

Jyotsna Kumar Mandal
Debika Bhattacharya *Editors*

Emerging Technology in Modelling and Graphics

Proceedings of IEM Graph 2018

Advances in Intelligent Systems and Computing

Volume 937

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 Hagras, School of Computer Science & Electronic Engineering,
University of Essex, Colchester, UK

László T. Kóczy, Department of Automation, Széchenyi István University,
Gyor, 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>

Jyotsna Kumar Mandal ·
Debika Bhattacharya
Editors

Emerging Technology in Modelling and Graphics

Proceedings of IEM Graph 2018

 Springer

Editors

Jyotsna Kumar Mandal
Department of Computer Science
and Engineering
University of Kalyani
Kalyani, West Bengal, India

Debika Bhattacharya
Department of Computer Science
and Engineering
Institute of Engineering and Management
Kolkata, West Bengal, India

ISSN 2194-5357

ISSN 2194-5365 (electronic)

Advances in Intelligent Systems and Computing

ISBN 978-981-13-7402-9

ISBN 978-981-13-7403-6 (eBook)

<https://doi.org/10.1007/978-981-13-7403-6>

© Springer Nature Singapore Pte Ltd. 2020

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, expressed 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

Preface

The 1st International Conference IEMGraph 2018—International Conference on Emerging Technology in Modelling and Graphics was held on 6–7 September 2018 in Kolkata, India. IEMGraph 2018 was an international and interdisciplinary conference covering research and development in the field of emerging technologies in intelligent systems and computing. It addressed all trending research topics and the latest research in emerging technology in modelling and graphics including image processing and analysis, image segmentation, digital geometry for computer imaging, image and security, biometrics, video processing, medical imaging, and virtual and augmented reality.

More than 250 pre-registered authors submitted their work in this conference. IEMGraph 2018 finally accepted and hosted 70 papers after a double-blind peer review process. The conference technical committee with the contribution of competent and expert reviewers decided about the acceptance of the submitted papers.

One of the primary objectives of IEMGraph 2018 was the investigation of information-based technological change and its adaptation in different industries and academic world. The conference tried to bridge the gap between industry demand and academic supply in research fields to address that demand. This annual event was addressed jointly to academics and practitioners and provided a forum for a number of perspectives based on either theoretical analyses or empirical case studies that foster the exchange of ideas. The conference offered a number of sessions under its patronage that was a valuable resource for scholars and practitioners, and also the conference workshop was organized to give hands-on experience to students in image processing, graphical modelling, and AI-related fields.

We would like to thank all participated in any way in the IEMGraph 2018 and team members of organizing committee and other committees for organizing the conference successfully. Also, thanks to the contributors of this volume for contributing their articles for publication. We express our sincere gratitude to the famous publication house Springer for their communication sponsorship and the

co-organizer “Indian Space Research Organisation (ISRO)”, India, for their technical sponsorship, and to the members of the technical committee who provided a significant contribution to the review of papers and all members of the organizing committee for their help, support, and spirit participation before, during, and after the conference.

This volume will be a state-of-the-art material for researchers, engineers, and students.

Kalyani, India
Kolkata, India

Jyotsna Kumar Mandal
Debika Bhattacharya

Contents

Identification of Noise Pollution Prone Regions in Mumbai Using Expectation-Maximization Clustering Technique	1
Arushi Agarwal, Pratibha Chaudhary, Rana Majumdar, Sunil Kumar Chowdhary and Abhishek Srivastava	
An Automated Segmentation Approach from Colored Retinal Images for Feature Extraction	9
Suchismita Goswami, Sushmita Goswami, Shubhasri Roy, Shreejita Mukherjee and Nilanjana Dutta Roy	
Epileptic Seizure Recognition Using Deep Neural Network	21
Anubhav Guha, Soham Ghosh, Ayushi Roy and Sankhadeep Chatterjee	
Graph-Based Supervised Feature Selection Using Correlation Exponential	29
Gulshan Kumar, Gitesh Jain, Mrityunjoy Panday, Amit Kumar Das and Saptarsi Goswami	
A Sentiment-Based Hotel Review Summarization	39
Debraj Ghosh	
Heuristic Approach for Finding Threshold Value in Image Segmentation	45
Sandip Mal and Ashiwani Kumar	
An Educational Chatbot for Answering Queries	55
Sharob Sinha, Shyanka Basak, Yajushi Dey and Anupam Mondal	
A Supervised Approach to Analyse and Simplify Micro-texts	61
Vaibhav Chaturvedi, Arunangshu Pramanik, Sheersendu Ghosh, Priyanka Bhadury and Anupam Mondal	

A Short Review on Different Clustering Techniques and Their Applications	69
Attri Ghosal, Arunima Nandy, Amit Kumar Das, Saptarsi Goswami and Mrityunjoy Panday	
An Efficient Descriptor for Gait Recognition Using Spatio-Temporal Cues	85
Sanjay Kumar Gupta, Gaurav Mahesh Sultaniya and Pratik Chattopadhyay	
Supervised Classification Algorithms in Machine Learning: A Survey and Review	99
Pratap Chandra Sen, Mahimarnab Hajra and Mitadru Ghosh	
Breast Cancer Diagnosis Using Image Processing and Machine Learning	113
Subham Sadhukhan, Nityasree Upadhyay and Prerana Chakraborty	
Using Convolutions and Image Processing Techniques to Segment Lungs from CT Data	129
Souvik Ghosh, Sayan Sil, Rohan Mark Gomes and Monalisa Dey	
Analyzing Code-Switching Rules for English–Hindi Code-Mixed Text	137
Sainik Kumar Mahata, Sushnat Makhija, Ayushi Agnihotri and Dipankar Das	
Detection of Pulsars Using an Artificial Neural Network	147
Rajarshi Lahiri, Souvik Dey, Soumit Roy and Soumyadip Nag	
Modification of Existing Face Images Based on Textual Description Through Local Geometrical Transformation	159
Mrinmoyi Pal, Subhajit Ghosh and Rajnika Sarkar	
A Neural Network Framework to Generate Caption from Images	171
Ayan Ghosh, Debarati Dutta and Tiyasa Moitra	
Remote Sensing and Advanced Encryption Standard Using 256-Bit Key	181
Sumiran Naman, Sayari Bhattacharyya and Tufan Saha	
Grasp-Pose Prediction for Hand-Held Objects	191
Abhirup Das, Ayon Chattopadhyay, Firdosh Alia and Juhi Kumari	
A Multi-level Polygonal Approximation-Based Shape Encoding Framework for Automated Shape Retrieval	203
Sourav Saha, Soumi Bhunia, Laboni Nayak, Rebeka Bhattacharyya and Priya Ranjan Sinha Mahapatra	

A Hand Gesture Recognition Model Using Fuzzy Directional Encoding of Polygonal Approximation	217
Sourav Saha, Soma Das, Shubham Debnath and Sayantan Banik	
Signature-Based Data Reporting in Wireless Sensor Networks	231
Monika Bhalla and Ajay Sharma	
Wine Quality Analysis Using Machine Learning	239
Bipul Shaw, Ankur Kumar Suman and Biswarup Chakraborty	
Generalized Smart Traffic Regulation Framework with Dynamic Adaptation and Prediction Logic Using Computer Vision	249
Vishal Narnolia, Uddipto Jana, Soham Chattopadhyay and Shramana Roy	
A Short Review on Applications of Big Data Analytics	265
Ranajit Roy, Ankur Paul, Priya Bhimjyani, Nibhash Dey, Debankan Ganguly, Amit Kumar Das and Suman Saha	
A Survey of Music Recommendation Systems with a Proposed Music Recommendation System	279
Dip Paul and Subhradeep Kundu	
Interactive Systems for Fashion Clothing Recommendation	287
Himani Sachdeva and Shreelekha Pandey	
Embedded Implementation of Early Started Hybrid Denoising Technique for Medical Images with Optimized Loop	295
Khakon Das, Mausumi Maitra, Minakshi Banerjee and Punit Sharma	
Stress Profile Analysis in n-FinFET Devices	309
T. P. Dash, S. Das, S. Dey, J. Jena and C. K. Maiti	
An Annotation System to Annotate Healthcare Information from Tweets	319
Nixon Dutta, Anupam Mondal and Pritam Paul	
Broad Neural Network for Change Detection in Aerial Images	327
Shailesh Shrivastava, Alakh Aggarwal and Pratik Chattopadhyay	
User-Item-Based Hybrid Recommendation System by Employing Mahout Framework	341
Sutanu Paul and Dipankar Das	
Gray Matter Segmentation and Delineation from Positron Emission Tomography (PET) Image	359
Abhishek Bal, Minakshi Banerjee, Punit Sharma and Mausumi Maitra	
A Joint Image Compression–Encryption Algorithm Based on SPIHT Coding and 3D Chaotic Map	373
Ramkrishna Paira	

Improved Multi-feature Computer Vision for Video Surveillance	383
Ashutosh Upadhyay and Jeevanandam Jotheeswaran	
Topic Modeling for Text Classification	395
Pinaki Prasad Guha Neogi, Amit Kumar Das, Saptarsi Goswami and Joy Mustafi	
Low Frequency Noise Analysis in Strained-Si Devices	409
Sanghamitra Das, Tara Prasanna Dash and Chinmay Kumar Maiti	
Employing Cross-genre Unstructured Texts to Extract Entities in Adapting Sister Domains	419
Promita Maitra and Dipankar Das	
Relation Extraction from Cross-Genre Unstructured Text	433
Promita Maitra and Dipankar Das	
An Efficient Algorithm for Detecting and Measure the Properties of Pothole	447
Amitava Choudhury, Rohit Ramchandani, Mohammad Shamoan, Ankit Khare and Keshav Kaushik	
Quantitative Interpretation of Cryptographic Algorithms	459
Aditi Jha and Shilpi Sharma	
Ransomware Attack: India Issues Red Alert	471
Simran Sabharwal and Shilpi Sharma	
An Automated Dual Threshold Band-Based Approach for Malaria Parasite Segmentation from Thick Blood Smear	485
Debapriya Paul, Nilanjan Daw, Nilanjana Dutta Roy and Arindam Biswas	
A Secured Biometric-Based Authentication Scheme in IoT-Based Patient Monitoring System	501
Sushanta Sengupta	
Automatic Facial Expression Recognition Using Geometrical Features	519
Tanmoy Banerjee, Sayantan De, Sampriya Das, Susmit Sarkar and Spandan Swarnakar	
A Review on Different Image De-hazing Methods	533
Sweta Shaw, Rajarshi Gupta and Somshubhra Roy	
An OverView of Different Image Algorithms and Filtering Techniques	541
Sumit Prakash, Abhas Somya and Ayush Kumar Rai	
Design of a Quantum One-Way Trapdoor Function	547
Partha Sarathi Goswami and Tamal Chakraborty	

Relation Estimation of Packets Dropped by Wormhole Attack to Packets Sent Using Regression Analysis	557
Sayan Majumder and Debika Bhattacharyya	
A Review on Agricultural Advancement Based on Computer Vision and Machine Learning	567
Abriti Paul, Sourav Ghosh, Amit Kumar Das, Saptarsi Goswami, Sruti Das Choudhury and Soumya Sen	
Transformation of Supply Chain Provenance Using Blockchain—A Short Review	583
Sekhar Kumar Roy, Runa Ganguli and Saptarsi Goswami	
A Framework for Predicting and Identifying Radicalization and Civil Unrest Oriented Threats from WhatsApp Group	595
Koushik Deb, Souptik Paul and Kaustav Das	
Flower Pollination Algorithm-Based FIR Filter Design for Image Denoising	607
Supriya Dhabal, Srija Chakraborty and Prosenjit Sikdar	
Image Enhancement Using Differential Evolution Based Whale Optimization Algorithm	619
Supriya Dhabal and Dip Kumar Saha	
Advanced Portable Exoskeleton with Self-healing Technology Assisted by AI	629
Piyush Keshari and Santanu Koley	
Crosstalk Minimization as a High-Performance Factor in Three-Layer Channel Routing	645
Sumanta Chakraborty	
A Human Intention Detector—An Application of Sentiment Analysis	659
Megha Dutta, Shayan Mondal, Sanjay Chakraborty and Arpan Chakraborty	
Survey on Applications of Machine Learning in the Field of Computer Vision	667
Himanshu Shekhar, Sujoy Seal, Saket Kedia and Amartya Guha	
Automatic Speech Recognition Based on Clustering Technique	679
Saswati Debnath and Pinki Roy	
A Study of Interrelation Between Ratings and User Reviews in Light of Classification	689
Pritam Mondal, Amlan Ghosh, Abhirup Sinha and Saptarsi Goswami	

A Study on Spatiotemporal Topical Analysis of Twitter Data	699
Lalmohan Dutta, Giridhar Maji and Soumya Sen	
A Secure Steganography Scheme Using LFSR	713
Debalina Ghosh, Arup Kumar Chattopadhyay, Koustav Chanda and Amitava Nag	
Crowd Behavior Analysis and Alert System Using Image Processing	721
Sayan Dutta, Sayan Burman, Agnip Mazumdar and Nilanjana Dutta Roy	
Review Article on Magnetic Resonance Imaging	731
Shatadru Majumdar, Rashmita Roy, Madhurima Sen and Mahima Chakraborty	
An Intelligent Traffic Light Control System Based on Density of Traffic	741
Kriti Dangi, Manish Singh Kushwaha and Rajitha Bakhthula	
Scheduling in Cloud Computing Environment using Metaheuristic Techniques: A Survey	753
Harvinder Singh, Sanjay Tyagi and Pardeep Kumar	
Crown Detection and Counting Using Satellite Images	765
Rebeka Bhattacharyya and Avijit Bhattacharyya	
Hand Segmentation from Complex Background for Gesture Recognition	775
Soumi Paul, Arpan Bhattacharyya, Ayatullah Faruk Mollah, Subhadip Basu and Mita Nasipuri	
A Study on Content Selection and Cost-Effectiveness of Cognitive E-Learning in Distance Education of Rural Areas	783
Anindita Chatterjee, Kushal Ghosh and Biswajoy Chatterjee	
An Automatic Method for Bifurcation Angle Calculation in Retinal Fundus Image	787
Suchismita Goswami and Sushmita Goswami	
Author Index	797

About the Editors

Jyotsna Kumar Mandal, M.Sc.(Ph.), JU, M.Tech.(CS), CU, Ph.D.(Eng.), JU Professor CSE, is former Dean of the FETM, and was KU for two consecutive terms. He has nearly 30 years of teaching and research experience, and has completed four AICTE and one state government project. He is a life member of the CSI, CRSI, a member of the ACM, and fellow of the IETE. He was also honorary vice chairman and chairman of the CSI. He has delivered over 100 lectures and organized more than 25 national and international conferences. He currently serves as an editorial board member and corresponding editor for the Proceedings of Science Direct, IEEE and other conferences, as well as guest editor of the MST Journal. He has published more than 400 research articles and six books.

Dr. Debika Bhattacharya obtained her B.Tech. and M.Tech. in Radiophysics and Electronics from Calcutta University, and her Ph.D. in Electronics and Telecommunication from Jadavpur University. She is currently Dean (of Academics) at the Institute of Engineering and Management. She has more than 22 years of teaching and research experience, has published many papers in prominent journals and conference proceedings, and has completed 2 AICTE and 1 DST project. In 2018 she received the Venus International Foundation's "Distinguished Leader" award in engineering.

Identification of Noise Pollution Prone Regions in Mumbai Using Expectation-Maximization Clustering Technique



Arushi Agarwal, Pratibha Chaudhary, Rana Majumdar,
Sunil Kumar Chowdhary and Abhishek Srivastava

Abstract Noise pollution is escalating at an alarming rate as a one of the critical outcomes of urbanization. This led to harmful effect on the health of human being as it can cause annoyance, hypertension, heart disease, and sleep disturbances. Despite all measures to control noise pollution that have been taken in Mumbai so far, those are prone to vulnerabilities. The differences in these vulnerability-inducing causes arise a need for an effective analysis. The motive of this paper is to have data mining to come to aid to create a model that provides the heterogeneity of the data by grouping similar objects together to find the noise pollution regions in the Mumbai state with respect to different factors.

Keywords Data mining · Cluster · Noise pollution · Root nodes

1 Introduction

From past few years, industries have been extended and numbers of transport vehicles on the roads have been increased due to rapid urbanization. These are the sources of environmental noise pollution. Noise pollution led to harmful effect on the physiological and physical health of human being. Noise pollution can cause annoyance,

A. Agarwal · P. Chaudhary · R. Majumdar (✉) · S. K. Chowdhary · A. Srivastava
Amity School of Engineering and Technology,
Amity University, Noida, Uttar Pradesh, India
e-mail: rana.majumdarwb@gmail.com

A. Agarwal
e-mail: arushiagwl@gmail.com

P. Chaudhary
e-mail: pratibha.chaudhary4991@gmail.com

S. K. Chowdhary
e-mail: skchowdhary@amity.edu

A. Srivastava
e-mail: asrivastava8@amity.edu

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_1

hypertension, heart disease, and sleep disturbances. As per the report by WHO (World Health Organization), several cities had exceeded the acceptable level of noise pollution. In India, in 2011, Central Pollution Control Board (CPCB) [1] has been set up for Real-time National Ambient Noise Monitoring Network that includes nine cities (Delhi, Lucknow, Chennai, Mumbai, Navi Mumbai, Thane, Kolkata, Hyderabad, and Bangalore). According to the statistical report generated by CPCB [1], Mumbai is on the top that exceeded maximum limits of prescribed noise level. Particularly, amid Ganeshotsav, the highest levels of noise are recorded on this day.

According to WHO standard [2], the community noise should not be more than 30 A-weighted decibels (dB (A), dB is used to measure noise) for sleeping at night and not be more than 35 dB (A) in classrooms for studying. The night noise should not be more than 40 dB (A) of annual average (light) of bedrooms to avoid adverse effects on health from night noise.

The main objective of this paper is to study the several regions of Mumbai against the contributing causes and draw conclusion in order to control noise pollution in the state. In this research work, we perform clustering techniques to group regions and identify causes of health risks in regions.

2 Literature Survey

The regular exposure to increased sound levels that may have adverse effects on the lives of humans or other living beings is called noise pollution. According to WHO study [1], sound levels should not be more than 70 db. Then, it cannot harm the living organisms. Exposure for 8 h and more to constant noise higher than 85 dB may be calamitous.

Noise free is thought to be a fundamental, essential of human well-being and prosperity. However, noise pollution is becoming a calamity to prosperity around the globe (WHO 2011). The study of noise pollution reveals that the effects of noise pollution are also hazardous as compare to other types of pollution, not only in Mumbai regions, but across the world. Many strategies have been introduced to monitor noise pollution in smart cities for different noise quality levels. The analysis also concluded that many measures have taken to reduce the effects and causes of noise pollution. There are many different techniques, along with data mining that can be used to monitor and analyze the noise pollution data. Table 1 summarizes some of the research studies where various techniques are used to understand its implication in noise pollution.

Clustering helps in the formation of data sets of similar type of group structures, thus dividing and arranging the data into subclasses. In this paper, the clustering technique used is expectation-maximization (EM) clustering using Gaussian mixture models (GMM). The data points are assumed to be in multivariate Gaussian distribution surface. Mean and standard deviation are the two important parameters describing the shape of the clusters. A single cluster is assigned an each Gaussian

Table 1 Techniques used

S. No.	Author	Research	Technique
1.	Panu Maijala (2018) [5]	Monitoring noise sources	Acoustic pattern classification algorithm
2.	Maisonneuve (2010) [6]	Noise pollution	Mobile phones

distribution. Expectation-maximization (EM), an optimization algorithm, is used to find the parameters. The further technique is preceded using GMMs.

3 Problem Conceptualizations

Mumbai region is one among the top five cities in India which have higher impact on lives due to the adverse effects created by the noise pollution. A certain high level of noise is also a deadly cause affecting the lives of people living in Mumbai. Many noise-monitoring schemas and noise-controlling plans have been directed by the authorities, but the implementation of the tasks hadn't created much difference to resolve the issue of noise pollution. We hardly find any silent zones in the Mumbai region which are safe away from the hands of noise pollution, as Central Pollution Control Board (CPCB) has also revealed that areas around hospitals, educational intuitions, courts, etc. didn't fall in the silence zones within the range of 100 m. In this paper, we have proposed a process which can be adopted for the analysis and identification of the highly prone noise pollution-affected regions in Mumbai using data mining approach of clustering techniques called expectation-maximization (EM) clustering using Gaussian mixture models (GMM), and also a decision tree can be made which shows the classification of the major causes of the noise pollution in the Mumbai region.

4 Proposed Methodology Model

The proposed methodology model is depicted in Fig. 1, and the flow of execution is from top to bottom. The steps involved are described in the following flowchart.

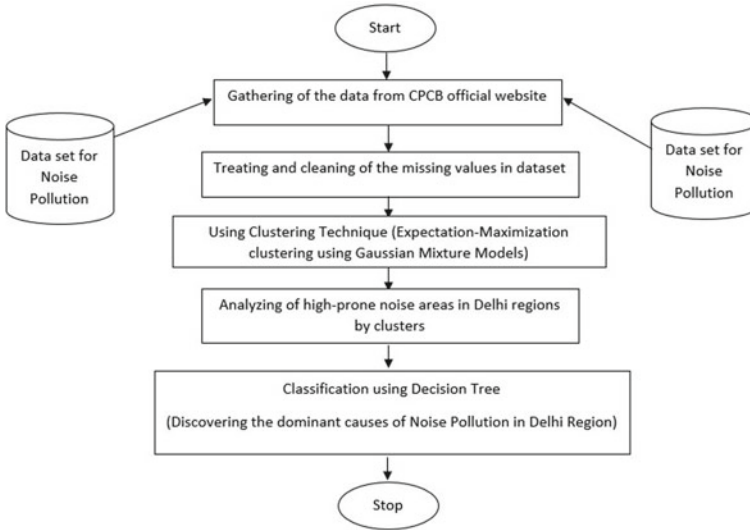


Fig. 1 Flowchart of applied techniques

4.1 Gathering of the Data from CPCB Official Website

The sample of noise monitoring data set of Mumbai for the year 2014 is collected from [3]. The data set consists of different regions of Mumbai affected by noise pollution in day time as well as night time. Figure 2 represents the chosen data set.

4.2 Treating and Cleaning of the Missing Values in Data Set

This is preprocessing step which involves the steps to remove the missing values (i.e., noisy values) from the data set, which in turn reduces the set of data (cleaning of data). Max–min normalization is also calculated, which converts the data in the normalized form for each year and combines the data into the Excel sheet.

4.3 Using Clustering Technique (Expectation-Maximization Clustering Using Gaussian Mixture Models)

The clustering technique to group similar object used is expectation-maximization clustering using Gaussian mixture models that is described in Fig. 3.

The data mining approach we proposed in this paper, used to form the clusters of the affected regions of Mumbai due to noise pollution, is expectation-maximization

Date	Thane		Vashi Hospital		Ashp		Bandra		MPCB HQ	
	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time	Day Time	Night Time
Standards	65	55	50	40	50	40	65	55	65	55
16/11/2014	62.13	56.17	68.95	78.22	64.07	59.77	68.89	66.09		
17/11/2014	63.95	55.03	70.03	58.04	65.17	59.19	69.62	65.31	72.28	70.58
18/11/2014	63.87	58.02	69.96	59.07	65.14	60.27	69.79	65.80	72.32	69.84
19/11/2014	63.89	56.35	69.86	57.13	65.29	60.77	69.85	65.68	72.33	70.27
20/11/2014	64.33	54.99	69.85	58.63	65.52	60.26	69.71	66.01	72.92	70.56
21/11/2014	63.65	56.43	69.98	57.93	65.62	60.80	69.83	65.91	72.51	70.37
22/11/2014	62.81	57.30	68.90	57.78	65.09	60.46	69.68	67.01	72.38	70.50
23/11/2014	62.57	56.32	67.57	58.12	63.32	60.18	68.79	66.64	70.62	68.95
24/11/2014	64.36	56.03	70.28	57.83	65.20	59.47	69.93	65.39	71.60	68.62
25/11/2014	64.26	56.22	69.87	57.66	64.99	60.13	69.86	66.11	71.99	70.27
26/11/2014	64.40	58.84	70.07	59.08	65.17	60.21	70.00	66.73	72.09	70.65
27/11/2014	64.09	55.82	69.90	56.81	65.64	60.20	70.12	66.32	72.13	71.11
28/11/2014	64.14	54.06	69.86	58.03	65.26	60.66	70.22	66.71	72.20	70.93
29/11/2014	63.94	57.24	69.08	59.23	65.66	60.12	70.04	65.43	71.85	70.94
30/11/2014	62.31	54.81	67.76	56.72	63.33	60.20	68.96	66.71	69.91	69.48
1/12/2014	64.99	57.07	70.06	57.55	64.89	59.91	70.26	65.87	71.89	69.33
2/12/2014	64.07	53.72	69.91	57.58	65.45	60.18	70.22	65.50	71.94	70.11
3/12/2014	64.44	54.22	70.12	59.66	65.24	60.72	70.14	66.24	71.99	70.71
4/12/2014	63.82	55.59	70.09	58.07	65.31	60.84	70.35	66.19	71.99	70.66

Fig. 2 The sample data set for several regions of Mumbai

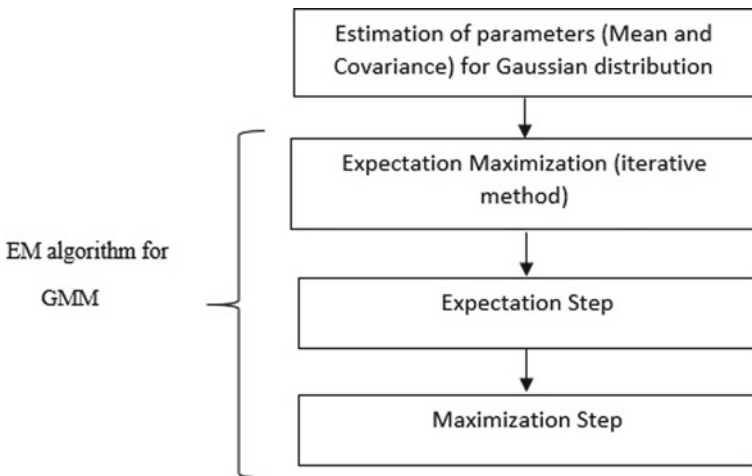


Fig. 3 Steps of clustering technique

(EM) clustering using Gaussian mixture models (GMM). The symbols used in the equations are described in the table. The steps involved are as follows:

1. Estimating Parameters (mean and covariance): Using the multivariate Gaussian distribution, we estimate the parameters (mean and covariance) of a distribution. In this step, the clusters of data are predicted to have an independent Gaussian distribution, each having their own mean and covariance matrices. The estimated parameters are as:

$$N(x|\mu, \Sigma) = \frac{1}{(2\pi|\Sigma|)^{1/2}} \exp\left\{-\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right\} \quad (1)$$

$$\text{Mean, } \mu_{ML} = \frac{1}{N} \sum_{n=1}^N x_n$$

$$\text{Covariance, } \Sigma_{ML} = \frac{1}{N} \left(\sum_{n=1}^N (x_n - \mu_{ML})(x_n - \mu_{ML})^T \right) \quad (2)$$

2. Expectation-Maximization: It is an iterative technique, which includes expectation and maximization step (explained in next steps). The expectation-maximization algorithm is used for the generation that maximizes the probability of output data of means and variance, for the distributional parameters of multi-mode data. The techniques involve the probabilistic models.
3. Expectation Step: In this step, we assign each point to a cluster. Using the estimated mean and covariance, we can predict and calculate the probability of each data point belonging to each cluster. We can calculate the fixed, data-independent parameters. Thus, for already calculated parameters we can calculate the approximation values of latent variable. The formula for depicting the responsibility (calculating data point for each cluster) in expectation step is:

$$r_{ic} = \frac{\pi_c N(x_i|\mu_c, \sigma_c)}{\sum_{j \in [0, k]} \pi_j N(x_i|\mu_j, \sigma_j)} \quad (3)$$

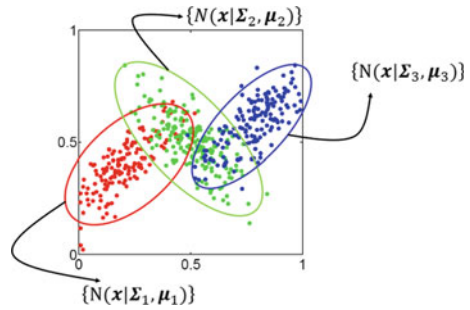
Suppose, if we have 80 data points with a mixture of 4 Gaussians, we need to calculate 320 numbers (matrix is of 80×4).

4. Maximization Step: Now in this step, we work on the improvement of guess of each curve mean, standard deviation, and weighting factor. The parameters that are known in above expectation step are fully determined, and maximization of parameters is done. This step updates the distributional class parameters based on the new parameters.

$$\text{New Mean, } \mu_c^{\text{new}} = \frac{1}{N_c} \sum_i r_{ic} x_i \quad (4)$$

$$\text{New Standard Deviation, } \sigma_c^{\text{new}} = \frac{1}{N_c} \sum_i r_{ic} (x_i - \mu_c^{\text{new}})^2 \quad (5)$$

Fig. 4 Example of Gaussian mixture models [4]



$$\text{Weighting factor, } \pi_c = \frac{N_c}{n}$$

Example of Gaussian mixture models forming clusters in expectation-maximization using the different parameters is shown in Fig. 4.

4.4 Analyzing of High-Prone Noise Regions in Mumbai by Clusters

After applying the technique discussed above, the clusters of the regions are formed according to the rate of noise pollution in the particular area. Now, the analysis of data is done to estimate the major causes of noise pollution in the clusters formed areas.

4.5 Classification Using Decision Tree

Decision tree is a tree-like structure consisting of root nodes, along with leaf and intermediate nodes. The sample data of the particular year are collected, along with the daytime and nighttime. The major and dominant reasons are identified with the help of classification methods which are responsible for the occurrence of noise pollution in the regions in certain year. The data set is classified on the basis of cluster analysis. The criterion used to make decision tree is gain ratio, which is calculated as (Table 2):

$$\text{Gain Ratio}(A) = \text{Gain}(A) / \text{Split Info}(A)$$

Table 2 Symbol table with its description

Symbols and description	
Symbol	Description
N	Sample of data points
μ	Mean of Gaussian distribution
Σ	Covariance of Gaussian distribution
x	Input vector
k	Possible curves
i	Data point
r_{ic}	Responsibility factor
c	Gaussian curve
π_c	Weighting factor
n	Number of data points in data set

5 Conclusion and Future Work

In this paper, the proposed model which shows how cluster analysis helps to determine the noise pollution prone regions of Mumbai. The decision tree is used to label the clusters which are classified to conclude the dominant factors responsible for the noise pollution. The advantages of used clustering technique are as follows:

- The choosing of component distribution is flexible.
- Availability of “soft” classification (EMM is sometimes considered as soft clustering technique).
- Density estimation for each cluster is obtained.

In future, we can extend this work by implementing proposed methodology using the data mining tools. In addition to this, the study of the data sets to determine the most-vulnerable situation that occurred due to the impact of noise pollution in Mumbai and its impact on the population of Mumbai.

References

1. <http://cpcb.nic.in>
2. <http://www.euro.who.int/en/health-topics/environment-and-health/noise/data-and-statistics>
3. <http://cpcb.nic.in/>
4. https://ibug.doc.ic.ac.uk/media/uploads/documents/expectation_maximization-1.pdf
5. P. Majjala, Z. Shuyang, T. Heittola, T. Virtanen, Environmental noise monitoring using source classification in sensors. *Appl. Acoust.* **129**, 258–267 (2018)
6. N. Maisonneuve, M. Stevens, B. Ochab, Participatory noise pollution monitoring using mobile phones. *Inf. Polity* **15**(1, 2), 51–71 (2010)
7. <http://ethesis.nitrkl.ac.in/5590/1/E-49.pdf>

An Automated Segmentation Approach from Colored Retinal Images for Feature Extraction



Suchismita Goswami, Sushmita Goswami, Shubhasri Roy, Shreejita Mukherjee and Nilanjana Dutta Roy

Abstract Segmentation from colored retina images plays a vital role in stable feature extraction for image registration and detection in many ocular diseases. In this study, the authors will look at the segmentation of the blood vessels from fundus images which will further help in preparation of digital template. Here, images are passed through the preprocessing stages and then some of the morphological operators for thresholding are applied on the images for segmentation. Finally, noise removal and binary conversion complete the segmentation method. Then, a number count on blood vessels around the optic disk is done as a feature for further processing. The authors will ensure whether the segmentation accuracy, based on comparison with a ground truth, can serve as a reliable platform for image registration and ocular disease detection. Experiments are done on the images of DRIVE and VARIA databases with an average accuracy of 97.20 and 96.45%, respectively, for segmentation, and a comparative study has also been shown with the existing works.

Keywords Segmentation · Fundus images · Blood vessels · Morphological operators

S. Goswami · S. Goswami · S. Roy · S. Mukherjee (✉) · N. D. Roy
Department of Computer Science and Engineering, Institute of Engineering and Management,
Kolkata, India
e-mail: shreejitamukherjee.m@gmail.com

S. Goswami
e-mail: suchismita.g24@gmail.com

S. Goswami
e-mail: sushmita.g24@gmail.com

S. Roy
e-mail: shubhasrir4@gmail.com

N. D. Roy
e-mail: nilanjanaduttaroy@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_2

1 Introduction

Automated analysis and segmentation from colored retinal images is a challenging task. It has a potential research impact in various areas like diagnosing of many diseases, person identification and security purposes due to significant advances in the field of digital image processing. Extraction of stable features from retina like bifurcation points, bifurcation angle, location of optic disk, width of the blood vessels, etc., needs special care. Thus, proper segmentation from colored retinal images is an essential task to accomplish the same. Manual segmentation of blood vessels is a cumbersome and time-consuming process. In a research setting, with proper technical support, automatic segmentation from retinal images provides better and faster solution. It also plays an important role in diagnosing many ocular diseases affecting retina [1, 2]. Blood vessels are indicative to many pathological changes caused due to hypertension, diabetes, arteriosclerosis, cardiovascular diseases. So observation on the vascular changes sometimes improves the disease diagnosis process. Apart from the medical diagnostics, the segmentation method has another potential application in person identification. The microvascular structure of human retina is unique for each individual and usually remains same during any one's lifetime except surgical issues. With some existing distinct features on retinal vascular structure, it could be used as a secure template for identification. Many algorithms are there in literature to segment the blood vessel from colored retinal images. In this paper, a best suited segmentation approach has been proposed using some of the morphological operators which will further help in macula detection in the future. The proposed segmentation algorithm is subdivided into three categories, preprocessing, thresholding and blood vessel segmentation and post-processing. Preprocessing, like channel conversion, enhancement and noise removal techniques, helps in improving the quality of the image as most of them suffer from poor local contrast compared to background. Thresholding and binarization help in actual segmentation, and finally, noise present on the images is removed by applying 2-D median filtering in a 3×3 scanning window. Later, a count on blood vessels around optic disk [3] is done from the segmented images by looking at the number of transitions on binary images. The approach proves sensitivity and specificity as 0.9632 and 0.9470, respectively, and accuracy as 97%. Due to the inadequacy of available resources, the experiments are done only on all the images of DRIVE [4] and VARIA [5] databases. A precise view of the segmentation method is described in Fig. 1.

