporate the notion of deep learning and integrate a knowledge base with our system so that any specific object can be detected from a real-time video. The resultant 3D view can then be used to observe and analyze frames that possess objects like guns or other important objects. This paper attempts to present a system that uses a novel cross-dimensional image processing technique to produce the 3D view of a 2D image. This approach can surely benefit the society and will help one think and understand differently, control his/her reflex rate and consider each scene from a bigger perspective by getting the 3D view of a 2D image.

## References

1. Varga M, Jadlovsky J (2016) 3D imaging and image processing—literature review. In: 16th scientific conference of young researchers of faculty of electrical engineering and informatics technical university of Košice (SYCR) 2016. Herľany, Slovakia, pp 12–15
2. Mathai S, Mathai PP, Divya KA (2015) Automatic 2D to 3D video and image conversion based on global depth map. In: 2015 IEEE international conference on computational intelligence and computing research (ICCIC). IEEE, Madurai, India, pp 1–4
3. Konrad J, Wang M, Ishwar P (2012) 2D-to-3D image conversion by learning depth from examples. In: 2012 IEEE computer society conference on computer vision and pattern recognition workshops. IEEE, Providence, RI, pp 16–22
4. Angot JL, Huang WJ, Liu KC (2010) A 2D to 3D video and image conversion technique based on a bilateral filter. In: Proceedings of SPIE-IS&T electronic imaging, international society for optics and photonics, vol 7526. SPIE-IS&T, pp 75260D1-10
5. Galabov MN (2015) A real time 2D to 3D image conversion techniques. Int J Eng Sci Innov Technol (IJESIT) 4(1):297–304
6. Cheng CC, Li CT, Chen LG (2010) A 2D-to-3D conversion system using edge information. In: 2010 digest of technical papers international conference on consumer electronics (ICCE). IEEE, Las Vegas, NV, pp 377–378
7. Gonzalez-Huitron V, Ramos-Diaz E, Kravchenko V, Ponomaryov V (2014) 2D to 3D conversion based on disparity map estimation. In: Bayro-Corrochano E, Hancock E (eds) Progress in pattern recognition, image analysis, computer vision, and applications (CIARP 2014), vol 8827. Lecture Notes in Computer Science. Springer, Cham, pp 982–989
8. Arduino Webpage, <https://www.arduino.cc/en/Main/ArduinoBoardMega2560?setlang=en>. Last accessed 15 Aug 2018
9. CircuitDigest Webpage, <https://circuitdigest.com/microcontroller-projects/arduino-tft-lcd-interfacing-tutorial>. Last accessed 15 Aug 2018
10. Arduino Webpage, <https://www.arduino.cc/en/Tutorial/JoyStick>. Last accessed 16 Aug 2018
11. Adams A (1980) The camera. The new Ansel Adams photography series/Book 1. New York Graphic Society, Boston
12. Autodesk Webpage, <https://www.instructables.com/id/3-Arduino-pins-to-24-output-pins/>. Last accessed 15 Aug 2018
13. IndiaMART Webpage, <https://www.indiamart.com/proddetail/sports-rubber-volleyball-17443322497.html>. Last accessed 15 Aug 2018

# Quantum-Inspired Bat Optimization Algorithm for Automatic Clustering of Grayscale Images



Alokananda Dey, Siddhartha Bhattacharyya, Sandip Dey, Jan Platos and Vaclav Snasel

**Abstract** This paper introduces a quantum-inspired framework with the bat optimization algorithm for automatic clustering of image datasets. The aim of this work is to find out the optimal number of clusters from an image dataset on a run. A comparison has been produced between the quantum-inspired bat optimization algorithm and its classical counterpart. As a result, it is seen that the quantum-inspired version outperforms its classical counterpart. Computational experiments have been conducted on four Berkeley image datasets.

**Keywords** Automatic clustering · Metaheuristic algorithm · Quantum computing · Bat optimization algorithm · Statistical test ( $t - test$ )

---

A. Dey (✉) · S. Bhattacharyya  
Department of Computer Application, RCC Institute of Information Technology,  
Kolkata, West Bengal, India  
e-mail: [alokananda\\_22@yahoo.co.in](mailto:alokananda_22@yahoo.co.in)

S. Bhattacharyya  
e-mail: [siddhartha.bhattacharyya@vsb.cz](mailto:siddhartha.bhattacharyya@vsb.cz)

S. Bhattacharyya · J. Platos · V. Snasel  
Faculty of Electrical Engineering and Computer Science,  
VSB Technical University of Ostrava, Ostrava, Czech Republic  
e-mail: [jan.platos@vsb.cz](mailto:jan.platos@vsb.cz)

V. Snasel  
e-mail: [vaclav.snasel@vsb.cz](mailto:vaclav.snasel@vsb.cz)

S. Dey  
Department of Computer Science and Engineering, OmDayal Group of Institutions,  
Howrah, India  
e-mail: [dr.ssandip.dey@gmail.com](mailto:dr.ssandip.dey@gmail.com)

© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_9](https://doi.org/10.1007/978-981-13-6783-0_9)

# 1 Introduction

Clustering [1, 2] can be considered as a process of partitioning a dataset of different objects into a meaningful group of similar objects. Several methods can be used for successful clustering, namely partitioning methods, fuzzy clustering, model-based clustering, hierarchical clustering, density-based clustering, etc. Among these methods, the K-means [3] performs very well. However, it requires an a priori knowledge about the present number of clusters in the dataset. Rather, automatic clustering is intended to discover the optimal number of clusters from a dataset without having the prior knowledge about the number of clusters, which may be useful during the segmentation and classification purpose. Basically, this requirement has dragged the researchers into the field of automatic clustering techniques.

Nature-inspired metaheuristic algorithms are extensively used for solving both simple and complex optimization problems. These types of algorithms are able to provide an optimal or near optimal solution within a minimum time frame. Some well-known nature-inspired metaheuristic algorithms include genetic algorithm (GA) [4], differential evolution (DE) [5], particle swarm optimization (PSO) [6], ant colony optimization (ACO) [7], bat algorithm [8], to name a few. Some nature-inspired metaheuristic algorithms have been successfully applied for automatic clustering techniques [10]. In spite of having such ability to find out a solution from a population, these algorithms sometimes may suffer from premature convergence. In order to overcome this problem, researchers have intended to incorporate the features of quantum computing within the nature-inspired metaheuristic algorithms. Nowadays, quantum-inspired techniques have been a flourishing research area. The field of quantum computation has emerged as a new computing paradigm of engineering insisting on quantum mechanical effects to solve computational problems [11]. Some quantum-inspired techniques merged with metaheuristic algorithms have been developed so far, and the betterment of the quantum version over the classical version of the same algorithms has been presented in terms of convergence, better fitness value and other parameters [11–13].

In this paper, a quantum-inspired metaheuristic algorithm has been introduced for solving the purpose of automatic clustering of image datasets. In this regard, the classical bat algorithm has been chosen as the metaheuristic algorithm. The superiority of the quantum-inspired bat algorithm over its classical counterpart has been presented and judged on the basis of the computed fitness value of the mean, standard deviation, standard error and the computational time. Finally, a statistical superiority test, namely the unpaired  $t$ -test, has been performed and the corresponding  $p$  value has been shown to prove that the outcome is in favor of the quantum-inspired method. During the computation, the fitness has been computed using the cluster validity index—the DB index [14].

This rest of the paper is organized as follows. The overview of the bat algorithm and quantum computing has been summarized in Sects. 2 and 3, respectively. Section 4 describes the functionality of the cluster validity index (DB index) [14]. The basic steps of the proposed quantum-inspired bat algorithm have been demonstrated in



Sect. 5. The experimental results and analysis have been presented in Sect. 6. Finally, the conclusion has been drawn in Sect. 7.

## 2 Overview of Bat Algorithm

A nature-inspired metaheuristic algorithm, viz., the bat algorithm, has been introduced by Yang in the year 2010 [8]. The bat algorithm is a population-based algorithm and can be considered as an efficient algorithm to discover the optimal solution from a complex optimization problem. In order to get the optimal results, bats use echolocation, by which they are able to sense the prey, avoid the obstacles and recognize their roosting crevices in the dark. The procedure of hunting strategies of the bats can be summarized as follows.

Initially, it is considered that  $n$  number of bats are flying randomly in the search space from the position  $P_i$  with the velocity  $V_i$ , frequency  $F_{\min}$ , pulse rate  $r_i$  and loudness  $L_0$  at time stamp  $t$ . At time  $t + 1$ , their next position and velocity should be changed and can be calculated as follows.

$$F_i = F_{\min} + (F_{\min} - F_{\max})\beta \quad (1)$$

$$V_i^{t+1} = V_i^t + (P_i^{t+1} - P_{\text{best}})F_i \quad (2)$$

$$P_i^{t+1} = P_i^t + V_i^{t+1} \quad (3)$$

where  $\beta \in [0, 1]$ . The minimum and the maximum frequency can be considered as  $F_{\min} = 0$  and  $F_{\max} = 2$ . Here,  $P_{\text{best}}$  represents the global best position of the bat from the population of  $n$  bats which has been chosen after evaluating the fitness of all solutions to the population. Here, a uniformly distributed random number between  $[0, 1]$  has been considered as  $\text{rand}_1$ . The new location can be obtained by using the following equation.

$$P_{\text{new}} = P_{\text{old}} + \varepsilon L^t \quad (4)$$

where  $\varepsilon \in [-1, 1]$  and  $L^t = \frac{1}{n} \sum_{i=1}^n L_i^t$  are the average loudness of  $n$  bats at the time stamp  $t$ . The pulse rate ( $r_i$ ) and the loudness ( $L_i$ ) of the bats will be increased and decreased, respectively, by the following equations.

$$r_i^{t+1} = r_i^0(1 - e^{-\gamma t}) \quad (5)$$

$$L_i^{t+1} = \alpha L_i^t \quad (6)$$

where  $\gamma(0 < \gamma < 1)$  and  $\alpha(0 < \alpha < 1)$  are constants and the initial pulse rate can be represented as  $r_i^0 \in [0, 1]$ . Then, the fitness of all the bats will be evaluated and the current  $P_{\text{best}}$  will be selected.

### 3 Overview of Quantum Computing

In case of quantum computing (QC), the quantum bit (*qubit*) can be considered as a unit of information like a classical bit. In case of classical bit, a single binary bit can be able to hold the value 0 or 1 but in contrast a single *qubit* can hold the value 0 and 1 at the same time and that is referred as quantum superposition states [11–13]. The benefit of using *qubit* is the fastest execution of process due to the coherent nature of *qubits* as they are able to execute multiple processes simultaneously. The superposition of the basis states in QC can be described as

$$|\Psi\rangle = \sum_{i=1}^n C_i |\mathcal{V}_i\rangle$$

$$|\Psi\rangle = C_1 |\mathcal{V}_1\rangle + C_2 |\mathcal{V}_2\rangle + \dots + C_n |\mathcal{V}_n\rangle \quad (7)$$

where  $\mathcal{V}_i$  refers to the  $i$ th states and  $C_i \in \mathbb{C}$ . In case of a two-state quantum bit, Eq. (7) can be rewritten as  $|\Psi\rangle = C_1|0\rangle + C_2|1\rangle$ . The state  $|0\rangle$  is known as the “ground state,” and the state  $|1\rangle$  is known as “excited state”;  $C_i$  is a complex number, and the sufficient condition for *quantum orthogonality* is represented as follows.

$$\sum_{i=1}^n C_i^2 = 1 \quad (8)$$

Coherence and decoherence are two striking features of quantum computing. In Eq. (7),  $|\Psi\rangle$  defines a linear superposition of the basis states and can be defined in terms of coherence. A forceful destruction of the above-mentioned linear superposition is referred to as decoherence. A quantum mechanical phenomenon, namely quantum entanglement, establishes a unique connection between the existing quantum systems by which the entangled *qubit* states are able to accelerate so that the computational capability may increase to a very large extent.

Some well-known quantum gates are NOT gate, C-NOT gate, rotation gate, Hadamard gate, Pauli  $-X$  gate, Pauli  $-Y$  gate, Pauli  $-Z$  gate, Toffoli gate, controlled phase shift gate, Fredkin gate, to name a few.

A rotation gate is responsible for updating the  $i$ th qubit value of  $(\alpha_i, \beta_i)$  and can be represented as follows

$$\begin{pmatrix} \alpha'_i \\ \beta'_i \end{pmatrix} = \begin{pmatrix} \cos \theta_i & -\sin \theta_i \\ \sin \theta_i & \cos \theta_i \end{pmatrix} \begin{pmatrix} \alpha_i \\ \beta_i \end{pmatrix} \quad (9)$$

Here, for each *qubit*,  $\theta_i$  represents the rotation angle. For a specific problem, the rotation angle is chosen accordingly.