2 Background

Segmentation of a blood vessel from the retinal images using tracking-based approach is proposed in [6]. In this approach, fuzzy model of a one-dimensional vessel profile drives the segmentation process. One drawback of these approaches is that this method of segmentation depends upon the methods for identifying the seed pixel

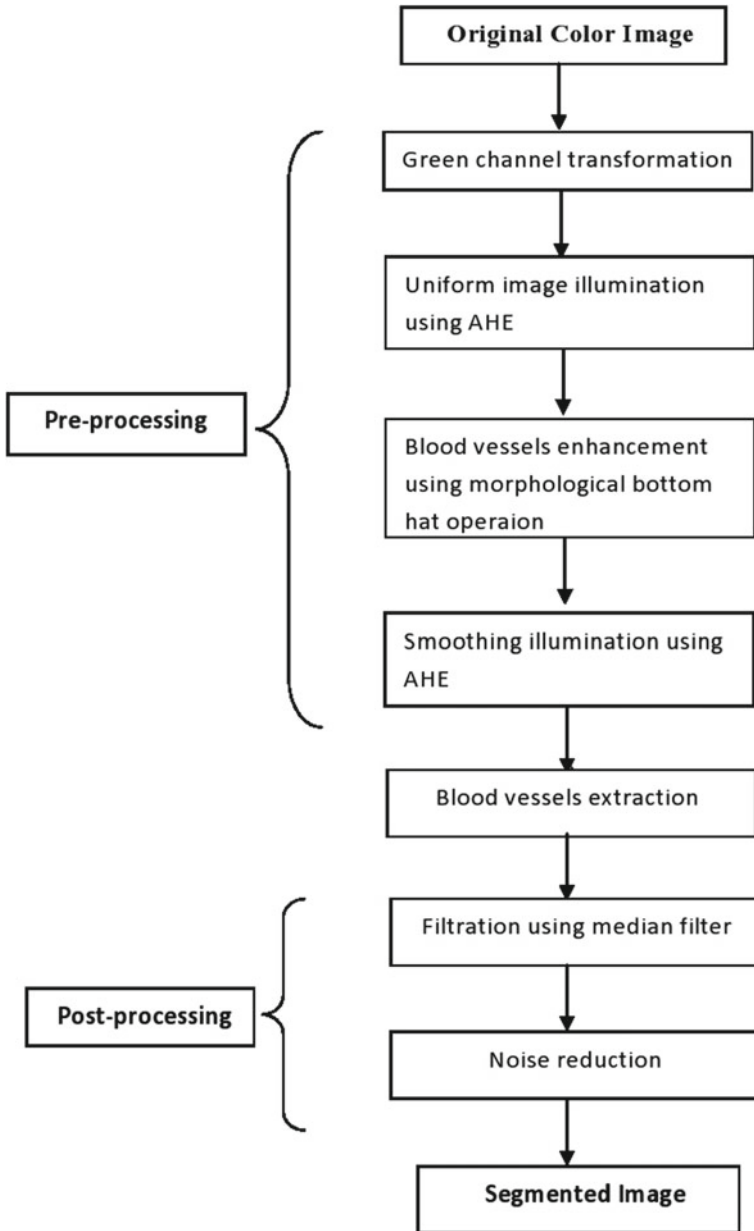


Fig. 1 Flow diagram of the segmentation method

which is either the optic disk or the branch points which are detected subsequently. Gabor filter-based image processing methods were proposed for retinal vessel extraction in [7–10]. For segmentation of the blood vessels in the retinal images, optimized Gabor filters with local entropy thresholding had been used in those approaches. The drawback of those segmentation methodologies is that optimized Gabor filter methods fail to detect vessel of different widths and may sometimes detect blood vessels falsely. Also, detection process fails in case of defected retinal image caused by various retinal diseases. An automated texture-based blood vessel segmentation has been proposed in the paper [11]. In this paper, they have used Fuzzy c-Means (FCM) clustering algorithm for the classification between vessels and non-vessels depending on texture properties. This algorithm is having 84.37% sensitivity and 99.19% specificity.

3 Proposed Method

The existing algorithms suffer from problem of non-uniform illumination of the background in the retinal fundus image. The proposed method shows a way to overcome the limitations and to improve the effectiveness, accuracy and computational time.

Broadly, there are three principal stages of the segmentation method, preprocessing, blood vessel extraction and post-processing. Preprocessing technique has been applied, to increase the accuracy in recognizing and extracting the blood vessels. Green channel conversion of RGB images initiates the process as the green channel provides the highest vessel background contrast. Next, the boundary of the green channel images is removed and contrast-limited adaptive histogram equalization (CLAHE) is applied over it to make a uniform illumination distribution. To this, morphological bottom hat operation with disk-shaped structuring element is applied to enhance the retinal blood vessels. Finally, to correct the illumination variation in the background of retinal image obtained from the previous stage, estimation of the background illumination and the contrast distribution is applied over it.

Image binarization and segmentation of the blood vessels are two essential tasks to be performed here with an empirically generated threshold value.

The primary task in post-processing stage is 2-D median filtering and morphological noise removal operations. This would remove the disconnected blood vessel and noise from the binary image and assist to accomplish the desired goal.

3.1 *Image Preprocessing and Segmentation*

The retinal images show non-uniform illumination of the background due to the presence of vitreous humor, which is a transparent gel that fills the interior of the eye. The blood vessels show variety in their thickness and contrast. So the darker

vessels are clearly seen and easily detected. But it is difficult to extract the thin vessels having low contrast. It is easy to segment if the blood vessels appear as dark structure in the brighter background. So, difference in the contrast between the retinal blood vessels and the background is desirable. Hence, preprocessing is applied to the original retinal image to eliminate these anomalies and to prepare the image for the next steps of segmentation.

Channel conversion It is necessary to convert color (RGB) images into green channel as the green channel provides the best contrast between blood vessels and background of the RGB representation. As the red channel has low contrast and the blue channel has poor dynamic range, we are focusing only on green channel, refer Fig. 2b.

Boundary removal In this step, the outer border of the green component of the image is removed by suppressing the structures that have lower contrast than the surroundings and that are connected to the image border.

Uniform image illumination using AHE Fundus images suffer from background intensity variation due to non-uniform illumination which further deteriorate the segmentation result. Due to this issue, background pixels sometimes have higher gray-level values than the vessel pixels. Because of the difference in gray-level intensity and false vessel's appearance, the global thresholding techniques cannot be applied in this phase. Contrast-limited adaptive histogram equalization (CLAHE) is applied here to enhance the contrast of the green channel retinal image, as shown in Fig. 2c. It redistributes the light value of the image within small regions of the image instead of the entire one. The process thus enhances the contrast of every smaller region of the image. This step also helps to reduce few false detection of blood vessels that would otherwise decrease the performance of blood vessel segmentation method.

Morphological transformation Sometimes, retinal blood vessels falsely appear as background due to poor intensity variation between the blood vessels and the image background. So to brighten the darker blood vessels in lighter background with an empirically tested disk-shaped structuring element of size 8×8 , morphological bottom hat operation is applied here. Bottom hat filtering subtracts the input image from the result to perform a morphological closing operation on the image which is dilation followed by erosion. Dilation is an operation that grows or thickens the object in binary image. Erosion operation shrinks the object by eroding away the boundaries of regions of foreground pixels. After applying bottom hat, the changes in the result are shown in Fig. 2d.

Smoothing Illumination The image produced after the bottom hat operation has poor distribution of lightness value. To improve the distribution of intensity, again CLAHE is applied, so that the image is well prepared for thresholding and segmentation stages; see Fig. 2e.

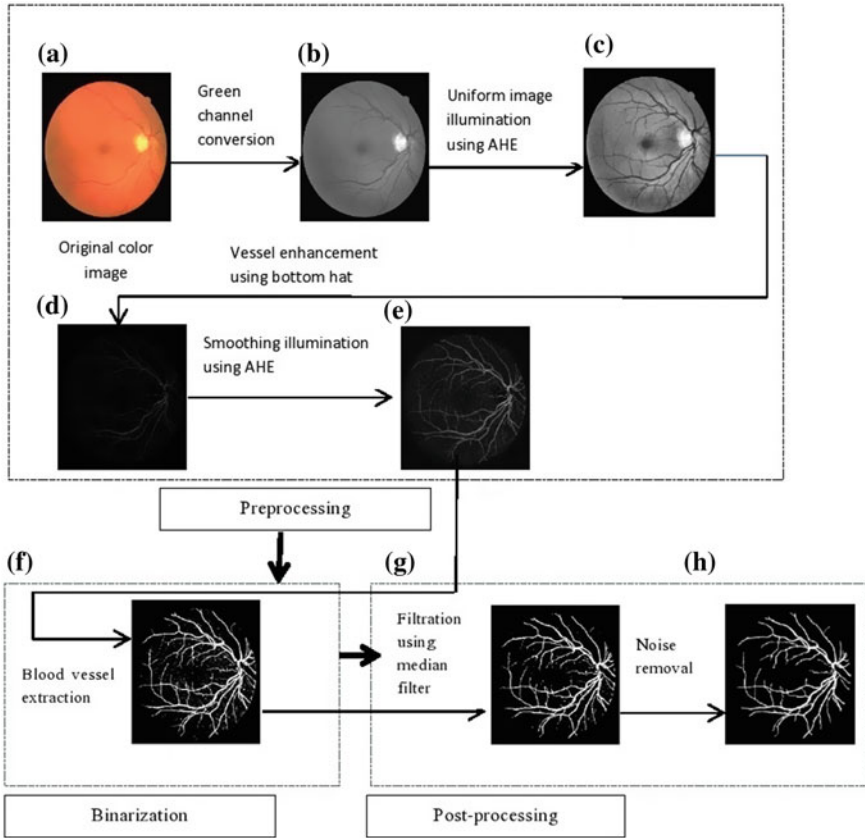


Fig. 2 Flow diagram of the segmentation method

3.2 Blood Vessel Segmentation

Preprocessing has enhanced the contrast of the original images. Now the images are being thresholded using Otsu’s algorithm by keeping the information below the threshold value and assigning rest of the image the same value as the threshold. Thus, cluster-based algorithm is used to perform clustering-based image thresholding and the preprocessed grayscale images are now reduced into the binary images (Fig. 2f).

3.3 Noise Removal

The clarity of the images obtained by the previous stages is not very good because of the presence of salt-and-pepper noise in it. Hence, a 2-D median filtering with 3×3

neighborhood size is applied to reduce the noise from the segmented image. Next, the small connected components, having lesser than 35 white pixels around it within the 8-connectivity window, are removed from the previously processed image. Desired image is shown in Fig. 2h.

3.4 Counting Vessels Around Optic Disk

Common parameters, like accuracy, specificity and sensitivity, are used to measure the strength of an algorithm and to compare the results with other segmentation algorithms. To compare the result of the proposed approach with gold standard images, count on the number of vessels is essential. Also, counting the number of blood vessels around optic disk helps in diagnosis of severe ocular disease called proliferative diabetic retinopathy (PDR). Abrupt changes in count on vessels are alarming for some unusual scenario on regular monitoring. Algorithm 1 explains the method of counting blood vessels in segmented images. Table 1 shows the number of blood vessels detected by the proposed algorithm on the images of DRIVE database.

Algorithm 1 Count on blood vessels around OD

```

1: count ← number of blood vessels around OD
2: for within a  $50 \times 50$  window around the optic disk do
3:   repeat
4:     for for each pixel along the window border do
5:       if a white pixel is found then
6:         repeat
7:           move along the border of the window by one pixel
8:         until white to black transitions occur
9:         count ← count + 1
10:        else
11:          move along the border of the window by one pixel
12:        end if
13:      end for
14:    until all the pixels along the four border of the specified window is traversed
15:  end for

```

Table 1 Count on number of blood vessels around OD

Image	Count	Image	Count	Image	Count
Image 1	14	Image 6	12	Image 11	10
Image 2	12	Image 7	16	Image 12	12
Image 3	16	Image 8	16	Image 13	14
Image 4	13	Image 9	14	Image 14	13
Image 5	10	Image 10	13	Image 15	12

4 Results and Performance Evaluation

A complete segmentation method, along with the intermediate steps, is shown in Fig. 2. To show the proficiency of the proposed algorithm, gold standard images which are manual segmented images from DRIVE [4] database have been used. Specificity, sensitivity and accuracy are calculated as measuring parameters to compare the results of the proposed algorithm with the other existing algorithms.

$$\text{Sensitivity} = \frac{TP}{(TP + FN)} \quad (1)$$

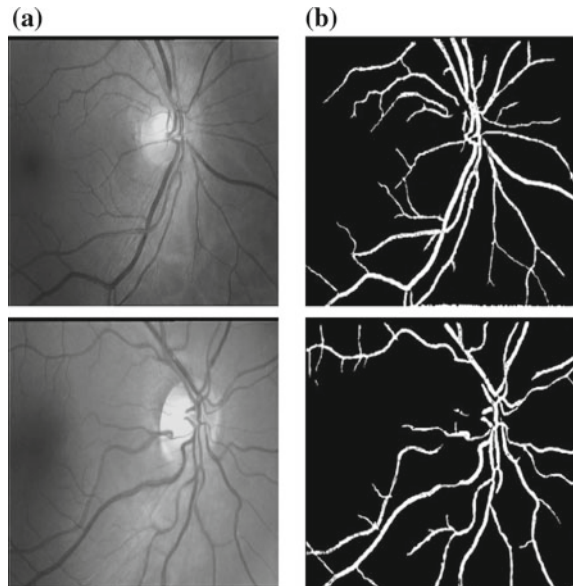
$$\text{Specificity} = \frac{TN}{(TN + FP)} \quad (2)$$

$$\text{Accuracy} = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (3)$$

where TP is correctly detected positive values; TN is correctly detected negative values; FP is feature is negative, but detected as positive; and FN is feature is positive, but cannot detect (Figs. 3 and 4).

Please refer Table 2 for the performance measurement of the proposed algorithm. Table 3 compares the performance of the proposed algorithm with others.

Fig. 3 Results on VARIA database **a** original images, **b** result of the proposed algorithm



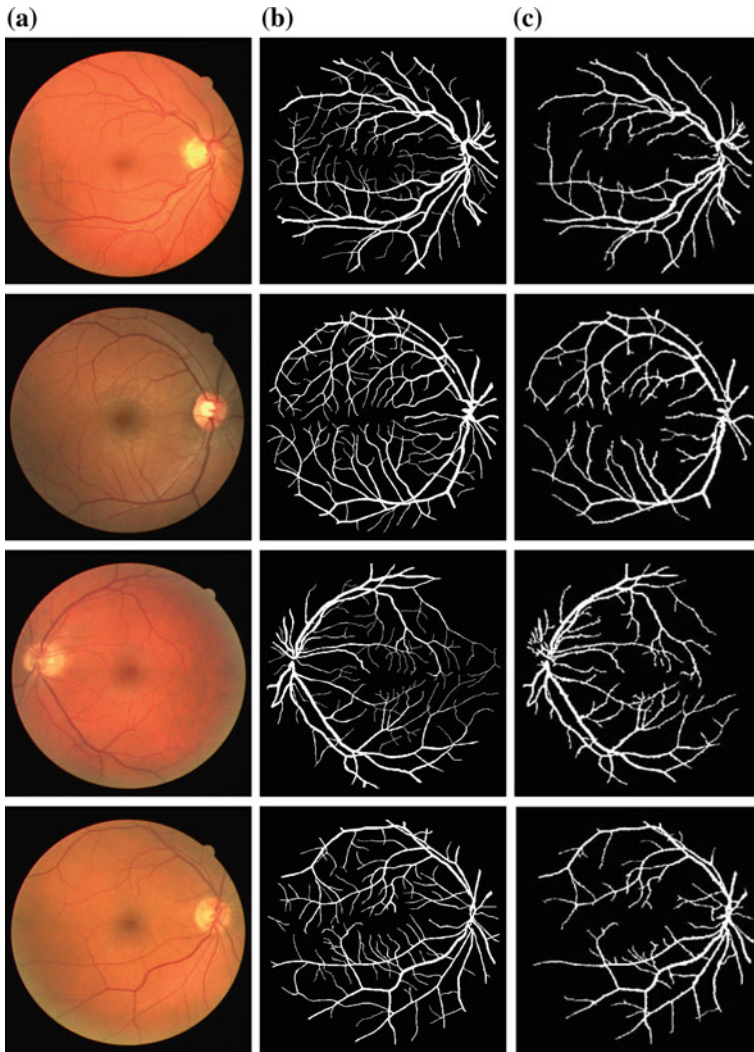


Fig. 4 Comparison with gold standard images from DRIVE **a** original image, **b** hand-driven ground truth images from DRIVE, **c** result of the proposed algorithm

Table 2 Performance measurement on DRIVE and VARIA databases

Database	Specificity	Sensitivity	Accuracy (%)
DRIVE	0.9470	0.9632	97.20
VARIA	0.9561	0.9812	98.34

Table 3 Comparison of the proposed algorithm with others

Methods	Database	Sensitivity	Specificity	Accuracy
Hand driven [4]	DRIVE	0.7763	0.9723	94.70
Mendonca et al. [12]	DRIVE	0.7344	0.9764	94.63
You et al. [13]	DRIVE	0.7410	0.9751	94.34
Marin et al. [14]	DRIVE	0.7067	0.9801	94.52
Villalobos-Castaldi et al. [15]	DRIVE	0.9648	0.9480	97.59
Proposed method	DRIVE	0.9632	0.9470	97.20

5 Conclusion

In this paper, we have presented an automated approach for blood vessel segmentation from retinal fundus images. Proposed algorithm has been tested on two publicly available databases, DRIVE [4] and VARIA [5]. We have executed the proposed algorithm on 32-bit MATLAB R2013a version on 32-bit Windows 7 running on Intel Core 2 dual processor with 2 GB RAM. We have used various contrast enhancement techniques and the morphological operators which aim to provide better segmentation results including both minor and major vessels. This segmentation algorithm works efficiently for low-contrast images with less space and time. This algorithm can speed up the ocular disease diagnosis process for ophthalmologists and can be helpful in designing biometric templates.

References

1. J.L. Boulnois, Photo physical processes in recent medical laser developments. *Lasers Med. Sci.* **1**, 47–66 (1986)
2. G.M. Bohigian, Lasers in medicine and surgery. *JA-MA* **256**, 900–907 (1986)
3. M.D. Abramoff, M. Niemeijer, The automatic detection of the optic disc location in retinal images using optic disc location regression, in *28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS '06* (2006)
4. The DRIVE database, Image sciences institute, university medical center utrecht, The Netherlands. <http://www.isi.uu.nl/Research/Databases/DRIVE/>. Last accessed 7 July 2007
5. VARIA Database, Department of Computer Science of the Faculty of Informatics of the University of Corua. <http://www.varpa.es/varia.html>
6. Y.A. Tolias, S.M. Panas, A fuzzy vessel tracking algorithm for retinal images based on fuzzy clustering. *IEEE Trans. Med. Imaging* **17**, 263–273 (1998)
7. B. Dizdarog, E. Ataer-Cansizoglu, J. Kalpathy-Cramer, K. Keck, M.F. Chiang, D. Erdogmus, Structure-based level set method for automatic retinal vasculature segmentation. *EURASIP J. Image Video Process.* (2014). (Springer)
8. P.C. Siddalingaswamy, K.G. Prabhu, Automatic segmentation of blood vessels in colour retinal images using spatial gabor filter and multiscale analysis, in *13th International Conference on Biomedical Engineering, IFMBE Proceedings*, vol. 23 (Springer, 2009), pp. 274–276
9. D. Wu, M. Zhang, J.C. Liu, W. Bauman, On the adaptive detection of blood vessels in retinal images. *IEEE Trans. Biomed. Eng.* **53**(2), 341–343 (2006)

10. M.M. Fraz, P. Remagnino, A. Hoppe, S. Velastin, B. Uyyanonvara, S.A. Barman, A supervised method for retinal blood vessel segmentation using line strength, multiscale Gabor and morphological features, in *IEEE International Conference on Signal and Image Processing Applications (ICSIPA)* (2011), pp. 16–18
11. A. Bhuiyan, B. Nath, J. Chua, R. Kotagiri, Blood vessel segmentation from color retinal images using unsupervised texture classification, in *IEEE International Conference, ICIP 2007*, vol. 5 (2007)
12. A.M. Mendona, A. Campilho, Segmentation of retinal blood vessels by combining the detection of centerlines and morphological reconstruction. *IEEE Trans. Med. Imaging* **25**, 1200–1213 (2006)
13. X. You, Q. Peng, Y. Yuan, Y. Cheung, J. Lei, *Segmentation of Retinal Blood Vessels Using the Radial Projection and Semi-supervised Approach* (Elsevier Science Inc., New York, NY, USA, 2011), pp. 2314–2324
14. D. Marin, A. Aquino, M.E. Gegundez-Arias, J.M. Bravo, A new supervised method for blood vessel segmentation in retinal images by using gray-level and moment invariants-based features. *IEEE Trans. Med. Imaging* **30**(1), 146–158 (2010)
15. F.M. Villalobos-Castaldi, E.M. Felipe-Rivern, L.P. Snchez-Fernndez, A fast, efficient and automated method to extract vessels from fundus images. *J. Visual.* **13**(3), 263–270 (2010). (Springer)
16. D.S. Fong, L. Aiello, T.W. Gardner, G.L. King, G. Blankenship, J.D. Cavallerano, F.L. Ferris, R. Klein, Diabetic retinopathy. *Diabetes Care* **26**, 226229 (2003)

Epileptic Seizure Recognition Using Deep Neural Network



Anubhav Guha, Soham Ghosh, Ayushi Roy and Sankhadeep Chatterjee

Abstract We introduce and assess a better machine learning way to deal with the problem of seizure detection by hosting a comparative study between different machine learning algorithms. This problem is a multiclass problem with overlapping attributes and thus making it demanding. The most crucial part of developing a highly efficient classifier was to identify the attributes that are necessary to distinguish seizure from other brain activities. We trained our model on 23.6 s of recorded brain activity from 500 patients, which detected 80% of 500 test cases with a F1 score of 71%. Information about the dataset gathered from UCI machine learning repository database, which we analyzed in this study, is also provided.

Keywords Deep neural network · Epileptic seizure · Perceptron · Detection

1 Introduction

Seizures are the result of sudden surge of electrical activity in the brain. This electrical activity is caused due to complex chemical imbalance that occurs in nerve cells in the brain. Although, seizures are not a disease themselves. On the contrary, it is the symptom of different disorders that affects the brain. While some seizures may disable a person, other types of seizures may not be identified at all. That is why a device capable of detecting a seizure and notifying a caregiver that medical

A. Guha (✉) · S. Ghosh · A. Roy · S. Chatterjee
Department of Computer Science & Engineering, University of Engineering & Management,
Kolkata, India
e-mail: anubhav.guha@yahoo.com

S. Ghosh
e-mail: meetsoham2050@gmail.com

A. Roy
e-mail: roy.ayushi@outlook.com

S. Chatterjee
e-mail: chatterjeesankhadeep.cu@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_3

attention is required could make dealing with epilepsy easier. At this moment, scalp electroencephalogram (EEG) test is the most widely recognized approach to diagnose epilepsy. EEG is a non-invasive, electrophysiological monitoring method to study electrical activity in the brain [1].

In this paper, we constructed detectors capable of detecting seizure quickly and with high accuracy using machine learning. The first step of seizure detection is to segment the EEG recording. So, we sampled the recorded time series into multiple distinct data points, each of which represents the value of the EEG recording at different points in time. We have EEG recordings of 500 individuals for 23.6 s, where each recording was sampled into 4097 data points. We have to train a seizure detector which signal is relevant or not. Hence, we opted to tackle the seizure detection problem using a supervised model. Even though the physiological activity is multi-class, we grouped them together and treated it as a binary classification problem. The reason behind doing such is that it is neither effortless nor reasonable for a medical expert to distinguish and label the classes comprising of the seizure and non-seizure states. On the contrary, classifying electrical brain activity into two classes, seizure and non-seizure, is considered to be an accepted clinical practice. A neural network plays a fundamental part in providing a high accuracy. They are loosely modeled to mimic the working mechanism of a human brain in order to label or classify raw input. Training on a labeled dataset is needed so that the model can classify unlabeled raw data based on the similarities to the training data.

2 Related Work

Gotman et al. [2] utilized sharp wave and spike recognition method. They additionally reinforced this procedure in Gotman [3], Gotman [4], Koffler et al. [5], Qu et al. [6].

Wilson et al. [7] tested a modified seizure detection algorithm which honed a new Reveal algorithm comprising of three methods: matching pursuit, small neural network rules, and a new connected object hierarchical clustering algorithm; all three of them being novel in their application to seizure detection. This study validated the Reveal algorithm as it performed better than the other methods.

Srinivasan et al. [8] suggested a neural network framework for automated epileptic EEG detection. As the input feature of the framework, it utilizes approximate entropy (ApEn) which is a logical parameter that determines the consistency of the present amplitude values of a physiological signal in light of its past amplitude values. It has been observed that the value of the approximate entropy decreases rapidly amid the onset of an epileptic seizure. Utilizing this information, the framework of the suggested model was developed. In this paper, two different neural networks were evaluated which were Elman and probabilistic neural network. The authors observed that they could achieve an accuracy of almost 100% using their proposed framework.

Shoeb et al. [1] assessed a machine learning way to deal with developing patient-particular classifiers that distinguishes seizure activity from other brain activities with the help of the analysis of the scalp EEG. In contrast to past endeavors, which

concentrated on grown-up EEG, the authors assessed their model on pediatric scalp EEG since it displays more fluctuation in seizure and non-seizure activity. The authors trained a support vector machine (SVM) to classify seizure and non-seizure activity. They generated nonlinear decision boundaries utilizing a RBF kernel as the seizure and non-seizure classes are generally not linearly divisible. The authors have claimed a higher accuracy and sensitivity than the Reveal algorithm, which is a patient non-specific detector using machine learning.

Guo et al. [9] introduced a unique technique for automated epilepsy diagnosis by utilizing an artificial neural network, along with the line length features in light of wavelet transform multiresolution decomposition. The authors achieved great accuracy which validated the effectiveness of the method.

Swami et al. [10] proposed an efficient technique for epileptic seizure detection which utilized dual-tree complex wavelet transform for feature extraction. The proposed system gave classification performance (accuracy, sensitivity, and specificity) that was through the roof. The authors claimed that this technique can be utilized for a quick and accurate detection of epileptic seizure.

3 Proposed Method

In our effort to build a system capable of detecting epileptic seizure, we have constructed a DNN classifier consisting of five hidden layers, where each layer comprises of 85 neurons. To train the classifier, it was provided with the necessary parameters, i.e., the features responsible for epileptic seizure detection. Finally, to evaluate the performance of the DNN classifier, we compared it with other models, such as multilayer perceptron (MLP) and k-nearest neighbor (K-NN).

MLP comprises of two layers—input layer and output layer—with one or multiple layers in-between. Each layer consists of neurons which are powered by a nonlinear activation function. Its numerous layers and nonlinear inducement are what makes it different from a linear perceptron. The model is trained on a dataset with the help of supervised learning.

KNN—k-nearest neighbor—is a nonparametric algorithm where a new variable is allocated to a class by measuring the distance (mainly Euclidian distance) of the new point with the existing points of different classes. The point is then categorized with the class which has the greatest number of points with the least possible distance, i.e., the maximum number of neighboring variables. This model is used for classification and regression purposes in machine learning.

The general steps performed by us for this experiment are as follows

1. Pre-processing: This is the first step that is needed to be performed before the classifier is developed. The following pre-processing steps were performed to the dataset.
 - (i) Feature Extraction—This step deals with the extraction of features most significant for solving the classification problem. The datasets are usually

composed of redundant attributes which serve no intrinsic part in solving the problem. So, it is essential that the unneeded attributes are filtered out such that the significant features can be provided the utmost attention.

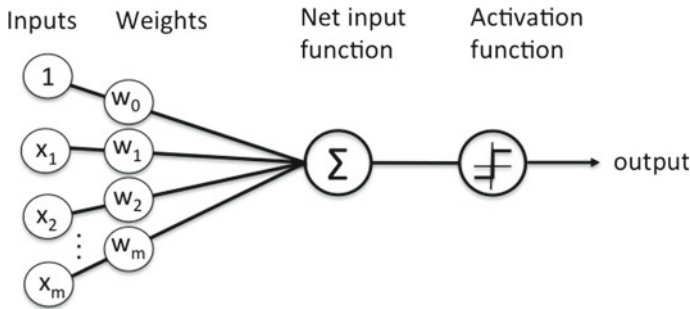
- (ii) Data Cleaning—In majority of the cases, the dataset contains missing or inconsistent values which may hinder the process of the training of the model. Such problems are dealt with in this step of pre-processing by getting rid of the inconsistencies.
 - (iii) Data Normalization—Usually, when the attribute values are ranged distinctly, the dataset is normalized to maintain a value range in between -1 and $+1$. By doing such, the distance among the attribute values is also reduced.
2. After the steps of pre-processing were over, the whole data set was split into two distinct sets, training and testing set, in a ratio of 7:3. To achieve the optimal result, both the data sets are shuffled.
 3. In the training phase of the DNN classifier, the model was provided with the training dataset. Through this process, the classifier is trained.
 4. After the training of the DNN classifier was over, the performance of the model was evaluated with the help of the testing data set. This step is the testing phase of the classifier.

After the completion of the training and testing phases, the accuracy of the model was calculated to evaluate the performance of the model.

4 Deep Neural Network

DNN is a subgroup of artificial intelligence which involves methods where it learns features directly from data as opposed to being hand engineered by the human. It causes a phenomenon called a feature hierarchy, where a layer trains on the output of the previous layer, and the more we advance through the neural network, and the complexity of decision making increases. The execution mechanism is in the form of graphs which helps it to execute the code in a distributed manner. The data comes into an input layer and flows across an output layer. The layers between input and output layers have hidden layers which are responsible for running calculations. Deep neural network consists of two or more layers; these are input layer and output layer. Deep neural networks are generally feedforward where the connection between the layers does not form a loop. Instead, the data flows forward such that the output of one layer is the input of the next layer. However, TensorFlow is an open-source library developed by Google which is compatible with many variants of machine learning. These networks are often powered by a gradient descent algorithm, which is one of the most popular optimization algorithms. The purpose of the gradient descent algorithm is to minimize the loss which is achieved through backpropagation at each iteration.

In a neural network, a layer is composed of numerous nodes which are similar to neurons. The neurons are responsible for receiving a set of inputs and processing the output with the assistance of a set of weights and an activation function. The output processed by these neurons is picked up by the next layer of neurons as their input. These inputs are further assigned an individual weight which denotes the significance of the input or the feature in accordance with the classification problem.



DNN can also be used to model sequences where the data points are not single images but rather temporally dependent. Deep learning also typically defines sets of predefined features and the data and they work to extract these features as part of their pipeline. The key differentiation point of deep learning is that it recognizes many practical situations.

Deep neural networks have the advantage of learning from massive load of unlabeled data which makes it a more favorable option over other learning algorithms. For it to be able to analyze unstructured data, first the neural network needs to be trained on a labeled dataset. However, the performance can be made exponentially better if the model can be provided huge quantities of training data to learn from. The more data the network will analyze, the more accurate it will become in its decision making.

5 Properties of the Scalp Electroencephalogram (EEG)

EEG is the continuous recording of the electrical brain activity. There are billions of brain cells called neurons, and the neurons have axons that relieve neurotransmitters and the dendrites they receive them. The axons of other neurons they cause electrical polarity change inside the neuron. This polarity change is what the EEG is recording. The brain gives off electromagnetic waves when there is electricity running in the brain, which produces brain waves that is amplified quite a lot.

Researchers are often interested in recordings from epilepsy patients because the information retrieved the EEG signals of epilepsy patients provide a particular outlook of the human brain. Such information grants an opportunity to the researchers to find something unprecedented.

6 Scalp EEG Dataset

The neural network was provided a dataset consisting of the recordings of the EEG signal recorded for 3.3 h and sampled at 173.61 Hz while the settings of the bandpass filter were set at 0.53–40 Hz (12 dB/oct.). Using a common frame of reference, the EEG readings were recorded with the help of a 128-channel amplifier system. Subsequently, the recorded data was sampled and written onto a disk using a 12-bit analog-to-digital conversion system [11].

The data set was divided into five sets, where each set consists of 100 single channel EEG segments lasting for a duration of 23.6 s. These sections were chosen and removed from the EEG dataset after analyzing for artifacts, e.g., because of activity in the muscular area or optic movement. Each of the sets consists of a distinct criterion which must be fulfilled by the sampled data. The data included in sets A and B is comprised of signal recorded from five volunteers, in their relaxed state, using the method of regulated electrode placement scheme. Data in set A included the data collected from the volunteers who had their eyes open, while in set B, they had their eyes closed. On the other hand, sets C, D, and E consisted of EEG recordings archived from presurgical diagnosis. While set D contained the EEG signals recorded from inside the epileptogenic zone, those in set C were recorded from the hippocampal formation of the opposite hemisphere of the brain. Sets C and D consist of EEG signals which represent the non-seizure activity. On the contrary, set E contains data when the subject was suffering from seizure. The data was recorded from the parts of the brain which exhibit ictal activity [12].

The EEG dataset which was evaluated in this paper was obtained from the UC Irvine Machine Learning Repository. It can be downloaded from the following website: <http://archive.ics.uci.edu/ml/index.php>.

7 Performance Metrics

After the training and the testing phases were over, the performance of the model was evaluated in terms of Accuracy, Precision, Recall, and F1 Score.

- Accuracy is denoted by the ratio of the number of observations predicted accurately by the model to the total number of observations made by the model.

$$\text{Accuracy} = (\text{True Positives(TP)} + \text{True Negatives(TN)}) / (\text{TP} + \text{False Positives(FP)} + \text{False Negatives(FN)} + \text{TN})$$

- Precision is the proportion of accurately anticipated positive observations to the aggregate anticipated positive observations.

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

- Recall is the ratio of correctly positive classifications from cases that are actually positive—yes.

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

- F1 Score is denoted by the mean of two components: precision and recall.

$$\text{F1 Score} = 2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$$

8 Results and Discussion

We evaluated our suggested model in accordance with the above-mentioned performance metrics and composed our result.

Table 1 illustrates the comparison study of the DNN model with the KNN and the MLP model. The table showcases that the accuracy of the DNN model is 80% which is comparatively better than the accuracy of the KNN model and the MLP model. However, precision is where the DNN model falls short, exhibiting only 64%, while the other two models displayed a precision of more than 70%. Additionally, the recall value of the DNN model (80%) is slightly better than that of the MLP model (79%), while the recall of the KNN model was only 76%. Another important performance metric is the F-measure. The DNN model displays a F-measure percentage of 71% with the KNN model at 72% and the MLP model at 78%.

Based on the study above, it can be inferred that the performance of the KNN model is by far the worst among the three. On the other hand, the DNN model exhibits the best signs of performance and can be concluded as the ideal classifier for this problem.

9 Conclusion

Thus, our study proposes the use of a deep-neural-network-based classifier instead of the traditional classifiers like MLP and KNN in detecting seizures of patients suffering from epilepsy. From the results obtained, it is seen that these traditional

Table 1 Performance comparison between different models

	DNN (%)	KNN (%)	MLP (%)
Accuracy	80	76	78
Precision	64	70	77
Recall	80	76	79
F-Measure	71	72	78

methods fall short when compared to deep neural networks. DNN yields a far better accuracy than MLP or KNN. Nevertheless, in the near future, focus should be on developing more efficient classifiers and in proper feature selection to yield the highest accuracies.

References

1. A. Shoeb, Application of machine learning to epileptic seizure onset detection and treatment. Ph.D. thesis, Massachusetts Institute of Technology, 2009
2. J. Gotman, J. Ives, P. Gloor, Automatic recognition of inter-ictal epileptic activity in prolonged EEG recordings. *Electroencephalogr. Clin. Neurophysiol.* **46**(5), 510–520 (1979)
3. J. Gotman, Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr. Clin. Neurophysiol.* **54**(5), 530–540 (1982)
4. J. Gotman, Automatic detection of seizures and spikes. *J. Clin. Neurophysiol.* **16**(2), 130–140 (1999)
5. D. Koffler, J. Gotman, Automatic detection of spike-and-wave bursts in ambulatory EEG recordings. *Electroencephalogr. Clin. Neurophysiol.* **61**(2), 165–180 (1985)
6. H. Qu, J. Gotman, Improvement in seizure detection performance by automatic adaptation to the EEG of each patient. *Electroencephalogr. Clin. Neurophysiol.* **86**(2), 79–87 (1993)
7. Scott B. Wilson, Mark L. Scheuer, Ronald G. Emerson, Andrew J. Gabor, Seizure detection: evaluation of the reveal algorithm. *Clin. Neurophysiol.* **115**, 2280–2291 (2004)
8. V. Srinivasan, C. Eswaran, N. Sriraam, Approximate entropy-based epileptic EEG detection using artificial neural networks. *IEEE Trans. Inf. Technol. Biomed.* **11**(3), 288–295 (2007)
9. L. Guo, D. Rivero, J. Dorado, J.R. Rabunal, A. Pazos, Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *J. Neurosci. Methods* **191**(1), 101–109 (2010)
10. P. Swami, T.K. Gandhi, B.K. Panigrahi, M. Tripathi, S. Anand, A novel robust diagnostic model to detect seizures in electroencephalography. *Expert Syst. Appl.* **56**, 116–130 (2016)
11. R.G. Andrzejak, K. Lehnertz, C. Rieke, F. Mormann, P. David, C.E. Elger, Indications of non-linear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys. Rev. E* **64**, 061907 (2001)
12. D. Rivero, J. Dorado, J. Rabuñal, A. Pazos, Automatic Recurrent ANN development for signal classification: detection of seizures in EEGs, in *ECAI 2008*, vol. 178 (2008), pp. 142–146

Graph-Based Supervised Feature Selection Using Correlation Exponential



Gulshan Kumar, Gitesh Jain, Mrityunjoy Panday,
Amit Kumar Das and Saptarsi Goswami

Abstract In this article, a graph-theoretic method for supervised feature selection using matrix exponential of pairwise correlation value has been illustrated. In machine learning, high-dimensional data sets have a enormous number of redundant and irrelevant features. The sum of mean and standard deviation of exponential matrix has been set as the threshold for selecting relevant features. Principles of vertex cover and independent set have then been used to remove redundant features. In the next step, mutual information value has been used to select relevant features that were initially rejected. The results show that this method has performed better than the benchmark algorithms when experimented on multiple standard data sets.

Keywords Feature selection · Feature redundancy · Graph-based visualization · Correlation exponentiation

1 Introduction

In the past few years, the number of high-dimensional data sets, i.e., data sets with a large number of features has increased rapidly, generated from different domains like computational biology, geospatial technologies, text and social data mining, etc.

G. Kumar (✉) · G. Jain · A. K. Das
Institute of Engineering and Management, Kolkata, India
e-mail: gulshankumar454@gmail.com

G. Jain
e-mail: giteshjain844@gmail.com

A. K. Das
e-mail: amitkumar.das@iemcal.com

M. Panday · S. Goswami
University of Calcutta, Salt Lake, Kolkata, India
e-mail: mrityunjoy@gmail.com

S. Goswami
e-mail: saptarsi007@gmail.com

The data from these domains are very large in size and have features in the range of thousands [1].

High-dimensional data sets are computationally expensive with respect to both space and time. To perform any task in finite time and limited storage space, we need optimization to scale down the number of features of the data set. All the features of a data set are not relevant [2]. Also, lot of features contribute the same information in predicting the results. These redundant features increase computational cost. The set of redundant features should be determined to reduce the overall size of the data sets. Few representative features should be selected from a set of redundant features. Hence, selecting a subset of features maximizes utilization of resources and minimizes execution time.

Feature selection is a combinatorial optimization problem [3] designed to optimize the number of features in the data sets without significant impact on output. An alternative is exhaustive search and selection of optimal feature subset, which is known to be NP-complete. During feature selection process, the complexity of the algorithm should also be taken care off. Optimizing feature subset has been found as NP-complete [4, 5]. Using approximation algorithm instead of exhaustive search reduces the order of the search space to $O(N^2)$ or below [6].

2 Related Work

Researchers have adopted several techniques to solve the feature selection problem with a significant number of them using graph-theoretic approach. In a survey paper on graph clustering [7], the possibility of converting feature vector data to a graphical format is mentioned. Another work has achieved the concept of a dominant set for feature selection [8]. In the algorithm DSFFC [9], dense subgraph and mutual information have been used to select optimal features. An algorithm was designed using a minimum spanning tree (MST)-based clustering algorithm called FAST (Fast Clustering-Based Feature Selection Algorithm) [10]. In the cluster-centric cluster (GCNC) algorithm [11], a graph theory method using community detection algorithms for feature selection is implemented.

Schaeffer [7] mentioned about the conversion of feature vector of a data set into a graph with features representing vertices and similarity representing edges. In the method used by Zhang and Hancock [8], first correlation matrix is calculated using the inter-feature mutual information, then the dominant set clustering is performed to cluster the feature vectors, and finally the optimal feature subset is selected from each dominant set using multidimensional interaction information criterion. This work has been used as a graph-based supervisory-based feature selection algorithm to compare with the proposed algorithm. Another algorithm [12] used the variant of normalized mutual information (NMI) to get the submap with the least number of features. In the first step, the two-step algorithm FAST [10] clusters the features into relatively independent groups. In the second step, it selects the features that most strongly represent each cluster.

Most of work on filter-based feature selection algorithms use pairwise metric, e.g., variations of correlation coefficient or mutual information, for computing redundancy. However, in this paper, we compute the expected value by using sum over all paths and then use minimum vertex cover and maximum independent set to identify a subset of the potentially redundant features. It also uses a local property of redundancy to derive a property based on global relationship structure.

The rest of this paper is organized as follows:

- Section 3 discusses some preliminary concepts.
- The proposed approach is discussed in Sect. 4.
- Methods and materials used are highlighted in Sect. 5 which includes CFS, mRMR, and DSCA as benchmarks.
- Finally, results and analysis are presented in Sect. 6.

3 Preliminary Concept

This section describes some fundamental concepts and information used.

3.1 Correlation

When applied to the population, Pearson's correlation coefficient is usually expressed in the Greek letter ρ and can be called the correlation coefficient or the Pearson correlation coefficient. The formula is:

$$\rho(X, Y) = \frac{\mathbf{Cov}(X, Y)}{\sqrt{\mathbf{Var}(X)\mathbf{Var}(Y)}}.$$

3.2 Expected Correlation

The expectation value is calculated as the integral or weighted sum over all values. The equivalent notion in graph is sum over all paths, which can be done using matrix exponential which converges due to normalization using Taylor Series expansion of e^A .

3.3 Mutual Information

Mutual information $I(X; Y)$ between two random variables measures how much information about one can be extracted through the knowledge of the other [13].

Mathematically, we define this as the difference between the entropy of X and the entropy of X conditioned on knowing Y .

$$I(X; Y) = H(x) - H(X|Y)$$

3.4 Vertex Cover and Independent Set

The minimum vertex cover and the maximum independent set are the graph theory principles used to derive the subgraph of the graph. Thebimore, it can be used to identify a representative subset of features from a complete feature set. The vertex cover of a graph is a set of vertices such that each edge of the graph is incident on at least one vertex of the set. Finding the minimum vertex cover, that is, the vertex cover with the fewest number of vertices, is an NP-complete problem. However, there are algorithms that can find the minimum vertex cover in polynomial time. An independent set is a set of vertices in a graph such that two vertices belonging to the set are not adjacent. Finding the maximum independent set problem is an NP-hard optimization problem. However, a list of all the maximal independent sets can be derived in polynomial time by an algorithm.

4 Proposed Approach

In the proposed method, correlation has been used for global attributes rather than local attributes. We try to represent each data set as an undirected graph. This has been done by the following steps:

- Step 1: Constructing base graph
The vertices of the graph represent the features of the data set. The similarity between two vertices is represented by the edge between them. We find out the correlation matrix based on the pairwise similarity between different features using Pearsons product-moment correlation coefficient. If correlation value is high, features are more similar.
- Step 2: Exponential of base matrix
The exponential of correlation matrix converts the local property of correlation between edges into global property. The sum of mean and standard deviation of the exponential matrix gives us a dynamic threshold. Features having correlation value higher than the threshold value form the feature graph. So, we have a subset of the features which are having a high probability of redundancy and thus can be eliminated resulting in a comparatively lesser number of vertices.
- Step 3: Identify representative features
The principles of minimal vertex cover and maximal independent sets have been used to find out reduced number of vertices. If two vertices are similar,

they have an edge between them in the feature association graph, either using minimal vertex cover or maximal independent set, we can take any one of those two vertices having higher degree of correlation. The selected vertex will act as the representative of its adjacent vertices. Thus from a group of similar vertices, we select a few and eliminate rest of the vertices which are the neighbors of the selected vertex. In this way, using minimal vertex cover or maximal independent set, we will get a subgraph of original feature association graph having lesser number of features with least redundancy.

- Step 4: Re-select relevant features

It may happen that a feature is not related to any other feature. So, it is not present in the feature graph because of low value of correlation coefficient with other features but contributes a lot in predicting the actual result. Thus, we can't ignore those features as they play a vital role in predicting the final output value. The property of mutual information has been used for this purpose. Suppose we have "m" features which are not the part of feature graph. The mutual information of these features with class labels has been found. The features have been ranked based on the calculated mutual information value, and some percentage of these m features are selected from top as per as their ranking. In this way, those relevant features which were initially ignored also gets selected.

The two subsets generated in steps 3 and 4 are combined and we get a feature set having lesser redundancy and also more relevant features, which is the desired goal of feature selection.

Algorithm: Feature Selection based on Correlation Exponential (FSCE)

Input: N-dimensional data set D_N , having original feature set $O = f_1, f_2, \dots, f_N$.

Output: S_{OPT1}, S_{OPT2} (Optimal Feature Subsets)

Begin

// Part 1 : Computing the similarity matrix and drawing FSCE

1: $M_{corr} \leftarrow |(correlation(DN1))|$ // M_{corr} = correlation matrix

2: $M_{exp} \leftarrow \text{expm}(M_{corr})$ // M_{exp} = exponential of correlation matrix

3: Threshold = mean (M_{exp}) + sd (M_{exp}) // sd = standard deviation

4: For each M_{corr} [col-count, row-count]

5: If col-count == row-count

6: Set M_{corr} [col-count, row-count] = 0

7: If M_{corr} [col-count, row-count] >= Threshold

8: Set M_{corr} [col-count, row-count] = 1

9: Else

10: Set M_{corr} [col-count, row-count] = 0

11: End If

12: Next

13: $G_{FSCE} \leftarrow \text{adjacency-graph}(M_{corr})$

```

14:  $G_{NFSC E} \leftarrow \text{column}(D) \text{ vertex}(G_{FSCE})$ 
// Part 2 : Deriving the subgraph of FSCE
15:  $V_{VERTCOV} \leftarrow \text{Vert-cover}(G_{FSCE})$ 
16:  $V_{INDSET} \leftarrow \text{Ind-set}(G_{FSCE})$ 
17:  $S_{VC} \leftarrow D[V_{VERTCOV}]$ 
18:  $S_{IS} \leftarrow D[V_{INDSET}]$ 
// Part 3 : Selecting the subset for non-FSCE features
19:  $D_{NFSC E} \leftarrow D[\text{vertex}(G_{NFSC E})]$ 
20: For each field in  $D_{NFSC E}$ 
21:  $F_{MI} \leftarrow \text{Mutual-Info}(\text{field}, \text{class})$ 
22: Rank ( $F_{MI}$ )
23:  $S_{NFSC E} \leftarrow \text{top}(\beta * \text{count}(F_{MI}) \text{ features})$ 
24: Next
// Part 4 : Selecting the final feature subset
25.  $S_{OPT1} \leftarrow S_{VC} + S_{NFSC E}$ 
26.  $S_{OPT2} \leftarrow S_{IS} + S_{NFSC E}$ 
End

```

5 Methods and Materials

5.1 Data Sets Used

For conducting the experiments, 16 standard data sets, details provided in Table 1, from the of UCI machine learning repository (University of California, Irvine) [14] have been used. Out of the 16 data sets, 14 data sets have number of features less than 100 while the remaining 2 data sets have number of features greater than 100. This allows to have an understanding of the relative performance of the algorithms in low as well as high-dimensional data sets. The data set is split between training and test data at a 7:3 ratio. The computing environment R along with its various libraries has been used for the experiments. All data sets that have been used have class information that allows verification of the classification results.

5.2 Competing Benchmark Algorithms

The proposed FSCE algorithm has been compared with three benchmark algorithms for supervised feature selection. These algorithms are:

- Correlation-based feature selection (CFS) [15], a correlation-based filtering method for identifying correlated features and redundancy between related features

Table 1 UCI data sets analyzed

Data set	# of features	# of classes	# of instances
Apndcts	7	2	106
Btissue	9	6	106
Cleaveland	13	5	297
CTG	34	10	2126
Ecoli	7	8	336
Glass	9	6	214
ILPD	10	2	579
Madelon	500	2	2000
Mfeat	649	10	2000
Pgblk	10	5	5473
Sonar	60	2	208
Texture	40	11	5500
Vehicle	18	4	846
Wbdc	30	2	569
Wine	13	3	178
Wiscon	9	2	682

- Minimum Redundancy Maximum Relevance (mRMR) [16] uses mutual information values between individual feature and class as the primary parameter to identify feature subset having minimal redundancy and maximal relevance.
- Dominant set clustering algorithm (DSCA) [8] based on graph-theoretic approach of dominant set clustering which also uses mutual information between features in the form of multidimensional interaction information as the feature selection criterion.

5.3 Evaluation Criteria

The experiments have been done to evaluate the performance of the proposed algorithm compared to the benchmark algorithms with respect it is a measure for the efficiency of any algorithm. For accuracy, the measurement has been done as follows: $\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{TN} + \text{FP} + \text{FN})$, where TP, FP, TN, and FN represent the number of true positives, false positives, true negatives, and false negatives, respectively. For measuring the classification accuracy with the derived feature subset from each algorithm, decision trees have been used. Decision tree is a basic classification model which can be used to test the efficacy of any feature selection algorithm.

6 Results and Analysis

The experiments are conducted with $\beta = 0.1, 0.2, 0.5,$ and $0.8,$ respectively. It was observed that at $\beta = 0.8,$ the performance is quite comparable with that obtained using state-of-the-art algorithms (Fig. 1).

In Table 2, the best precision values obtained by FSCE at $\beta = 0.8$ and the accuracy values using CFS, mRMR, and DSCA have been introduced. For FSCE minimum

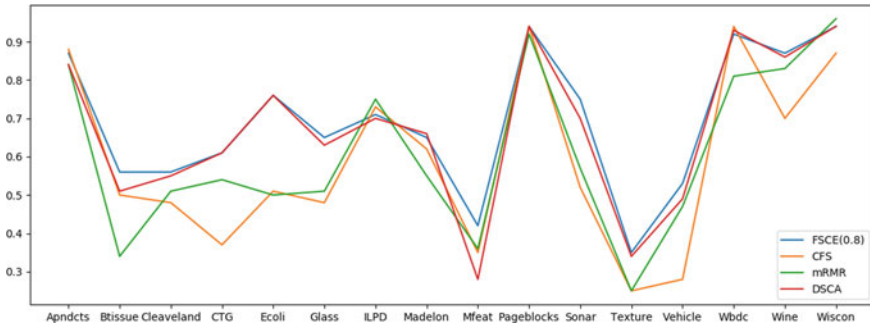


Fig. 1 Comparison of performance of algorithms for different data sets

Table 2 Comparison of accuracy values

Dataset	FSCE	Winning approach	CFS	mRMR	DSCA
Apndcts	0.85	MVC	0.88	0.84	0.84
Btissue	0.52	MVC	0.5	0.34	0.51
Cleaveland	0.56	MIS	0.48	0.51	0.55
CTG	0.62	Both	0.37	0.54	0.61
Ecoli	0.77	MIS	0.51	0.5	0.76
Glass	0.65	MVC	0.48	0.51	0.63
ILPD	0.71	MIS	0.73	0.75	0.7
Madelon	0.65	MVC	0.62	0.55	0.66
Mfeat	0.38	MVC	0.35	0.36	0.28
Pgblk	0.95	MIS	0.94	0.92	0.94
Sonar	0.71	MIS	0.52	0.57	0.7
Texture	0.33	MIS	0.25	0.25	0.34
Vehicle	0.54	MIS	0.28	0.47	0.49
Wbdc	0.93	MVC	0.94	0.81	0.93
Wine	0.87	MIS	0.7	0.83	0.86
Wisconsin	0.95	MIS	0.87	0.96	0.94
Mean accuracy	0.69	–	0.59	0.61	0.68

vertex cover (MVC) and another method using the maximal independent set (MIS), the result of the winning approach has been captured in Table 2.

It is observed, from Table 2 that:

- For 10 of the 16 data sets, the performance of the FSCE is better than the three benchmark algorithms CFS, mRMR, and DSCA.
- The rest, in five data sets, the performance of FSCE is better than at least two of all features and benchmark algorithms—CFS, mRMR, and DSCA.
- Even in the case of the remaining 1 data set, the performance of the FSCE is better than at least one of the benchmark algorithms—CFS, mRMR, and DSCA.
- The average accuracy over all data sets obtained from FSCE is better than the benchmark algorithms—CFS, mRMR, and DSCA.
- Among the 6 data sets, MVC is the winning method, and for 9 data sets, MIS is the winning method and 1 has the tie.

7 Conclusion

In this paper instead of using pairwise correlation, sum over all paths is used for supervised feature selection, using matrix exponential, along with data-driven thresholds. We found that this works better than benchmarks on certain data sets, showing the value of the expected correlation value, which is also intuitively better than using a simple pairwise value. Future work will explore the optimal threshold for feature selection.

References

1. E.R. Dougherty, J. Hua, W. Tembe, Performance of feature-selection methods in the classification of high-dimension data. *Pattern Recogn.* **42**, 409–424 (2009)
2. M. Dash, H. Liu, Feature selection for classification. *Intell. Data Anal.* **1**, 131156 (1997)
3. G.H. John, R. Kohavi, Wrappers for feature subset selection. *Artif. Intell.* **97**, 273–324 (1997)
4. P.N. Koch, T.W. Simpson, J.K. Allen, F. Mistree, Statistical approximations for multidisciplinary design optimization: the problem of size. *J. Aircr.* **36**(1), 275286 (1999)
5. G. Li, S.-W. Wang, C. Rosenthal, H. Rabitz, High dimensional model representations generated from low dimensional data samples. I. mp-Cut-HDMR. *J. Math. Chem.* **30**(1), 130 (2001b)
6. G.H. John, R. Kohavi, K. Peger, Irrelevant features and the subset selection problem, in *ICML* (1994)
7. S.E. Schaeffer, Graph clustering. *Comput. Sci. Rev.* **1**, 2764 (2007)
8. Z. Zhang, E.R. Hancock, A graph-based approach to feature selection, in *GBRPR* (2011)
9. S. Bandyopadhyay, T. Bhadra, P. Mitra, U. Maulik, Integration of dense subgraph mining with feature clustering for unsupervised feature selection. *Pattern Recogn. Lett.* **40**, 104112 (2014)
10. Q. Song, J. Ni, G. Wang, A fast clustering-based feature subset selection algorithm for high-dimensional data. *IEEE Trans. Knowl. Data Eng.* **25**, 114 (2013)
11. P. Moradi, M. Rostami, A graph theoretic approach for unsupervised feature selection. *Eng. Appl. Artif. Intell.* **44**, 3345 (2015a)

12. A. Strehl, J. Ghosh, Cluster ensembles a knowledge reuse framework for combining multiple partitions. *J. Mach. Learn. Res.* **3**, 583617 (2002)
13. T.M. Cover, J.A. Thomas, *Elements of Information Theory* (Wiley, New York, USA, 2012)
14. M. Lichman, UCI machine learning repository (2013). <http://archive.ics.uci.edu/ml>
15. M.A. Hall, Correlation-based feature selection for machine learning
16. H. Peng, F. Long, C.H. Ding, Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **27**, 1226–1238 (2005)
17. A.K. Das, S. Goswami, B. Chakraborty, A. Chakrabarti, A graph-theoretic approach for visualization of data set feature association, in *ACSS* (2016)
18. M. Dash, H. Liu, Feature selection for clustering, in *PAKDD* (2000)
19. N. Deo, *Graph Theory with Applications to Engineering and Computer Science* (PHI, 1979)
20. S. Goswami, A. Chakrabarti, B. Chakraborty, An efficient feature selection technique for clustering based on a new measure of feature importance. *J. Intell. Fuzzy Syst.* 112
21. S. Goswami, A.K. Das, A. Chakrabarti, B. Chakraborty, A feature cluster taxonomy based feature selection technique. *Expert Syst. Appl.* **79**, 7689 (2017)
22. P. Moradi, M. Rostami, Integration of graph clustering with ant colony optimization for feature selection. *Knowl. Based Syst.* **84**, 144161 (2015b)
23. Z. Zhang, E.R. Hancock, Hypergraph based information-theoretic feature selection. *Pattern Recogn. Lett.* **33**, 19911999 (2012)
24. E. Estrada, J.A. Rodriguez-Velzquez, Subgraph centrality in complex networks. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **71**(5 Pt 2), 056103 (2005)

A Sentiment-Based Hotel Review Summarization



Debraj Ghosh

Abstract Nowadays, with the growth of information and data provided in Internet, it becomes too difficult for a user to read and understand all the reviews from huge amount of reviews. In today's world, we purchase products, book movie tickets, book train tickets, book hotel rooms, and buy products from different websites. Users also share their views about product, hotel, news, and other topics on Web in the form of reviews, blogs, etc. We can found some basic reviews in user review and also can find user own opinions about the experience with various products. Many users read the reviews of the information given on the Web to take decisions such as buying products, watching movie, going to restaurant, etc. It is difficult for Web users to read and understand the contents from a large number of reviews. The important and useful information can be extracted from the reviews through opinion mining and summarization process. We obtained about 78.2% of accuracy of hotel review classification as positive or negative review by machine learning method. The classified and summarized hotel review information helps the Web users to understand the review contents easily in a short time.

Keywords Summarization · Products · Results · Review

1 Introduction

In today's world to find any products or services from Internet, users create a huge amount of online reviews. Reviews can be text or rating for a particular product. Sometimes the criteria of users are mentioned in user's review, so there is a chance that many information which is provided by a user is mentioned in the reviews but may not include in a user's explanation. These user reviews have large contribution for making today's Internet world and e-commerce a huge successful, but users are facing the problem when they have to take their own view for making a right decision

D. Ghosh (✉)
Institute of Engineering and Management, Kolkata, India
e-mail: hellodebraj@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_5

from a large scale of reviews which holds both regular basis comments about market production as well as various opinions and sentiments of reviews.

Now the problem is to create an automatic analysis from a large scale of review sets to generate a summarized review system which will understand by user effortlessly and quickly. Now, we are going to discuss various issues about the review summarization in accommodation domain. Automatic summarization is an activity of drawing out the applicable information from an origin to generate the precipitated version, occasionally prejudiced toward particular persons and problems. Summarization can be categorized into two parts. Extractive summarization means content reduction is directed by selection, and abstractive summarization means compression is completed through generalization using most appropriate information from various origins.

2 Related Work

Opinions which are extracting from text can present many natural language processing challenges. One of the most challenging tasks of natural language processing is sentiment analysis. Many techniques have been introduced and widely studied for natural language processing. Wei et al. proposed hierarchical learning process with a defined sentiment ontology tree (HL-SOT) approach to label a product's attributes and their associated sentiments in product reviews. Detail work on sentiment analysis has been developing on binary classification of positive and negative point of views. The aspect rating inference has been previously analyzed by Pang and Lee [13], Goldberg and Zhu [11], Leung et al. [12]. The inference procedure of multi-aspect ratings where reviews containing numerical ratings from mutually dependent aspects has been extended by Snyder and Barzilay [16]. Snyder and Barzilay [16] introduced that modeling of the dependencies among various aspect ratings in the same review can help to reduce the rank loss [9]. Mahajan et al. [10] have also attempted summarization of restaurant reviews by determining high-information snippets inside a review.

3 Problem Identification

While summarization of well-structured technical details is well-embedded skill in today's e-commerce, summarization of review sets from user reviews is a new tendency. In these conditions, the job may be structured imperfectly, and the information and contents from authors (e.g., age, sex) full of subjective matters indicated through opinion, metaphor, or cultural references. Extracting result from a given database illustrates the diversity of different reviews for the same source. With the huge uses of the Web, nowadays online review is becoming more useful and important resource

for all type of users. Generally, different Web sites (like makemytrip.com, yatra.com) ask their users to give their reviews and share their experience.

4 Proposed Method

Opinion words present in reviews collected after feature extraction and pruning will be used by the system to generate appropriate summarized reviews for hotels for the ease of users. It summarizes huge amount of hotel reviews available online and provides users an easy way to access those shortened reviews. This also helps service providers to improve their resources with better summarized feedback.

4.1 Methodology

Feature Extraction Process

For extracting the features, word distribution model wherein the words along with their frequency of occurrence are stored has been used.

Algorithm

1. Collection of hotel reviews from online Web site <https://www.tripadvisor.in/> by the means of Web scraping.
2. Matching sentiment scores like good or bad in the reviews with relative position against topics like room, service, food, etc.
3. Analyzing results based upon scores like positive, negative, neutral, or compound. Using analysis results to calculate a threshold value.
4. Computing graph with sentences satisfying that value.
5. Finding out most common sentences in the review for summary generation.
6. Generating response to user-based queries in form of natural language.

Summaries generated are pretty good in terms of readability. User gets an overall glimpse of hotel based on only ten lines of reviews generated by the system. Even occurrence of one or two negative remarks out of ten sentences would give a fair idea that the actual review text contains about 300–500 negative comments. Efficiency of current algorithm is still low (~78.2% on average) based on the average weighted F-score value calculated using ROUGE. Some sentences which are most common of all occur more than once in summary, which needs to be eliminated. Some of the summaries contain two or more lines with same meanings but in different composition, which may be avoided.

4.2 Formulae Used

For a review r , with sentences pairs as sent_1 and sent_2 , an edge is connected between the nodes representing those sentences when the weight is more than or equal to a threshold value (0.07).

$$\bar{x} = \frac{\sum_{i=1}^n (X_i * W_i)}{\sum_{i=1}^n W_i}$$

Calculation of weights of each node when the whole directed graph is formed is done by advanced PageRank technique, where the weight of a node is determined by the weights of edges pointing to that node. This gives a list of all sentences and their corresponding weights.

$$\text{Page Rank of site} = \sum \frac{\text{Page Rank of inbound link}}{\text{Number of links on that page}}$$

The efficiency of a summary is calculated against 100 reference summaries generated by previous computational algorithms based on text summarization. This is calculated by ROUGE metric tool implementation in Java, using recall, precision, and F-score values.

$$\text{Precision} = \frac{\text{tp}}{\text{tp} + \text{fp}}$$

$$\text{Recall} = \frac{\text{tp}}{\text{tp} + \text{fn}}$$

5 Results and Discussion

5.1 Improving Accuracy

The efficiency can be improved by generating more than one summaries using different threshold weights while building graph and then finding sentences with most common relativity between those summaries. While generating a summary, if a node occurs which contains the same text or short title, already encountered it can be skipped for avoiding collision. If two sentences contain the same lexical inclination, i.e., they have common words with more frequency, one might be omitted.

6 Conclusion

Review summarization provides a better understanding to users about what they are looking for in the hotel. We can also rate the hotels with the help of reviews, i.e., how is the hotel in terms of the location and other qualities, which can help the users to take his/her own decision of which hotel to choose as their criteria. That is also helpful to the hotel management because they will now be well informed of which areas they need to improve and what their strong points are. This can be further extended to services like movies, colleges, and online products and helps service providers to improve their resources with better feedback. Users can get all the required information about hotels and they can save their time. This can also be used to generate more personalized reviews relating to a particular topic like room, service, food, and also list of hotels by their price and rating.

Websites

1. <http://www.aaai.org/Papers/AAAI/2004/AAAI04-119.pdf>
2. <http://ieeexplore.ieee.org/document/1250949/>
3. <https://arxiv.org/abs/1603.06042>
4. <https://arxiv.org/abs/1609.03499>
5. ilpubs.stanford.edu/422/1/1999-66.pdf

References

6. D. Jurafsky, J.H., Martin Speech Language Processing
7. Globally Normalized Transition-Based Neural Networks by
8. WaveNet: A Generative Model for Raw Audio by
9. K. Crammer, Y. Singer, Pranking with ranking, in *Neural Information Processing Systems: Natural and Synthetic (NIPS)*, ed. by T.G. Dietterich, S. Becker, Z. Ghahramani (MIT Press, Vancouver, British Columbia, Canada, 2001), pp. 641–647
10. M. Mahajan, P. Nguyen, G. Zweig, Summarization of multiple user reviews in the restaurant domain. Technical Report (MSR-TR-2007-126), Sept 2007
11. A.B. Goldberg, J. Zhu, Seeing stars when there aren't many stars: graph-based semisupervised learning for sentiment categorization (2006)
12. C.W. Leung, S.C. Chan, F. Chung, Integrating collaborative filtering and sentiment analysis: a rating inference approach, in *Proceedings of The ECAI 2006 Workshop on Recommender Systems*, Riva del Garda, vol. I (2006), pp. 62–66
13. B. Pang, L. Lee, Seeing stars: exploiting class relationships for sentiment categorization with respect to rating scales, in *Proceedings of the Association for Computational Linguistics (ACL)* (2005), pp. 115–124
14. B. Pang, L. Lee, S. Vaithyanathan, Thumbs up? Sentiment classification using machine learning techniques, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2002), pp. 79–86
15. K. Shimada, T. Endo, Seeing several stars: a rating inference task for a document containing several evaluation criteria, in *Advances in Knowledge Discovery and Data Mining, 12th Pacific*

- Asia Conference, PAKDD 2008*, volume 5012 of Lecture Notes in Computer Science (Springer, Osaka, Japan, 2008), pp. 1006–1014
16. B. Snyder, R. Barzilay, Multiple aspect ranking using the Good Grief algorithm, in *Proceedings of the Joint Human Language Technology/North American Chapter of the ACL Conference (HLT-NAACL)* (2007), pp. 300–307
 17. P. Turney, Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews, in *Proceedings of the Association for Computational Linguistics (ACL)* (2002), pp. 417–424
 18. H. Yu, V. Hatzivassiloglou, Towards answering opinion questions: separating facts from opinions and identifying the polarity of opinion sentences, in *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2003)
 19. R. Malloy, Internet and Personal Computing Abstracts: IPCA, Volume 22 Issues, Information Today, Incorporated (2001)
 20. J. Pan, S. Chen, N. Nguyen, Intelligent information and database systems, in *4th Asian Conference, ACIIDS, Proceedings Part 2*, Kaohsiung, Taiwan (2012)

Heuristic Approach for Finding Threshold Value in Image Segmentation



Sandip Mal and Ashiwani Kumar

Abstract Image segmentation is a major challenge in the field of image processing. Proper segmentation of medical image is difficult because of the size and position of the inhomogeneity in the tissues, blood flow and similar contrast to the wall. Thus, an automated approach is required to develop proper segmentation of the region of interest (ROI). This paper describes a *Heuristic Image Segmentation Algorithm* based on genetic algorithm optimization. OTSU algorithm is used for optimization. Threshold value of proposed method can be used for finding the ROI. A comparison has been made with other thresholding algorithms for supporting the better results.

Keywords Image segmentation · Medical images · Feature extraction · Heuristic · Genetic · Optimization · Static · Dynamic · OTSU

1 Introduction

Segmentation process of a digital image separates the image into multiple segments, i.e., the set of pixels [1, 2]. The main objective of segmentation is to set up the representation of an image into a more meaningful way for easy analysis. Image segmentation is generally applied to identify the region of interest, objects, and boundaries like lines, curves, corners, etc. of an image. In another way, this process allocates some label to each pixel of an image. Research on medical images is most important now a day for assisting health care and allows health care to provide a better diagnosis. Therefore, medical images segmentation is a challenging part of Medical Image Analysis. The segmentation of medical images requires different algorithms and different procedure to segment and classify of an image [7–10]. For a given set of goal, heuristic approach is useful to get a guaranteed optimal solution but have some

S. Mal (✉)
VIT Bhopal University, Bhopal, India
e-mail: sandip.mal@vitbhopal.ac.in

A. Kumar
Central University of Jharkhand, Ranchi, India

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_6

difficulties. It is an experience-based technique for problem-solving, learning, and discovery. Genetic algorithm (GA) is an optimization technique [11] based on the evolution of a population of solutions. GA operates on of constant size population which is formed by chromosomes. Populations are the solution of a problem and chromosomes are the representation of coding of the problem. Genetic operators like selection, crossover, and mutation are applied to create a new population. Selection process selects the most pertinent candidates. In the crossover, two new chromosomes find from two old or parent chromosomes. Mutation is the process of inversion from one or several genes into a chromosome.