## 4 Cluster Validity Index

In this paper, the Davies–Bouldin (DB) index [14] has been used as a cluster validity index. The minimum value of *DB* index is the criteria of discovering the optimal number of clusters from a dataset. Let us consider two different clusters are  $C_i$  and  $C_j$  and a cluster similarity measure is  $CSM_{ij}$ . Therefore, the cluster similarity measure  $CSM_{ij}$  can be defined as follows.

$$CSM_{ij} = \frac{DM_i + DM_j}{d_{ij}} \quad (10)$$

Here,  $DM_i$  is the dispersion measure of  $i$ th cluster and can be defined as

$DM_i = \left[ \frac{1}{no_i} \sum_{k=1}^{no_i} \|dp_k^{(i)} - cc_i\|^2 \right]^{\frac{1}{2}}$ , such that  $no_i$  represents the total number of objects and  $cc_i$  represents the  $i$ th cluster center in cluster  $C_i$  and  $\forall dp_k^{(i)} \in C_i$ .

The distance between the clusters  $C_i$  and  $C_j$  can be defined as

$$d_{ij} = \|cc_i - cc_j\|.$$

Now, the Davies–Bouldin (DB) index can be defined as

$$DBI = \frac{1}{n_c} \sum_{i=1}^{n_c} I_i \quad (11)$$

where  $I_i = \max_{j=1, \dots, n_c, i \neq j} (CSM_{ij})$ ,  $i = 1, \dots, n_c$ .

## 5 Basic Steps in Quantum-Inspired Bat Algorithm

### 5.1 Principle of Operation of the QIBAT Algorithm

In this section, a quantum-inspired bat algorithm has been presented. The proposed method has been applied on image dataset to determine the optimal number of cluster from the given dataset.

Initially, the population of  $P_s$  number of bats has been initialized (line 1). Now, in order to get the quantum states the population  $P_s$  has been encoded to produce  $P^\alpha$ . The encoding process has been carried out in the following way.

Initially,  $\theta_{i,j} (0 \leq \theta_{i,j} \leq 2\pi)$  has been generated randomly for each of the solutions to the population. Then, generate  $P_{i,j}^\alpha$ , where  $P_{i,j}^\alpha = \cos \theta_{i,j}$ ,  $i = (1, 2, \dots, Ps)$  and  $j = (1, 2, \dots, Ln)$ . Now, the  $P^\beta$  has been generated from  $P^\alpha$  to achieve quantum orthogonality by using Eq. (8) (lines 2–3). Now, the search process has been guided by the state generated from  $P_{i,j}^\alpha$  and  $P_{i,j}^\beta$  by using Eq. (12). The active cluster centroids have been identified by using the information of the excited states (line 7). Now, the fitness of each individual solution has been computed to find out the fittest solution from the population. The rotation gate has been used [Eq. (9)] to generate another set of populations  $P_{new}^\alpha$  and  $P_{new}^\beta$ . Update the values of  $P^\alpha$  and  $P^\beta$  by their previous values or by the newly generated values  $P_{new}^\alpha$  and  $P_{new}^\beta$  depending upon the fitness value. If the newly generated values can be able to provide lesser fitness, then choose the new one; otherwise choose the old set of values. Using this mechanism, a noble quantum population can be created (lines 5–13). Then, the velocity ( $\mathcal{V}_i$ ), frequency ( $\mathcal{F}_i$ ), pulse rate ( $\mathcal{R}_i$ ) and loudness ( $\mathcal{A}_i$ ) for each quantum solution to the quantum population have been initialized (lines 14–15). In each of the iterations of the main loop, the quantum positions, velocity and fitness have been updated (lines 18–19). Then, improvement of the best quantum solutions has been done (lines 20–22). Next, the new quantum solutions have been evaluated (line 23). The best quantum solutions have been determined and saved (lines 24–27). Finally, determine the best quantum solution (line 28). Finally, the optimal value for the number of cluster has been evaluated from the best quantum solution.

Algorithm: Quantum-Inspired Bat

Input:

Maximum iteration number:  $MxIt$

Population size:  $Ps$

Output:

Optimal number of cluster:  $\mathcal{N}_C$

Optimal fitness value:  $\mathcal{F}_T$

1. Initially, a population  $P$  is created with  $Ps$  number of particles by choosing the normalized value of the intensity of the image, where each particle is considered as a solution to the problem. Let us consider the length of each particle is  $Ln$  where  $Ln$  has been chosen as the square root of the maximum value of the intensity of the input image.
2. Now, each element of  $P$  is encoded to produce  $P^\alpha$  using an encoding scheme by using Eq. (9), described in Sect. 3.
3. Now to establish the feature of *quantum orthogonality*, each particle of  $P^\alpha$  is participating to produce  $P^\beta$  by using Eq. (8).
4. In order to produce  $I_{normal}$ , the normalization has been carried out on the input image to convert the pixel intensity values between the ranges 0 and 1.
5. for  $i = 1$  to  $Ps$  do
6. for  $j = 1$  to  $Ln$  do
7. Create  $\mathcal{N}_{C_i}$  number of unique cluster points from  $P_i$  by satisfying the condition  $P_{ij}^\beta > P_{ij}^\alpha$ , and store the cluster points in  $CP_i$ .

8. Now compute the fitness  $\mathcal{F}_{T_i}$  from  $CP_i$  by using a cluster validity index, named DB index [Eq. (11)].
9. *end for.*
10. *end for.*
11. Find out the best fitness value  $\mathcal{F}_T$  among all  $\mathcal{F}_{T_i}$  along with its corresponding number of cluster point  $\mathcal{N}_C$ .
12. Now apply a small rotation  $\Delta\theta$  on each of the elements of  $P^\beta$ , and again establish *quantum orthogonality* by generating corresponding value of  $P^\alpha$ .
13. Continue the steps 5–12 until satisfying a predefined condition to achieve the best fitness value  $\mathcal{F}_T$ , and preserve its corresponding cluster points in CP, total number of cluster points in nc and also the new value of  $P^\alpha$  and  $P^\beta$  for further use.
14. Initialize the frequency ( $\mathcal{F}_i$ ) and velocity ( $\mathcal{V}_i$ ) for each solution to the population.
15. Loudness ( $A_i$ ) and the pulse rates ( $R_i$ ) have been initialized.
16. *for I = 1 to MxIt do*
17. *for J = 1 to Ps do*
18. Update the frequency, velocity and location of each particle in  $P_J^\beta$  using Eqs. (1), (2), (3).
19. Update each value of  $P_J^\alpha$  using Eq. (8) to ensure *quantum orthogonality*.
20. *if* ( $\text{Rand}(0, 1) > R_{ij}$ ) *then*
21. Select the best solution by computing the fitness from the updated positions, and generate a local solution by using Eq. (4).
22. *end if.*
23. Now, new solution has been evaluated.
24. *if* ( $\text{Rand}(0, 1) < A_i$  &  $\mathcal{F}_{T_{\text{new}}} < \mathcal{F}_{T_{\text{Best}}}$ ) *then*
25. New solution has been accepted.
26. Increase  $R_i$  and decrease  $A_i$  by using Eqs. (5) and (6).
27. *end if.*
28. The best solution has been chosen.
29. *end for.*
30. *end for.*

Finally, the optimal number of cluster  $\mathcal{N}_C$  with its corresponding fitness value  $F_T$  is reported.

## 6 Experiment and Analysis of Result

This paper is intended to propose a quantum-inspired algorithm along with the working methodology of bat algorithm to achieve the optimal number of clusters from a given image dataset. During the experiment, the DB index [14] has been chosen as an objective function of which the minimum value is indicative of the optimal result. The following subsections provide the experimental results obtained.

## 6.1 Cluster Centroid Representation Scheme

The optimal number of cluster from an image dataset has been achieved by identifying the cluster centroids from the given dataset on a run. During the experiments, the cluster centroids have been chosen from a solution by satisfying the following condition.

$$\delta_i = \begin{cases} 1 & \text{if } \beta_i > \alpha_i \\ 0 & \text{Otherwise} \end{cases} \quad (12)$$

where the activation threshold  $\delta_i (1 < i < Ln)$  has been used to identify the number of cluster centroids from a single solution by observing the value of  $\delta_i$ . A value of 1 for  $\delta_i$  represents the activated cluster center.

## 6.2 Experimental Result

See Tables 1, 2 and 3.

## 6.3 Simulation of Work

The experimental implementation has been done on Python environment. The proposed method has been applied on two Berkeley images of size  $80 \times 120$  and other two Berkeley images of size  $120 \times 80$ . A machine on Windows 7 with configuration Dell Intel(R) Core(TM) i3, 2.00 GHz, 4.00 GB RAM has been used as the development environment.

The proof of the superiority of the proposed method over its classical counterpart has been shown on the basis of the mean value of the fitness, standard error, standard deviation and minimum computational time. Moreover, a statistical superiority test, namely the unpaired t-test, has been performed to prove the efficiency of the quantum-inspired bat algorithm over the classical bat algorithm. The optimal number of cluster ( $\mathcal{N}_C$ ), optimal fitness value ( $\mathcal{F}_T$ ) and the optimal execution time ( $\mathcal{T}_E$ ) for both the algorithms QIBAT and CBAT have been reproduced in Table 1. The mean value of the fitness ( $\mathcal{F}_\mu$ ), standard deviation ( $\mathcal{F}_\sigma$ ) and standard error ( $\mathcal{S}_E$ ) for both the algorithms have been demonstrated in Table 2. Finally, the result of unpaired  $t$  – test has been shown in Table 3 which has been done with 95% confidence level. This test basically checks whether the ( $\mathcal{P}$ -value) is less than 0.05 or not. If the value of  $\mathcal{P}$  is less than 0.05, then null hypotheses will be rejected against the alternative hypothesis.

**Table 1** Optimal number of cluster ( $\mathcal{N}_C$ ), optimal fitness value ( $\mathcal{F}_T$ ) and optimal execution time ( $\mathcal{T}_E$ ) in second for QIBAT and CBAT

	#86000			#92059			#94079			#97017		
	$\mathcal{N}_C$	$\mathcal{F}_T$	$\mathcal{T}_E$	$\mathcal{N}_C$	$\mathcal{F}_T$	$\mathcal{T}_E$	$\mathcal{N}_C$	$\mathcal{F}_T$	$\mathcal{T}_E$	$\mathcal{N}_C$	$\mathcal{F}_T$	$\mathcal{T}_E$
QIBAT	5	0.201196	12.67	4	0.191752	22.31	4	0.203669	21.08	5	0.301434	17.15
CBAT	5	0.239731	38.93	4	0.226732	63.77	5	0.241101	49.37	5	0.337640	40.03

**Table 2** Mean ( $\mathcal{F}_\mu$ ), standard deviation ( $\mathcal{F}_\sigma$ ) and standard error ( $S_\mathcal{E}$ ) of fitness for QIBAT and CBAT

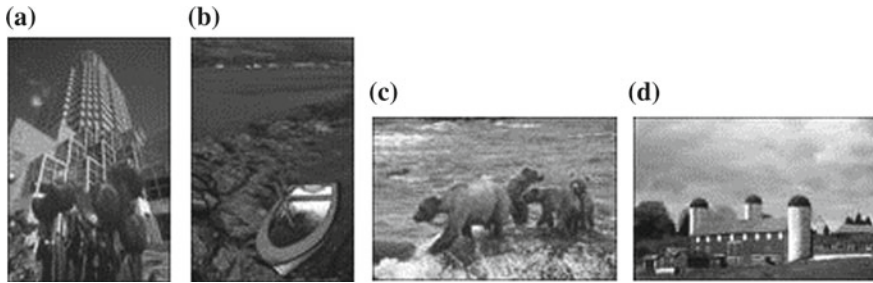
	#86000			#92059			#94079			#97017		
	$\mathcal{F}_\mu$	$\mathcal{F}_\sigma$	$S_\mathcal{E}$	$\mathcal{F}_\mu$	$\mathcal{F}_\sigma$	$S_\mathcal{E}$	$\mathcal{F}_\mu$	$\mathcal{F}_\sigma$	$S_\mathcal{E}$	$\mathcal{F}_\mu$	$\mathcal{F}_\sigma$	$S_\mathcal{E}$
QIBAT	0.217981	0.021745	0.006556	0.221743	0.025413	0.008471	0.228776	0.045434	0.007138	0.315690	0.038735	0.007521
CBAT	0.289976	0.023456	0.007136	0.298700	0.031889	0.009183	0.270012	0.062098	0.010017	0.373454	0.043451	0.009976



**Table 3** Result of unpaired *t* – test (*P*-value) between QIBAT and CBAT

	#86000		#92059		#94079		#97017	
	<i>P</i> -value)	Significant level	<i>P</i> -value)	Significant level	<i>P</i> -value)	Significant level	<i>P</i> -value)	Significant level
QIBAT and CBAT	<0.0001	1	0.0456	2	0.0456	2	< 0.0001	1

Significant level: 1— extremely significant; 2— significant



**Fig. 1** Images used for testing: **a** #86000 ( $80 \times 120$ ), **b** #92059 ( $80 \times 120$ ), **c** #94079 ( $120 \times 80$ ), **d** #97017 ( $20 \times 80$ )

## 6.4 Dataset Used

The experiment has been performed on the following images, and these image datasets have been normalized between (0,1) during the execution of the program (Fig. 1).