2 Related Work

Several segmentation techniques have been used in literature to study the heart disease and diagnosis of different heart ailments. Medical image segmentation can be classified in following approaches [3–5]. It includes

- Threshold-based segmentation by selecting a threshold value.
- Edge-based and region-based approach separates the regions.
- Pixel-based classification separates the image based on intensities of pixels.
- Atlas-guided approaches.
- Active shape and appearance model (ASM/AAM).

Above first three techniques are the weak prior image-driven technique which involves a limited framework. These segmentation techniques are used very less and low-level information. Atlas-guided and active shape and appearance approaches provide details, versatile framework can be used for weak or strong prior or both. These segmentation techniques keep strong and high-level information. Atlas-guided segmentation is used to set of manually segmented images. Region-based image segmentation requires a threshold value to partition the regions. This method is simplest but most powerful technique for identifying the characteristics of the image [6]. Basic concept of this technique identifies the lighter intensity pixels of the image than the darker background or vice versa. Another basic and simplest segmentation technique is static thresholding. This segmentation technique is applied on grayscale image and converts the image into black or white based on static threshold value [12]. Dynamic thresholding process finds global threshold value and attempts to calculate the optimal threshold value from the gray-level values of the image. Line intersection approach for thresholding finds the highest intensity pixel point and second highest pixel point of images. OTSU's thresholding algorithm finds the threshold value that minimizes or maximizes the weight within class variance [13–15]. The algorithm works on directly on the grayscale image for first computation (once the histogram is computed).

Steps for Algorithm:

1. Draw histogram and find probability of each intensity level.

2. Compute the mean of the intensity of the pixels from histogram.
3. Compute cross-class variance (maximum), i.e., $V(T)$

$$\begin{aligned}
 V(T) = & \text{Probability of one intensity level} \\
 & [\text{Mean of one intensity level} - \text{Total mean}]^2 \\
 & + \text{Probability of remaining intensity level} \\
 & [\text{Mean of remaining intensity level} - \text{Total mean}]^2
 \end{aligned}$$

A limitation with this algorithm is that it has to take every intensity level to find mean and probability which becomes very complex to get the maximum cross-class variance. Hence need to implement a feasible algorithm who gives us result to use suitable intensity level for OTSU’s Thresholding.

3 Proposed Work

Proposed work involved multiple phases. Initially, color image is taken as input into the system. The input image has been preprocessed by cropping, resizing and converting into grayscale image in the next phase. Thirdly, fitness function is calculated for using in GA tool. The output of a fitness function produces threshold value. The grayscale image has been segmented using threshold value to find out the desired output. Figure 1 clearly shows the flow of work.

Proposed work contains the implementation of genetic algorithm (GA) in OTSU algorithm to find the optimum minimum interclass variance. The solution of GA provides the threshold value and the minimum value from fitness function. For this method, the fitness function for GA is defined as follows

$$y = \frac{1}{\text{prob_left}(\text{global_mean} - \text{left_mean})^2 - \text{prob_right}(\text{global_mean} - \text{right_mean})^2}$$

Consider T as threshold value of histogram of gray image. prob_left is the probability of intensity level in left of T . prob_right denotes the probability of remaining intensity levels. global_mean indicates total mean of intensity of pixels. left_mean is the mean of intensities of pixels from left of T and right_mean is the mean of intensity

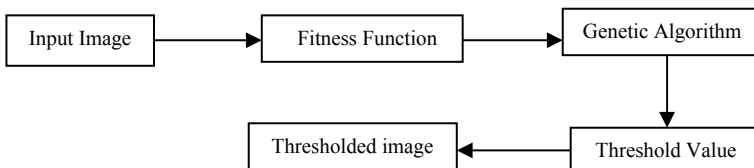


Fig. 1 Flow chart of proposed method

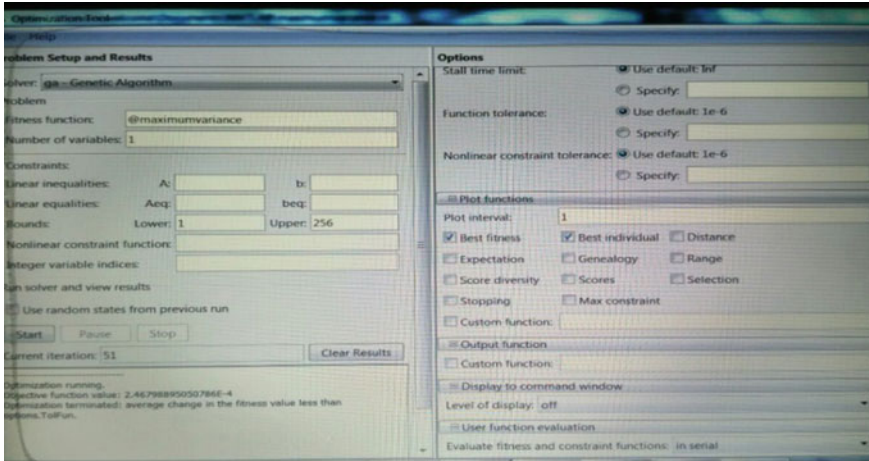


Fig. 2 GA output which give objective function value for default iteration

of pixels from right of T . “ y ” is the threshold value after using the GA optimization tool in MATLAB and threshold value for the above fitness function. The optimum threshold value has been taken as input for the OTSU algorithm which gives the maximum interclass variance. OTSU algorithm checks the complexity of maximum interclass variance for each gray value of histogram. But the proposed method produced a threshold value for which there is minimum value for the function. Therefore, value of y is the optimum minimum interclass variance. So, maximum interclass variance can be produced by taking inverse of OTSU equation. The minimum value of y actually denotes the maximum value for OTSU algorithm. Figure 2 shows the result which displayed the value of y and threshold value T for the given fitness function for images.

The equation of y is named as maximum_variance, i.e., the fitness function for GA tool in MATLAB and shown in Fig. 2. Only threshold value T is unknown for function maximum_variance. Therefore, the range of intensity value of single known variable of the grayscale image is from 1 to 256. Other parameters in GA tool are set to default value. The objective function value and corresponding graph of each best fitness and best individual can be obtained clearly in Fig. 3. The graph is used to find out best individual value of required threshold. Based on the threshold value, image is segmented.

3.1 Result of GA Optimization

For all default values in GA tool, the result will such as.

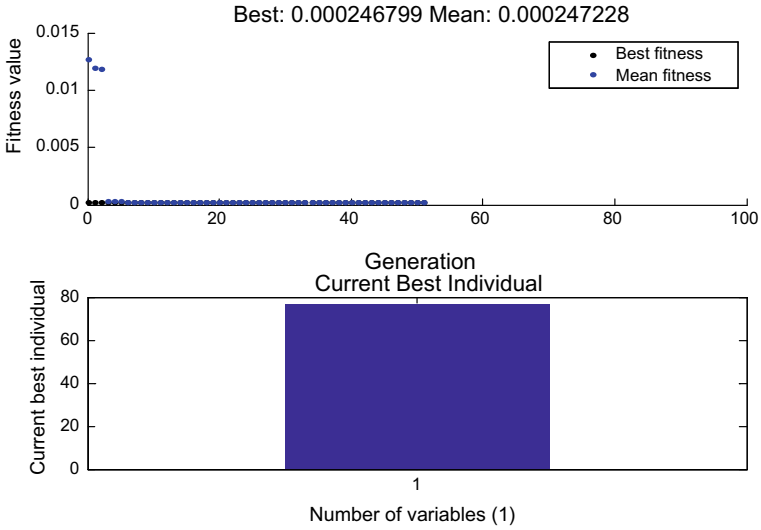


Fig. 3 GA output which give fitness value and best value for variable, i.e., for Threshold (T)

4 Result and Discussion

Ten medical images which are shown in Fig. 4 have been considered for our experiment. This section clearly shows segmentation results for each of the methods discussed above and the proposed method on the medical images. Table 1 shows the threshold values obtained by different algorithms. Existing algorithm and proposed algorithm has been applied on the collected image. Five output images are shown in Figs. 5, 6, 7, 8 and 9. In first row, first column is the grayscale image 1, second column is the output image by applying static thresholding, and third column is the output image by applying dynamic thresholding. In second row, first column is the output image by applying line intersection method thresholding, second column is the output image by applying OTSU thresholding, and third column is the output image by applying proposed thresholding method. Results conclude that proposed method for thresholding is the better algorithm because of producing the good segmentation. However, running time of genetic algorithm takes more time to find out the threshold value, but it guarantees to produce better threshold value for segmentation of images. Here, we see that proposed algorithm result is more accurate compared to other algorithms for thresholding.

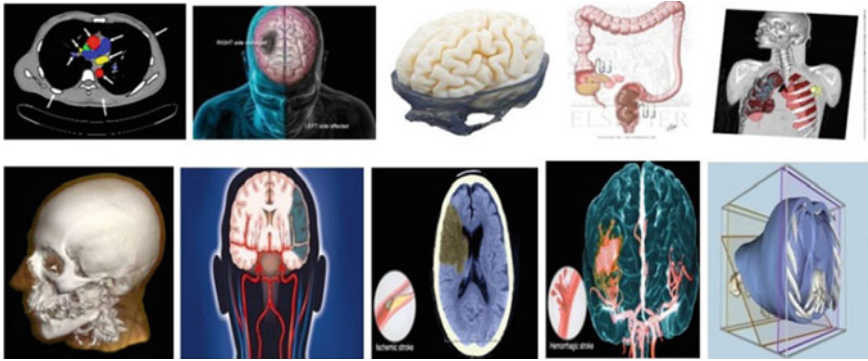


Fig. 4 List of medical images for segmentation

Table 1 Comparison of all used image segmentation algorithms

Image No.	Static threshold value (T)	Dynamic threshold value (T)	Line intersection method threshold value (T)	OTSU threshold value (T)	Proposed method threshold value (T)
1	128	101	158	96	86
2	128	75	26	64	157
3	128	163	254	160	162
4	128	204	254	192	100
5	128	114	101	128	114
6	128	101	2	96	116
7	128	124	136	128	118
8	128	86	30	96	154
9	128	106	52	96	126
10	128	158	199	160	73

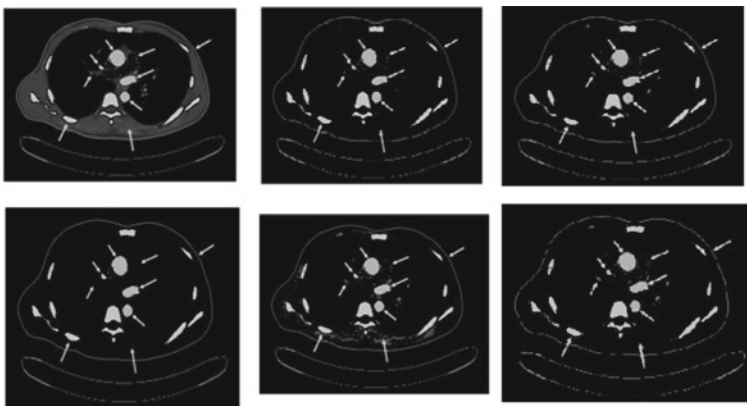


Fig. 5 Output image set 1 for input image number 1

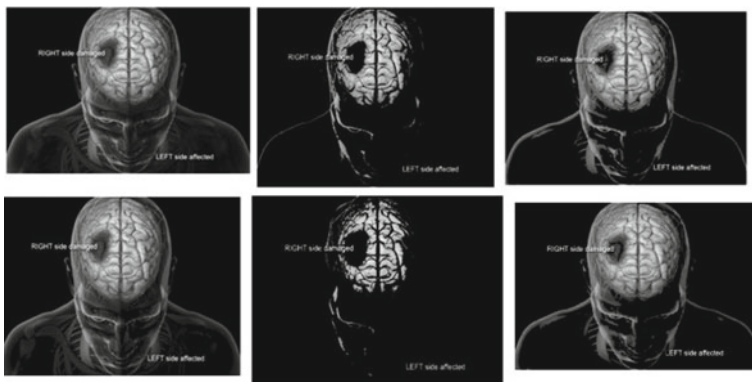


Fig. 6 Output image set 2 for input image number 2

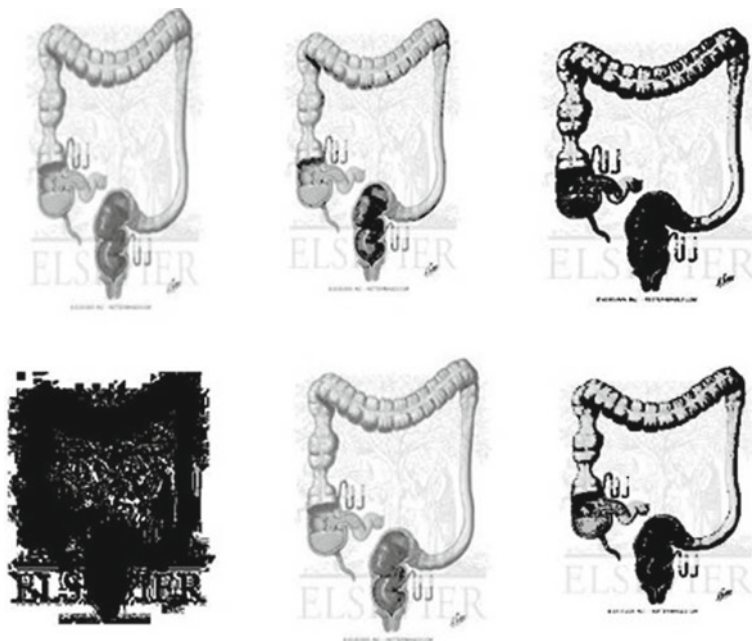


Fig. 7 Output image set 3 for input image number 4

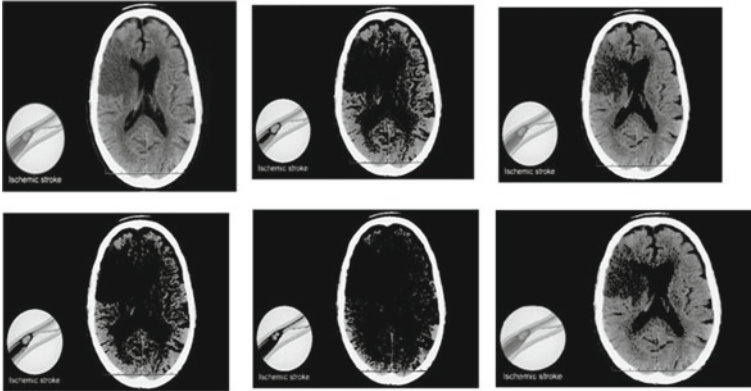


Fig. 8 Output image set 4 for input image number 7

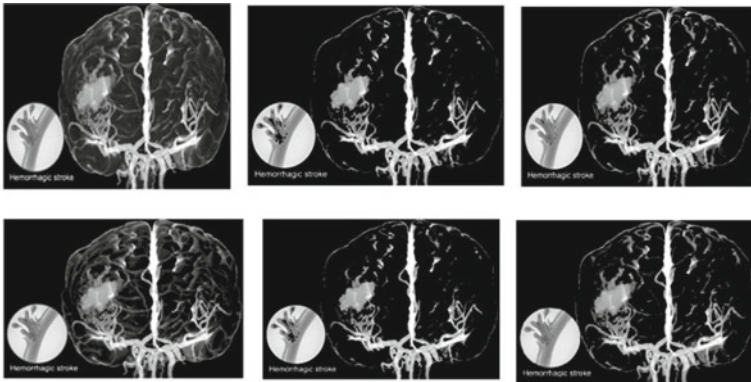
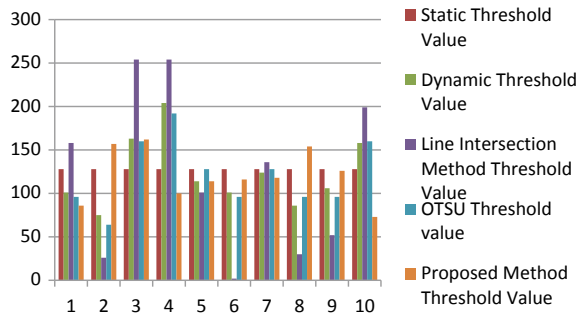


Fig. 9 Output image set 5 for input image number 8

5 Conclusion and Future Work

This paper demonstrates the modification in image thresholding. The dynamic thresholding can be used to fulfill the requirements of large-scale performance evaluation of the image segmentation. The result gives a clear probabilistic interpretation by proposed method. Results are correctly normalized to find out similarity in a set of real images and can be computed efficiently for many numbers of images. After successfully segmenting the image, the contours of objects can be extracted using edge detection. Image segmentation techniques are extensively used in similarity searches. Figure 10 shows the comparative graph of different threshold values found from different technique. The proposed work can be applied on the robustness of color images and segmentation of the color images based on red, green and blue color. Combination of three threshold images (based on red, green, and blue) may be used to find out a better segmented image.

Fig. 10 Graph showing the change in threshold value for image segmentation algorithms



References

1. R.C. Gonzalez, R.E. Woods, Digital image processing. Image Segmentation (2008), pp. 589–658
2. R.C. Gonzalez, R.E. Woods, S.L. Eddins, Digital image processing using MATLAB (2009), pp. 338–339
3. M.R. Kaus, S.K. Warfield, A. Nabavi, P.M. Black, F.A. Jolesz, R. Kikini, Automated segmentation of MR images of brain tumors. *Radiology* **218**, 586–591 (2001)
4. M. Tabb, N. Ahuja, Unsupervised multiscale image segmentation by integrated edge and region detection. *IEEE Trans. Image Process.* **6**(5), 642–655 (1997)
5. D.L. Pham, C. Xu, J.L. Prince, Current methods in medical image segmentation. *Annu. Rev. Biomed. Eng.* **2**, 315–337 (2000)
6. A.K. Jain, M.N. Murty, P.J. Flynn, Data clustering: a review. *ACM Comput. Surv.* **31**(3), 264–323 (1999)
7. H. Khotanlou, J. Atif, O. Colliot, I. Bloch, 3D brain tumor segmentation using fuzzy classification and deformable models, in *WILF* (2005), pp. 312–318
8. M. Shasidhar, V.S. Raja, B.V. Kumar, MRI brain image segmentation using modified fuzzy C-means clustering algorithm, in *CSNT* (2011), pp. 473–478
9. R.J. Oweis, M.J. Sunna, A combined neuro fuzzy approach for classifying image pixels in medical applications. *J. Electr. Eng.* **56**, 146–150 (2005)
10. S. Murugavalli, V. Rajamani, An improved implementation of brain tumor detection using segmentation based on neuro fuzzy technique. *J. Comput. Sci.* 841–846 (2007)
11. M. Sharma, S. Mukharjee, Brain tumor segmentation using genetic algorithm and artificial neural network fuzzy inference system (ANFIS), in *Advances in Intelligent Systems and Computing* (2012), pp. 329–339
12. K. Zuiderveld, Contrast limited adaptive histogram equalization, in *Graphic Gems IV* (Academic Press Professional, San Diego, 1994), pp. 474–485
13. L. Dongju, Y. Jian, OTSU method and k-means, in *Hybrid Intelligent Systems, 2009, HIS '09, Ninth International Conference*, vol. 1 (2009), pp. 344–349
14. V.B. Langote, D.S. Chaudhari, Segmentation techniques for image analysis. *IJAERS I*(II) (2012) (January–March)
15. A. Fabijanska, A survey of thresholding algorithms on yarn images, in *MEMSTECH, International Conference* (2000), pp. 23–26

An Educational Chatbot for Answering Queries



Sharob Sinha, Shyanka Basak, Yajushi Dey and Anupam Mondal

Abstract The fast progress in the development of communication and information has made people very diverse in knowledge improvement, education, and learning methods. In this paper, we show how we can convert documents into the knowledge of a chatter robot. It helps users to make profit out of it by asking and answering queries, using electronic documents which are integrated with the simulated system. Hence, we are motivated to develop an educational chatbot system which provides a virtual assistant. Its main function is to streamline and to automate manual and administrative tasks while supporting other course-related activities. The aim of this research work is to develop a system which is automated and can provide an answer to a question asked by a user on behalf of a person, for educational purposes. We have focused on the local as well as Web databases to make the model user-friendly, interactive, and scalable.

Keywords Chatbot · Education · Machine learning · Question answering

1 Introduction

The development in the fields of communication and information technology has made Artificial Intelligence (AI) systems more complicated. Nowadays, the artificial intelligence systems are trying to imitate human actions such as taking decisions,

S. Sinha (✉) · S. Basak · Y. Dey · A. Mondal

Department of Computer Science & Engineering, Institute of Engineering & Management,
Y-12, Salt Lake Electronics Complex, Sector-V, Kolkata 700091, India

e-mail: sharobsinha@gmail.com

S. Basak

e-mail: shyankabasak@gmail.com

Y. Dey

e-mail: deyyajushi@gmail.com

A. Mondal

e-mail: anupam.mondal@iemcal.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,

https://doi.org/10.1007/978-981-13-7403-6_7

performing daily chores. In the field of artificial intelligence, there are some hybrid and some adaptive methods available which are making the systems complex.

By reading all the articles available on the Internet, these complex systems often known as Internet answering engines can learn by themselves and renew their knowledge. The users can enquire the system like they usually enquire another human being. Besides these Internet answering engines, presently, many applications have been introduced which are known as chatter robots or chatbots. These are often used for customer services in language learning, museum guides, various e-commerce Web sites, or chatting for entertainment purposes. It is like a virtual friend that are integrated into instant messengers, applications, or Web sites. The work is not very complex as the knowledge is already pre-programmed. To implement and design a natural intuitive human-computer interaction domain, a chatbot represents a virtual assistant which performs an automated task, using artificial intelligence to carry out near-natural conversation with users especially over the Internet. In the present time, chatbots understand users well and in turn, they can give responses like a fellow human being. Simple chatbots work on the basis of keywords that have been written in advance and that they understand. The developer has to write these commands separately by using regular expressions or any other form of string analysis. The chatbot then matches the input string from the speaker or user with that pattern that already exists in the database. Each of these patterns is then compared with the dataset of the chatbot. If the user asks a question but he/she does not use a single keyword, the bot fails to understand it and instead, replies like “Sorry, I did not understand”. However, a smart machine-based chatbot inherits its capabilities from artificial intelligence and cognitive computing and adapt its behavior based on the user interactions. The proposed system may assist in designing various applications such as relationship extraction and category-based classification systems in future [1–4].

The rest of the paper presents a detail background study related to chatbot application in Sect. 2. Sections 3 and 4 describe the proposed system and the obtained result from this system in details. Finally, Sect. 5 illustrates the concluding remarks and future scopes in this research.

2 Related Work

Bayan Abu Shawar and Eric Atwell, “A Comparison Between Alice and Elizabeth Chatbot Systems”, School of Computing, University of Leeds research report series is the first chatbot for designing to emulate a psychotherapist [5]. Thereafter, numerous chatbot systems were developed like ALICE1. Imran Ahmed and Shikha Singh also incorporate speech recognition to design modern chatbots.

Imran Ahmed et al., “AIML Based Voice Enabled Artificial Intelligence Chatbot” in an International Journal. The authors have employed speech recognition of Microsoft to detect voice and response of the chatbot [6]. Also, gaming system is used for designing chatbots but most of chatbots are not supported “knowledge learning mechanism” well.

The approach of Jizhou Huang et al., “Extracting Chatbot Knowledge from Online Discussion Forums” in an international conference has been discussed learning mechanism using extracting knowledge which not be able to answer for the question [7]. This approach aims to resolve the problem of time, limitation of knowledge, and the problem of typing by developing the system. This system is automatically extracting text for answering the question by itself.

In order to resolve this problem, the researchers have introduced natural language processing (NLP) which is presented as an important research arena in the domain of artificial intelligence [8–10]. The rule-based system has been introduced with the help of modern linguistics is Chomsky’s concept of generative grammar by describing all possible sentences [11–13]. This system has been developed using a lot of literatures on syntactic formalisms and Chomsky grammar notation-based parsing algorithms [14]. Finally, these parsers and grammars of sentences are decomposed into grammatical elements for extracting knowledge. Since, have a problem to understand the parsing output such as NP (Nouns), VP (Verbs), subjects, and objects for the system, which is very easy to understand by the human.

J. Weizenbaum designed ELIZA, a program which makes certain conversation between human and computer using natural language processing [15]. This program employed rule-based patterns to figure out the replies of the human’s statements. We have observed that this technique is still widely applied to build the modern chatbots. T. Winograd has proposed the concept of natural language understanding to develop a chatbot in his project “SHRDLU” [16].

3 Methodology

A chatbot is a service, powered by rules and sometimes artificial intelligence, that makes user interaction via a chat interface. The service could be any number of things, ranging from functional to fun, and it could live in any major chat product.

There are mainly two types of chatbots, one function based on a set of rules, and the other more advanced version uses machine learning.

Chatbot that functions based on rules:

- This bot is limited. It only responds to very specific commands. If the queries of the user do not match with any of the keywords in the chatbot’s dataset, the chatbot will fail to understand anything and will not give any reply.

Chatbot that functions using machine learning:

- This bot has an artificial brain (artificial intelligence). Not only can it understand commands but also language.
- This bot gets smarter day by day, learning from the previous conversations it has with various users.

In order to design the chatbot, we have used a machine learning algorithm over simple keyword or similar word pattern matching from a predefined database. The following subsections discuss the data preparation, feature extraction, and model building in details.

3.1 Dataset Preparation

A label data is the most important thing in the process of building an automated chatbot application, which we could present as a question answering system. Initially, we have collected around 1000 educational query-based conversations from different Web sites and by preparing manually. Since the data collected from various Web sites was unstructured, we have preprocessed them using various packages such as regex under python environment. The preprocessed data has been presented as an experimental dataset to build and validate the proposed chatbot. In the following subsection, we have discussed the feature extraction steps to develop the educational chatbot from the experimental dataset.

3.2 Feature Extraction

The feature extraction plays a crucial role in the process of developing a chatbot. We have extracted various features as Wh-words (e.g., what, when, why, where, and how), number of sentimental words, number of nouns, number of adjectives, etc. from 1000 pairs of questions and answers.

Thereafter, we have split the experimental dataset into two parts namely, training and test dataset. The training dataset contains 800 pairs of questions and answers whereas the rest of the 200 pairs of questions and answers are used as a test dataset. The extracted features of the training dataset have been applied through unsupervised machine learning to build the proposed chatbot system.

3.3 Module Building

In order to develop the educational chatbot, we have employed a well-known unsupervised clustering approach namely K-means. K-means helps to identify groups of data, which is extracting similar information. This approach assists in learning the module using a proper pair of questions and answers, which helps to predict a particular answer for a user-provided question. The following algorithm illustrates the module building steps in detail.

STEP-1: Initially, we have extracted various features (F) = [F1, F2, F3, ..., Fn] from the experimental dataset.

STEP-2: Thereafter, we have employed K-Means approach on the training dataset to assign the proper groups of questions and answers.

STEP-3: Initially, we have set the value of K as 2 and incremented the range to 5 due to the nature of our training dataset.

STEP-4: On the other hand, the test dataset has been applied to the learned module to validate the proposed chatbot.

Finally, the chatbot is able to provide the following output as Chatbot on the user-provided question as Visitor in a single environment.

Visitor: Is there any placements regarding with this course?
Chatbot: Yes, we provide complete assistance for the placements.

4 Result Analysis

In order to validate the proposed chatbot, we have used the K-means clustering approach. We have varied the K value from 2 to 5 and measure the accuracy in the form of standard evaluation matrices such as precision, recall, and F-Measure score. Table 1 presents the distribution of precision, recall, and F-measure score for the proposed system.

We have observed that among all the values of K, when the value of K = 4, the chatbot provides a better output with respect to different K-values. We have also noticed that the F-Measure score is varying in the range of 0.572–0.601, which is not so much impressive. This score motivates to enhance the range of our experimental dataset as well as features in future experiments. The proposed chatbot presents as a baseline system to develop an advanced chatbot in the educational domain.

Table 1 A comparative result analysis for various K-values under K-means

K-values	Precision	Recall	F-measure
2	0.562	0.583	0.572
3	0.569	0.588	0.578
4	0.600	0.602	0.601
5	0.597	0.599	0.598

5 Conclusion

The research was primarily focused on developing a chatbot in the educational domain. To prepare the chatbot, we have prepared a training data, which contains around 1000 pairs of questions. After that, we have identified various features from our dataset and processed through K-means clustering algorithm. The chatbot is able to respond most of the queries with an accuracy of 60.10%. The queries which are not answered by chatbot are answered by a human through e-mail.

In the future, this system can be extended for the banking sector by providing an additional layer of security or in foreign language instruction or in education for giving career advice to students seeking information about different universities.

References

1. A. Mondal, E. Cambria, D. Das, A. Hussain, S. Bandyopadhyay, Relation extraction of medical concepts using categorization and sentiment analysis. *Cogn. Comput.* **10**, 1–16 (2018)
2. A. Mondal, D. Das, S. Bandyopadhyay, Relationship extraction based on category of medical concepts from lexical contexts, in *14th International Conference on Natural Language Processing (ICON 2017)*, Kolkata, India (2017), pp. 212–219
3. A. Mondal, E. Cambria, A. Feraco, D. Das, S. Bandyopadhyay, Auto-categorization of medical concepts and contexts. In *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (IEEE, 2017), pp. 1–7 (November)
4. C. Sanli, A. Mondal, E. Cambria, Tracing linguistic relations in winning and losing sides of explicit opposing groups (2017). arXiv preprint [arXiv:1703.00317](https://arxiv.org/abs/1703.00317)
5. B. Shawar, E. Atwell, *A Comparison Between Alice and Elizabeth Chatbot Systems* (2002)
6. I. Ahmed, S. Singh, AIML based voice enabled Artificial Intelligence Chatterbot. *IJUNESST* **8**, 375–384 (2015). <https://doi.org/10.14257/ijunesst.2015.8.2.36>
7. J. Huang, M. Zhou, D. Yang, Extracting chatbot knowledge from online discussion forums in *20th International Joint Conference on Artificial intelligence* (2007), pp. 423–428
8. J. Allen, *Natural Language Understanding* (Benjamin/Cummins Publishing Inc., 1995)
9. R. Dale, et al. (eds.), *Handbook of Natural Language Processing* (Marcel Dekker Inc., 2000)
10. L.M. Iwanska, et al., *Natural Language Processing and Knowledge Representation: Language for Knowledge and Knowledge for Language* (AAAI Press/MIT Press, 2000)
11. N. Chomsky, Three models for the description of language. *IRE Trans. PGIT* **2**, 113–124 (1956)
12. N. Chomsky, *Aspects of the Theory of Syntax* (MIT Press, Cambridge, 1965)
13. N. Chomsky, *Topics in the Theory of Generative Grammar* (Mouton & Co. N.V. Publishers, 1969)
14. P. Tapanainen, T. Järvinen, A non-projective dependency parser, in *Proceedings of the 5th Conference on Applied Natural Language Processing* (ACL, Washington, D.C., 1997)
15. “ELIZA”, J. Weizenbaum, *Computer Power and Human Reason: From Judgment to Calculation* (W.H. Freeman and Company, New York, 1976), pp. 2, 3, 6, 182, 189. ISBN 0-7167-0464-1
16. T. Winograd, *Understanding Natural Language* (Edinburgh University Press, 1972)

A Supervised Approach to Analyse and Simplify Micro-texts



Vaibhav Chaturvedi, Arunangshu Pramanik, Sheersendu Ghosh,
Priyanka Bhadury and Anupam Mondal

Abstract Analysis of micro-text presents a challenging task due to the incompleteness of its corpus in the domain of Natural Language Processing (NLP). Primarily, micro-text refers to the limited textual content in the form of letters and words, collected from various web-based resources. In the present paper, we are motivated to build a supervised model for analysing micro-text. The model assists in simplifying the texts and extracting the important knowledge from unstructured corpora. Additionally, we have prepared an experimental dataset to validate the proposed model. The validation process offers 94% accuracy to identify the micro-text from the unstructured corpus like Twitter. The proposed model helps to design various applications such as annotation system and prediction system for micro-texts in the future.

Keywords Micro-text · Machine learning · Natural language processing · Dependency matching

V. Chaturvedi · A. Pramanik (✉) · S. Ghosh · P. Bhadury · A. Mondal
Department of Computer Science & Engineering, Institute of Engineering & Management,
Y-12, Salt Lake Electronics Complex, Sector-V, Kolkata 700091, India
e-mail: arunangshu01@gmail.com

V. Chaturvedi
e-mail: vaibhav1579@hotmail.com

S. Ghosh
e-mail: shibrata15@gmail.com

P. Bhadury
e-mail: priyankabhadury5@gmail.com

A. Mondal
e-mail: anupam.mondal@iemcal.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_8

1 Introduction

Micro-text refers to a microfilmed or micro-photographed¹ textual information. This information is widely availed by several users in various social media platforms to express their emotions [1, 2]. The micro-text analysis task is challenging due to the presentation of texts in an impromptu manner. This presentation also introduced other difficulties such as identification of knowledge-based information and proper meaning recognition from unstructured corpora. Primarily, the users often make use of a various form of micro-texts at the time of communication through social media platforms like Facebook, Twitter, YouTube, Google+, WhatsApp, Instagram, and LinkedIn [3]. Also, they are preferred to use abbreviation of texts and other SMS languages patterns to convey their sentiments or emotions very fluently to other [4, 5]. We have observed that the micro-texts are generated by digital short messages between the range of 2–700 characters along with unconventional grammar and style [3]. It is also hybridisation between informal and traditional expressions with threading characteristics. Micro-text is sufficiently divergent for a progenitor to necessitate unique study. The researchers have observed that the traditional long-text techniques do not translate well to micro-text due to unstructured and semi-structured nature of the corpus. Additionally, the micro-text corpora contain a minute-level time-stamp and a source attribution.

In this paper, we have developed an analysis system for simplifying the micro-text. The simplified form assists in converting the structural corpus from the unstructured and semi-structured corpora. In order to build this system, we have employed various well-known machine learning classifiers such as Logistic Regression, Decision Tree, and Support Vector Classifier (SVC) on the top of the prepared experimental dataset. Finally, we have validated the proposed system using test dataset as a part of experimental dataset, which provides an adequate output. The output may assist in designing various domain-specific applications such as annotation, relationship extraction, and concept-network systems in the future [6, 7].

The rest of the paper presents a detail background study related to micro-text analysis in Sect. 2. Sections 3 and 4 describe the proposed system and the result analysis of the mentioned system in details. Finally, Sect. 5 illustrates the concluding remarks and future scopes in this research.

2 Related Work

Universal in the present world, micro-text has refined the way of communication with an effortless technique. It redefines new defiances for Natural Language Processing (NLP) tools, which are usually designed for well-written text. In some cases, authors present a novel supervised method to translate Chinese abbreviations, which extract the relation between a full-form phrase and its abbreviations from monolingual cor-

¹<https://www.merriam-webstar.com/>.

pora, and induce translation entries for the abbreviation by using its full-form as a bridge. Other works based on this are Topic Detection in IRC chat-rooms which follows the approach of TF/IDF with temporal augmentation which impacted to the achievement of 71.5% accuracy on a system modelled to detect the speaker's intent to flirt using a spoken corpus of speed dates [1].

In order to develop an analysis system for micro-texts, the researchers have applied primarily two different types of approaches, namely lexicon-based and machine learning-oriented approaches [8, 9]. Additionally, they have introduced sentiment analysis and phonetic-based approach to analyse the micro-texts [8, 10, 11].

Generally, it follows a dictionary-based approach where acronyms and emoticons are found and extracted from various online sources which ended up forming a table of 1727+ acronyms and 512+ emoticons. The task performed for this proposed model follows the same polarity classification as used for sentiment analysis. The use of POS-tags is more frequent in subjective texts which can be hypothesised by dependency types [8].

The use of SO-dictionaries was challenged to be dogmatic. Hence, the researchers have employed supervised and unsupervised learning modules after extracting various features like parts of speech (POS) and sentiment words, etc. to build and validate the micro-text analysis systems [4].

Due to the origin of human life on Earth, people have been considered as social animals exploitable to opinions as practically all vows and conducts are influenced by them. Generally, when decisions are to be made, individuals and organisations frequently go for other's perspectives. Perspectives, opinions and its associated concepts like sentiments, emotions, attitudes, etiquettes, evaluations comprise of the sentiment analysis. Gradual upsurge of Web 2.0, people express their views upon various matters and on certain issues. The economic benefits from this can be derived from the knowledge are pretty decent that the market has proposed solutions for analysis of these views. Sentiment analysis is a branch of effecting computed research that tends to assimilate the text into either positive, negative, neutral or mixed expressions. The task required in this field is polarity classification which determines the above objective. No standardisation is followed about polarity categories, but the results give analysis of binary or ternary classification. The task has been incorporated from two different perspectives supervised machine learning (ML) approaches and non-supervised semantic-based methods. Statistical approaches have proved to be subdued as statistical text classifiers only work with adequate precision when given a satisfactorily vast input text. Concept-level sentiment analysis deals with large semantic text analysis in scientific community as well as the business world [12].

Common textual languages in phones generally SMS languages have made more significance on the monotony of a common man's life. Re-imagining and reconstructing a large word into consequential formatted small words have made work and communication more likeable and lovable. The phonetic-based approach follows a simple but tactful algorithm to manipulate micro-text. Soundex is the most famous algorithm which is used effectively to group similar sounding letters together, and

each group unit is assigned to a letter of the numeric characters. The main objective is to use homophones for encoding [8].

The provided background assists in developing the proposed system for micro-texts, which is described in the following section.

3 Methodology

In order to build the proposed system for micro-text analysis, we have initially prepared an experimental dataset which has been collected from Twitter.² This experimental dataset has been processed through three different types of classifiers such as Logistics Regression, Decision Tree, and Support Vector Classifier (SVC) for improving the accuracy of the proposed simplification system [13, 14]. In the following subsections, we describe, (a) how we have prepared the dataset? (b) selection process of machine learning classifiers, and (c) the design steps of the proposed model in details.

3.1 Dataset Preparation

Initially, we have collected a dataset from a Twitter Repository,³ contains around 2500 textual tweets. Besides, we have preprocessed the crawled tweets and identified a number of 1000 tweets as an experimental dataset. All the tweets of the experimental dataset have been satisfied the length of the micro-text as 2–120 characters. These tweets have been labelled manually by a group of Internet users to assign the tweets as a micro-text or general tweets. Thereafter, we have split the experimental dataset into two parts such as training and test dataset. The training dataset helps to learn the classifiers whereas test dataset assists in validating the proposed system. Training dataset contains 800 number of tweets where rest of 200 tweets presented as a test dataset. In the following subsection, we have discussed the classifier selection approach.

3.2 Classifier Selection

Logistic regression is a classified algorithm mainly used for Machine Learning Statistical Analysis. It constructs a statistical model to apply a binary dependent variable. The variable contains a coded data as 1 which indicates a success or 0 for failure. The

²<https://twitter.github.io/>.

³<https://twitter.github.io/>.

algorithm is generally used for the prediction of the probability of the variable. In simple words, the model generates a variable say P as a function of another dependent variable X .

Decision tree is a one-dimensional regression analysis which is used to place a sine curve in accordance with and addition of a noisy observation. If the maximum depth of the tree is plotted to be high enough, the regression impacts to learn fine details of the trained data and from the incurred noise. It breaks the dataset into tiny datasets whereas at the same time, the decision tree is hierarchically formed.

Support vector classifiers, commonly known as support vector machines or networks, are trained supervised learning models with algorithms in association that examines data used for classification as well as regression. However, to use SVM for analysis of sparse data predictions, it must fit the dimensions properly.

3.3 Proposed Algorithm

Thereafter, we have applied the following algorithm to identify the micro-text from the unstructured corpus.

Step-1: Initially, we have collected a dataset from the Twitter repository and preprocessed them.

Step-2: Prepared an experimental dataset after manually labeled (L) the crawled micro-texts tweets.

Step-3: Extract various features like capital words, alphanumeric characters, etc for the experimental dataset in the form $(X) = X_1, X_2, \dots, X_n$.

Step-4: The prepared experimental dataset is split into two parts as training and testing datasets.

Step-5: The extracted features (X) and their corresponding label (L) of the training dataset have been processed through three different classifiers namely Logistic Regression (M_{LR}), Decision Tree (M_{DT}), and SVC (M_{SVC}) to build the model.

Step-6: Thereafter, we have merged these classifiers M_{LR} , M_{DT} , and M_{SVC} with the help of Equation 1 to build another approach (M_{Merged}) under the proposed module.

$$M_{Merged} = M_{LR} \cup M_{DT} \cup M_{SVC} \quad (1)$$

Step-7: Finally, the test dataset has been applied on the above classifiers to validate the proposed micro-text analysis system.

In the following section, we have discussed the validation process and obtained output for the proposed system.

4 Estimated Results

Besides, to validate the proposed system, we have employed the test dataset and processed through all the classifiers individually. The accuracy of these classifiers has been measured using standard evaluation matrices like precision, recall, and F-Measure. Table 1 presents a comparative analysis between all the mentioned classifiers to the process of designing micro-text analysis system.

Additionally, we have generated the confusion matrix for all classifiers. Table 2 shows a confusion matrix for Logistic Regression Classifier.

The result shows that the combined classifier provides an adequate output for the proposed system. We have also observed that the combined classifier offers 94% accuracy to identify the micro-text from the corpus.

Table 1 An evaluation of the proposed system using precision, recall, and F-measure for all classifiers

Classifiers	Precision	Recall	F-measure
M_{LR}	0.948	0.929	0.939
M_{DT}	0.921	0.909	0.915
M_{SVC}	0.933	0.921	0.927
M_{Merged}	0.954	0.942	0.948

Table 2 A confusion matrix representation for identifying micro-text using Logistic Regression Classifier

Samples		Predicted		Total
		Micro-text	General-texts	
Original	Micro-text	146	8	154
	General-texts	11	35	46
Total		157	43	

5 Conclusion and Future Scopes

This paper aims towards deciphering information from the micro-text, widely used in different social media platforms. Here, we have adopted the technique of Machine Learning and applied it to the abbreviated texts and the various small phrases containing alphanumeric characters to interpret them correctly. The concepts of Logistic Regression, Decision Tree, and SVC have been applied to achieve adequate accuracy for simplifying the micro-texts. The simplified micro-texts may assist in designing various social media applications as a platform to interact and thus it is gaining its importance in market understanding where skilful and strategic planning is required. It can thus be predicted that studies and enhances development on micro-text analysis on big data platform for customer relation management and various other aspects are going to get increasing attention in near future.

References

1. J. Ellen, All about microtext-a working definition and a survey of current micro-text research within artificial intelligence and natural language processing, in *ICAART 2011*, vol. 1 (2011), pp. 329–336
2. J. Young, C.H. Martell, P. Anand, P. Ortiz, H.T. Gilbert IV, A microtext corpus for persuasion detection in dialog, in *Analyzing Microtext* (2011)
3. J.C. Mallery, Semantic content analysis: a new methodology for the RELATUS natural language environment, in *Artificial Intelligence and International Politics* (1991), pp. 347–385
4. D. Vilares, Sentiment analysis for reviews and microtexts based on lexico syntactic knowledge, in *Proceedings of 5th BCS-IRSG symposium on future directions in information access*, pp. 38–43. 2013
5. P. Smith, M.G. Lee, A CCG-based approach to fine-grained sentiment analysis in microtext, in *AAAI Spring Symposium: Analyzing Microtext*, vol. 13 (2013), p. 1
6. A. Mondal, E. Cambria, D. Das, S. Bandyopadhyay, Mediconceptnet: an affinity score based medical concept network, in *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference, FLAIRS 2017* (pp. 22–24)
7. A. Mondal, E. Cambria, D. Das, A. Hussain, S. Bandyopadhyay, Relation extraction of medical concepts using categorization and sentiment analysis. *Cogn. Comput.* 1–16 (2018)
8. R. Satapathy, C. Guerreiro, I. Chaturvedi, E. Cambria, Phonetic-based microtext normalization for twitter sentiment analysis, in *2017 IEEE International Conference on Data Mining Workshops (ICDMW)* (IEEE, 2017), pp. 407–413
9. E. Bejek, P. Strank, P. Pecina. Syntactic identification of occurrences of multiword expressions in text using a lexicon with dependency structures, in *Proceedings of the 9th Workshop on Multiword Expressions* (2013), pp. 106–115
10. A. Mondal, R. Satapathy, D. Das, S. Bandyopadhyay, A hybrid approach based sentiment extraction from medical context, in *SAIIP@ IJCAI*, vol. 1619 (2016), pp. 35–40
11. A. Mondal, E. Cambria, D. Das S. Bandyopadhyay, Employing sentiment-based affinity and gravity scores to identify relations of medical concepts, in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)*, Honolulu, HI (2017), pp. 1–7
12. G. Garzone, Textual analysis and interpreting research (2000)
13. Matthew Shardlow, A survey of automated text simplification. *Int. J. Adv. Comput. Sci. Appl.* 4(1), 58–70 (2014)
14. A. Siddharthan, Syntactic simplification and text cohesion. *Res. Lang. Comput.* 4(1), 77–109 (2006)

A Short Review on Different Clustering Techniques and Their Applications



Attri Ghosal, Arunima Nandy, Amit Kumar Das,
Saptarsi Goswami and Mrityunjoy Panday

Abstract In modern world, we have to deal with huge volumes of data which include image, video, text and web documents, DNA, microarray gene data, etc. Organizing such data into rational groups is a critical first step to draw inferences. Data clustering analysis has emerged as an effective technique to accurately accomplish the task of categorizing data into sensible groups. Clustering has a rich association with researches in various scientific domains. One of the most popular clustering algorithms, k-means algorithm was proposed as early as 1957. Since then, many clustering algorithms have been developed and used, to group data in various commercial and non-commercial sectors alike. In this paper, we have given concise description of the existing types of clustering approaches followed by a survey of the fields where clustering analytics has been effectively employed in pattern recognition and knowledge discovery.

Keywords Clustering · Applications of clustering · Review · Types of Clustering

A. Ghosal · A. Nandy (✉) · A. K. Das
Institute of Engineering & Management, Salt Lake, Kolkata, India
e-mail: arunimanandy27@gmail.com

A. Ghosal
e-mail: attrig9@rediffmail.com

A. K. Das
e-mail: amitkumar.das@iemcal.com

S. Goswami · M. Panday
Cognizant Technology Solutions, Kolkata, India
e-mail: saptarsi007@gmail.com

M. Panday
e-mail: mrityunjoy@gmail.com

1 Introduction

As the world steadily becomes more connected with an ever-increasing number of electronic devices, data are set to grow rapidly. 90% of the data in the world have been collected in the last two years alone. Data output in 2017 alone has been roughly 2.5 quintillion bytes per day [1]. With such huge amount of data in hand, we need advanced and well-equipped methods that can automate the process of analyzing, understanding, and summarizing the data thus recognizing viable patterns and extracting useful knowledge. In a scenario primarily concerned with pattern recognition, predictive modeling is the most relevant method adopted for data analysis.

Learning problems can be segmented into two distinct categories:

- **Supervised:** where all training data are labeled
- **Unsupervised:** where input data are unlabeled

Clustering, a prominent part of unsupervised learning is a more difficult and challenging problem than classification [2]. The technique of data clustering, also referred to as cluster analysis, aims at discovering natural groupings within a given set of records. Cluster analysis, as mentioned in Merriam-Webster Online Dictionary (2008), is a statistical classification technique for discovering whether the individuals of a population fall into different groups by making quantitative comparisons of multiple characteristics.

The main objective of a clustering algorithm is to develop a technique that will identify the natural groupings in the unlabeled data. Data clustering has been used for the following main purposes:

- To gain useful knowledge from data, i.e., generate hypotheses, detect anomalies, and identify salient features within given data
- To identify the degree of similarity among forms, organisms, or points that comprise the data
- As a method for organizing the data and summarizing it through cluster prototypes

Clustering can be divided into two categories:

- **Hard Clustering:** In this method, the object belongs strictly to a single group or to none
- **Soft Clustering:** In this method, an object may simultaneously belong to more than one cluster

In this paper, we make a survey of the various fields where data clustering has been successfully used to harness new, relevant, and useful information that has been used to achieve organized results within the said fields. Section 2 focuses on categorization of algorithms, highlights their key details, and provides few examples. Section 3 comprises of various fields where data clustering analysis has been implemented with success. Finally, Sect. 4 summarizes the points presented in the paper in the form of conclusion.

2 Types of Clustering Algorithms

The goal of clustering is subjective. A particular clustering algorithm is oriented toward a particular set of applications. Every algorithm follows a different methodology to specify the closeness of data points. Till date, over 100 clustering algorithms have been proposed and studied [3]. All these can be grouped under five distinct subsets [4]. Figure 1 gives a clear picture of the taxonomy of clustering algorithms.

2.1 Partitional Clustering

This is an iterative approach which finds similarity among intra-cluster points with respect to their distances from the cluster-centroid. It takes into consideration two assumptions [4]:

- Each cluster must contain at least one data-point
- Each data-point must be assigned to at least one cluster

In this method, cluster-centers are initialized first. Based on a particular metric, the distance of a data-point from all the centers is calculated. The particular data-point is assigned to the cluster whose centroid is at the least distance from the point and the centroid of the group is reassigned in the modified form.

Examples: K-Means, K-Medoids, K-Modes, etc.

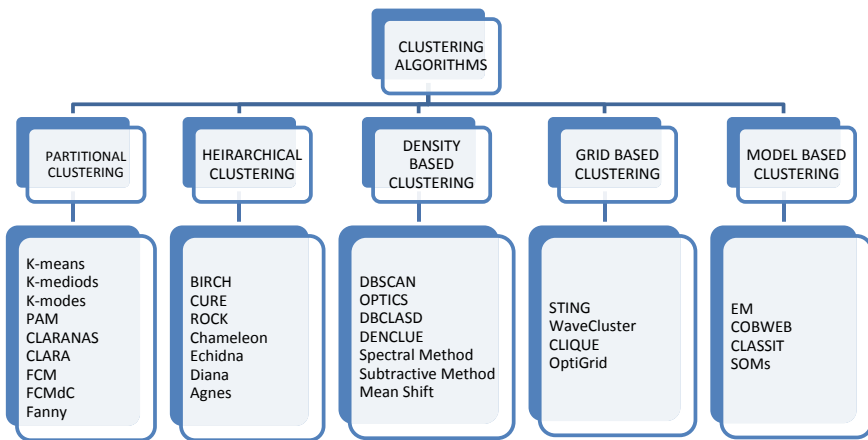


Fig. 1 Categorization of clustering algorithms

2.2 Hierarchical Clustering

This model follows two approaches: **agglomerative (bottom-up)** and **divisive (top-down)**. In the first approach, each data-point is considered to be a separate cluster. Then by selecting a particular distance metric, the proximity between two points is calculated and the closest pairs are clubbed together into a single cluster. This process continues iteratively, until all the data-points are combined together to form a single cluster.

The second approach starts with a single cluster containing all the data-points. Subsequently, they are divided into separate groups as their distance increases. The underlying hierarchical structure of the complete dataset is represented by a dendrogram as shown in Fig. 2.

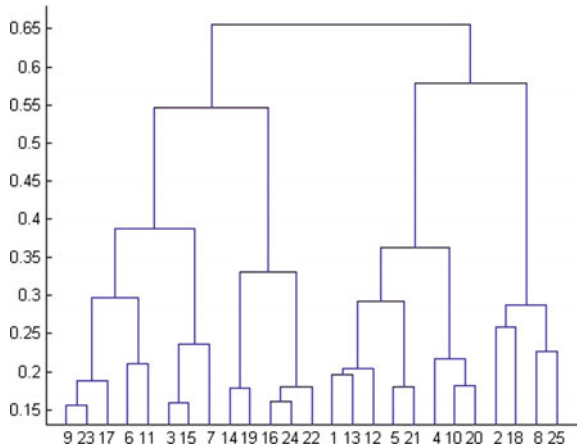
Examples: BIRCH, CURE, ROCK, Chameleon, Echidna, etc.

2.3 Density-Based Clustering

This algorithm uses density of the data-points in the data space to form clusters. Regions of higher density are separated as clusters and regions of very low density are used as partition. In this way, it provides a barrier against outliers or noise data. It starts with arbitrary data-points that have not been visited yet and checks its neighborhood. If there are sufficient number of points within a particular distance ‘ ϵ ’, only then a cluster is formed. Else, the data-point is marked as an outlier [5]. This operation is performed iteratively for all sets of non-visited points.

Examples: DBSCAN, OPTICS, DBCLASD, DENCLUE, etc.

Fig. 2 A dendrogram



2.4 Grid-Based Clustering

The grid-based algorithms do not use the database directly. It forms a uniform grid by collecting data from the database by using statistical methods. The performance of the algorithm depends on the size of the grid and not on the size of the actual data space. As the size of the grid is much less than the actual database, the time taken by this algorithm is much less. After forming the grid, it computes the density of each cell in the grid. If the density of the cell is below the threshold value, the cell is discarded. Finally, the clusters are formed from contiguous groups of dense cells [6].

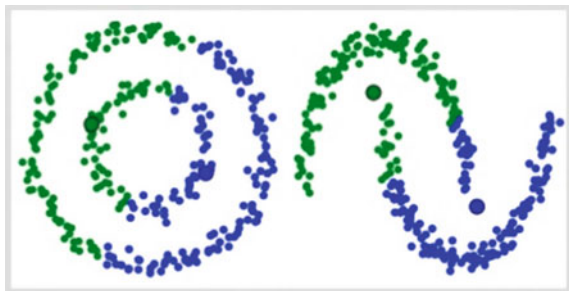
Examples: STING, CLIQUE, Wavecluster, OptiGrid, etc.

2.5 Model-Based Clustering

The image depicted in Fig. 3 shows the two cases where k-means fails. Since the centers of the two clusters almost coincide, the k-means algorithm fails to separate the two clusters. This is due to the fact that k-means algorithm uses only a single parameter, i.e., the mean and hence can find only circular clusters. The model-based algorithms make use of many pre-defined statistical or mathematical models to form clusters. The number of clusters may be pre-specified, though it is not necessary. This algorithm works on the mixture of underlying probability and creates clusters on the basis of it [7].

Examples: Expectation Maximization (using GMM), COBWEB (uses hierarchical conceptual clustering), Neural Network (weighted connected input/output feature map), etc.

Fig. 3 Two cases of failure for the K-Means



3 Applications

Cluster analysis has been successfully implemented in various fields to extract useful patterns in data. In this section, we explore the various fields where clustering technique has been used successfully.

3.1 Banking

Banking industry is one of the most proliferating industries. It spearheads the drive for digitization around the world. But with advancements in the digital banking industry, a number of threats have emerged to hinder the progress of the digitization drive. Clustering analysis can provide a handy solution for these threats.

The biggest threat is money laundering, and to overcome this, the DBSCAN clustering algorithm has been implemented in AMLRAS (Anti Money Laundering Regulatory Application System). Here, DBSCAN is used to detect and report suspicious financial transactions. AMLRAS has been successfully tested on large financial data from where it could identify possible fraudulent transactions thus preventing laundering [8].

For every banking establishment, it is important to acknowledge the threats that can arise from customers' ignorance of protocols and thus handing out sensitive information like the security pin to unknown individuals leading to massive banking frauds. The banks can implement a clustering analysis to identify such customers and warn them to be extra cautious apart from broadcasting general warnings. The k-means method and k-means++ method of clustering have both achieved success in assessing share of cardholders susceptible to bank frauds including the likes of fishing, vishing, skimming, etc. [9].

Apart from mitigating threats from the banking system, cluster analysis can also serve certain mainstream applications like choosing locations for optimum business expansion. Any bank would prefer to widen their area of service to draw in more customers thus increasing the establishments profit margins. Such expansions would possibly include installing ATMs and e-corners at strategic locations so as to cover maximum economical area of relevance as well as setting up branches at desired stations that experiences a high demand. As we see, the choice of location plays a key role in economic expansion and GFDBSCAN clustering algorithm provides a handy solution for this task [10].

Talking of profit maximization and clientele expansion, customer churn management needs a special mention and data clustering techniques like Subtractive Clustering Method (SCM), Semantic Driven Subtractive Clustering Method (SDSCM) and fuzzy c-means (FCM) play major roles in this section [11].

The modern world is largely dependent on industrial correlation, and a financial debacle is the ultimate nightmare for any nation, big and small alike. Therefore, long-term bank failure prediction poses itself as a major challenge. Fortunately, we have

yet another clustering technique—Fuzzy refinement domain adaptation to provide a fruitful solution [12].

3.2 Segmentation Engines and Recommendation Engines

The tasks of segmentation and recommendation go hand in hand as the former is necessary for the latter action to be completed with success. Segmentation is used in various fields to create clusters based on certain attributes. Perhaps the most classic example of segmentation is the market analysis. Every business needs to provide a superlative customer service in order to thrive in today's cut-throat competition, and therefore, it becomes very important for them to understand the demands of their customers as well as their socioeconomic background. K-means clustering function has been extensively applied in this field to make customer segmentation with each cluster having customers with certain common traits exhibited in their marketing habits [13]. Same can be done for the aspiring tourists and travelers. Indeed, everyone would not have a similar taste when it comes to visiting places and events, and thus, it is important for the tourism industry to segregate potential clients into clusters according to their aesthetic tastes. A bagged clustering technique with a fuzzy C-medoids clustering algorithm (FCM_dC) has been used to execute tourist market segmentation [14]. Crime domain documentation is a major task which is essential in segmenting criminal records to identify similar modus operandi of criminals. These records are further referred to in order to get a clear cut idea of unsolved cases and patterns in criminal activities. Both K-means and k-means++ algorithms have contributed significantly in this area and have helped build a viable crime database [15].

Segmentation also includes segregating all the information available online and creating possible clusters with materials that contribute to a particular topic, and this makes the task of recommending easier because all it needs to be done is identifying the information seeker's area of interest. Two very interesting techniques need a special mention here—text mining and web mining. These techniques mainly use K-means, fuzzy C-means, and hierarchical clustering algorithms. Text mining can be implemented in various walks as in drawing patterns in social media usage [16], handling all the side-information or metadata within any document [17]. Text mining is also essential in the working of web search engines. An interesting example would be the working of the likes of Semantic Scholar and Google Scholar where searching for relevant research papers on a particular topic is made easy with the help of text mining approach [18]. Applying this concept on a larger scale, web mining approach proves itself as potentially able to recommend relevant information to any seeker from the vast repository of knowledge available online [19].

Segmentation and recommendation also help in the field of education. K-means and hierarchical clustering algorithms both have been used to create groups of students based on their academic performance [20, 21]. This information can be used further to recommend special programs to students to enhance their learning skills.

3.3 *Healthcare*

Health care has been a prevalent field that needs upgradation along with advancement of modern civilization. It is one of those key fields that directly impact the masses in general. Recently, clustering analysis has been heavily relied upon to modernize the health care services. Diagnosis is the first step in treating any disease, and clustering algorithms are known to contribute significantly in this field. Clustering algorithms have simplified the process of detection of ocular diseases by segmenting retinal blood vessels [22]. Multivariate m-Medoids-based classifier has been used to detect neurovascularization in retinal images [23], and K-means clustering algorithm has been implemented in detection of tumors [24].

Medical imaging is a major part of medical science that not only contributes to direct treatment of a patient but also toward medical research work to study various aspects of human anatomy and understand the effects caused by a certain ailment. Mean shift clustering has been used in blood oxygen level dependent functional MRI activation detection [25]. Brain image segmentation has been done using semi-supervised clustering [26]; inhomogeneous medical image segmentation has been executed successfully using a hybrid method based on fuzzy clustering [27]. Clustering of medical X-ray images has been accomplished by various clustering techniques [28]. Medical image analysis also uses spectral clustering for a clear study [29, 30].

A treatment methodology is unique to every patient. The hype in modern medicine is creating a personalized treatment system for an individual based on their genetic composition. Recently, there has been extensive research in the field of genetics and gene expressions, and clustering analysis is not far behind in aiding this process [31–34].

3.4 *Urban Development*

Modern society is rapidly adopting smart technology. In such a scenario, it is apt that this technology is not only limited to electronic gadgets but also becomes a part of our daily lives. To that end, we have to be careful about the very basics such as positioning of towns, building efficient power supplies and providing other basic amenities. Data clustering analysis is capable of playing a key role in achieving this efficiently. K-means clustering based on Ant Clustering Method has been known to play a decisive role in fixing stations for setting up of industrial towns and economic hubs along highways [35]. Further, geographic positions for setting up amenities such as banks, schools etc. within these strategically placed cities can be decided using GFDBSCAN Clustering algorithm [10]. k-medoids algorithm has helped in transformer fault diagnosis [36], and k-means has contributed significantly toward identifying patterns in online cable-discharge monitoring [37].

The hectic process of population census can be eased by using satellite imaging. But high resolution of the satellite images causes a spatial mismatch with ground

reference data. This problem is overcome by using convolutional expectation maximization [38].

3.5 Data Transfer Through Network

These days, we share large amount of data online through various social networks and forums. In order to achieve high-speed data transfers and avoid delays, we need an efficient transport system. Applying clustering analysis helps in such a scenario. K-means, DBSCAN, and AutoClass algorithms, and all can be used to detect groups of similar traffic using only transport layer statistics based on distinct characters of applications as they communicate through any given network [39]. This aids to speed up data transfer by transporting clusters of similar traffic along network at once.

For certain environmental data collected by sensors and communicated across sensor network to a processing head, it is desired that the process be accomplished with spending of minimum energy. To this end, a hierarchical clustering algorithm has effectively contributed to this task reducing energy consumption significantly [40]. This works by grouping sensors in a wireless network into clusters.

3.6 Privacy Protection

In modern virtual world, privacy and data security is one of, if not the most, the important concerns that need attention. Responsible and secured data sharing can be achieved by partitioning data into clusters and providing a single relevant cluster to any information seeker. This can be done using k-means algorithm [41] and distributed clustering [42]. Along with securing shared data, it is also important to protect data that is publicly available on Web sites. Presence of certain criminally replicated Web sites has been observed, and a combined clustering method has been successfully developed that can link scam Web sites together thus making their marking and elimination easier [43].

Identification of malware families and generating signatures of anti-virus systems is another challenging task. When malware differs from regular data considerably, they can be considered as outliers and thus excluded from clusters of regular data. Both regular clustering algorithms like k-means and x-means and co-clustering technique have been used to detect anomalies in networks [44]. Behavioral malware clustering has been achieved using single linkage hierarchical clustering [45] to protect data from malware attack. K-means algorithm has been around for some time now for anomaly based intrusion detection system, but a modification of Opti-Grid Clustering and a cluster labeling algorithm together has a better scope in succeeding in the field of high-performance intrusion detection [46]. Clustering algorithms have contributed generously in the fields of computer security. Utilizing data clustering can

significantly put down cyber-criminals and enhance the coexistence of data mining and privacy.

3.7 *Image Segmentation*

Image segmentation defined as the process of partitioning image pixels into super-pixels with the goal of simplifying the representation of a digital image thus making it precise and easier to analyze. Certain fields that make extensive use of image segmentation are:

- Object-class segmentation
- Scene segmentation
- Surface layout labeling
- Single view 3D reconstruction

Image segmentation can be achieved by a number of clustering algorithms among them fuzzy algorithms being the best choice, but they are susceptible to misclassification of image pixels. There has been extensive research in this field, and some modifications have been proposed to overcome this difficulty without much success. However, the use of Guided Filters clubbed with Fuzzy C-means algorithm and Generalized Gaussian Density-based agglomerative fuzzy segmentation algorithm do provide a sustainable solution to degraded edges in image segments [47, 48]. Also use of spatial clustering—Fuzzy C-means with edge and local information (FELICM) reduces noisy edges by edge weighting [49]. Again Higher-order correlation clustering contributes significantly toward deciphering higher-order relationships between super-pixels [50]. Hierarchical image segmentation is accomplished by correlation clustering method [51] for extraction of local information, and Hierarchical pixel clustering has been done by k-means method and verified by Otsu Method and Mumford-Shah model for detection and segregation of sub-images or objects with respect to color image segmentation [52, 53].

Image segmentation also finds application in medical investigation and research. Anatomical images are studied extensively to find affected areas that need treatment; discover important features as well as for observing varied side effects of certain drugs. To this end, brain image segmentation done using semi-supervised clustering [26] and tumor diagnosis using k-means clustering [24] needs special mention.

As summary of all areas of application along with which algorithm has been used is listed in Table 1.

In Table 2, a quick snapshot of most commonly used algorithm along with their area of application has been highlighted.

Table 1 Area of application versus algorithm used

Area of application	Algorithm used
<p>1. In Banking:</p> <ul style="list-style-type: none"> ● Anti Money Laundering Regulatory System ● Broadcasting warning messages against bank frauds ● Installing ATMs and e-corners at strategic locations ● Long-term bank failure prediction ● Profit maximization and clientele expansion, customer churn management 	<ul style="list-style-type: none"> – DBSCAN – K-means, K-means++ – DBSCAN – Fuzzy Refinement domain adaptation – Subtractive Clustering Method(SCM), Fuzzy c-means(FCM)
<p>2. In segmentation engines and recommendation engines:</p> <ul style="list-style-type: none"> ● Market analysis ● Tourist market segmentation ● Education ● Crime domain documentation ● Drawing patterns in social media usage, handling all the side-information and metadata within any document, working of web search engines 	<ul style="list-style-type: none"> – K-means – Fuzzy C-medoids – K-means and Hierarchical clustering – K-means, k-means++ – k-means, fuzzy C-means and hierarchical clustering algorithms
<p>3. In Health Care:</p> <ul style="list-style-type: none"> ● Detection of neurovascularization in retinal images ● Detection of tumors ● Blood oxygen level dependent functional MRI activation detection ● Brain image segmentation ● Inhomogeneous medical image segmentation ● Medical image analysis 	<ul style="list-style-type: none"> – Multivariate m-Medoids based classifier – K-means – Mean shift algorithm – Spectral clustering – Fuzzy clustering – Spectral clustering
<p>4. In Urban Development:</p> <ul style="list-style-type: none"> ● Positioning of towns, building efficient power supplies ● Setting up amenities such as banks, schools etc. ● Transformer fault diagnosis ● Identifying patterns in online cable-discharge monitoring ● Population estimation with satellite imagery 	<ul style="list-style-type: none"> – K-means clustering based on Ant Clustering Method – DBSCAN – k-medoids – k-means – Expectation Maximization
<p>5. In Network analysis:</p> <ul style="list-style-type: none"> ● Speed up data transfer by transporting clusters of similar traffic along network at once ● Reducing energy consumption in sensors across sensor network 	<ul style="list-style-type: none"> – K-means, DBSCAN and AutoClass Algorithms – Hierarchical clustering algorithm
<p>6. In Privacy protection:</p> <ul style="list-style-type: none"> ● Responsible and secured data sharing ● Elimination of scam websites ● Behavioral malware clustering ● Anomaly based intrusion detection 	<ul style="list-style-type: none"> – k-means and distributed clustering – Combined clustering method – Single linkage hierarchical clustering – k-means, Opti-Grid Clustering, cluster labeling
<p>7. In Image segmentation:</p> <ul style="list-style-type: none"> ● Partitioning image pixels into super-pixels ● Hierarchical image segmentation ● Hierarchical pixel clustering ● Medical image analysis 	<ul style="list-style-type: none"> – FCM, FELICM – Correlation clustering method – k-means method – k-means algorithm

Table 2 Most popular algorithms used versus areas of application

Most popular algorithms	Areas of application
K-Means	<ul style="list-style-type: none"> ● Broadcasting warning messages against bank frauds ● Market analysis, crime domain documentation ● Drawing patterns in social media ● Web search ● Detection of tumors ● Identifying patterns in online cable-discharge monitoring ● Speed up data transfer ● Responsible and secured data sharing ● Medical image analysis
Hierarchical Clustering Algorithms	<ul style="list-style-type: none"> ● Education ● Drawing patterns in social media ● Web search engines ● Handling side information and metadata within any document ● Reducing energy consumption in sensors across sensor network ● Behavioral malware clustering ● Hierarchical image segmentation
DBSCAN	<ul style="list-style-type: none"> ● Anti Money Laundering Regulatory System ● Profit maximization and clientele expansion ● Customer churn management ● Managing network traffic ● Setting up amenities such as banks, schools etc. ● Installing ATMs and e-corners at strategic locations
Mean-shift Algorithm	<ul style="list-style-type: none"> ● Blood oxygen level dependent functional MRI activation detection ● Image processing ● Medical Image analysis
Expectation Maximization	<ul style="list-style-type: none"> ● Population estimation with satellite imagery ● Image reconstruction ● Auto fill function in web search engines

4 Conclusion

This paper provides a comprehensive review of the different clustering algorithms and their applications. Due to the large quantity of data maintained in several domains, clustering analysis has become indispensable in extracting patterns from bulk data thus aiding the process of useful knowledge discovery. Studying the various

clustering algorithms and their uses in different fields, we gather that out of several algorithms published till date, some are more popular than the others. Top five most commonly used clustering algorithms along with their fields of application have been highlighted upon in this paper. While some algorithms like the k-means and DBSCAN have been well explored and exploited, others like the Sparse Subspace Clustering Algorithm are in their nascent stage and are undergoing extensive research. A noteworthy point is that no single clustering algorithm has been found to dominate all areas of implementation. Thus, it is safe to say that though clustering algorithms have found many applications, there still exists much scope for research and development. With time, we will see more extensive applications of the data clustering approach in diverse application scopes.