## 7 Conclusion

This paper envisages an automatic image clustering technique based on bat algorithm combined with the principles of quantum computing. This proposed procedure has been found superior over its classical counterpart as it is capable to decide the optimal number of clusters automatically from an image dataset with a lesser amount of computation time. So far, the proposed procedure has been applied only to gray-level images with satisfying only one objective at a time.

There remains a scope of research on finding the optimal number of clusters from a true color image, and in the future the quantum-inspired algorithms will open up the door to resolve the multi-objective optimization problem efficiently within a short time frame.

## References

1. Jain A, Dubes R (1988) Algorithms for clustering data. Prentice Hall
2. Jain AK, Murty MN, Flynn PJ (1999) Data clustering: a review. *ACM Comput Surv* 31(3):264–323
3. Yadav J, Sharma M (2013) A review of k-mean algorithm. *Int J Eng Trends Technology (IJETT)* 4(7)
4. Narayanan A, Moore M (1996) Quantum-inspired genetic algorithms. In: *Proceedings IEEE evolutionary computation*, pp 61–66

5. Storn R, Price K (1995) Differential evolution—a simple and efficient heuristic for global optimization over continuous spaces. Technical Report TR-95-012, ICSI
6. Hui-liana F, Xian-lib L (2011) Discrete particle swarm optimization for TSP based on neighborhood. *Appl Res Comput* 2:030
7. Dey S, Bhattacharyya S, Maullik U (2018) Quantum-inspired automatic clustering technique using ant colony optimization algorithm. In *Quantum-inspired intelligent systems for multimedia data analysis*. <https://doi.org/10.4018/978-1-5225-5219-2.ch002>
8. Yang X-S (2010) A new metaheuristic bat-inspired algorithm. *Studies in Computational Intelligence* 284:65–74
9. Alihodzic A, Tuba M (2014) Improved hybridized bat algorithm for global numerical optimization. In: *Proceedings of the 16th IEEE international conference on computer modelling and simulation (UKSim-AMSS '14)*, pp 57–62
10. Das Swagatam, Abraham Ajith, Konar Amit (2008) Automatic clustering using an improved differential evolution algorithm. *IEEE Trans Syst Man Cybern Part A Syst Hum* 38:218–237. <https://doi.org/10.1109/TSMCA.2007.909595>
11. Mcmohan D (2008) *Quantum computing explained*. Wiley Inc, Hoboken, New Jersey
12. Dey S, Saha I, Bhattacharyya S, Maulik U (2014) Multi-level thresholding using quantum inspired meta-heuristics. *Knowl-Based Syst* 67:373–400
13. Dey A, Dey S, Bhattacharyya S, Snasel V, Hassanien AE (2018) Simulated annealing based quantum inspired automatic clustering technique. In: Hassanien A, Tolba M, Elhoseny M, Mostafa M (eds) *The international conference on advanced machine learning technologies and applications (AMLTA2018)*. AMLTA 2018. *Advances in intelligent systems and computing*, vol 723. Springer, Cham
14. Maulik U, Bandyopadhyay S (2002) Performance evaluation of some clustering algorithms and validity indices. *IEEE PAMI* 24:1650–1654

# A Wavelet-Based Blind Digital Image Watermarking using Dynamic LSB Replacement (DLSBR) and Symmetric Key Cryptography



Ananya Ghosh, Subhadeep Koley and Sagnik Acharyya

**Abstract** With advancement of technology and ease of access of digital information and its creation and transference, protection of copyrights for information security is of the prime concern. Digital watermarking gives copyright protection of digital images by impregnating additional information in the original cover image. Thus, watermarking prevents illegal copying by providing a method of enacting the ownership of a redistributed copy. In this paper, we have proposed a wavelet-based dynamic LSB replacement method, which serves the purpose by providing excellent robustness and imperceptibility. Moreover, the watermark is encrypted by a symmetric key cryptographic algorithm to make it secure from further eavesdropping.

**Keywords** Digital watermarking · LWT · LSB replacement · Symmetric Key Cryptography

## 1 Introduction

Enormous usage of cloud computing-based systems and transmission of massive multimedia information over the Internet have created a prominent need for state-of-the-art information security frameworks [1]. Steganography, cryptography, and digital watermarking are the methods developed over the years to ensure secure and seamless transmission of data from transmitter to receiver. With the aid of digital watermarking, it is feasible to protect the ownership of any multimedia content and to defend intruders against any sort of attack [2–4]. Digital watermarking can be

---

A. Ghosh  
National Institute of Technology, Durgapur, India  
e-mail: [ananyag81196@gmail.com](mailto:ananyag81196@gmail.com)

S. Koley (✉) · S. Acharyya  
RCC Institute of Information Technology, Kolkata, India  
e-mail: [subhadeepkoley@gmail.com](mailto:subhadeepkoley@gmail.com)

S. Acharyya  
e-mail: [sagnik.acharyya@gmail.com](mailto:sagnik.acharyya@gmail.com)

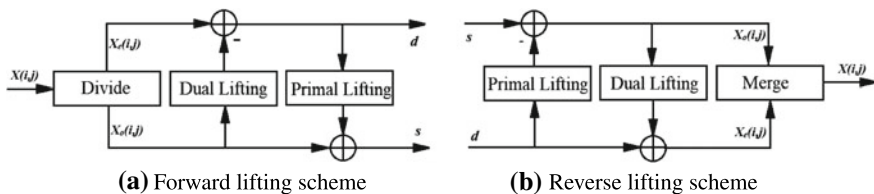
© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_10](https://doi.org/10.1007/978-981-13-6783-0_10)

mostly classified into three domains, namely frequency domain, spatial domain, and hybrid domain. Hybrid and frequency domain watermarks are robust and have high computational complexity, whereas watermarking in the spatial domain has added advantages of simplicity and low computational complexity [2]. However, nowadays techniques in wavelet domain are becoming more popular because of the accurate human visual system (HVS) modelling [5]. Recently, outstanding progress has been made in the area of digital image watermarking. Nayak et al. proposed a visual saliency-based LSB replacement method, which has the advantage of increased imperceptibility but at the cost of less robustness [3]. Basu et al. proposed another visual attention model-based pixel adaptive LSB replacement method, which is good in terms of robustness but has a minimal peak signal-to-noise ratio (PSNR) [2]. Verma et al. proposed a novel watermarking scheme based on dynamic stochastic resonance (DSR) and lifting wavelet transform (LWT). In this technique, the attacked image is subjected to DSR to enhance the watermark coefficients to achieve maximum correlation [6]. Tian et al. gave another HVS-based approach of the unique dual watermarking for simultaneous copyright protection and tamper detection [4]. This technique is very robust for all kinds of single and combined attacks but has higher computational complexity. We have incorporated the integer-to-integer LWT instead of conventional discrete wavelet transform (DWT) to segregate the low-frequency component of the image, and then the binary watermark is infused inside the cover image via dynamic LSB replacement (DLSBR). However, only watermarking is not enough to protect the ownership information inside an image. The watermark logo must be encrypted with some sort of encryption in order to attain added security [7]. Therefore, the original watermark logo is transformed into a random chaotic image using symmetric key cryptography (SKC) based on a private key. The same private key is needed at the time of decryption as used for encryption. The rest of the paper contains some other sections like proposed methodology followed by result analysis and conclusion.

## 2 Proposed Methodology

### 2.1 Lifting Wavelet Transform

The LWT proposed by W. Sweldens is the second-generation wavelet transform, an advancement of DWT [8]. It is faster and requires lesser memory space as compared to the traditional wavelet [9]. It generates integer coefficients for all sub-bands unlike the conventional first-generation scheme, which generates floating point coefficients, which may be truncated at any point of time during further processing [9]. LWT transforms the original image to four sub-bands, namely *LL*, containing low-frequency components, and most of the image energy and *LH*, *HL*, and *HH* comprising of high-frequency detailed information. Therefore, we use *LL* sub-band to embed the watermark data. Lifting scheme in general consists of three steps, namely divide, dual lifting, and primal lifting (Fig. 1).



**Fig. 1** Lifting wavelet transform

1. Divide Operation: The sample image  $X(i, j)$  is divided into two subsets, i.e. even sample set  $X_e(i, j)$  and odd sample set  $X_o(i, j)$  as described in Eqs. 1, and 2.

$$X_e(i, j) = X(i, 2j) \tag{1}$$

$$X_o(i, j) = X(i, 2j + 1) \tag{2}$$

2. Dual Lifting: The odd sample set  $X_o(i, j)$  is estimated from the local neighbourhood even coefficients with a prediction operator  $P$ . The error in odd sample prediction is used to generate the high-frequency coefficients  $h(i, j)$  as described in Eq. 3.

$$h(i, j) = X_o(i, j) - P[X_e(i, j)] \tag{3}$$

From Eq. 4, we can recover the odd sample set as shown below

$$X_o(i, j) = h(i, j) + P[X_e(i, j)] \tag{4}$$

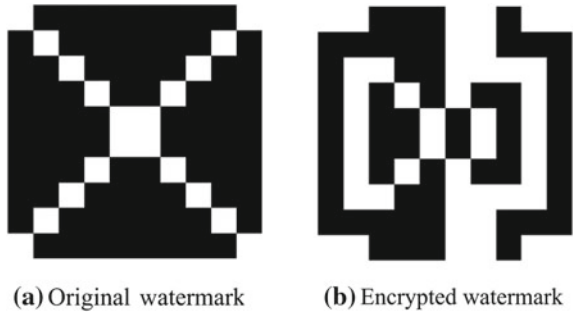
3. Primal Lifting: The low-frequency coefficients  $l(i, j)$  are generated by revising the even sample with the updating value  $U_h(i, j)$  as shown in Eq. 5.

$$l(i, j) = X_e(i, j) + U_h(i, j) \tag{5}$$

## 2.2 Symmetric Key Cryptography

The robustness of the proposed scheme is further enhanced by addition encryption to the watermark through symmetric key cryptography, which covers the inner information from eavesdroppers. The encryption process includes the following steps

**Fig. 2** Watermark before and after encryption



1. Segregate all rows of the binary watermark in separate arrays.
2. Convert the user-given key to 16-bit binary sequence.
3. Perform bit-wise XOR between the 16-bit binary sequence and the segregated rows of the watermark.
4. Perform Step 3 until all the rows are processed.
5. Concatenate all the processed rows to get the encrypted watermark.

The decryption process is exactly the same as the encryption process, as performing *XOR* again between the segregated rows and the binary version of the user-given decryption key will yield the original unprocessed rows. Figure 2 depicts the original as well as the encrypted watermark.

### 2.3 Watermark Encoder and Decoder

Let  $I$  be the original host image of size  $m \times n$ . The image is represented as

$$I = X(i, j) \mid 0 \leq I < m, 0 \leq I < n, X(i, j) \in \{0, 1, 2, \dots, 255\} \quad (6)$$

Also,  $W$  is the original binary logo of size  $p \times q$  and characterized as

$$W = V(i, j) \mid 0 \leq I < p, 0 \leq I < q, V(i, j) \in \{0, 1\} \quad (7)$$

After encryption, the original logo  $W$  is changed to  $W_k$  and represented as

$$W_k = A_{\text{enc}}(W, k) \quad (8)$$

where  $A_{\text{enc}}$  is the encryption operator and  $k$  is the private encryption key. The decrypted watermark  $W_d$  is represented as

$$W_d = A_{\text{dec}}(W_d, k) \quad (9)$$

where  $A_{\text{dec}}$  is the decryption operator.