References

1. T. Hale, How much data does the world generate every minute? (2018, June 18). Retrieved July 5, 2018, from <http://www.ifsscience.com/technology/how-much-data-does-the-world-generate-every-minute/>
2. A.K. Jain, Data clustering: 50 years beyond K-means. ECML/PKDD (2008)
3. S. Kaushik, An introduction to clustering & different methods of clustering (2016, December 10). Retrieved July 5, 2018, from <https://www.analyticsvidhya.com/blog/2016/11/an-introduction-to-clustering-and-different-methods-of-clustering/>
4. A. Fahad, N. Alshatri, Z. Tari, A. Alamri, I. Khalil, A.Y. Zomaya, S. Foufou, A. Bouras, A survey of clustering algorithms for big data: taxonomy and empirical analysis. *IEEE Trans. Emerg. Top. Comput.* **2**, 267–279 (2014)
5. G. Seif, The 5 clustering algorithms data scientists need to know (2018, February 5). Retrieved July 5, 2018, from <https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>
6. Data Clustering: Theory, Algorithms, and Applications. (n.d.). Retrieved July 12, 2018, from <https://epubs.siam.org/doi/abs/10.1137/1.9780898718348.ch12>
7. F. Chamroukhi, Robust EM algorithm for model-based curve clustering, in *The 2013 International Joint Conference on Neural Networks (IJCNN)* (2013), pp. 1–8
8. Y. Yang, B. Lian, C. Chen, P. Li, DBSCAN clustering algorithm applied to identify suspicious financial transactions, in *2014 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery* (2014), pp. 60–65
9. S.S. Alkhasov, A.N. Tselykh, A.A. Tselykh, Application of cluster analysis for the assessment of the share of fraud victims among bank card holders, in *SIN* (2015)
10. N.R. Kisore, C.B. Koteswaraiah, Improving ATM coverage area using density based clustering algorithm and voronoi diagrams. *Inf. Sci.* **376**, 1–20 (2017)
11. W. Bi, M. Cai, M. Liu, G. Li, A big data clustering algorithm for mitigating the risk of customer churn. *IEEE Trans. Ind. Inform.* **12**, 1270–1281 (2016)
12. V. Behbood, J. Lu, G. Zhang, Fuzzy refinement domain adaptation for long term prediction in banking ecosystem. *IEEE Trans. Ind. Inform.* **10**, 1637–1646 (2014)
13. C.P. Ezenkwu, S. Ozuomba, C. Kalu, Application of K-Means algorithm for efficient customer segmentation: a strategy for targeted customer services (2015)
14. P. D’Urso, L.D. Giovanni, M. Disegna, R. Massari, Bagged Clustering and its application to tourism market segmentation. *Expert Syst. Appl.* **40**, 4944–4956 (2013)
15. B. Aubaidan, M. Mohd, M. Albared, Comparative study of k-means and k-means++ clustering algorithms on crime domain. *JCS* **10**, 1197–1206 (2014)

16. C. Yang, N. Benjamasutin, Y. Chen-Burger, Mining hidden concepts: using short text clustering and wikipedia knowledge, in *2014 28th International Conference on Advanced Information Networking and Applications Workshops* (2014), pp. 675–680
17. R.E. Thomas, S.S. Khan, Improved clustering technique using metadata for text mining, in *2016 International Conference on Communication and Electronics Systems (ICCES)* (2016), pp. 1–5
18. E.A. Calvillo, A. Padilla, J.M. Arteaga, J.C. Gallegos, J.T. Fernandez-Breis, Searching research papers using clustering and text mining, in *CONIELECOMP 2013, 23rd International Conference on Electronics, Communications and Computing* (2013), pp. 78–81
19. G. Kazeminouri, A. Harounabadi, S.J. Mirabedini, E. Kazemi, Web personalization with web usage mining techniques and association rules (2015)
20. O.J. Oyelade, O.O. Oladipupo, I.C. Obagbuwa, Application of k Means clustering algorithm for prediction of students academic performance (2010). CoRR, abs/1002.2425
21. S. Rana, R. Garg, Application of hierarchical clustering algorithm to evaluate students performance of an institute, in *2016 Second International Conference on Computational Intelligence & Communication Technology (CICT)* (2016)
22. A. Waheed, M.U. Akram, S. Khalid, Z. Waheed, M.A. Khan, A. Shaukat, Hybrid features and medioids classification based robust segmentation of blood vessels. *J. Med. Syst.* **39**, 1–14 (2015)
23. M.U. Akram, S. Khalid, A. Tariq, M.Y. Javed, Detection of neovascularization in retinal images using multivariate m-Medioids based classifier. *Comput. Med. Imaging Graph. Official J. Comput. Med. Imaging Soc.* **37**(5–6), 346–57 (2013)
24. P.M. Patel, B.N. Shah, V. Shah, Image segmentation using K-mean clustering for finding tumor in medical application (2013)
25. L. Ai, X. Gao, J. Xiong, Application of mean-shift clustering to Blood oxygen level dependent functional MRI activation detection. *BMC Med. Imaging* **14**, 6 (2014)
26. S. Saha, A.K. Alok, A. Ekbal, Brain image segmentation using semi-supervised clustering. *Expert Syst. Appl.* **52**, 50–63 (2016)
27. M. Rastgarpour, J. Shanbehzadeh, H. Soltanian-Zadeh, A hybrid method based on fuzzy clustering and local region-based level set for segmentation of inhomogeneous medical images. *J. Med. Syst.* **38**, 1–15 (2014)
28. I. Ziedan, A. Zamel, A.A. Zohairy, Clustering of medical X-ray images by merging outputs of different classification techniques, in *CLEF* (2015)
29. T. Schultz, G.L. Kindlmann, Open-box spectral clustering: applications to medical image analysis. *IEEE Trans. Visual. Comput. Graph.* **19**, 2100–2108 (2013)
30. C. Kuo, P.B. Walker, O.T. Carmichael, I. Davidson, Spectral clustering for medical imaging, in *2014 IEEE International Conference on Data Mining* (2014), pp. 887–892
31. M.K. Aouf, L. Lyanage, S. Hansen, Review of data mining clustering techniques to analyze data with high dimensionality as applied in gene expression data (June 2008), in *2008 International Conference on Service Systems and Service Management* (2008, June), pp. 1–5
32. S. Datta, S. Datta, Comparisons and validation of statistical clustering techniques for microarray gene expression data. *Bioinformatics* **19**(4), 459–66 (2003)
33. T.L. Bailey, C. Elkan, Fitting a mixture model by expectation maximization to discover motifs in biopolymers, in *Proceedings. International Conference on Intelligent Systems for Molecular Biology*, vol. 2 (1994), pp. 28–36
34. J. Oyelade, I. Isewon, F. Oladipupo, O. Aromolaran, E. Uwoghiren, F. Ameh, M. Achas, E. Adebisi, Clustering algorithms: their application to gene expression data. *Bioinform. Biol. Insights* (2016)
35. Y. Meng, X. Liu, Application of K-means algorithm based on ant clustering algorithm in macroscopic planning of highway transportation hub, in *2007 First IEEE International Symposium on Information Technologies and Applications in Education* (2007), pp. 483–488
36. Z. Zhou, G. Si, J. Chen, K. Zheng, W. Yue, A novel method of transformer fault diagnosis based on k-medioids and decision tree algorithm, in *2017 1st International Conference on Electrical Materials and Power Equipment (ICEMPE)* (2017), pp. 369–373

37. X. Peng, C. Zhou, D.M. Hepburn, M.D. Judd, W.H. Siew, Application of K-means method to pattern recognition in on-line cable partial discharge monitoring. *IEEE Trans. Dielectr. Electr. Insul.* **20**, 754–761 (2013)
38. A. Pomente, D. Aleandri, Convolutional expectation maximization for population estimation, in *CLEF* (2017)
39. J. Erman, M.F. Arlitt, A. Mahanti, Traffic classification using clustering algorithms, in *MineNet* (2006)
40. S. Bandyopadhyay, E.J. Coyle, An energy efficient hierarchical clustering algorithm for wireless sensor networks, in *INFOCOM* (2003)
41. J. Vaidya, C. Clifton, Privacy-preserving k-means clustering over vertically partitioned data, in *KDD* (2003)
42. Z. Erkin, T. Veugen, T. Toft, R.L. Lagendijk, Privacy-preserving distributed clustering. *EURASIP J. Inf. Secur.* **2013**, 4 (2013)
43. J. Drew, T. Moore, Automatic identification of replicated criminal websites using combined clustering, in *2014 IEEE Security and Privacy Workshops* (2014), pp. 116–123
44. M. Ahmed, A.N. Mahmood, J. Hu, A survey of network anomaly detection techniques. *J. Netw. Comput. Appl.* **60**, 19–31 (2016)
45. B. Biggio, K. Rieck, D. Ariu, C. Wressnegger, I. Corona, G. Giacinto, F. Roli, Poisoning behavioral malware clustering, in *AISec@CCS* (2014)
46. M. Ishida, H. Takakura, Y. Okabe, High-performance intrusion detection using optigrd clustering and grid-based labelling, in *2011 IEEE/IPSJ International Symposium on Applications and the Internet* (2011), pp. 11–19
47. L. Guo, L. Chen, C.L. Chen, Image guided fuzzy clustering for image segmentation, in *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (2016), pp. 004271–004276
48. S. Choy, S.Y. Lam, K.W. Yu, W.Y. Lee, K.T. Leung, Fuzzy model-based clustering and its application in image segmentation. *Pattern Recognit.* **68**, 141–157 (2017)
49. N. Li, H. Huo, Y. Zhao, X. Chen, T. Fang, A spatial clustering method with edge weighting for image segmentation. *IEEE Geosci. Remote Sens. Lett.* **10**, 1124–1128 (2013)
50. S. Kim, C.D. Yoo, S. Nowozin, P. Kohli, Image segmentation using higher-order correlation clustering. *IEEE Trans. Pattern Anal. Mach. Intell.* **36**, 1761–1774 (2014)
51. A. Alush, J. Goldberger, Hierarchical image segmentation using correlation clustering. *IEEE Trans. Neural Netw. Learn. Syst.* **27**, 1358–1367 (2015)
52. M.V. Kharinov, Hierarchical pixel clustering for image segmentation (2014). CoRR, <http://arxiv.org/abs/1401.5891>
53. M.V. Kharinov, Pixel clustering for color image segmentation. *Program. Comput. Softw.* **41**, 258–266 (2015)

An Efficient Descriptor for Gait Recognition Using Spatio-Temporal Cues



Sanjay Kumar Gupta, Gaurav Mahesh Sultaniya
and Pratik Chattopadhyay

Abstract In this paper, we introduce a new efficient feature for gait recognition, namely the boundary energy image (BEI). To construct this feature, the contours of the binary silhouettes in a gait sequence are extracted, and average of these contours is computed. Finally, the dimensionality of the feature set is reduced by applying PCA, and LDA is used to classify the reduced feature set. Evaluation of the proposed approach on CASIA B and TUMGAID data sets has shown significantly accurate results. Comparative study with existing techniques also show that our approach outperforms existing features like gait entropy image, active energy image and gait energy image by a substantially high margin both in terms of accuracy and response time.

Keywords Gait recognition · Boundary energy image (BEI) · Visual surveillance

1 Introduction

Gait refers to the walking pattern of a subject, and the process of identifying a subject from his gait signatures is known as gait recognition. Existing biometric features such as fingerprint, iris and face recognition require detail texture information, due to which these are not reliable for recognizing subjects at a distance. On the other hand, a person cannot hide his/her own gait. A gait recognition method is a promising tool in visual surveillance, monitoring, tracking, forensics, etc., because it provides reliable and efficient means of identification and verification [1, 2].

There are broadly two types of gait recognition methods: model-based and appearance-based. Model-based methods are usually view and scale invariant, but

S. K. Gupta · G. M. Sultaniya · P. Chattopadhyay (✉)
CSE Department, Indian Institute of Technology (Banaras Hindu University), Varanasi, India
e-mail: pratik.cse@iitbhu.ac.in

S. K. Gupta
e-mail: sanjaykrgupta.rs.cse17@itbhu.ac.in

G. M. Sultaniya
e-mail: gauravm.sultaniya.cse15@iitbhu.ac.in

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_10

these require good-quality silhouettes as well as suffer from high computational cost, e.g., [3, 4]. On the other hand, appearance-based gait recognition methods [1, 2, 5–11] directly use the silhouettes from gait sequences for feature extraction and is computationally much more efficient. These approaches are suitable for functional applications because they operate on binary silhouettes which could be of low quality and no texture/color information is needed. Spatio-temporal method is a subclass of appearance-based gait recognition method that deals with both spatial and temporal domain information [2]. In this work, we introduce a new appearance-based gait recognition method for human identification, namely boundary energy image (BEI). The BEI is constructed by averaging the contours of the binary silhouettes over a gait cycle. Thus, similar to popular appearance-based approaches such as those in [1, 7–9], BEI also represents human motion condensed into a single image, but it does so in time as well as space-efficient manner.

The rest of the paper is structured as follows: in Sect. 2, we review the related work, while the BEI feature construction method is detailed in Sect. 3. Section 4 presents experimental results, and Sect. 5 concludes the paper with insights into some future research directions.

2 Related Work

In recent years, a number of techniques have been proposed for human recognition from gait signatures, of which majority are appearance based that compute features representing the overall shape information. For example, the motion silhouette image (MSI) [6] is a gray-level image where each pixel intensity is proportional to the temporal history of motion of that pixel. Han and Bhanu later extended this idea to construct a feature named gait energy image (GEI), which represents overall shape changes of the silhouette over a gait cycle [1]. An extension of GEI is given in [7] which has been seen to perform robustly in the presence of noise. Since GEI compresses an entire gait cycle into a single image, it loses dynamic characteristics of gait to a certain extent. As an improvement, the pose energy image (PEI) is proposed in [2], which divides the gait cycle into a fixed number of divisions and compute gait features corresponding to each division. PEI preserves both static and kinematic aspects of gait. To capture the dynamic information accurately, another feature, namely the gait history image (GHI), is developed. Like PEI, GHI also retains temporal information along with spatial information [8].

Later, gait moment image (GMI) was proposed in [9] which computes gait probability image at every important key moment of all gait cycles. The number of important key moments is a few dozens. However, it is not easy to select important key moments from the gait cycles with different periods. In the search for better representation of dynamic information, some other methods are developed by modifying directly the basic GEI, namely enhanced GEI (EGEI) [12], frame difference energy image (FDEI) [10], active energy image(AEI) [11], gait entropy image (GENI) [13], optical flow image [5] and chrono-gait image (CGI), CGI template first extracts the

contour from each gait frame, followed by encoding each of the gait contour frame in the same gait sequence with a multi-channel mapping function and aggregating them into a single CGI image [14]. These papers have reported better performance than that of GEI.

A number of effective frontal gait features by making use of depth data provided by Kinect have been developed recently. Important among these are gait energy volume [15] and pose depth volume [16], which are essentially three-dimensional versions of GEI and PEI, respectively. However, the sensing range of Kinect is small (only about 4 m), and it works only in an indoor environment. Hence, deployment of Kinect is not very suitable in real-world surveillance sites. In contrast, RGB cameras have larger viewing range and work effectively even in outdoor environment. In this work also, we consider that gait videos are captured by RGB cameras. Hence, no further discussions on existing depth-based gait recognition techniques have been done in the literature review.

The ability of deep convolution neural network (CNN) in estimating complex nonlinear functions has been utilized in [17] in developing a view-invariant gait recognition approach. To solve the visual-angle change problem and measure the similarity, the data is mapped to another metric space so as to maximize the distance between the gait signatures from different persons. Two fusion strategies are adopted: (a) score-level fusion and (b) feature-level fusion (person re-identification by fusion strategy). In spite of accurate performance, it is not a practical solution in real-life scenarios, since the response time of deep learning-based solutions is significantly high.

We focus on developing an accurate and efficient gait recognition system. Motivated by the success of GEI and its variants in predicting person identity accurately from his/her gait signatures, we propose a new appearance-based gait recognition method, termed as the *boundary energy image* (BEI), by analyzing the variation in the outer contour of human silhouettes over a gait cycle.

3 Proposed Approach

As mentioned before, we propose an efficient gait feature extraction technique using gait videos from RGB cameras. Figure 1 presents a block diagram of our approach. As seen in the block diagram, as a pre-processing step, silhouettes are extracted from the background, and the bounding box containing each silhouette is normalized to a fixed height and width. Subsequently, a complete cycle is detected and the BEI feature is constructed by aggregating information over the cycle. Finally, feature dimension is reduced by applying principal component analysis and classification is carried out for identifying a subject. Each of the steps in the block diagram is explained in the following sections.

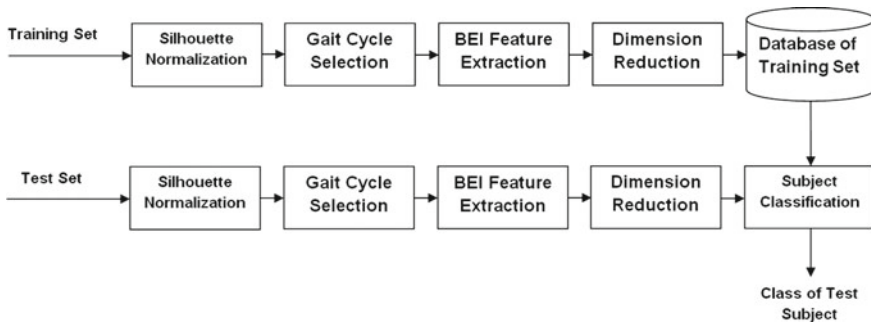


Fig. 1 Block diagram for gait recognition using BEI feature

3.1 Cycle Selection Through Correlation

We select a frame and compute the correlation score of all frames in the input sequence. Figure 2 shows a plot depicting the variation of this correlation score. Peak values in the plot correspond to frames having similar appearance as that of the starting frame. Since, any pose appears twice during a gait cycle, the frames between any two peaks correspond to a half gait cycle.

Let us consider that the gait sequence of a subject contains N silhouette frames, denoted by $f_1, f_2, f_3, f_4, \dots, f_N$. Suppose, each f_i is represented as a column vector, $i = 1, 2, \dots, N$. Correlation between two vectors f_i and f_j is computed as follows:

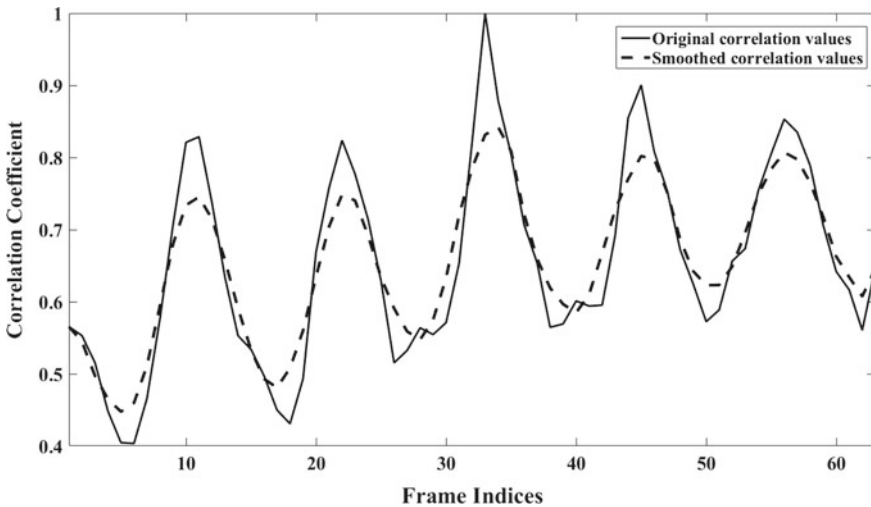


Fig. 2 Correlation coefficient vector graph before and after smoothing

$$r = \frac{N \sum(f_i f_j) - (\sum f_i)(\sum f_j)}{\sqrt{((N \sum(f_i^2) - (\sum f_i)^2)(N \sum(f_j^2) - (\sum f_j)^2))}}$$

Figure 3 shows an example of complete gait cycle by applying the above algorithm on an input gait sequence. Any complete gait cycle contains two half cycles, and in each of these half cycles, similar poses get repeated, once with right limb forward and next with left limb forward. Since it is not possible to distinguish between left and right limbs in binary silhouettes (refer to Fig.3), it can be argued that in the absence of a complete gait cycle, we need at least one half of a gait cycle to extract reliable gait features. In the results section (i.e., Sect.4), we will explain this topic further along with suitable experimental results.

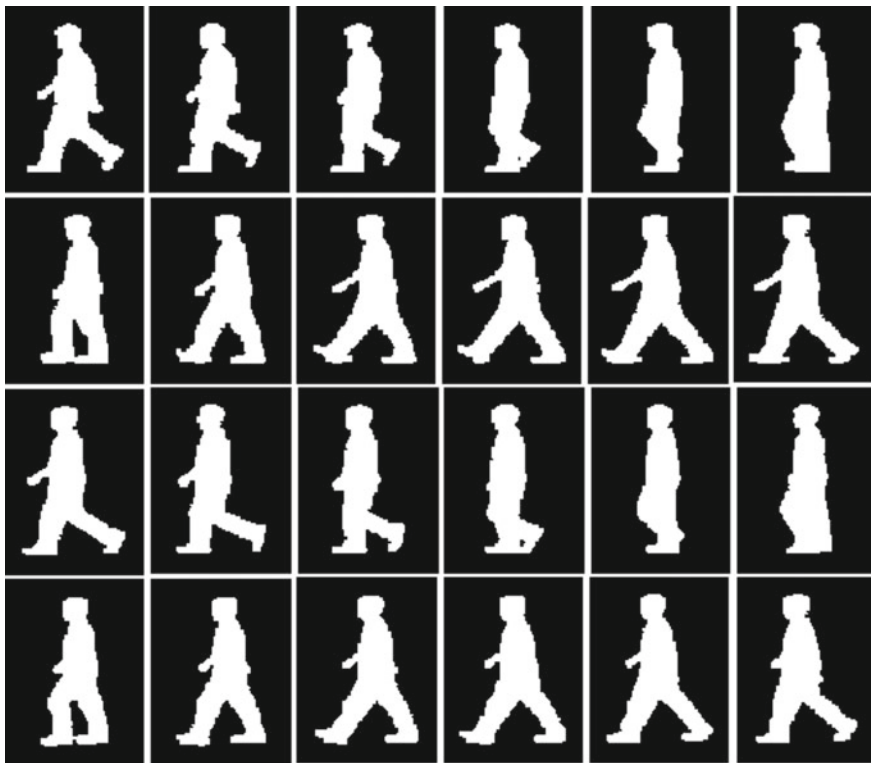


Fig. 3 Silhouette frames after cycle selection

3.2 Boundary Energy Image Representation

In this section, we describe the construction of our proposed gait feature, namely the boundary energy image. The first step is to obtain the outer contour of each silhouette in an input sequence. This is done by applying morphological dilation operation on each binary silhouette by employing a disk-structuring element. We term the absolute difference between the dilated image and the original binary silhouette as the *boundary image*.

Let B_i represents the boundary image corresponding to frame f_i and $B_i(x, y)$ be the pixel intensity at coordinate (x,y) in B_i . The *boundary energy image* (BEI) is computed from the boundary images using the following equation:

$$BEI(x, y) = \frac{1}{N} \sum_{i=1}^N B_i(x, y). \tag{1}$$

Figure 4 shows the sample of boundary images in a gait cycle from two people, and the rightmost image is the corresponding BEI. It can be visually observed that BEI captures accurate silhouette shape and shape variation over a gait cycle. A pixel with high-intensity value in BEI corresponds to those regions in the height normalized bounding box that undergo maximum changes during the gait cycle. Once BEI is computed, principal component analysis is used for projecting the original templates in a low-dimensional subspace and finally applies linear discriminant analysis for classifying an unknown person into one of the subjects in the gallery.

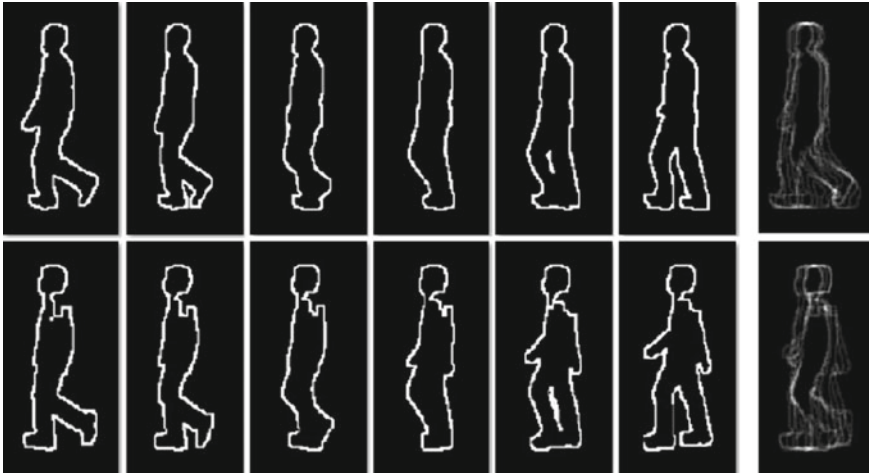


Fig. 4 First six frames in each row show boundary images in the sequences of two different subjects, last frame in each row shows the corresponding BEI

3.3 Dimension Reduction by Principal Component Analysis (PCA) & Classification by Linear Discriminant Analysis (LDA)

Principal component analysis (PCA) [18] is a well-known unsupervised data analysis technique that aims to identify some linear trends and simple patterns in a group of samples. Let us consider n sample points, x_1, x_2, \dots, x_n . PCA tries to find a subspace so that variability among the sample points projected in this space is maximum. If we denote sample mean with symbol of \bar{x} , variance of the projected data on the subspace described by direction vector v will be as follows:

$$\frac{1}{n} \sum_{i=1}^n (v^T x_i - v^T \bar{x})^2 = \frac{1}{n} \sum_{i=1}^n (v^T (x_i - \bar{x}))^2 = v^T \frac{\sum_{i=1}^n (x_i - \bar{x})^T (x_i - \bar{x})}{n} v$$

The linear discriminant analysis (LDA) is a fundamental data analysis approach originally proposed by R. Fisher for discriminating between different types of clusters. The inspiration behind the approach is to determine a subspace of lower dimension, compared to that of the original data sample, in which the data points of the original problem are separable. One of the main advantages of LDA is that the solution can be achieved by solving a generalized eigenvalue system [18]. This is made for fast and massive processing of data samples.

Let us consider $x_1, x_2, \dots, x_p \in R^m$ to be a set of p data samples which belongs to two different class sets A and B . If mean of class A and B is \bar{x}_A and \bar{x}_B respectively, and the number of samples in class A and B are N_A and N_B , respectively, then we can define the positive semi-definite scattered matrices as follows:

$$S_A = \sum_{x \in A} (x - \bar{x}_A)(x - \bar{x}_A)^T, \quad S_B = \sum_{x \in B} (x - \bar{x}_B)(x - \bar{x}_B)^T.$$

Each of these matrices represents the sample variability in a class. Ideally, we would like to find a hyperplane, for which if we project the data samples, their variance would be minimal. Diagonal linear discriminant analysis (DLDA) has been used for classification.

4 Experiments and Results

We have carried out on a system with Intel(R) Core(TM) i7-7700 3.60GHz CPU with 16.0 GB RAM. Two publicly available gait data sets CASIA B [19] and TUMGAID [20] have been used to evaluate our approach. The CASIA B data consists of 106 subjects and six gait sequences for each subject were captured under the following situations: (a) six sequences with normal walk (nm-01 to nm-06), (b) two sequences

with carrying bag (bg-01 and bg-02), (c) two sequences with wearing coat (cl-01 and cl-02). There are 11 different visual guidelines in each gait sequence, the gap with 18° between 0 and 180° , between each of the two closest visual directions. Here we only use data obtained from 90° (fronto-parallel/side view of walking sequences).

The second data set used in the study is the TUMGAID gait database which consists of 304 subjects, and for each subject, six gait sequences under normal circumstances, two sequences with people walked carrying bag and two sequences with people wearing coating shoes are captured. These are tagged as n1, n2, n3, n4, n5, n6, b01, b02, s01, s02, respectively.

In the first experiment, the training set is formed with six normal walking sequences n1–n6 from the TUMGAID data. Similarly, walking sequences nm-01–nm-06 form the training set corresponding to the CASIA B data. Each of the remaining sequences form the test set in case of both the data sets. A disk-structured element is used to dilate each silhouette and henceforth obtain the boundary image. In Tables 1 and 2, we present the accuracy of the individual test sets for both CASIA B and TUMGAID data sets. For each test set, we vary the radius of the structuring element from 1 to 12 and determine the radius for which best performance is obtained.

We observe from Tables 1 and 2 that BEI with disk-structured element of radius three provides the best recognition accuracy, in general. Hence, each of the following experiments has been performed by considering the radius of the structuring element as three.

Table 1 Table showing the percentage accuracy of our method on the different test sets in CASIA B data for different sizes of the structuring element

Probe frame name	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	$r = 8$	$r = 9$	$r = 10$	$r = 11$	$r = 12$
bg-01	72.64	81.13	85.84	83.01	79.24	75.47	73.58	77.35	73.58	72.64	73.58	66.98
bg-02	72.11	78.84	86.53	85.57	82.69	78.84	75.96	78.84	75.00	71.15	72.11	69.23
cl-01	57.84	79.80	59.80	53.92	51.96	52.94	51.96	50.98	50.98	47.05	50.00	48.03
cl-02	61.00	63.00	63.00	58.00	56.00	55.00	55.00	56.00	56.00	55.00	56.00	53.00
Average	65.90	70.69	73.79	70.12	67.47	65.56	64.12	65.79	63.89	61.46	62.92	59.31

Table 2 Table showing the percentage accuracy of our method on the different test sets in TUMGAID data for different sizes of the structuring element

Probe frame name	$r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$	$r = 8$	$r = 9$	$r = 10$	$r = 11$	$r = 12$
b01	6.57	25.98	26.97	22.69	19.07	16.44	13.81	12.82	12.17	10.52	11.84	12.17
b02	4.30	26.49	28.80	21.19	13.24	10.59	7.61	6.62	6.62	5.62	4.96	4.30
s01	11.66	41.00	51.66	52.66	53.33	60.66	60.66	60.66	60.33	64.00	63.66	65.66
s02	12.41	50.00	58.38	57.38	61.40	63.75	64.42	65.43	67.78	66.77	67.78	65.77
Average	8.74	35.86	41.45	38.48	36.76	37.86	36.63	36.38	36.72	36.73	37.06	36.97

Next, we study the performance of our algorithm on the CASIA-B gait data set with the increase in the volume of training data. We test with sequences containing half gait cycles and multiples of half gait cycle and study its impact on the recognition accuracy. Results are shown in Fig. 5 considering the training set to be the six normal walking sequences, namely nm-01, nm-02, nm-03, nm-04, nm-05, nm-06, and the test sets to be bg-01, bg-02, cl-01 and cl-02. It can be seen from Fig. 5 that four half cycles (or equivalently, two full cycles) provide the best result. If the number of half gait cycles is too less, then the accuracy suffers due to the presence of insufficient information. On the other hand, if a higher number of gait cycles is considered, accuracy decreases possibly due to the presence of noisy silhouettes. We could not carry out similar experiments with the TUMGAID data since most of the sequences in this data contain at most one complete cycle which is equivalent to two half cycles. Rest of the experimental results reported in this paper consider four half gait cycles (or two full cycles) for the CASIA B data and two half gait cycles (or one full cycle) for the TUMGAID data.

In our next experiment, we compare the performance of the proposed approach with other popular gait recognition techniques, namely those using GENI [13], AEI [11], GEI [1] and GFI [5] features. Results are shown in Table 3 for CASIA B data, and in Table 4 for TUMGAID data.

It can be seen from the tables that GFI has the best recognition performance among all the features used in the study. However, the response time of GFI has been found to be significantly high (approximately, 588 s). Although BEI usually outperforms most of the other features by a high margin, accuracy of AEI has been found to be comparable with that of BEI in case of the TUMGAID data. But BEI significantly outperforms AEI in case of CASIA B data, and the overall average performance of BEI has also been seen to be better than that of AEI.