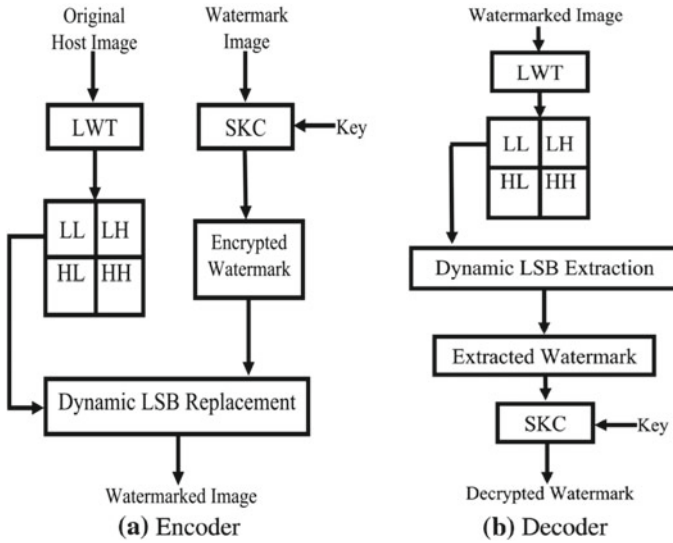


Fig. 3 Watermark encoder and decoder block diagram

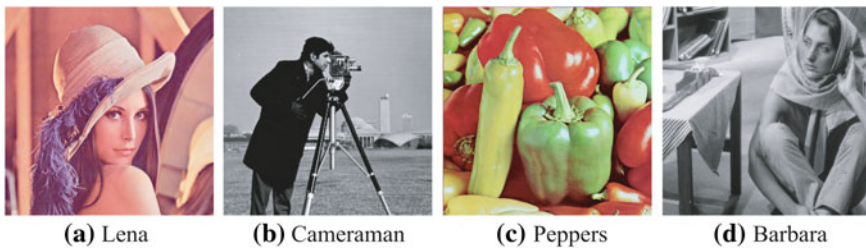
Figure 3 represents the watermark encoding and decoding procedures. Here, the ‘approximation’ or the ‘LL’ component is isolated from the image via LWT. After the watermark encryption process, based on the LL coefficient value each watermark bit is infused inside the LL block of the cover image via a novel method called the *dynamic least significant bit replacement (DLSBR)*. In DLSBR, the LSBs of the LL block coefficients are replaced dynamically with the watermark bits depending upon the LL coefficient values. If the LL coefficient value is less than 100 then last 3 bits, if coefficient value is less than 200 and greater than 100 then last 2 bits, otherwise last 1 bit of the LL coefficient is replaced with the watermark bit. This dynamic substitution has proven to be more robust than conventional single-bit replacement. Being a *blind* watermarking process, the decoder does not need the original host image in the time of decoding. The decoding process is depicted in Fig. 3b. In the decoding process, each watermark bit is extracted from the LL coefficients using dynamic LSB extraction. To decrypt the extracted watermark, one should provide the same private key as used in the time of encryption. However, the robustness against attacks and the visual imperceptibility of the proposed algorithm must be evaluated using some well-established quality metrics, which is the topic of discussion in the next subsection.



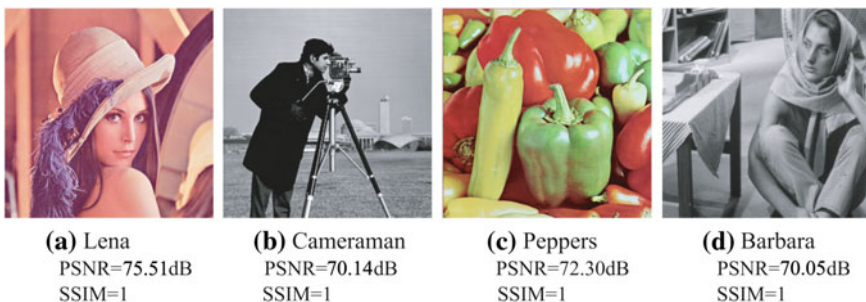
### 3 Result Analysis and Discussion

To test the imperceptibility and the robustness of the scheme proposed, we have taken several standard  $512 \times 512$  grayscale and colour images and one binary watermark of dimension  $10 \times 10$ . Figure 4 depicts four out of the various cover images we have used. Figure 5 shows the corresponding watermarked image along with corresponding *PSNR* and *structural similarity index (SSIM)*.

From Fig. 5, it is evident that the watermark image has no visible hint of the hidden logo and also all of the images possess high *PSNR* and *SSIM* value. To prove the imperceptibility of the proposed scheme, Table 1 has been presented with some renowned quality metrics [2]. Metric values in S. No. 1–2, 5–9, 12–15 depict premier quality as higher metric value denotes better imperceptibility [2]. On the other hand, metric values in S. No. 3–4, 10–11, also illustrate that the hidden logo is completely invisible as lower metric value denotes higher imperceptibility [2]. Hence, the proposed algorithm is proven to be beneficial in every test of imperceptibility. ‘Robustness’ deals with the proficiency of the watermarking algorithm in correctly detecting the watermark in the cover image after common geometric and signal processing attacks [1]. Figure 6 is presented with some attacked watermarked images along with the corresponding retrieved watermarks. It is evident from Fig. 6 that the watermarked image can sustain the watermark even after strong signal processing and geometric impairments.



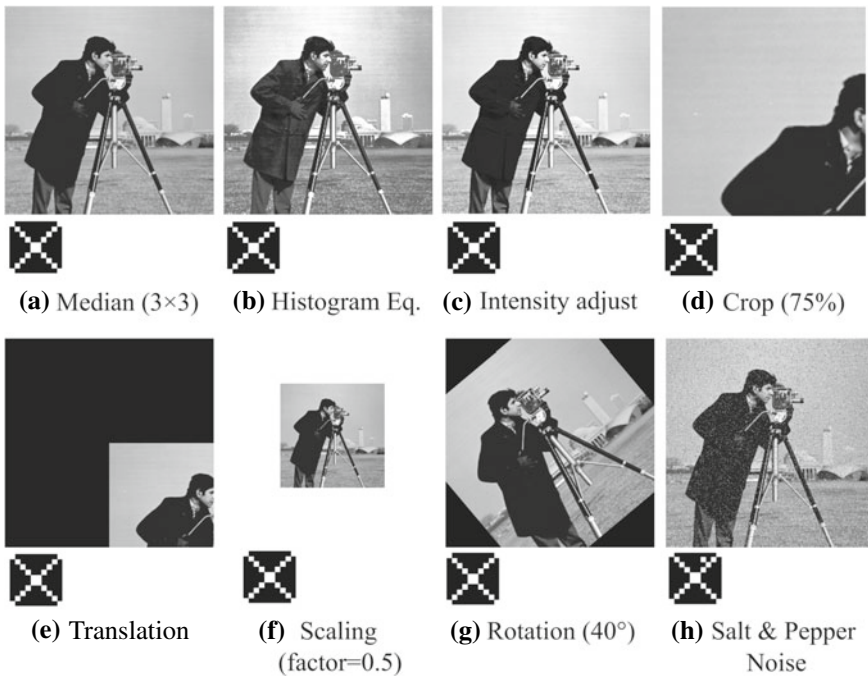
**Fig. 4** Original cover images



**Fig. 5** Watermarked images and their *PSNR* and *SSIM* values

**Table 1** Result of various imperceptibility metrics on *Lena* image

S. No.	Metrics	Value
1	Peak signal-to-noise ratio	75.51 dB
2	Structural similarity index	1.0000
3	Average difference	0
4	Mean square error	0.000182
5	Structural content	1.0000
6	Signal-to-noise ratio	65.6533 dB
7	Image fidelity criterion	7.438
8	Universal image quality index	1.0000
9	Normalized cross-correlation	1.0000
10	Normalized absolute error	0.000002
11	Laplacian mean square error	0.000019
12	Weighted signal-to-noise ratio	92.3797 dB
13	Visual information fidelity	0.9992
14	Pixel-based visual information fidelity	0.9999
15	Multi-scale structural similarity index	1.0000



**Fig. 6** Attacked watermarked images and corresponding extracted watermarks

**Table 2** Imperceptibility comparison with various state-of-the-art methods

S. No.	Method	PSNR (dB)	SSIM
1	<b>Proposed scheme</b>	<b>75.51</b>	<b>1.0000</b>
2	Visual attention model-based LSB replacement [2]	42.27	0.9490
3	Visual saliency-based watermarking approach [3]	54.00	0.9670
4	DCT-based compression sensing watermarking [10]	47.19	0.9958
5	Bit pairs matching and cryptographic watermarking [11]	52.99	0.9854
6	Image region detector-based spatial watermarking [12]	53.35	0.9930

Bold signifies the highest metric value among all the competing method and also the result of the proposed method

*Normalized cross-correlation (NCC)* has been taken as the primary metric to test our approach. In most of the cases, we got the NCC value of 1 or almost 1. Other metric values like *SSIM*, *MSE*, *PSNR* are also in support of the superior ruggedness of the algorithm. To establish the superiority of the proposed framework, a competitive analogy with various state-of-the-art schemes has been drawn in Table 2. It can be observed that, among all models, the projected approach gives highest PSNR as well as SSIM.

## 4 Conclusion

A wavelet domain, blind watermarking, and secure watermarking framework have been presented in this paper. This scheme impregnates the watermark by substituting the LSBs of the *LL* block coefficients with the watermark bits. Furthermore, the original watermark logo is encrypted with SKC for added security. Experimental results establish the fact that the projected approach has excellent imperceptibility and is robust against most forms of signal processing impairments. In near future, we aim to make the algorithm faster for effective hardware implementation.

## References

1. Cox IJ, Kilian J, Leighton T, Shamoon T (1997) Secure spread spectrum watermarking for multimedia. *IEEE Trans Image Process* 6(12):1673–1687
2. Basu A, Sarkar SK (2013) On the implementation of robust copyright protection scheme using visual attention model. *Inf Secur J Glob Perspect* 22(1):10–20

3. Nayak MR, Tudu B, Basu A, Sarkar SK (2015) On the implementation of a secured digital watermarking frame work. *Inf Secur J Glob Perspect* 24(1):1–9
4. Tian L, Zheng N, Xue J, Li C, Wang X (2011) An integrated visual saliency-based watermarking approach for synchronous image authentication and copyright protection. *Signal Process Image Commun* 26(8–9):427–437
5. Reddy AA, Chatterjii BN (2005) A new wavelet based logo watermarking scheme. *Pattern RecognLett* 26(7):1019–1027
6. Verma VS, Jha RK (2017) LWT-DSR based new robust framework for watermark extraction under intentional attack conditions. *J Franklin Inst, Elsevier* 354(14):6422–6449
7. Menezes AJ, van Oorschot PC, Vanstone SA (1996) *Handbook of applied cryptography*. CRC Press, Boca Raton, FL
8. Sweldens W (1998) The lifting scheme: a construction of second generation wavelets. *Soc Ind Appl Math* 29(2):511–546
9. Lee S, Yoo CD, Kalker T (2007) Reversible image watermarking based on integer-to-integer wavelet transform. *IEEE Trans Inf Forensics Secur* 2(3):321–330
10. Thanki R, Dwivedi V, Borisagar K (2017) A hybrid watermarking scheme with CS theory for security of multimedia data. *J King Saud Univ Comput Inf Sci* 1–10
11. Bal SN, Nayak MR, Sarkar SK (2018) On the implementation of a secured watermarking mechanism based on cryptography and bit pairs matching. *J King Saud Univ Comput Inf Sci* 1–10
12. Abraham J, Paul V (2017) An imperceptible spatial domain color image watermarking scheme. *J King Saud Univ Comput Inf Sci* 1–10

# Fuzzy-Based Classification for Fusion of Palmprint and Iris Biometric Traits



Akram Alsubari, Preeti Lonkhande and R. J. Ramteke

**Abstract** The fusion of palmprint and iris biometric traits is implemented in this paper. The region of interest (ROI) of palm is detected and segmented based on the valley detection algorithm and the ROI of iris is extracted based on the Neighbor-Pixels Value Algorithm (NPVA). The statistical-local binary pattern (SLBP) is used for extracting the local features from the iris and palm. For enhancing the palm features, the combination of discrete cosine transform (DCT) and histogram of oriented gradient (HOG) are applied. The Gabor–Zernike moment is used to extract the iris features. This experimentation was carried out in the identification system. The fuzzy Gaussian membership function was used as classification in the matching stage for the fusion system of palm and iris. The CASIA datasets of palm and iris were used in this research work. The proposed system accuracy was found to be satisfactory.