Finally, we plot a rank-based improvement in average classification accuracy on all the test sets in CASIA B and TUM GAID data. Results for the two data sets are shown

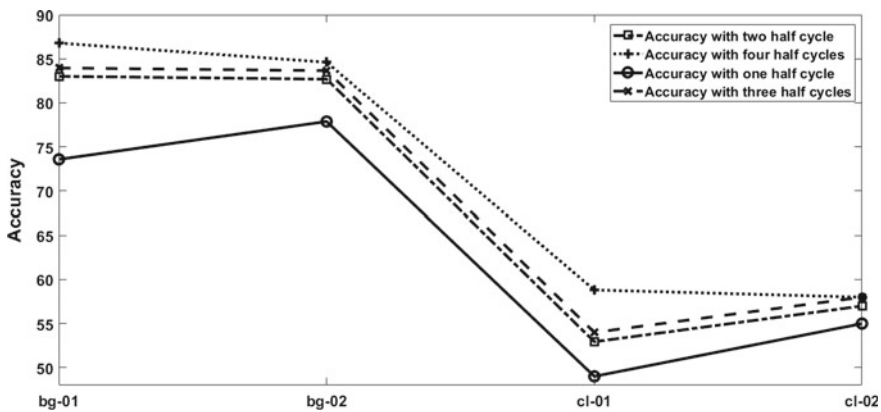


Fig. 5 Improvement in gait recognition accuracy with increase in the number of half cycles on CASIA B data

Table 3 Table showing comparative performance analysis of our approach with that of existing approaches, namely those using GEnI [13], AEI [11], GEI [1] and GFI [5] features on CASIA B data

Probe frame name	BEI	GEnI	AEI	GEI	GFI
bg-01	85.84	68.86	73.58	71.69	83.96
bg-02	86.53	70.19	72.19	62.50	87.50
cl-01	59.80	46.07	51.96	49.01	64.70
cl-02	63.00	50.00	52.00	53.00	64.00
Average	73.79	58.78	61.93	59.92	75.04

Table 4 Table showing comparative performance analysis of our approach with that of existing approaches, namely those using GEnI [13], AEI [11], GEI [1] and GFI [5] features on TUMGAID data

Probe frame name	BEI	GEnI	AEI	GEI	GFI
b01	26.97	28.28	29.60	12.17	27.96
b02	28.80	21.85	30.46	5.96	24.50
s01	51.66	49.66	51.00	66.66	62.66
s02	58.38	55.70	58.05	69.12	69.12
Average	41.45	38.87	42.28	38.48	46.06

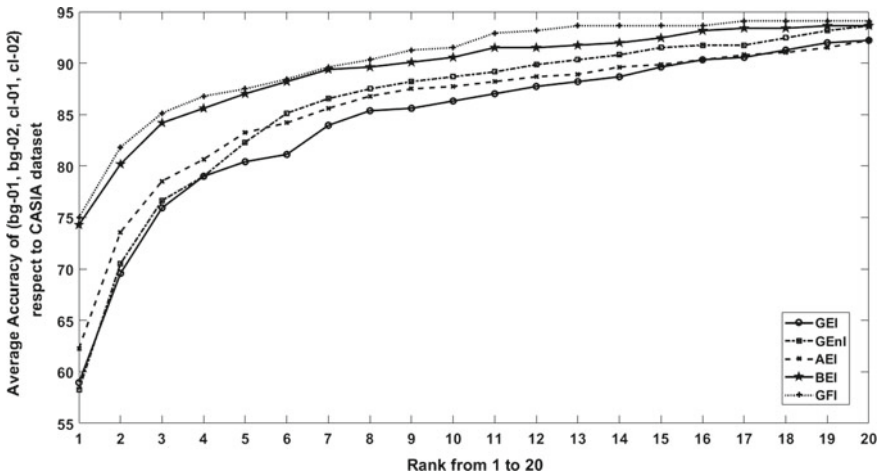


Fig. 6 Rank-based performance comparison between average accuracy of BEI with that of GEnI [13], AEI [11], GEI [5] and GFI [13] using CASIA B data set

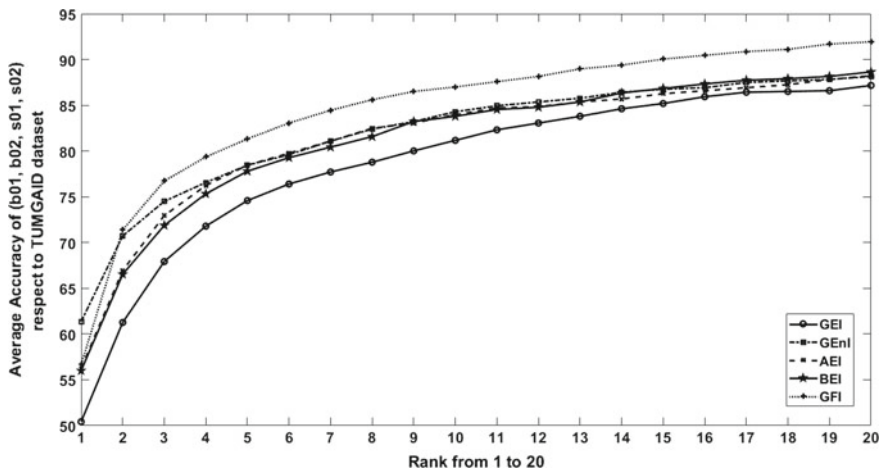


Fig. 7 Rank-based performance comparison between average accuracy of BEI with that of GEnI [13], AEI [11], GEI [5] and GFI [13] using TUMGAID data set

in Figs. 6 and 7, respectively, by means of cumulative match characteristic curves. As expected, the classification accuracy of each of the gait recognition approaches increases with increment in the value of rank. Our method achieves the 85.61% accuracy mark within a rank of 4 for CASIA B data, which is quite remarkable. Accuracy of each of the methods on TUM GAID data is slightly less. In general, our method achieves a better recognition rate than most of the other features, except GFI [13] for all values of rank. However, construction of the GFI feature is time-intensive due to extensive mathematical calculations and is thus not suitable for practical purposes. Average feature construction time of BEI for a given gait sequence is 12.26 s, whereas the corresponding time for GFI is 588.53 s. From our experiments, we conclude that BEI is indeed an accurate and efficient feature for gait recognition from complete cycles.

5 Conclusions and Future Scope

The manuscript introduces a new appearance-based gait feature termed as the *boundary energy image* (BEI), for recognizing a person from gait videos captured by an RGB camera. BEI represents human motion condensed into single image. Unlike GEI, which captures the spatio-temporal changes of the entire silhouette over a gait cycle, the present feature preserves information about the spatio-temporal variations in the silhouette boundary over a gait cycle. Recognition using this feature has been seen to be very accurate as well as time-efficient.

Comparison with existing features like gait entropy image, active energy image and gait energy image on two data sets shows that the proposed feature outperforms

each of the other features by a substantially high margin. In future, the method can be extended to construct a key-pose-based recognition technique similar to that using pose energy image. Finally, it can be tested whether it can be fused with a suitable occlusion handling strategy to perform gait recognition in the presence of occlusion.

Acknowledgements Authors sincerely acknowledge Dr. Prasenjit Banerjee, Nalanda Foundation Mentor, for his support. Authors also share their sincere gratitude to HP Innovation Incubator program for giving them the opportunity toward initiation of this project.

References

1. J. Man, B. Bhanu, Individual recognition using gait energy image. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(2), 316–322 (2006)
2. A. Roy, S. Sural, J. Mukherjee, Gait recognition using pose kinematics and pose energy image. *Sig. Process.* **92**(3), 780–792 (2012)
3. Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, Gait recognition using a view transformation model in the frequency domain, in *European Conference on Computer Vision* (Springer, 2006), pp. 151–163
4. S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, Robust view transformation model for gait recognition, in *2011 18th IEEE International Conference on Image Processing (ICIP)* (IEEE, 2011), pp. 2073–2076
5. T.H. Lam, K.H. Cheung, J.N. Liu, Gait flow image: a silhouette-based gait representation for human identification. *Pattern Recogn.* **44**, 973–987 (2011). (Elsevier)
6. T.H. Lam, R.S. Lee, A new representation for human gait recognition: Motion silhouettes image (MSI) (2006), pp. 612–618
7. H. Lee, S. Hong, I.F. Nizami, E. Kim, A noise robust gait representation: motion energy image. *Int. J. Control Autom. Syst.* **7**(4), 638–643 (2009)
8. J. Liu, N. Zheng, Gait history image: a novel temporal template for gait recognition, in *2007 IEEE International Conference on Multimedia and Expo* (IEEE, 2007), pp. 663–666
9. Q. Ma, S. Wang, D. Nie, J. Qiu, Recognizing humans based on gait moment image, in *Eighth ACIS International Conference on Software Engineering, Artificial Intelligence, Networking, and Parallel/Distributed Computing, 2007. SNPD 2007*, vol. 2 (IEEE, 2007), pp. 606–610
10. C. Chen, J. Liang, H. Zhao, H. Hu, J. Tian, Frame difference energy image for gait recognition with incomplete silhouettes. *Pattern Recogn. Lett.* **30**(11), 977–984 (2009)
11. Y. Qi, Gait recognition based on active energy image and parameter-adaptive kernel PCA, in *Processing of Information Technology and Artificial Intelligence Conference*, pp. 156–159
12. L. Chunli, W. Kejun, A behavior classification based on enhanced gait energy image, in *2010 2nd International Conference on Networking and Digital Society (ICNDS)*, vol. 2 (IEEE, 2010), pp. 589–592
13. K. Bashir, T. Xiang, S. Gong, Gait recognition using gait entropy image (2009)
14. C. Wang, J. Zhang, L. Wang, J. Pu, X. Yuan, Human identification using temporal information preserving gait template. *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(11), 2164–2176 (2012)
15. S. Sivapalan, D. Chen, S. Denman, S. Sridharan, C. Fookes, Gait energy volumes and frontal gait recognition using depth images, in *2011 International Joint Conference on Biometrics (IJCB)* (IEEE, 2011), pp. 1–6
16. P. Chattopadhyay, A. Roy, S. Sural, J. Mukhopadhyay, Pose depth volume extraction from RGB-D streams for frontal gait recognition. *J. Vis. Commun. Image Represent.* **25**(1), 53–63 (2014)
17. Z. Liu, Z. Zhang, Q. Wu, Y. Wang, Enhancing person re-identification by integrating gait biometric. *Neurocomputing* **168**, 1144–1156 (2015)

18. P. Xanthopoulos, P.M. Pardalos, T.B. Trafalis, Linear discriminant analysis, in *Robust Data Mining* (Springer, 2013), pp. 27–33
19. I. Bouchrika, M. Goffredo, J. Carter, M. Nixon, On using gait in forensic biometrics. *J. Forensic Sci.* **56**(4), 882–889 (2011)
20. M. Hofmann, J. Geiger, S. Bachmann, B. Schuller, G. Rigoll, The tum gait from audio, image and depth (gaid) database: multimodal recognition of subjects and traits. *J. Vis. Commun. Image Represent.* **25**(1), 195–206 (2014)

Supervised Classification Algorithms in Machine Learning: A Survey and Review



Pratap Chandra Sen, Mahimarnab Hajra and Mitadru Ghosh

Abstract Machine learning is currently one of the hottest topics that enable machines to learn from data and build predictions without being explicitly programmed for that task, automatically without human involvement. Supervised learning is one of two broad branches of machine learning that makes the model enable to predict future outcomes after they are trained based on past data where we use input/output pairs or the labeled data to train the model with the goal to produce a function that is approximated enough to be able to predict outputs for new inputs when introduced to them. Supervised learning problems can be grouped into regression problems and classification problems. A regression problem is when outputs are continuous whereas a classification problem is when outputs are categorical. This paper tries to compare different types of classification algorithms precisely widely used ones on the basis of some basic conceptions though it is obvious that a complete and comprehensive review and survey of all the supervised learning classification algorithms possibly cannot be accomplished by a single paper, but the references cited in this paper hopefully cover the significant theoretical issues and our survey has been kept limited to the widely used algorithms because the field is highly growing and not possible to cover all the algorithms in a single paper. One more point to be mentioned here that any study of complex procedure like neural networks has not been included as it has been tried to keep the content as much simple as possible.

Keywords Machine learning · Classification algorithm · Supervised learning · Accuracy

P. C. Sen (✉) · M. Hajra · M. Ghosh
Department of Computer Science and Engineering, Institute of Engineering
and Management, Kolkata, India
e-mail: pchsen97@gmail.com

M. Hajra
e-mail: mahimarnab2014@gmail.com

M. Ghosh
e-mail: mitadrugosh100@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling
and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_11

1 Introduction

This paper will focus on summarizing the key advantages of different, widely renowned, and most frequently used machine learning algorithms used for classification task and to do a comparative study on them to find out which algorithm works best based on different parameters so that while doing any classification task, one can know when to use which algorithm to get best result. Classification can be performed on both the unstructured or structured dataset. Members of a dataset are classified according to some given label or category and for new input instances, the class or label that will be assigned to it is predicted by this technique. Some terms frequently used in this field are also introduced here with explanation. A classifier algorithm is an algorithm that learns from the training set and then assigns new data point to a particular class. A classification model concludes some valid mapping function from training dataset and predicts the class label with the help of the mapping function for the new data entry. An attribute or feature is a parameter found in the given problem set that can sufficiently help to build an accurate predictive model. There are different types of classification task.

Binary classification is a classification with two possible outcomes. For example, weather forecast (it will rain or not), spam or fraud detection (predict whether an email is spam or not).

Multi-label classification is a classification task with more than two possible outcomes. For example, classify academic performance of students as excellent or good or average or poor.

In classification, a sample can even be mapped to more than one tag labels. For example, a sample of news article can be labeled as sport article, an article about some player, and an article about a certain venue at the same time.

A classification model can be built by following steps:

1. Collect and clean the dataset or data preprocessing.
2. Make the classifier model initialized.
3. Split the dataset using cross-validation and feed the classifier model with training data. Python-based scikit-learn package has inbuilt methods named $\text{fit_transform}(X, Y)/\text{fit}(X, Y)$ that map the input data member set X and corresponding label set Y to prepare the classifier model.
4. Predict the label for a new observation data. There is also a method $\text{predict}(X)$ that returns the mapped label Y for the input instance X .
5. Evaluate error rate of the classifier model on the test dataset (Fig. 1).

So, the very first step is to collect the dataset. The most relevant attributes/fields/features are to be figured out next, which is called feature extraction. “Brute-force” method is the simplest one which isolates the most relevant/informative attributes by measuring everything available. Besides, there is a process named feature subset selection that identifies and eliminates as many redundant, irrelevant, and unnecessary attributes as possible.¹ Secondly, it is required to do some data prepro-

¹See Ref. [28].

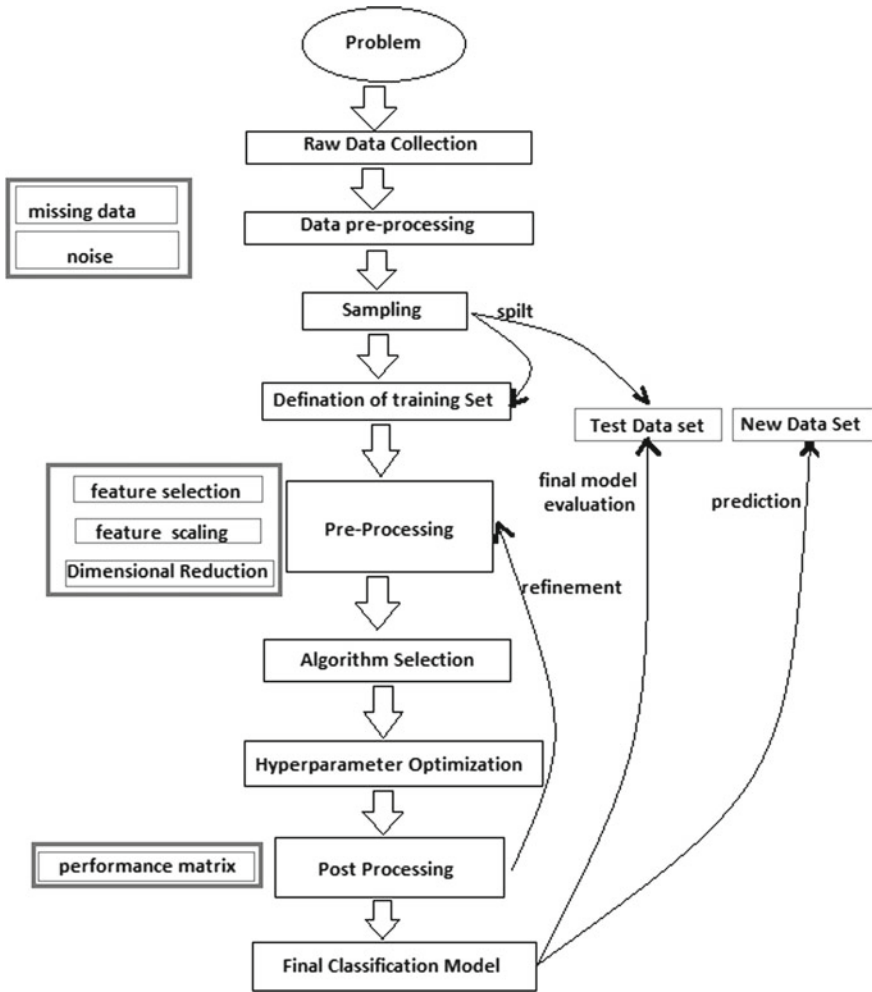


Fig. 1 Working flowchart of supervised classification model

cessing.² The dataset often contains noise (outliers) and missing feature values and some categories that needs to be converted into dummy variable. Various methods are available to deal with missing data³ which should be paid attention. Thus, it requires a well data preprocessing.

²See Ref. [5].

³See Ref. [27].

2 General Issues of Supervised Classification

There are contemporary techniques by Batista and Monard⁴ to detect outliers. All the data analysis is done on significant samples taken from survey dataset and there available a few number of approaches for sampling purpose. It leads the dataset toward significant dimensionality reduction and makes the different data mining algorithms capable of working more efficiently and also faster. Due to inter-dependencies of many attributes, often results inordinate influences on the accuracy level of these classification models heavily. Feature transformation is an efficient technique that constructs new features or dummy features from the originally given feature set to stand against these unduly effects on model by feature inter-dependencies. This method performs the classification task with the help of already classified/categorized training data. The input point under consideration and the corresponding desired output is already known for the training dataset. When the supervised learning algorithms for classification purpose are fed with data that is when it is trained with the already known training dataset, it become capable to generalize the new unseen data and predict corresponding class. Here, the dataset is about the *salary of employees of a company, attributes: 7 and instances: 48,842*. From the census bureau database,⁵ this data sample is extracted.

The next section describes the basic definition and working method of most widely used supervised classification machine learning algorithms with a brief review so that the survey explanation can be well understood. Those algorithms learn through different ways and based on that we can classify them.

2.1 Logically Learning Algorithm

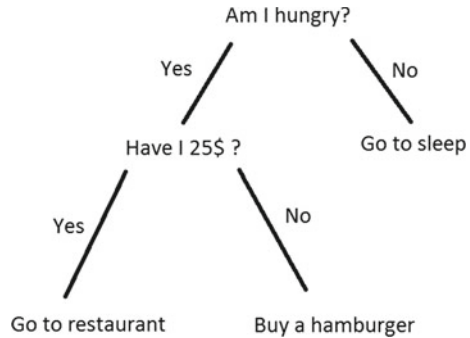
Logic-based algorithms deal with the problem with step by step data streaming with functioning a logic in each step. Here **decision tree** has been overviewed as a classical example of logic-based algorithm. It is statistical technique used in regression and classification. Given a data of features along their correct labels, decision tree is an example of logically learning supervised algorithm mainly used for classification that generates a set of decision sequences that if followed will lead to predicting the label of an unlabeled data.

- A. *Definition* A decision supported tool that is represented as a logical tree having a range of conditions and conclusions as nodes and branches that connects the conditions with conclusions. Decision tree takes some decision in every stage of proceeding and considers some alternative choices of action, and among them it selects the most relevant alternatives. A decision tree is a mathematical model used to make decisions. It uses estimates and probabilities to calculate

⁴See Ref. [13].

⁵<https://www.census.gov/DES/www/welcome.html>.

Fig. 2 A decision tree to decide how to have food



likely outcomes and helps to decide whether the net gain from a decision is worthwhile.

- B. *Method* An object defined by a set of features is an input here, output is a decision for the corresponding input. In the tree structure, each internal node tests an attribute and each of the nodes is assigned to a classification. Following is an example of how a decision tree looks like (Fig. 2).
- C. *Details* Like SVM, decision tree also works fine for both categorical and continuous dependent instances. This method actually constructs a sequential decision stream-based model on actual values of features in the dataset. Decisions are split into tree-like structure. A decision is made at every node of this tree unless a prediction is made for a certain input data item. Decision trees are trained on data for classification problems. This algorithm works very fast often with satisfactory accuracy. It is a big favorite and widely used in machine learning and also works efficiently on comparatively less amount of dataset. This algorithm splits the set of data items into two or more homogeneous sets based on most significant attribute to make as distinct groups as possible. To split up the data into different groups, various techniques like information gain, chi-square, Gini, entropy, etc. are used. Entropy has a well discriminatory power for classification. By the name it is known, it represents some kind of randomness. Entropy defines the amount of randomness in features and measures discriminatory capability of an attribute for the classification problem. Information gain ranks the attributes for the purpose of filtering at given node.
- D. *Examples of decision tree algorithms* Widely used algorithms are *ID3*, *C4.5*, *CART*.
- E. *Advantage* Decision tree is a simple method, easy to understand and visualize, fast, requires less data preprocessing, and can deal with both categorical and numerical data.
- F. *Disadvantage* Sometimes this algorithm may lead to a complex tree structure not generalized enough, besides it is quite unstable model.
- G. *Application* It has applications in various fields like *determining galaxy counts*, *control system*, *financial analysis*, etc.

In the next session, the latest method of class prediction has been introduced named as support vector machine.

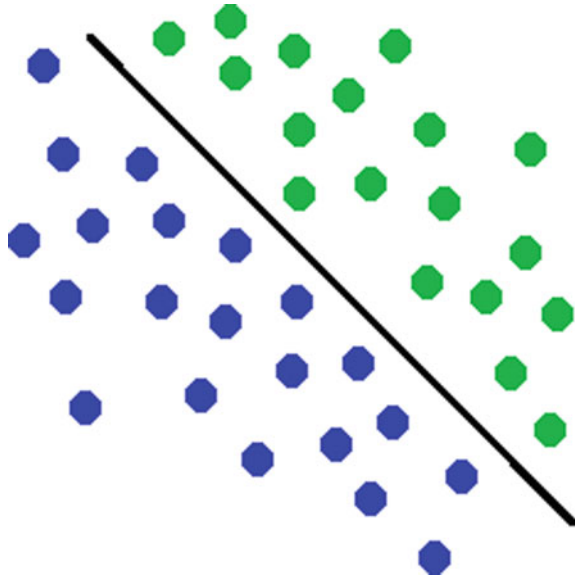
2.2 Support Vector Machine

Support vector machine or in short SVM is an advanced supervised algorithm that can deal with both regression and classification tasks though it is more favorable for the classification. It can handle multiple continuous and categorical instances.

- A. *Definition* It represents the dataset items or records each having “ n ” number of features plotted as points in a n -dimensional space segregated into classes by a clear margin widest possible known as hyperplane. Into that same n -dimensional space, data items are then mapped to get the prediction of the category they belong to, based on the side of hyperplane they fall.
- B. *Method* The coordinates of these data items are actually termed as “support vectors.” If there is a linear type hyperplane present between these classes, the problem becomes very simple. But, another question arises is the manual addition of these attributes to have a hyperplane is actually necessary or not. The answer is “Not Really” as this algorithm uses a technique known as “kerneltrick.” Kernels are nothing but a group of functions which transforms the low-dimensional input space into a higher-dimensional space. In simple words, what it does is, by performing some typical data transformations that are actually complex in nature, does some conversions and non-separable problem becomes separable problem. After that, it figures out the procedure to segregate the data according to the defined labels.
- C. *Details* It is the newest supervised classification technique.⁶ An example is depicted in the illustration below. Here, two kinds of objects are considered either belonging to class GREEN or to class BLUE. The segregating boundary has all GREEN objects on the right side of it and on the left, all the BLUE objects lie. A new observation object when plotted in the n -dimensional vector space is labeled as GREEN if it lies in the right side of the separating boundary or hyperplane, otherwise it is tagged with the label BLUE (Fig. 3).
The shown figure depicts a linear classifier. Most of the classification tasks are not really simple like this one and often leads to complex operations that should be executed in order to do an optimal segregation in case of some complex structures. SVM can be classified in two categories: C-SVM classification and nu-SVM classification.
- D. *Advantages* It shows a noticeable hike in performance where the “ n ” of the n -dimensional space is greater than the total size of sample set. Thus, it is a good choice to pick this algorithm while dealing high-dimensional data. If the hyperplane is well built, it shows high performance. This is also memory efficient

⁶See Ref. [12].

Fig. 3 A depiction of SVM classification with hyperplane



because of the use of subset training points in the decision function (priorly known as support vectors).

- E. *Disadvantages* One flaw present here is that training time is quite high compared to others so if the dataset is very large then the prediction task turns out to be slow noticeably. When target classes are overlapping that is the dataset has more noise, its performance also decreases. In case of probability estimation calculation, a n -fold cross-validation is used which is again quite expensive for computation time.
- F. *Applications* This is used in stock marketing for various predictions besides it has various applications in *bioinformatics, face detection, classification of images, handwriting recognition, classification of images, etc.*

Unlike the last one discussed, there is a branch of algorithm that learns predicting task through broad statistical description and this one has been explained next.

2.3 *Statistics-Based Algorithm*

Statistics-based algorithms generalize problems with help of distributive statistics and look into the distribution structure to continue the predicting task. Here, *Naïve Bayes* has been explained as a popular example of statistics-based algorithm Naive Bayes is a simple but surprisingly powerful algorithm for predictive modeling. This is a collection of classification techniques based on Bayesian theorem of probability.

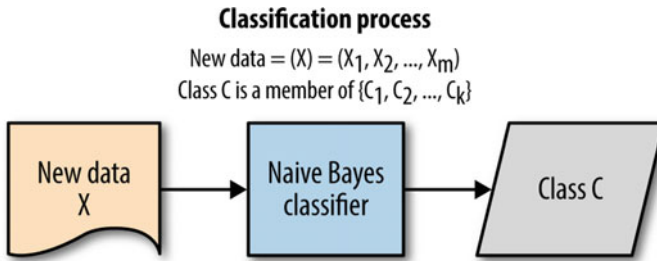


Fig. 4 Classification model of Naïve Bayes

- A. *Definition* Naive Bayes classifier produces the probabilities for every case. Then it predicts the highest probability outcome. A naive assumption is made that the features are independent, classification of every pair of features is independent of each other.
- B. *Details* These classifiers are capable of handling an arbitrary number of independent continuous and categorical variables efficiently. Let us consider a set of variables, $X = \{x_1, x_2, x_3, \dots, x_t\}$; it is required to find out the posterior probability for the event C_j from the sample space set $C = \{c_1, c_2, c_3, \dots, c_t\}$. Simply, the predictor is X and C is the set of categorical levels present in the dependent variable. Applying Bayes' rule:

$$P(C_j | x_1, x_2, x_3, \dots, x_t) \cdot P(x_1, x_2, x_3, \dots, x_t | C_j) P(C_j)$$

where $P(C_j | x_1, x_2, x_3, \dots, x_t)$ is the posterior probability that is the probability of the event X belonging to C_j is indicated. In Naive Bayes, there is an assumption that the conditional probabilities of the independent variables have statistical independence. Using Bayes' rule, a new case X is labeled with a class level C_j that accomplishes the highest posterior probability. Although this naive assumption that the predictor variables are independent of each other is not always accurate actually. This assumption makes the classification process simpler, as it allows the class conditional densities $P(x_d | C_j)$ to be calculated for each variable separately and thus a multidimensional task is reduced to some one-dimensional tasks. More precisely, it converts a high-dimensional density estimation task to a one-dimensional kernel density estimation. Classification task remains unaffected as this assumption does not greatly affect the posterior probabilities, mainly in regions located closely around the decision boundaries (Fig. 4).

- C. *Different ways of Naïve Bayes Modeling* Naïve Bayes can be modeled in several different ways including normal, gamma, and Poisson density functions.
- D. *Advantages* Naive Bayes algorithm does not require huge dataset, for estimation of required parameters, a small-sized training data is good enough here. It also performs explicit probability calculation for hypothesis. A useful prospect to

understand various learning algorithm is also provided by this method. In the comparison with other subtle methods, this classifier is also intensely fast. It can solve diagnostic problems efficiently.

- E. *Disadvantage* This method is comparatively known to be a bad estimator.
- F. *Application* It has various applications like *recommendation system, text classification, spam filtering, real-time prediction*, etc.

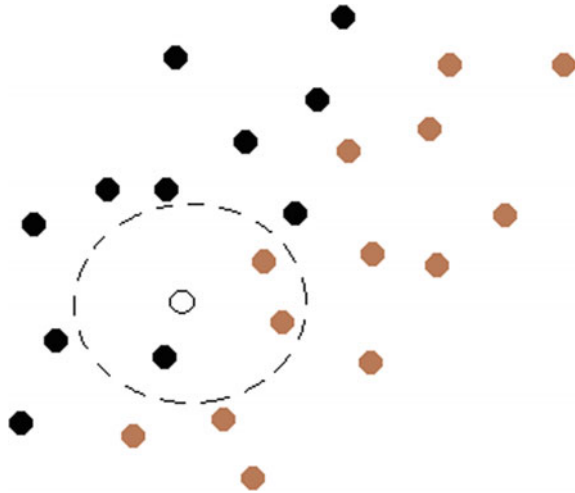
One more renowned process of supervised machine learning is lazy learning which is also known as instance-based learning which has been described in next section.

2.4 Lazy Learning Algorithm

There is another renowned category that falls under the title of statistical methods. This method in general known as “Instance-based learning” which delays the process of generalization until the classification task is performed resulting the naming of this algorithm with the tag of “lazy” and widely known as lazy learning algorithm. Here, **kNN** or K-nearest neighbor algorithm has been discussed as a fine example of instance-based learning or lazy learning. kNN is one of the easiest to understand and very simple classification algorithms available. Despite of simplicity, it can give highly competitive results. It can deal with both of the classification and regression types predictive problems. However, it is more concisely used to perform and execute classification task.

- A. *Definition* K-nearest neighbor stores all available records and predicts the class of a new instances giving attention to similarity measurements from the nearest neighbors in likelihood. This classification technique is known to be lazy learning method because it keeps the data members stored simply in efficient data structures like hash table by virtue of which computation cost becomes less to check and apply the appropriate distance function between the new observation and all k number of different data points stored and then come to any conclusion about the label of the new data point, without constructing a mapping function or internal model like other classification algorithms. Result is obtained from a simple majority support of the k number of nearest neighbors of each new data point.
- B. *Details* The “K” in k-NN algorithm is “k” number of nearest neighbors whose vote is taken to predict label for a new record around which those neighbors are situated. Let us say $K = 3$. Then a circle with new data item as center will be visualized just as big as to enclose only three nearest neighbor data points on the plane and the virtue of the distance between the record and each of the neighbors will decide the label of the new record. For a given K-value, boundaries of each class can be built. These boundaries can efficiently segregate one class from another. It is observed that with increasing value of K, the boundary becomes smoother. If K increases to a very high vale and finally tends to infinity, it all becomes finally one class or the one with the total majority.

Fig. 5 An example of K-nearest neighbor algorithm approach



There are two parameters: the validation error rate and training error rate that need to be accessed and tested on different K -value (Fig. 5).

Here, a new member has been shown by a uncolored circle and its class has to be predicted based on the distance from its three nearest neighbors as we are considering three nearest neighbors so the value of k is 3 here and so a circle having center at the new member has been drawn covering only three previous members. Now, by judging the shortest distance, it is resulted that the new member will be black. Thus, we can predict the label of an unclassified record considering its distance from nearest members, nearer the member more the effect of its label on the new record.

- C. *Advantages* Powerful against noisy training data, works effectively with large training data, besides implementation is simple.
- D. *Disadvantage* One big flaw of this algorithm is for every new instance, all the distance from K neighbors needs to be calculated again and again which leads to high computational time consumption. The value of K needs to be determined correctly for lower error rate.
- E. *Applications* It is often seen that kNN is used in search applications where “similar” items are searched by the user; that is, when the task is some kind of the form of “find items similar to this one⁷”. It is highly recommended *in e-discovery packages* and for *recommended system* also.

In the next section, the comparative study of above-mentioned algorithms has been done with relative to some important parameters and in the end, the overall accuracy of these algorithms has been given and this test has been done on the earlier mentioned dataset.

⁷<https://www.quora.com/Industry-applications-of-the-K-nearest-neighbor-algorithm>.

Table 1 A comparative study on widely used four supervised classification algorithm

Comparison parameters	Decision tree	Naive Bayes	k-NN	SVM
Learning speed	Average	Best	Best	Worst
Classification speed	Best	Best	Worst	Best
Performance with presence of missing value	Average	Best	Worst	Good
Performance with non-relevant features	Average	Good	Good	Best
Noise tolerance	Good	Average	Average	Good
Performance on discrete/binary attributes	Good	Average	Average	Worst
Tolerance with parity problems	Good	Worst	Worst	Average
Clarity on Classification prediction	Best	Best	Average	Worst
Handling of model parameter	Average	Best	Average	Worst
Overall accuracy	Good (84.13%)	Worst (80.14%)	Good (83.65%)	Best (84.94%)

3 Comparison Table

So, here is the relative comparison between widely used and most popular supervised classification algorithms. Accuracy is measured using confusion matrix. Here, accuracy is defined as ratio of correct predictions to all observations. It is the most instinctive performance measure. The comparison of accuracy is done by applying the algorithms on earlier mentioned dataset. This paper describes different schemes of supervised learning and their parameters which influence to achieve accuracy rather than just detailed study over them (Table 1).