**Keywords** Palmprint · Iris · Valley detection · SLBP · HOG · DCT · Zernike moment · Fuzzy

## 1 Introduction

The pattern recognition is to classify the data based on the knowledge which is already gained in the cognition process [1]. The pattern recognition is a process to understand the object/pattern to which class it belongs. Those patterns/objects can be 1D, e.g., signal or 2D, e.g., image. The biometric is one application of the pattern recognition. The biometric system is to analyze and measure the characters of human body. There are three types of biometrics physical, behavior, and cognitive. The physical biometrics are the bodily parts of the human body suchlike face, fingerprint, and palm. The behavior biometrics are the activities of the human body suchlike gait,

---

A. Alsubari (✉) · R. J. Ramteke  
School of Computer Sciences, KBC North Maharashtra University, Jalgaon, India  
e-mail: [akram.alsubari87@gmail.com](mailto:akram.alsubari87@gmail.com)

P. Lonkhande  
MGAH University, Wardha, India

© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_11](https://doi.org/10.1007/978-981-13-6783-0_11)

keystroke, and voice. The cognitive biometric means every nervous tissue of human has a different response to the signal likes electrocardiogram (ECG), electromyogram (EMG), electroencephalogram (EEG), and electrodermal Response (EDR). In this paper, two physical biometrics were experimented (palmprint and iris).

The iris looks like a ring inside the eye, and it controls the intensity of light which is entering through the pupil. So, the radius of the iris is increasing and decreasing according to the illumination of light. There are more features on the iris such as tiny crypts, nevi anterior border layer collarette, and texture features [2]. The right iris has different features than the left iris for the same human. Also, the iris is one of the touchless biometrics. The palmprint is another trait of physical biometrics which is located between the root-finger and wrist of the hand. Wei and David [3] were the first authors who describe the palm as one type of biometric traits in August 1998.

The paper organization as the next section describes the previous research regarding to the fusion of palm and iris. Section 3 describes the methodologies of the fusion system based on iris and palm. Section 4 discusses the implementation and the results. At the end of the paper will be the conclusion and the references.

## 2 Literature Survey

Jai et al. [4] have presented a method to extract the palm features and it is named as the histogram of oriented lines (HOL). Indeed, the HOL method is derived from the HOG technique [5] which was presented by N. Dalal and B. Triggs in 2005 for the human detection. The HOL is not sensitive to light changing. Also, the HOL is used to extract the palm lines along with the orientation features. Wei Jai et al. have presented Euclidean distance in the verification and identification system for the palmprint trait. In the verification mode, the error rates were computed, so 0.31 and 0.64% are obtained as error rate in the PolyU M\_B and PolyU II databases, respectively. In the identification system, the recognition rates were calculated based on the matching samples from the testing templates. Thus, 99.97 and 100% were the recognition rates in the PolyU II and PolyU M\_B databases, respectively.

Fathi et al. [6] have presented the face recognition system which is extracted the global and local features. In the global features, the Zernike moment is applied on the output of the Gabor filter. For extracting the local features of face image, the histogram of oriented gradient (HOG) is used to extract the face features. Thus, the global and local features were combined for producing a feature vector of the face image. The nearest neighbor classification is used for the matching process, and this classification is based on the Euclidean distance. The accuracies were 97.8, 98, and 97.7% in Yale, ORL, and AR databases, respectively.

Mahesh and Raj [7] were proposed the magnitudes of the Zernike moments for generating the feature vector of the face images. They have experimented in 2D and 3D face images. The features were selected on the basis of the order in the Zernike moments, so the length of feature vector was varied from 16 to 42 features. The ORL dataset was into 60% for training 40% for testing. The authors have used the neural

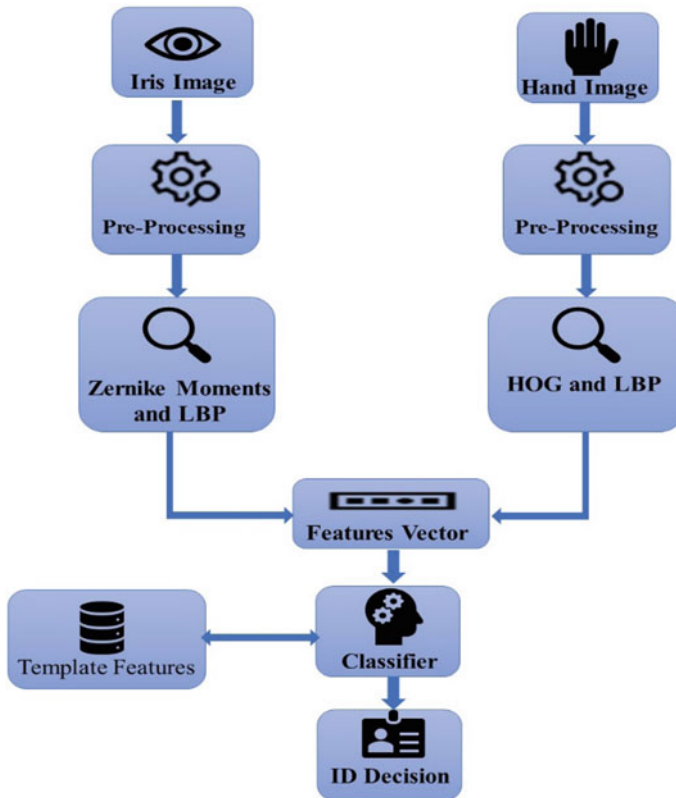
network tools in MATLAB (nprtool) for the matching process. In the ORL database, the highest recognition rate is 99.50% with 36 features by using the radial basis function neural network (RBFNN). In 3D face database, the best achievement was 99.71% with 42 features by using multilayer perceptron neural network (MLPNN).

Gayathri and Ramamoorthy [8] were extracted the iris and palmprint features by applying Gabor filter with four orientations and three spatial frequencies. They have fused the palm and iris after the features were described by Gabor filter. The output of Gabor filter for the palm and iris images was integrated in single matrix by using the second level of wavelet transform. In other words, the second wavelet transform was applied on the palm and iris images, then it was combined in one matrix for generating the feature vector. The fusion system accuracy was 99.2% by using the KNN classifier.

Kihal et al. [9] were presented a multimodal of iris and palmprint by extracting the global features using the wavelet transforms. Thus, the iris features were extracted by using the Haar wavelet transform at four levels. For extracting the palmprint features, the Daubechies wavelet transform was applied at four levels as well. The Hamming distance is utilized in the matching phase for the iris modal and palmprint modal separately. The researchers have proposed three different levels of fusion for the iris and palmprint biometric traits features level, score level, and decision level. The authors have compared the results between the three levels of fusion. In case at the feature level, the feature vectors of iris and palm were combined in single vector for generating a feature vector of the iris and palmprint. In the score-level fusion, the feature vectors of iris and palmprint were evaluated separately/individually and based on each feature vector score, the decision is made by using a weighted sum rule. Finally, at the decision level, the fusion was based on the error rate from the feature vectors of iris and palmprint. Khachane et al. [10] were presented the fuzzy rule-based classification for identifying the normal sperm from the abnormal.

### 3 Methodologies

In this paper, the fusion of palmprint and iris biometric traits was integrated at the feature level as shown in Fig. 1. The palm is segmented from the rest part of hand by using the proposed algorithm from the previous work of authors [11]. The iris was localized by using the Neighbor-Pixels Value Algorithm (NPVA) [12]. The texture features of iris and palm were extracted by calculating the mean and standard deviation of the local binary pattern (LBP) filter. The DCT and HOG were applied together to extract palm texture features. The Gabor filter with eight orientations and five scales were applied on the ROI of iris for generating forty sub-images. Then, the Zernike Moment (ZM) is applied on each sub-image to generate feature vector of the iris. The Gaussian fuzzy membership function is used as classifier in the identification system.



**Fig. 1** Proposed system of palmprint and iris fusion system

### 3.1 Preprocessing

The region of interest (ROI) of palm was detected and segmented by applying the valley detection algorithm [13] and thresholding segmentation. For extracting the iris region, the Neighbor-Pixels Value Algorithm (NPVA) was proposed by the author in [12] for calculating the radius and center values of the region of iris. After the iris detection, the iris region is converted into rectangular form by using the rubber sheet normalization [14]. The dimensionality of iris images was resized to  $64 \times 384$  and palm images were resized to  $64 \times 64$ .

Further, three of enhancement techniques were applied on the ROI of the palm: contrast-limited adaptive histogram equalization (CLAHE), Sobel filter, and Gaussian filter mask. The contrast-limited adaptive histogram equalization (CLAHE) is used to enhance the intensity contrast of the ROI of palm and iris [15]. The histogram equalization (HE) is applied on the ROI of iris for the enhancing the contrast of the iris pixels. The histogram equalization is distributed the intensity values of pixels in



the range between 0 and 255, and it is powerful technique in the image processing techniques.

### 3.2 Feature Extraction

The concept of feature extractor in pattern recognition in the field of image processing is converting the matrix (image) into a set of mathematical parameters which will be in the form of vector (feature vector) or template features. Those features can be edges, lines in palm, minutia in fingerprint, neighbor pixels-based (local binary pattern features), texture features, or shape-based features, etc. In this experiment, there are three methods were used to extract the features of the palm and iris.

#### 3.2.1 SLBP

The local binary pattern (LBP) was introduced in 2002 by Ojala [16]. In this research, the LBP is used as filter for the palm and iris images. The LBP is applied in each pixel of an image except the first and last row and column. The LBP filter is to compare the pixel with the eight-neighbor values. If the pixel intensity value for the neighbor is greater than or equal to the central pixel, then it will be assigned as one; otherwise, it will be assigned as zero. From the LBP filter, the output of the previous process will be eight binary numbers, so those numbers will convert to decimal number, and it will be replaced with the center pixel as shown in Fig. 2 [17].

For obtaining the SLBP, the mean and standard deviation were calculated from the LBP matrix. Those values were composed for creating the feature vector of SLBP. SLBP method is the common feature extraction techniques for palm and iris.

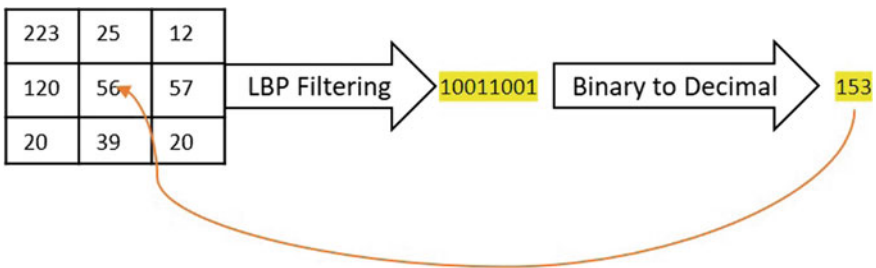


Fig. 2 LBP filtering

### 3.2.2 Composite of DCT and HOG

Discrete cosine transform (DCT) is used for conversion several signals/pixels into a frequency domain. In addition to transform an image, it can be extracted the edges from the palm image. When the palm has different features (principal lines and wrinkles etc.) of the palm, the composite of DCT and HOG was proposed for extracting the features of palm.

Based on following equations, discrete cosine transform (DCT) is computed:

$$f(a) = w(a) \sum_{n=1}^N y(n) \cos\left(\frac{\pi}{2n}(2n-1)(a-1)\right) \quad (1)$$

$$w(a) = \begin{cases} \frac{1}{\sqrt{N}}, & a = 1 \\ \sqrt{\frac{2}{N}}, & 2 \leq a \leq N \end{cases} \quad (2)$$

where  $a = 1, 2, 3, \dots, N$ , and  $N$  is the length of  $y$ .  $a$  and  $y$  are the same size. The  $a$  and  $N$  are indexed from  $a = 1, N = 1$  because in MATLAB Vectors runs from 1. The output of DCT is having negative and positive values, and thus the negative values will be converted into positive values.

The histogram of oriented gradient (HOG) is introduced and proposed on 2005 by Dalal and Triggs [5] for the purpose of human detection. In this experiment, the HOG is used to extract the palm texture feature. Thus, the combination of DCT and HOG was proposed to extract the features of palm. The DCT is applied to analyze and describe the edges from the palm image. So, the discrete cosine transform was applied in the palm region images. Then, the HOG is used on the DCT output.

The following steps are describing the HOG algorithm:

**Step. 1.** The region of palm image is divided into cells and the four groups of cells are considered as blocks, where the cell is  $8 \times 8$  pixels and the size of block is  $2 \times 2$  cells which means  $16 \times 16$  pixels.

**Step. 2.** 50% overlapping was assigned from the next block as shown in Fig. 3. So, the number of blocks for each palm image is  $7 \times 7 = 49$  Blocks.

**Step. 3.** Nine directions were selected from the range between  $0^\circ$  and  $180^\circ$ . The orientation and gradient magnitude were computed by using the following equations:

$$dx = I(x+1, y) - I(x-1, y) \quad (3)$$

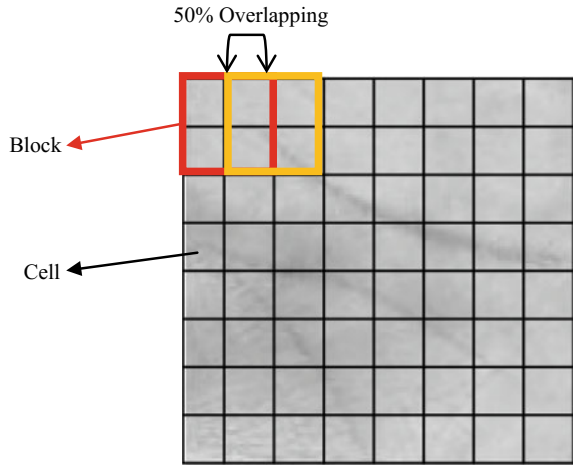
$$dy = I(x, y+1) - I(x, y-1) \quad (4)$$

$$m(x, y) = \sqrt{dx^2 + dy^2} \quad (5)$$

**Step. 4.** The histograms of each cell with respect to nine bins direction were computed.

$$\text{Total no. of features} = \text{NB} \times \text{CB} \times P \quad (6)$$

**Fig. 3** Blocks and cells in the palm image



where NB is the total of blocks number in the palm image. CB is assigned as four because each block in palm image is divided into four cells.  $P$  is the bins orientation which is assigned by 9. From Eq. 9, the total number of HOG features is  $49 \times 4 \times 9 = 1764$ .

### 3.2.3 Gabor–Zernike Moments

The idea behind this technique is to combine the Gabor filter with the Zernike moments for extracting the texture features of an iris image. The combination of Zernike moments and Gabor filter was achieved a satisfactory result in the face and iris traits. In this implementation, the Gabor filter [18] was used on the iris image with respect to eight orientations and five scales, thus forty sub-images were generated from the Gabor filter. Therefore, the Zernike moments [19] technique was applied to each sub-image of iris with different orders and repetitions. Four features were selected from each sub-image. The feature vector length of this technique is 160 features. The Zernike moments of order  $n$  with repetition  $m$  are obtained by the following equations:

$$Z_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}(x, y), \quad x^2 + y^2 \leq 1 \tag{7}$$

where Zernike moment  $Z_{nm}$  on rotated image has the same magnitudes. Therefore,  $|Z_{nm}|$  is the rotation invariance of features.  $V_{nm}(x, y)$  is a Zernike polynomial on a unit circle  $x^2 + y^2 \leq 1$ .

$$V_{nm(x,y)} = V_{nm}(\rho, \theta) = R_{nm}(\rho) \exp(jm \theta) \tag{8}$$

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s! \left(\frac{n-|m|}{2} - s\right)! \left(\frac{n-|m|}{2} - s\right)!} \rho^{n-2s} \quad (9)$$

where  $n$  is a nonnegative integer which represents the order of Zernike moment and  $m$  represents the repetition. Satisfying the condition  $n - m$  is Even and  $|m| \leq n$ . Also  $\rho = \sqrt{x^2 + y^2}$  and  $\theta = \tan^{-1} \frac{y}{x}$ .

### 3.3 Fusion

The fusion is to integrate the palm and iris biometric traits. In this experiment, the fusion was done in the feature level. The feature vectors of iris and palm were composited to generate a single vector. Before composite the vectors, the max normalization was adopted for making the values of the vector in the same range [0–1]. The max normalization steps as follows:

**Step. 1.** It is to convert the feature vector values into integer values.

**Step. 2.** Find out the maximum value of the feature vector of iris and palm.

**Step. 3.** Divide the feature vector of iris and palm with the maximum values of iris and palm, respectively.

**Step. 4.** Composite the feature vector of iris and palm in single vector.

### 3.4 Fuzzy-Based Classification

The feature vectors are classified by using the Gaussian fuzzy membership function [20, 21], and this function gives output in the range between 0 and 1. The following steps are the algorithm of fuzzy classification based on the Gaussian fuzzy membership function:

**Step. 1.** Enroll the all feature vectors of the biometric traits as matrix.

**Step. 2.** Divide the feature matrix into training and testing. Where, 50% of the feature matrix was assigned for training model and the rest for testing the system accuracy.

**Step. 3.** Calculate the average and standard deviation for each training sample of the specific label separately.

$$\mu(H) = \frac{\sum H_i}{n} \quad (10)$$

$$\sigma(H) = \sqrt{\frac{\sum |H_i - \mu|^2}{n}} \quad (11)$$

where  $H$  is a features vector and the  $n$  is the length of the  $H$  vector. The  $\mu$  is the mean or the average of the specific features.  $\sigma$  is the standard deviation.

**Step. 4.** Apply the Gaussian fuzzy membership function to the testing samples based on the mean and standard deviation of the training samples.

$$F(H_w) = e^{-\frac{(H_i - \mu)^2}{2\sigma^2}} \tag{12}$$

where  $e$  is a mathematical constant and the approximate value is 2.7183.  $H_w$  is the Gaussian fuzzy membership values which are in the range between zero and one.  $H_i$  is the testing vector.  $\mu$  and  $\sigma$  are the mean and the standard deviation, respectively. The average of each observation ( $H_w$ ) of the testing vector is stored in another matrix for computing the accuracy.

### 4 Implementation and Results

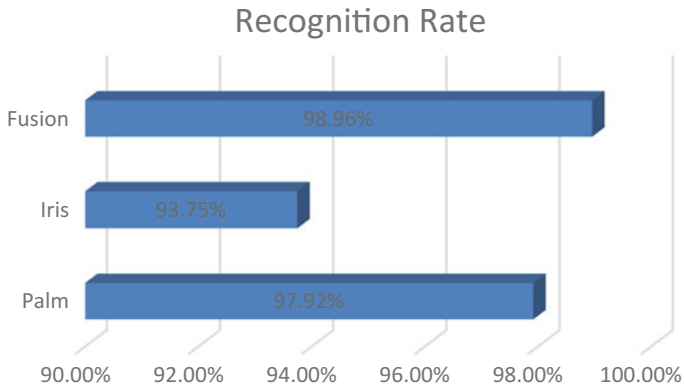
The methodologies and algorithms of this paper were developed by using MATLAB software. The CASIA-iris and CASIA-palm databases [22] were used in the implementation. In this experiment, 24 subjects with eight samples  $8 \times 24 = 192$  subjects were selected randomly from each database of CASIA-Palmprint Database and CASIA-Iris-Interval Database. A total number of selected samples are  $8 \times 24 + 8 \times 24 = 384$  images.

The feature vectors of the iris and palm biometric traits were classified based on the Gaussian fuzzy membership function. The feature vectors were divided into 50% as training and 50% as testing. The mean and standard deviation were computed for the training features and it stored as two vectors for each person. The testing vector is evaluated by using the Gaussian fuzzy membership function with respect to the two vectors which were stored (mean and standard deviation vectors).

From Table 1, the accuracies in unimodal biometrics of palm and iris were 97.92 and 93.75%, respectively. The average accuracy of unimodal biometric is  $((97.92 + 93.75)/2) = 95.83\%$ . So, the recognition rate was increased in the multimodal biometric by 3.125% (Fig. 4).

**Table 1** Identification accuracies based on fuzzy classifier

Classification method	Palm	Iris	Fusion
Recognition rate (%)	97.92	93.75	98.96
Time in second (s)	(5.0109)	(0.9434)	(7.47215)



**Fig. 4** Results comparison

## 5 Conclusion

The comparison between the unimodal and multimodal biometrics system based on fuzzy classifier was developed by using the MATLAB software. The ROI of palm and iris was detected and extracted based on the valley detection algorithm and NVPA, respectively. SLBP was used for extracting the local features of iris and palm. For enhancing the features, the composite of DCT and HOG was applied on the palm images. In case of iris image, the composite of Gabor filter and Zernike moment was applied to extract the iris features. The max normalization was applied to make the feature vectors in the same range, then the feature vectors of palm and iris are integrated in single vector. The fuzzy-based classifier was developed for this system. The recognition rate of the multimodal is 98.96% and approximate time of classification is 7.47 s. The recognition rate of this system is found to be satisfactory.

**Acknowledgements** The research work is supported and sponsored by SAP DRS-II (No.: F.4-7/2018/DRS-II(SAP-II)), UGC New Delhi, India.

## References

1. Narasimha M, Susheela Devi V (2011) Pattern recognition: an algorithm approach. Springer Science & Business Media, Berlin, pp 1–6
2. Wildes RP (1997) Iris recognition: an emerging biometric technology. In: Proceeding of the IEEE, vol 85, issue no. 9, September 1997
3. Shu W, Zhang D (1998) Automated personal identification by palmprint. In: SPIE [s0091-3286(98)01908-4]
4. Jia W, Hu R-X, Lei Y-K, Zhao Y, Gui J (2014) Histogram of oriented lines for palmprint recognition. IEEE Trans Syst Man Cybern 44(3)

5. Dalal N, Triggs B (2005) Histogram of oriented gradients for human detection. In: IEEE computer society conference on Computer Vision and Pattern Recognition (CVPR'05),1063-6919/05
6. Fathi A, Alirezazadeh P, Abdali-Mohammadi F (2016) A new Global-Gabor-Zernike feature descriptor and its application to face recognition. *J Vis Commun Image Represent* 38:65–72
7. Mahesh VGV, Raj ANJ (2015) Invariant face recognition using Zernike moments combined with feed forward neural network. *Int J Biom* 7(3):286–307
8. Gayathri R, Ramamoorthy P (2012) Feature level fusion of palmprint and iris. *Int J Comput Sci Issues (IJCSI)* 9(4):194–203
9. Kihal N, Chitroub S, Meunier J (2004) Fusion of iris and palmprint for multimodal biometric authentication. In: 2014 4th international conference on Image Processing Theory, Tools and Applications (IPTA). IEEE, New York, pp 1–6
10. Khachane MY, Manza RR, Ramteke RJ (2015) Fuzzy rule based classification of human spermatozoa. In: 2015 international conference on Electrical, Electronics, Signals, Communication and Optimization (EESCO). IEEE, New York, pp 1–5
11. Alsubari A, Ramteke RJ (2016) Extraction of palmprint texture features using combined DWT-DCT and local binary pattern. In: IEEE 2nd international conference on Next Generation Computing Technologies (NGCT), Dehradun, pp 748–753
12. Alsubari A, Ramteke R (2018) Multimodal of face and iris based on local binary pattern and Gabor-Zernike moments. *Int J Adv Res Comput Sci* 9(1)
13. Michael GKO, Connie T, Teoh ABJ (2008) Touch-less palmprint biometrics: novel design and implementation. *Image Vis Comput* 26:1551–1560 (Elsevier)
14. Masek L (2003) Recognition of human iris patterns for biometric identification. M.S. Dissertation, the University of Western Australia
15. Zuiderveld K (1994) Contrast limited adaptive histogram equalization. *Graphic Gems IV*. Academic Press Professional, San Diego, pp 474–485
16. Ojala T, Pietikäinen M, Mäenpää T (2002) Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans Pattern Anal Mach Intell* 24(7):971–987
17. Alsubari A, Satange DN, Ramteke RJ (2017) Facial expression recognition using wavelet transform and local binary pattern. In: 2017 2nd international conference for Convergence in Technology (I2CT). IEEE, New York, pp 338–342
18. Štruc V, Pavešić N (2010) The complete Gabor-Fisher classifier for robust face recognition. *EURASIP J Adv Sign Process* 2010(1)
19. Khotanzad A, Hong YH (1990) Invariant image recognition by Zernike moments. *IEEE Trans Pattern Anal Mach Intell* 12(5):489–497
20. Ramteke RJ, Mehrotra SC (2006) Feature extraction based on moment invariants for handwriting recognition. In: 2006 IEEE conference on cybernetics and intelligent systems. IEEE, New York, pp 1–6
21. Wang LX, Mendel JM (1992) Fuzzy basis functions, universal approximation, and orthogonal least-squares learning. *IEEE Trans Neural Networks* 3(5):807–814
22. CASIA iris image database and CASIA palm image database. <http://biometrics.idealtest.org/>

# Employing FPGA Accelerator in Real-Time Speaker Identification Systems



Omran Al-Shamma, Mohammed A. Fadhel and Haitham S. Hasan

**Abstract** In the most recent years, numerous approaches have been accomplished in the discipline of human voice recognition for building speaker identification systems. Frequency and time domain techniques are widely used in extracting human voice features. This paper presents the most robust and most popular Mel-frequency cepstral coefficient (MFCC) technique to parameterize voices and to be used later in the voiced/unvoiced different feature extraction process methods. In addition, the direct classical techniques for human voice feature extraction purposes are used. For the purpose of the processing time consumption and to speed up the system performance for use in real-time applications, a field programming gate array (FPGA) is utilized. Its type is Altera DE2 Cyclone II.

**Keywords** Speech recognition · MFCC · Human voice identification · FPGA · Real-time classification

## 1 Introduction

Human voice recognition is attacking our lives. It is manufactured into our mobiles, automated our homes, helping handicapped people, automated and enhanced operator services in telephone networks, and buying trash bags or ordering your lunch (e-marketing)—just by getting a wonderful box (Amazon Echo Dot) [1].

The main objective of human voice recognition is to develop systems and techniques for human voice input to the computer. Computerized human voice processing depends on a wide spectrum of disciplines and their techniques. For a successful

---

O. Al-Shamma · M. A. Fadhel (✉) · H. S. Hasan  
University of Information Technology and Communications, Baghdad, Iraq  
e-mail: [Mohammed.a.fadhel@uoitc.edu.iq](mailto:Mohammed.a.fadhel@uoitc.edu.iq)

O. Al-Shamma  
e-mail: [o.al\\_shamma@uoitc.edu.iq](mailto:o.al_shamma@uoitc.edu.iq)

H. S. Hasan  
e-mail: [haitham@uoitc.edu.iq](mailto:haitham@uoitc.edu.iq)

© Springer Nature Singapore Pte Ltd. 2019  
S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,  
Advances in Intelligent Systems and Computing 922,  
[https://doi.org/10.1007/978-981-13-6783-0\\_12](https://doi.org/10.1007/978-981-13-6783-0_12)



progress in its application areas, a wide knowledge as possible of these disciplines is needed [2]. Two main tasks arise in computerized human voice, which are identification and verification. The aim of the first task (identification) is to identify the mysterious voice from a set of  $N$  identified voices. An identity state, for example, a speaker name is given to the recognized voice, otherwise no further action. The aim of the second task (verification) is to pass or stop the recognized voice based on the designed system. More specifically, identification task can be classified into two categories: open-set and closed-set. The open-set means that the examined voice is from outside of the training set, while the closed-set means that the examined voice is always one of the training set [3].

However, speaker recognition was developed for decades, but why it is just now more widely spread unconventionally? The answer is the development of the parallel processing technologies (e.g., FPGA) and the deep learning techniques, which made speaker recognition faster (real time) and improved the accuracy up to 99%, respectively. This paper presents the conventional Mel-frequency cepstral coefficient (MFCC) for feature extraction along with FPGA hardware for a real-time human voice recognition system.

## 2 Speaker Identification System

In general, the speaker identification system comprises two stages: database creation stage and recognition stage. In the first stage, speaker samples are gathered from the human to train their models. In the second stage, the unknown human voice sample is examined with the speaker database. The core task of the first stage is the feature extraction step which is the extraction of the speaker-dependent characteristics of his/her voice. A numerous techniques are available for feature extraction [4], such as:

- **Linear Predictive Coding (LPC):** It is founded on the fundamentals of the sound production. The main drawback is that the performance degrades with noise presence.
- **Linear Predictive Cepstral Coefficient (LPCC):** It is better than LPC due to its stable representation, and it gives smoother spectral envelope. Its main drawback is because of the linearly spaced frequency bands.
- **Mel-Frequency Cepstral Coefficient (MFCC):** Due to the Mel-spaced filter banks, it has more information at lower frequencies than the higher frequencies. It is preferred to the other techniques, because its behavior is more like a human ear. Hence, it is used in this paper.

Conversely, numerous MFCC enhancement techniques have been developed which are varied in their environmental conditions and simplicity. On the whole, the simplicity leads to decrease in the recognition accuracy. And, most researches indicate that increasing the extracted features leads to important usefulness in the noisy voice, but it will increase very significantly the computation complexity. Hence,

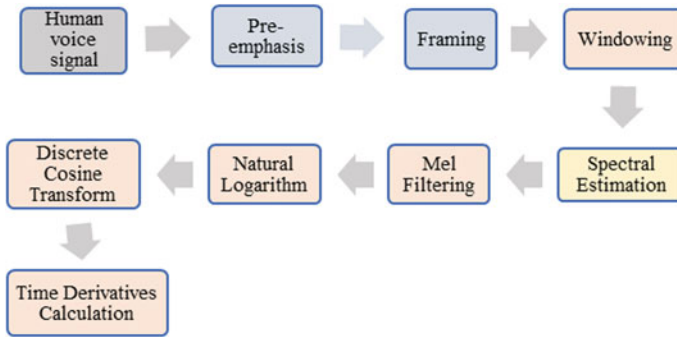
it should be a comparison between the required accuracy and the number of features. Conversely, the conventional MFCC encounters many challenges. These challenges include the weak performance in the noisy environment, its robustness, and its complexity. This paper focuses on the conventional MFCC-technique-based FPGA platform to speed up the processing time. Also, it will lead to the growing necessity to develop an advanced and better confidence MFCC extracted features with more accuracy up to 99% in speaker identification system [6].

## 2.1 Mel-Frequency Cepstral Coefficient Technique (MFCC)

In general, the speaker identification system primarily concentrates on training the system to identify the characteristics of the human voice. Thus, the most common technique for feature extraction is the MFCC. It is more effective, easy to implement, and more applicable under different conditions [5]. One of the notable drawbacks of utilizing MFCC is noise-sensitive, since it depends on the spectral form.

However, MFCC procedure is summarized in the following steps [6]:

1. **Pre-emphasis:** This step includes pre-emphasizing the signal spectrums (usually, the digitized voice signal is spectral flattened by applying a first-order filter).
2. **Framing:** In this step, the voice signal is segmented into frames (small duration blocks). This is because the human voice signal is time-varying slowly, so it can be handled like a quasi-static process. For speech recognition, 20–30 ms is the almost common frame length and 10 ms is for frame shift [7].
3. **Windowing:** This step is required for reducing the effect of discontinuity produced by the framing step. Each frame is multiplied by a window function (Hanning window) to reduce the frequency resolution, and to decrease the sample of the voice signal to zero on both start and end of each frame.
4. **Spectral estimation:** In this step, the fast Fourier transform (FFT) is employed to convert the time-domain signal into frequency-domain signal. It is applied to each frame, to compute the spectral coefficients, which are complex numbers. In addition, each number consists of two parts. The first part represents the spectral coefficient magnitude. The second part represents the phase information, which is usually discarded.
5. **Mel filtering:** The human voice spectrum is multiplied by a set of 20 triangular band-pass filters, so as to develop soft magnitude spectrum. This method is entitled Mel filtering [8]. The human ear, which utilizes a number of overlapped key bands, synthesizes the voice spectrum in groups. These overlapped bands are scattered in a specific way. This way can be described as follows: In the high-frequency region, the frequency resolution is low, and it is high in the low-frequency region.
6. **Natural logarithm:** In this step, the relationship, between the voice intensity and the loudness perception, is approximated by using the natural algorithm [9]. In addition, the multiplication relation between parameters is exchanged with



**Fig. 1** Common stages of MFCC technique

addition relation by the natural algorithm [10]. For instance, the complication of distortions (such as the microphone filtering effect) and the soft voice amplification are converted to addition relationship after applying this step [11].

7. **Discrete cosine transform:** This step includes the obtaining of the cepstral coefficients by applying the discrete cosine transform [12]. Note that the lower-order cepstral coefficients indicate the vocal tract shape. And, the higher-order cepstral coefficients stand for the excitation information [13]. In general, the speaker identification systems only use the lower-order coefficients, so that the dimension reduction is attained. Hence, it is one of the DCT advantages.
8. **Time derivatives:** This additional step is used to add additional features to the original cepstral features. It includes the calculation of the first delta (velocity) and the second delta (acceleration). The feature dimension will be 26 with only the first time derivative (velocity) added. Moreover, it becomes 39 with both derivatives (velocity + acceleration) added [14].

Figure 1 shows the block diagram of the common stages of the MFCC technique.

## 2.2 FPGA Implementation

The FPGA is symbolized for “Field Programmable Gate Array.” It is a larger device than the standard gate array. The “Field” word refers to the capability to program the gate array for a certain task by the user rather than by the device manufacturer. And the “Array” word refers to gates ordered in a series of rows and columns. In the FPGA architecture, the PLB stands for the programmable logic block. For instance, the PLB is identified as logic array block (LAB for Altera company) or configurable logic block (CLB for Xilinx company). However, looking inside the FPGA, it contains three main elements: input/output block (IOB), PLB, and lines to interconnect resources. The IOB offers a programmable interface between the pins of the external device and the internal PLB. And, the PLB executes a given user task

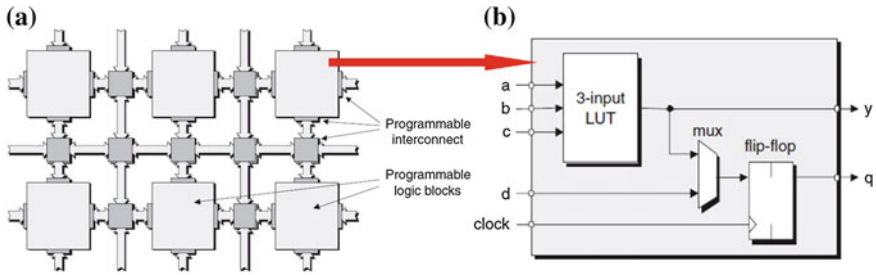


Fig. 2 a FPGA block structure and b inside logic block [16]

[15]. Figure 2a illustrates the organization of the FPGA blocks, while Fig. 2b shows the inside PLB [16].

From the configuration side view, the FPGA can be classified into three main types: SRAM, antifuse, and EEPROM/flash. In the SRAM configuration, an external memory is employed to carry out the logic function by storing the configuration data. The antifuse configuration needs a special in-chip programmer circuitry, and it is non-volatile. The programming in the EEPROM/flash-based is similar to the SRAM configuration, but it is non-volatile [17].

Now, for implementing the speaker identification system on the FPGA platform (Altera DE2, Cyclone II board), the SD card-in/out, microphone-in, and line-out are used for database storage, human voice, and system speakers, respectively. Figure 3 shows the FPGA hardware implementation of the speaker identification system. The SD card driver (on the DE2 board) is employed to read/store the data from/in the SD card. Note that the Altera family utilizes a 32-bit embedded processor called Nios II. It is implemented completely in the memory and programmable logic blocks. One of the important Nios II features is that it can be tailored by the user, based on his/her particular design necessities, due to its soft-core nature. Therefore, the user can add custom peripherals, defining practice instructions, and a predefined memory management unit, to the Nios II basic functionality [18].

In this paper, it is employed to read data from the SD card and to process these data by the Wolfson WM8731 audio CODEC. In general, several CODECs have been installed on the FPGA board. It supports digital audio input word of length 16–32 bits and sampling rates of 8–96 kHz. It is configured in the “Master” mode, which allows the device to generate the left/right channel clock (LRCK) and BCK (AD/DA serial bit clock). The WM8731 is managed through 2-wire serial interface, which is called the inter-integrated circuit (I2C) [19]. The I2C is proposed to let several “Slave” chips to interface with one or more “Master” chips. Furthermore, it only needs 2-wire for exchanging information, one for clock signal and the other for the data signal, as illustrated in Fig. 4. The “Reset” push button is employed to reconfigure the audio gains of the CODEC through the I2C interface, which selects one of the ten pre-identified volume levels [20].

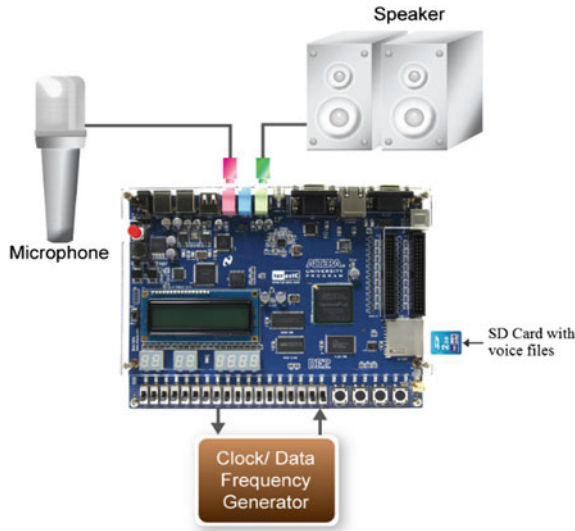


Fig. 3 FPGA implementation of the speaker identification system

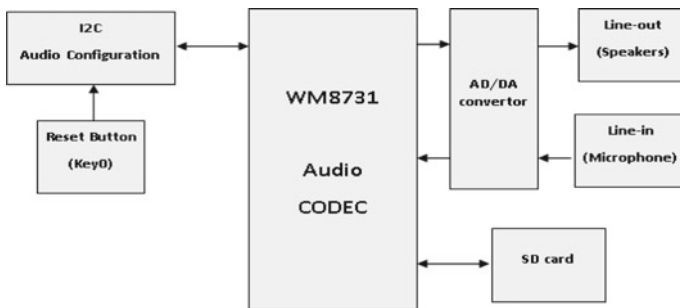


Fig. 4 Block diagram of the speaker identification system

### 3 Results and Discussion

Initially, the utilization of MATLAB and Simulink together offers a best simulation environment that helps the designer to combine both textual and graphical programming to implement the MFCC technique [21]. Rather than physically writing numerous lines of code, it automatically generates a high-quality HDL code which behaves similar to the created MFCC technique model. Then, install it directly on the FPGA board.

The interface between MATLAB or Simulink tools and the FPGA board is through the DSP Builder. It helps the designer to develop the optimized hardware description language (HDL) of his/her DSP design with an automatically generating performance and an algorithm-friendly development environment. More specifically, the

DSP Builder is a software tool that combining, all together, the FPGA development tool, MATLAB tools, simulation, and the algorithm development [22]. It creates make-able and high-quality Verilog code based on Simulink models and MATLAB functions [23].

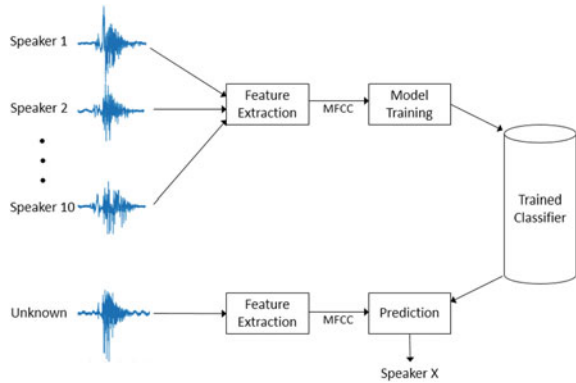
As mentioned earlier, the procedure of the system consists of two stages. In the first stage, the database is constructed. The human voice for each predefined words picks up by the microphone, converted to digital signals through the AD/DA converter, processed with the MFCC technique, and finally stored in the SD card. Each identified word has only a period of three seconds recording time, so that all stored words have equal file size each. The stored word files are collected into a single directory with unique name that identifies the speaker's sound. This stage is repeated several times based on the number of the identified persons needed to form the database. Due to one utterance human voice recorded for each predefined word, it is called one-dimensional identification database.

On the other hand, feelings, health condition, surrounding effects, and different situations are all factors make the human voice non-stationary, which in turn reduces the system recognition performance significantly. Therefore, the strategy used previously has to be modified for covering the idea. More specifically, two or more of different uttered human voices of each predefined word should be stored in the database. Hence, the modified strategy is called multidimensional identification database.

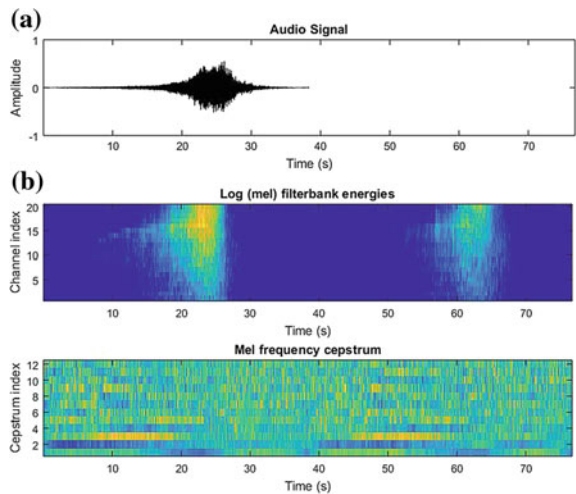
Establishing multidimensional database means storing the same word multitudes in different modes (e.g., normal, happy, sad, angry, ill, etc.) for each person. Although the database volume will be very large, the processing time will increase significantly, but the system performance and accuracy are much better. The problem of database volume is solved by using an SD card with the required memory. And the processing time consumption problem is also solved by using the FPGA accelerator. Figure 5 shows the block diagram of the speaker identification system. The upper part represents the first stage, i.e., the creating and storing the voice words for different number of persons. As in the figure, ten subdatabases are created. The bottom part is the testing (working) stage. When unknown person utters a word, the uttered voice signal is processed through MFCC technique and compared with the stored database, through the FPGA accelerator, to recognize the speaker. If it matched with one of the stored database, then the speaker is identified and the action is on; otherwise, no action is taken place.

Looking inside the MFCC technique, Fig. 6a shows the human voice signal in time-domain basis, while Fig. 6b shows the signal after processing in frequency-domain basis. Figure 6b consists of two parts. The top part is the signal after applying the Mel filtering process (step 5, Sect. 2.1). The bottom part shows the signal, which represents the Mel Frequency Cepstrum plot, after the final step (step 8, Sect. 2.1). Figure 7 obtains detailed information about the person under test, as well as, the analysis parameters of the MFCC technique. This procedure can help the examiner to have an idea about the voice signal that is being displayed on the screen for the specific speaker.

**Fig. 5** Block diagram of the speaker identification system



**Fig. 6** **a** Time domain of a database voice word signal and **b** the frequency domain of a database voice word signal



An interactive software toolbox system was built using MATLAB version R 2018a, to perform; maintain (such as list and/or hear the recorded voices), test and modify/update (such as delete word and add new word), create a subdatabase for a new person, and display technical and personal information related to each stored word, as illustrated in Fig. 7.

## 4 Conclusions

The following conclusions can be drawn from the research work.

1. Human voices are so much time-varying that one recorded voice signal of a short time can never convey to distinguish speaker identification almost (100%). On

```
Speaker Identification Database  
Record Number (1, 1) of length 20  
Name: 'Mohammed A Fadhel'  
Sex: 'Male'  
Birth: '16/10/1987'  
Address: 'UOITC'  
Telephone: 'none'  
Speech: [30000x1 double]  
MFCC: [150x14 double]  
Pitch: 82.2553  
Jitter: 0.4216  
MeanMFCC: 7.4326  
Stddev: 3.2125  
Minspeech: -0.1552  
Maxspeech: 0.0697  
MFCC1: [88x16 double]  
MeanMFCC1: 9.9815  
Stddev1: 4.0910  
MedianMFCC1: 7.9915  
MedianMFCC: 6.1510  
Press any key to continue
```

Fig. 7 Human sound database information record display

the contrary, as seen with the DSP builder on the FPGA engine at least 1 s of continues speech is required from a person for the recognition process to be able to recognize his/her speech.

2. On the speaker identification system, the recorded voices were for males and female persons, grown-ups, and under ages. A problem was encountered when under age voices were mixed with grown-up voices on the same database used for speaker identification.
3. Therefore, for a better performance of the speaker identification system, it is wise to separate speakers on different subdatabases: one for male grown-up, another for male under ages, third for female grown-up, and fourth for female under ages.
4. For a better speaker identification final target, it is important to understand that the surrounding recording environment has a major effect on the decision efficiency the system provides as a final result.
5. For a better decision making of a speaker identification system, it is not the number of features extractions that counts, but the decision efficiency of the human sound features extraction that effects.
6. It is well known that speech wave files need large storage capacity in the memory (SD card). The FPGA seems to be quite suitable for dealing with such files.



7. High performance by digital signal processing (DSP) builder on the FPGA because of the parallel processing feature.

## References

1. Gaikwad SK (2010) A review on speech recognition technique. *Int J Comput Appl* (0975–8887) 10(3)
2. Bachu RG, et al (2008) Separation of voiced and unvoiced using zero crossing rate and energy of the speech signal. In: American Society for Engineering Education (ASEE) zone conference proceedings
3. Kaur G, Singh D, Kaur G (2015) A survey on speech recognition algorithms. *Int J Emerg Res Manage Technol* 4(5):289–298
4. Shrawankar U, Thakare V (2010) Techniques for feature extraction in speech recognition system: a comparative study. *Int J Comput Appl Eng Technol Sci (IJCAETS)* 2(1):412–418
5. Poonkuzhali C, Karthiprakash R, Valarmathy S, Kalamani M (2013) An approach to feature selection algorithm based on ant colony optimization for automatic speech recognition. *Int J Adv Res Electr Electron Instrum Eng* 11(2) (2013)
6. Majeed S, Husain H, Abdulsamad S, Idbeaa T (2015) Mel Frequency Cepstral Coefficients (MFCC) feature extraction enhancement in the application of speech recognition: a comparison study. *J Theor Appl Inf Technol* 79(1):38–56
7. Rabiner LR, Juang BBH (1993) *Fundamentals of speech recognition*. Prentice Hall, Upper Saddle River
8. Huang X, Acero A, Hon H-W (2001) *Spoken language processing*, vol 15. Prentice Hall, Upper Saddle River
9. Von Békésy G, Wever EG (1960) *Experiments in hearing*, vol 8. McGraw-Hill, New York
10. Quatieri TF (2002) *Discrete-time speech signal processing*. Pearson Education, London
11. Xiong X (2009) *Robust speech features and acoustic models for speech recognition*. PhD. Thesis, Nanyang Technological University, Singapore
12. Rabiner LR, Schafer RW (2010) *Theory and applications of digital speech processing*. Pearson, London
13. Ephraim Y, Rahim M (1999) On second-order statistics and linear estimation of Cepstral coefficients. *IEEE Trans Speech Audio Process* 7:162–176
14. Singh P, Rani P (2014) An approach to extract features using MFCC. *IOSR J Eng* 4(8):21–25
15. Churiwala S, Hyderabad I (2017) *Designing with Xilinx® FPGAs*. Springer International Publishing, Cham
16. Bailey DG (2011) *Design for embedded image processing on FPGAs*. Wiley, New York
17. Ehkan P, Allen T, Quigley SF (2011) FPGA implementation for GMM-based speaker identification. *Int J Reconfig Comput* 3
18. Chu PP (2012) *Embedded SoPC design with NIOS II processor and Verilog examples*. Wiley, New York
19. Khan ARM, Thakare AP, Gulhane SM (2010) FPGA-based design of controller for sound fetching from codec using Altera DE2 Board. *Int J Sci Eng Res* 1(2)
20. Altera DSP (2003, July) *Builder–reference manual*. Altera Corporation
21. Mathworks Homepage. <https://www.mathworks.com/products/simulink.html>. Last Accessed 1 Sep 2018
22. Nurmi J et al (eds) (2015) *GALILEO positioning technology*. Springer, Berlin
23. Tlelo-Cuautle E, de la Fraga LG, Rangel-Magdaleno J (2016) *Engineering applications of FPGAs*. Springer, Berlin

# Author Index

## A

Acharyya, Sagnik, 103  
Adhikari, Arunabha, 31  
Al-Shamma, Omran, 125  
Alsubari, Akram, 113  
Alva, Srinivas, 43  
Anjaneyulu, L., 53  
Ashraf, Rehan, 43

## B

Bhattacharyya, Siddhartha, 23, 89

## C

Chatterjee, Anulekha, 1  
Chaudhuri, Dipta, 63

## D

Das, Dipnarayan, 79  
Das, Indrajit, 1  
Dey, Alokannand, 89  
Dey, Sandip, 89

## F

Fadhel, Mohammed A., 125

## G

Gangopadhyay, Indrasom, 1  
Ghosh, Ananya, 103  
Ghosh, Koushik, 63  
Gupta, Sumit, 79

## H

Hasan, Haitham S., 125  
Hemanth, D. Jude, 13

## I

Iftikhar, Ayesha, 43

## J

Jana, Ranjan, 23

## K

Khondekar, Mofazzal H., 63  
Koley, Subhadee, 103

## L

Lonkhande, Preeti, 113

## M

Mukherjee, Moloy, 63  
Mukherjee, Soumen, 31

## P

Pardeshi, Rajmohan, 43  
Platos, Jan, 89  
Pranitha, B., 53

## R

Ramteke, R. J., 113  
Roy, Madhusudan, 31

## S

Saeed, Asim, 43  
Sandip Dey, 89  
Snael, Vaclav, 89

## W

Winston, J. Jenkin, 13