References

1. J.H. Friedman, Regularized discriminant analysis. *J. Am. Stat. Assoc.* **84**(405), 165–175 (1989)
2. N. Friedman, D. Geiger, M. Goldszmidt, Bayesian network classifiers. *Mach. Learn.* **29**, 131–163 (1997)
3. N. Friedman, D. Koller, Being Bayesian about network structure: A Bayesian approach to structure discovery in Bayesian networks. *Mach. Learn.* **50**(1), 95–125 (2003)

4. R.G. Cowell, Conditions under which conditional independence and scoring methods lead to identical selection of Bayesian network models, in *Proceedings of 17th International Conference on Uncertainty in Artificial Intelligence*
5. R.L. De Mantaras, E. Armengol, Machine learning from examples: inductive and lazy methods. *Data Knowl. Eng.* **25**(1–2), 99–123 (1998)
6. D. Heckerman, C. Meek, G. Cooper, A Bayesian approach to causal discovery, in *Computation, Causation, and Discovery*, ed. by C. Glymour, G. Cooper (MIT Press, Cambridge, 1999), pp. 141–165
7. N. Japkowicz, S. Stephen, The class imbalance problem: a systematic study. *Intell. Data Anal.* **6**(5), 429–449 (2002)
8. D.E. Rumelhart, G.E. Hinton, R.J. Williams, Learning internal representations by error propagation, in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, vol. 1, ed. by D.E. Rumelhart, J.L. McClelland et al. (MIT Press, Cambridge, MA), pp. 318–362.
9. A. Roy, On connectionism, rule extraction, and brain-like learning. *IEEE Trans. Fuzzy Syst.* **8**(2), 222–227; L. Breiman, Bagging predictors. *Mach. Learn.* **24**, 123–140
10. I.J. Good, *Probability and the Weighing of Evidence* (London, Charles Grin)
11. N.J. Nilsson, *Learning Machines* (McGraw-Hill, New York)
12. B. Cestnik, I. Kononenko, I. Bratko, Assistant 86: a knowledge elicitation tool for sophisticated users, in *Proceedings of the Second European Working Session on Learning*, pp. 31–45
13. B. Cestnik, Estimating probabilities: a crucial task in machine learning, in *Proceedings of the European Conference on Artificial Intelligence*, pp. 147–149
14. T. Cover, P. Hart, Nearest neighbor pattern classification. *IEEE Trans. Inf. Theor.* **13**(1), 21–27 (1967)
15. W. Cohen, Fast effective rule induction, in *Proceedings of ICML-95*, pp. 115–123
16. J.M. Kalyan Roy, Image similarity measure using color histogram, color coherence vector, and sobel method. *Int. J. Sci. Res. (IJSR)* **2**(1), 538–543 (2013), India Online ISSN: 23197064
17. A. Smola, S. Vishwanathan, *Introduction to Machine Learning* (United Kingdom at the University Press, Cambridge, 2010)
18. R.G. Cowell, Conditions under which conditional independence and scoring methods lead to identical selection of Bayesian network models, in *Proceedings of 17th International Conference on Uncertainty in Artificial Intelligence* (2001)
19. W. Gerstner, *Supervised learning for neural networks: a tutorial with JAVA exercises*
20. R. Olshen L. Breiman, J.H. Friedman, “Classification and regression trees.” Belmont CA Wadsworth International group, 1984. B. C. U. P.E.tgoff, “Multivariate decision trees: machine learning,” no. 19, 1995, pp. 45–47.
21. T. Dietterich, M. Kearns, Y. Mansour, Applying the weak learning framework to understand and improve C4. 5 (Sanfrancisco, Morgan), pp. 96–104
22. Kufmann, in *Proceeding of the 13th International Conference on Machine Learning* (1996)
23. K.M.A. Chai, H.L. Chieu, H.T. Ng, Bayesian online classifiers for text classification and filtering, in *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (August 2002), pp. 97–104
24. T. Elomaa, The biases of decision treepruning strategies (Springer, 1999), Lecture Notes in Computer Science, vol. 1642, pp. 63–74
25. A. Kalousis, G. Gama, On data and algorithms: understanding inductive performance. *Mach. Learn.* **54**, 275–312 (2004)
26. P. Brazdil, C. Soares, J. Da Costa, ranking learning algorithms: using IBL and meta-learning on accuracy and time results. *Mach. Learn.* **50**, 251–277 (2003)
27. G. Batista, M.C. Monard, An analysis of four missing data treatment methods for supervised learning. *Appl. Artif. Intell.* **17**, 519–533 (2003)
28. J. Basak, R. Kothari, A classification paradigm for distributed vertically partitioned data. *Neural Comput.* **16**(7), 1525–1544 (2004)
29. A. Blum, Empirical support for winnow and weighted-majority algorithms: results on a calendar scheduling domain. *Mach. Learn.* **26**(1), 5–23 (1997)

30. A. Bonarini, *An Introduction to Learning Fuzzy Classifier Systems* (2000), Lecture Notes in Computer Science, vol. 1813, pp. 83–92
31. R. Bouckaert, Choosing between two learning algorithms based on calibrated tests, in *Proceedings of 20th International Conference on Machine Learning* (Morgan Kaufmann, 2003), pp. 51–58
32. R. Bouckaert, *Naive Bayes Classifiers that Perform Well with Continuous Variables* (2004), Lecture Notes in Computer Science, vol. 3339, pp. 1089–1094
33. P. Brazdil, C. Soares, J. Da Costa, ranking learning algorithms: using IBL and meta-learning on accuracy and time results. *Mach. Learn.* **50**, 251–277 (2003)
34. L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees* (Wadsworth International Group, 1984)
35. L. Breiman, Bagging predictors. *Mach. Learn.* **24**, 123–140 (1996)
36. L.A. Breslow, D.W. Aha, Simplifying decision trees: a survey. *Knowl. Eng. Rev.* **12**, 1–40 (1997)
37. H. Brighton, C. Mellish, Advances in instance selection for instance-based learning algorithms. *Data Min. Knowl. Disc.* **6**, 153–172 (2002)

Breast Cancer Diagnosis Using Image Processing and Machine Learning



Subham Sadhukhan, Nityasree Upadhyay and Prerana Chakraborty

Abstract Breast cancer is one of the second leading causes of cancer death in women. Despite the fact that cancer is preventable and curable in primary stages, a huge number of patients are diagnosed with cancer very late. Conventional methods of detecting and diagnosing cancer mainly depend on skilled physicians, with the help of medical imaging, to detect certain symptoms that usually appear in the later stages of cancer. Therefore, we present here the computerized method for cancer detection in its early stage within a very short time. Here, we have used Machine learning to train a model using the predicted features of the nuclei of cells. A comparative study of two different algorithms KNN and SVM is conducted where the accuracy of each classifier is measured. After this, we analyze a digital image of a fine needle aspirate (FNA) of breast tissue using image processing to find out the features of nuclei of the cells. We then apply the feature values to our trained model to find whether the tumor developed is benign or malignant.

Keywords Machine learning · Fine needle aspirate (FNA) · Image processing · Benign · Malignant

1 Introduction

Breast cancer has evolved as the leading cause of death among women in developed countries [1]. The most effective way to reduce breast cancer deaths is to have an early diagnosis. It requires an accurate and reliable diagnostic procedure that allows physicians to distinguish benign breast tumors from malignant ones without directly going for surgical biopsy. The objective of these predictions is to assign patients to either a “benign” group that is non-cancerous or a “malignant” group that is cancerous.

S. Sadhukhan · N. Upadhyay (✉) · P. Chakraborty
Department of Computer Science & Engineering,
Institute of Engineering & Management, Kolkata, India
e-mail: u.nityasree95@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_12

The two most commonly used screening methods for breast cancer detection are physical examination and mammography. However, they can offer an approximate likelihood that a lump is malignant and can also detect some other lesions, like a simple cyst. As these screening methods are inconclusive, a healthcare provider can perform a procedure known as fine needle aspiration, or fine needle aspiration and cytology (—FNAC) by removing a sample of the fluid in the lump for analysis under a microscope. This helps to establish the diagnosis.

A fine needle aspiration biopsy is a quite simple and quick procedure to diagnose malignancy [2]. The procedure involves the collection of a sample of cells or fluid from a cyst followed by examination of the cells under a microscope. If the lump cannot be felt, then imaging is required to find the exact location. This can be done with ultrasound or stereotactic mammogram. In ultrasound technique, the surgeon watches the needle on the ultrasound monitor and guides it to the area, and in stereotactic mammogram (for the breast), two mammograms are used at different angles and with the help of a computer exact coordinates are created. However, all of these methods are very time-consuming. Our challenge is to reduce this time gap by using image processing and machine learning techniques. In the following subsections, we propose methods of analysis of image of an FNA slide to obtain the feature values of infected cell. Following this is using a model to predict whether the tumor-infected cell is cancerous or not using the feature values obtained in the previous step.

The following subsections contain related works, methodology, results, conclusion, and references.

2 Related Works

In [3], Sau Loong Ang et al. made an attempt for improving the Naïve Bayes by Tree Augmented Naive Bayes (TAN). They had applied General Bayesian Network (GBN) with the hill-climbing learning approach. They had used seven nominal data sets with the absence of missing values to measure the performance of General Bayesian Network against the Naive Bayes and Tree Augmented Naive Bayes, for comparative purposes. The data sets were taken from the UCI Machine Learning Repository [4] and were fed into the Naive Bayes, GBN, and TAN. They had used ten-fold cross-validation in WEKA software which contains 286 instances. 71.68% accuracy was obtained in Naïve Bayes, and 69.58% for TAN, 74.47% for GBN.

In [5], Diana Dumitru applied Naive Bayes classifier to the Wisconsin Prognostic Breast Cancer (WPBC) data set, which contains information of 198 patients and a binary decision class: among which non-recurrent events have 151 instances and recurrent events have 47 instances. The testing accuracy was about 74.24%.

In [6], K. Shivakami made breast cancer prediction using DT-SVM Hybrid Model. This study was performed using the Wisconsin Breast Cancer Data set (WBCD) which contained 699 instances, among which 458 cases were diagnosed to fall under category of “benign” that is non-cancerous; remaining 241 cases fell under malignant class. Each record in the database had nine attributes. In case of DT-SVM,

the accuracy obtained was 91% with an error rate of 2.58%. Other classification algorithms had also been applied like IBL, SMO, and Naïve Bayes. For IBL, the accuracy obtained was 85.23% with an error rate of 12.63%. For SMO, the accuracy was 72.56% with an error rate of 5.96%. For Naïve Bayes, the accuracy obtained was 89.48% with an error rate of 9.89%. So this comparative study revealed that DT-SVM performed better than any other classification algorithms.

In [7], Shweta Kharya et al. used Naïve Bayes classifier to build up a system that predicted accurately about the presence of breast cancer. The boon of the system was that it could be implemented in remote areas like countryside or rural regions, for human diagnostic expertise for treatment of cancer disease. The system is user-friendly and reliable as model was already developed. The model used Wisconsin data sets containing 699 records with nine medical attributes; for training. For Testing, 200 records were taken. About 65.5% cases were diagnosed to be benign and remaining 34.5% malignant cases. 93% accuracy was achieved.

In [8], breast cancer diagnosis is developed by Afzan Adam. He used back-propagation neural network to reduce the diagnose time and as well as accuracy. Two different cleaning processes on the data set were used. In Set A, the records which had missing values were eliminated, and Set B was trained with normal statistical cleaning process. Set A gave 100% of highest accuracy percentage and Set B gave 83.36% of accuracy.

The objective of this paper [9] by Haowen You et al. was to provide a comparative analysis using Bayesian network on the problem by a benchmark data set that contained several numeric feature values, obtained from cells that were preprocessed from fine needle aspiration biopsy image of cell slides. The benchmark data set in this research was obtained from the UCI machine learning repository which was classified as malignant (M) or benign (B). The data set contained 357 benign and 212 malignant observations, respectively. Here, Naïve Bayes classifier gave an accuracy of 89.55%.

In [10], survival rate of breast cancer was analysed using 20 variables present in the seer data. Algorithms used for this purpose were artificial neural networks, decision tree classifier, and logistic regression. However, it was observed that, former two algorithms predicted accurately for 91.2% cases and 93.6% cases, respectively, which was better compared to logistic regression which predicted accurately for 89.2% cases.

The survey above shows that all the researches conducted aims to find an algorithm that will accurately predict if a tumor is benign or malignant, only when the machine is provided with certain feature values of nuclei of tumor affected cells. There is no predefined method specified that can actually extract these feature values. The research conducted here aims to overcome these shortcomings. Here, in this paper we propose a new methodology to predict the feature values by analyzing the image of a FNA slide using image processing techniques. We have also trained a model using the Wisconsin data set and applied machine learning algorithms to find the algorithm with best accuracy. Here, we have then applied the feature values obtained

in the previous step, to the model built. Finally, we predict if the tumor affected cell is cancerous (malignant) or not (benign) using the predictions obtained from the model.

3 Methodology

Many challenges were faced while conducting the research work. We have tried to overcome the limitations of diagnosis by proposing algorithms for finding out the features and have applied various machine learning algorithms to predict the malignancy.

First section of the methodology includes the various techniques that we have proposed for the calculation of features, and the second part includes the techniques of classification that are used for prediction. In the first part using various image processing techniques, we have tried to build our own algorithm by using these concepts.

The proposed methodology for processing the images of FNA slides is discussed in detail in the upcoming sections. The features that are in the breast cancer prediction data set are calculated by using this method.

Proposed Methodology

First, the steps for tracing the contours [11] of the nuclei are discussed for extracting the feature values. The steps for contour tracing are shown in Fig. 1.

FNA Image

FNA image is the image of the slide that has the cancer cells as obtained by fine needle aspiration. Following image is the FNA image of breast cancer cells (Fig. 2).

On this image, we perform the image processing steps to obtain the contours of the nuclei of the cell.

Grayscale

Grayscale [12] involves removal of the hue from the image and retaining only the luminance. As a result, the image has different intensities at different locations based on the luminance. The entire image is divided into various levels of gray color, white,



Fig. 1 Contour tracing

Fig. 2 FNA Image to be processed

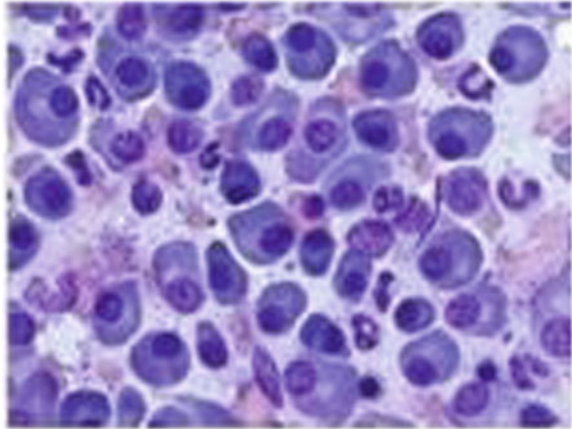
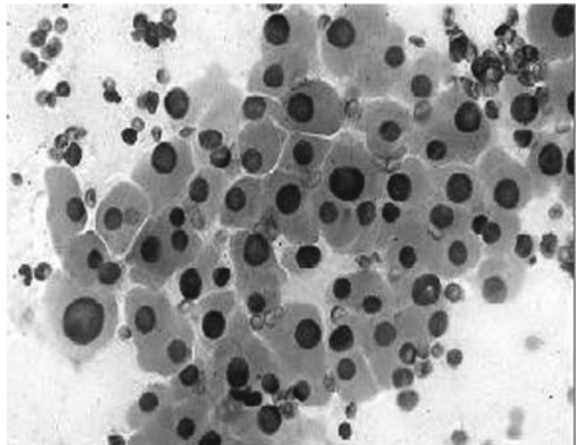


Fig. 3 Grayscaled image



and black. Here, the image has been grayscaled for two reasons. First one is to get a thresholded image based on the intensities, and the second one is to find out the texture of the nuclei that is calculated using the grayscale values (Fig. 3).

The nuclei have least luminance in the image. Thus, they can be separated using their intensities which will be comparatively lower than the cells.

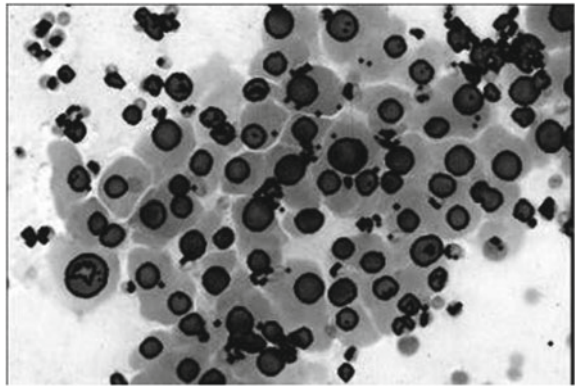
Thresholding

Once gray scaling of the FNA image is over we proceed to thresholding [13]. It involves segmenting the image in binary color that is black and white based on the intensities. It creates a foreground and background. It improves the contrast of the image that helps in the next process that is contour detection of the nuclei. A thresholded value is given, and all the intensities that are above this value are converted to white. The intensity values below the threshold value are converted to black (Fig. 4).

Fig. 4 Image after thresholding



Fig. 5 Contour detection



The threshold values have been set in such a way that the nuclei are segmented the image and converted to black color and the other details have been converted to white. Thus, here the nuclei act as the foreground and the other details as the background. As the nuclei are distinctly segmented, now contour detection can be done easily on it.

Contour detection

Having done with the thresholding of the FNA image, we proceed to contour detection. Contours can be explained simply as a curve joining all the continuous points along the boundary, having same color or intensity. The contours are a useful tool for shape analysis and object detection and recognition. The nuclei in the thresholded image are of same intensity. So contours of these nuclei are easily found out and they are traced on the grayscale image. The contours have been detected to find various dimensions of the nuclei like radius, area, and perimeter (Fig. 5).

The above steps were the preprocessing techniques employed for tracing the contours on the grayscale image. Now comes the processes involved for calculation of the dimensions like radius, area, and perimeter that are necessary for calculating the features present in the Wisconsin data set.

The calculation of various feature values is discussed below.

Radius Calculation

For the calculation of radii of the contours, we approximate circles around them and calculate their radii. It has to be the minimum enclosing circle to get an accurate result.

Then, the mean, standard error, and maximum of the radii values are calculated.

For perimeter and area calculation, there is a predefined function for calculation of the length and the area of the contours. Thus using this mean radius, mean perimeter, mean area, standard error of radius, perimeter, and area can be calculated.

Smoothness Calculation

Calculation of 'smoothness' feature, involves calculating the local variation of radii of each of the contours, which are not perfect circles. So there will be a variation of radii for each contour. This is calculated easily by a predefined function that calculates local standard deviation.

Then, its mean, standard error, and maximum value can be calculated by using known formulas. This gives a measure of how much the nuclei is circular, that is the level of uniformity of the nuclei.

Compactness Calculation

Compactness is calculated using the following formula

$$\text{perimeter}^2/\text{area} - 1.0$$

As we have already calculated the perimeter and area, this can be calculated easily. This tells us how much dense is the nuclei.

Texture Calculation

A few steps following contour detection is required for texture calculation. Texture is basically the standard deviation of grayscale values. Following are the steps to find the grayscale values of the nuclei. After the contours are traced on the grayscale image, we create a masked image of the same shape as the original image and we draw the contours on this masked image (Figs. 6 and 7).

The grayscale value of the background is 0 and that of the contour regions is 255. After that the masked image and the grayscale image are bitwise added to have the original nuclei images in this masked image (Fig. 8).

Then, we get the grayscale values of this image. As the grayscale value for black is zero, whatever value we get is the grayscale value for the nuclei. Then, we find the standard deviation of grayscale values for each of the nuclei. After that the mean, standard error, and maximum value can be calculated using the formulae.

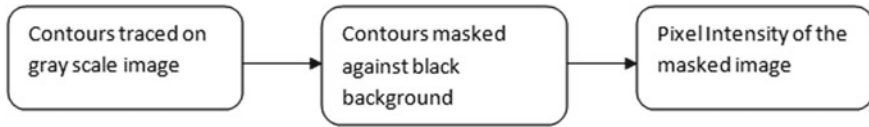


Fig. 6 Texture calculation

Fig. 7 Contours on masked image

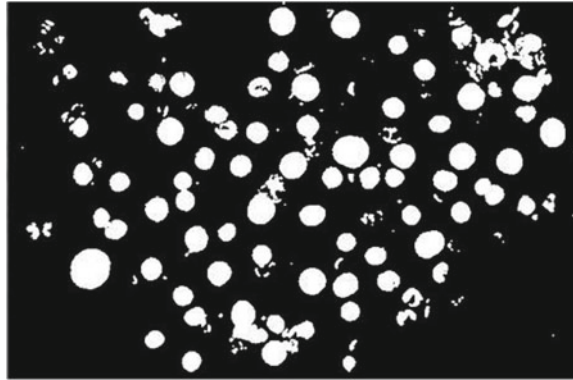
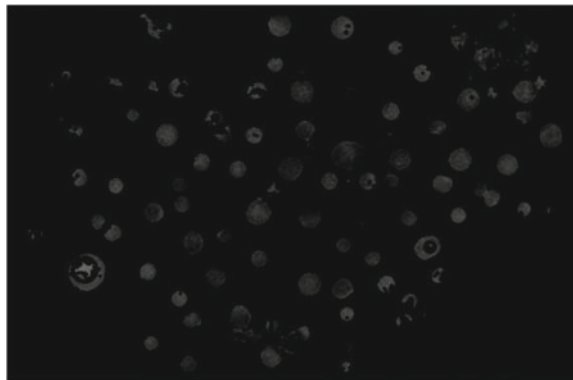


Fig. 8 Bitwise addition of masked and grayscale image



Contour Severity

Here again the approximated minimum enclosed circles come into play. The contours have several concave portions. The severity of these concave portions is calculated. The area of this enclosed circle is calculated followed by the area of the contour.

Severity is given by the formula

$$\text{Area of the contour} / \text{Area of the circle enclosing it.}$$

Then its mean, standard error, and maximum value are calculated. After applying these steps, we are able to extract the required features from the FNA image.

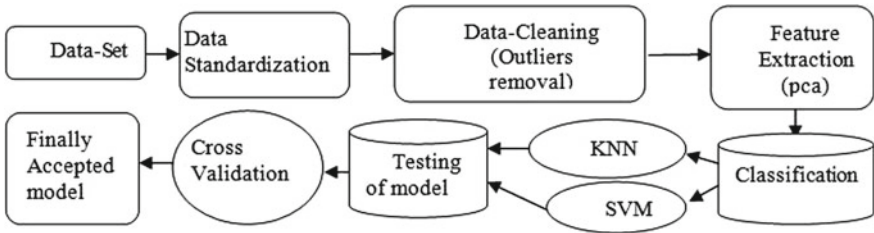


Fig. 9 Flowchart for machine learning process

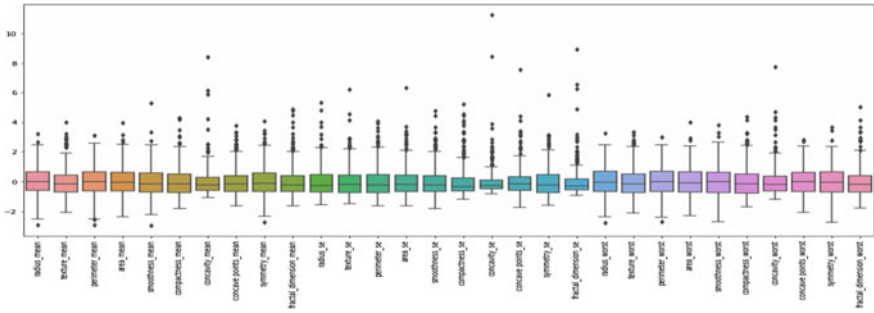


Fig. 10 The data set with outliers

Thus after obtaining all the features from the image, we put it in the machine learning model. Following is the methodology for building up that model for prediction.

Building up of model

Today’s real-world databases are highly vulnerable to noisy, missing, and inconsistent data due to their typically massive size and their likely origin from multiple, miscellaneous sources. Hence, data preprocessing is required before applying any machine learning algorithm.

The steps involved in machine learning are as follows (Fig. 9):

The breast cancer Wisconsin (diagnostic) data set used here contains 31 columns and 569 entries (rows). It requires data cleaning. The categorical values of the column “diagnosis” have been changed to the binary values, that is “malignant” (cancerous) tumor has been changed to “1” and “benign” (non-cancerous) tumor has been changed to “0,” respectively. The data set is being normalized to bring its mean to 0 and standard deviation to 1. The outliers have been removed to get an evenly distributed data set (Figs. 10 and 11).

Now the data set is ready to apply KNN [14] and SVM [15] algorithm, respectively.

Description of the classifiers used

Study of previous research works shows that better accuracy is obtained while predicting the type of cancer by applying KNN and SVM models on the data set,

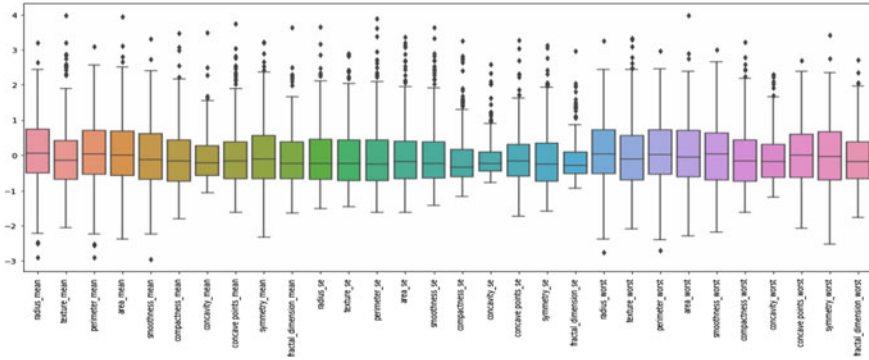
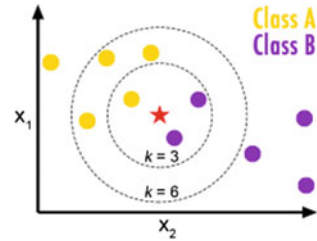


Fig. 11 Data set without outliers

Fig. 12 K-nearest neighbors



respectively. So we have tried to apply both these algorithms to the data set and finally find the one with better accuracy.

The following subsection contains a brief description of both these algorithms.

KNN: KNN is a supervised learning technique that deals with classification problems. The algorithm in its initial steps requires choosing a suitable value of “K.” The next step of the algorithm states to choose “K” number of nearest neighbors to the new data-point that requires classification. Normally, the nearest neighbors are chosen according to ascending order of the square of the distance of its neighbors to the required data-point. The required data-point is classified to either of the classes depending on which class has more number of neighbors to the data-point.

Figure 12 depicts KNN classification for red data-point using $K = 3$ and $K = 6$, respectively.

SVM:- Support vector machine is a classification algorithm that actually draws a hyperplane separating the two classes of data widely apart. Hyperplane actually runs through the line of symmetry of the plane formed by two support vectors (Fig. 13).

Analysis of model

The model has been trained using 70% of the entries, and the rest of the data set has been used for testing the model. The level of effectiveness of the classification model is calculated by using cross-validation. We present here a comparative study to find out the least number of features that can be used to obtain the best accuracy. In

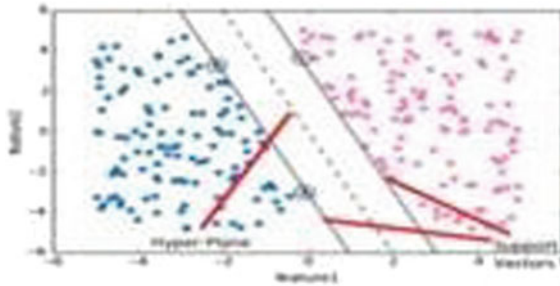


Fig. 13 Support vector machines. Source: https://www.ft.com/_origami/service/image/v2/images/raw/https%3A%2F%2Fftalphaville-cdn.ft.com%2Fwp-content%2Fuploads%2F2017%2F10%2F26114150%2Flinear.png?source=Alphaville

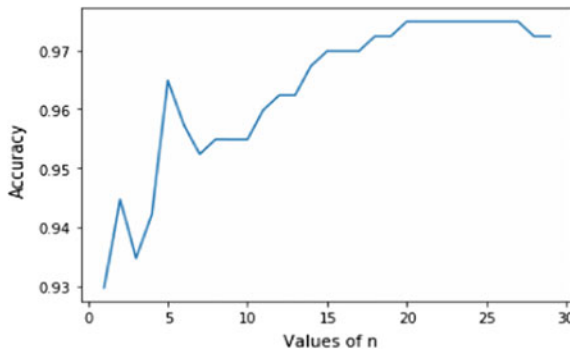


Fig. 14 Graph of number of features versus accuracy score for SVM model

the graph (Fig. 15), the results are observed by taking first dominant feature at first, then taking first two dominant features, after those first three dominant features and continue this procedure until the combination of 30 features. Maximum accuracy has been obtained at $n = 19$, and the accuracy is 97.273% using SVM model (Fig. 14).

KNN algorithm has been initially performed on the data set given to find the value of “K” for which there is less error. The graph (Fig. 16) shows a plot of “K” against “error rate,” depicting the least error for the value of “K” = 9.

Figure represents the plot of accuracy obtained by applying KNN model to the data set obtained each time by varying the total number of features in principal component analysis. The plot represents maximum accuracy of 97.4893% obtained by taking 13 features for data analysis.

Here, confusion matrix is used which is a table that is often applied to describe the performance of a “classifier” or classification model on a collection of test data for which the true values are known. The level of effectiveness of the classification model is calculated with the number of incorrect and correct classification in each possible value of the variable being classified in the confusion matrix (Tables 1 and 2).

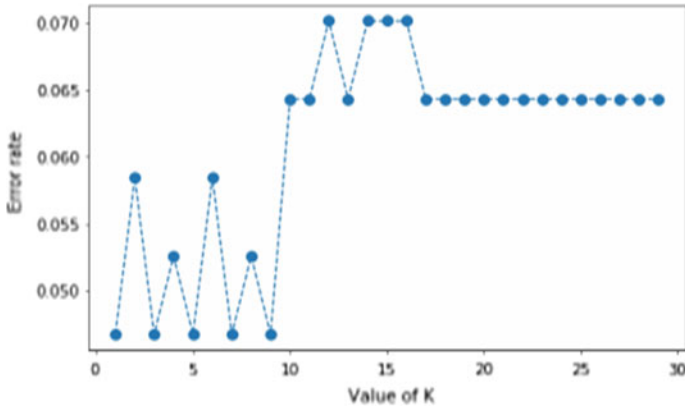


Fig. 15 Error rate versus K

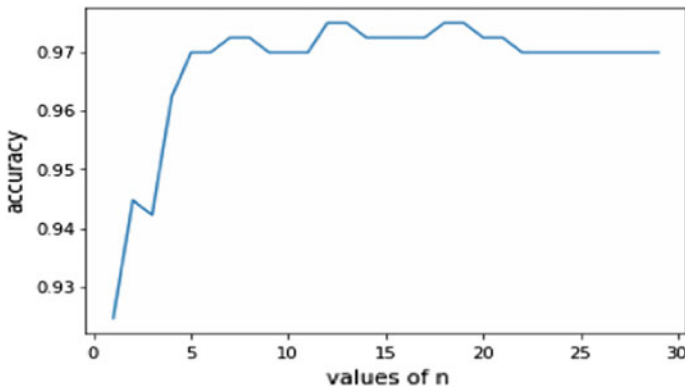


Fig. 16 Graph of number of features versus accuracy for KNN model

Table 1 Confusion matrix for SVM classifier

	Predicted benign	Predicted malignant
Actual benign	104	1
Actual malignant	3	63

Table 2 Confusion matrix for KNN Classifier

	Predicted benign	Predicted malignant
Actual benign	103	2
Actual malignant	6	60

Table 3 Comparison for accuracies of SVM and KNN using different PCA values

	PCA for value = 19	PCA for value = 13
SVM	97.273	96.234
KNN	97.489	97.493

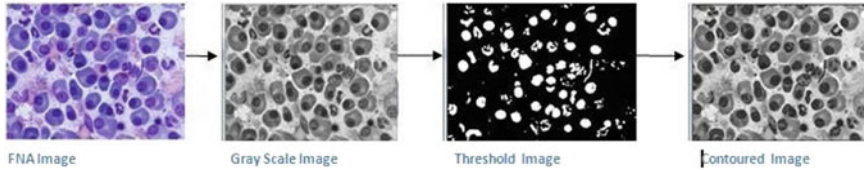
**Fig. 17** Transition from FNA image to contoured image

Table 3 same accuracy is obtained for both cases of PCA application in case of KNN model. This makes the KNN algorithm more preferable. Moreover, highest accuracy is obtained in KNN using less number of features. Thus, KNN model is finally accepted.

From the above comparison table, it is observed that KNN model with PCA value = 13 performs to be the best classifier.

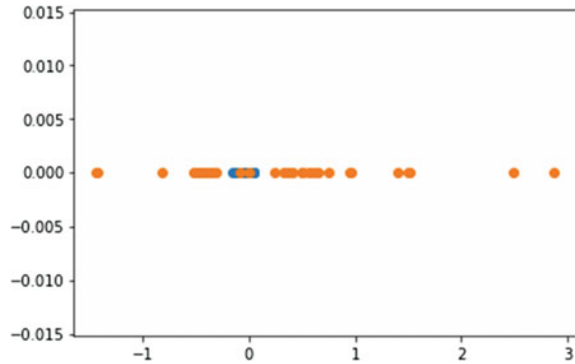
4 Results

The following is obtained for tracing the contours (Fig. 17).

It is observed that after thresholding the nuclei are segmented. After that finding the contours and tracing them on the grayscale image, the nuclei are distinctly separated. After getting the nuclei, we apply the above steps to obtain the features values.

We find a scatter plot where we take y-axis to be set to zero x-axis has been taken to be the feature values obtained. Following scatter plot shows a plot using two values for each feature: one by using the mean of all values of all columns and the second one by using the values obtained by image processing techniques. It is observed that the graph is almost overlapping except for at two points (shown by orange and blue color, respectively). This confirms that the methodology used here is almost accurate (Fig. 18).

Hence, we finally apply the obtained feature values to the proposed model. This confirms that the affected tumor is “BENIGN,” that is “*non-cancerous*.”

Fig. 18 Scatter plot

5 Conclusion

Comparing to all other cancers, breast cancer is one of the major causes of death in women. So, the early detection of breast cancer is needed in reducing life losses. This early breast cancer cell detection can be predicted with the help of modern machine learning techniques. The efficiency (97.489%) of this entire model is quite high. Thus in future software can be developed where the FNA slide images can be uploaded and it will automatically process the image to get the feature values and will automatically apply it to the model to predict the result. Thus, it will lead to a very fast diagnosis of breast cancer.

References

1. Breast Cancer Facts. <http://www.nationalbreastcancer.org/breast-cancer-facts>
2. Fine Needle Aspiration Biopsy. <https://www.myvmc.com/investigations/fine-needle-aspiration-biopsy-fna/>
3. S.L. Ang, H.C. Ong, H.C. Low, Classification using the general Bayesian network. *Pertanika J. Sci. Technol.* **24**(1), 205–211 (2016)
4. Wisconsin Breast Cancer Dataset- <https://archive.ics.uci.edu/ml/datasets/Breast+Cancer+Wisconsin+%28Diagnostic%29?fbclid=IwAR1seM6TeFolsHOxjeyHNwEIDZDIVyN9mkW2qSWcl6xi35P2bsIg2A-m8jM>
5. D. Dumitru, Prediction of recurrent events in breast cancer using the Naive Bayesian classification. *Ann. Univ. Craiova-Math. Comput. Sci. Ser.* **36**(2), 92–96 (2009)
6. K. Sivakami, Mining big data: breast cancer prediction using DT-SVM hybrid model. *Int. J. Sci. Eng. Appl. Sci. (IJSEAS)* **1**(5), 418–429 (2015)
7. S. Kharya, S. Agrawal, S. Soni, Naïve Bayes classifiers: probabilistic detection model for breast cancer. *Int. J. Comput. Appl.* **92**(10), 0975–8887 (2014)
8. A. Adam, K. Omar. *Computerized Breast Cancer Diagnosis with Genetic Algorithms and Neural Network*. fitm.mu.edu.my/caiic/papers/afzaniCAIET.pdf
9. <https://dialnet.unirioja.es/descarga/articulo/4036558.pdf>
10. D. Dursun, W. Glenn, K. Amit, Predicting breast cancer survivability: a comparison of three datamining methods. *Artif. Intell. Med.* **34**, 113–127 (2005)

11. A.G. Ghuneim. *Contour Tracing*. http://www.imageprocessingplace.com/downloads_V3/root_downloads/tutorials/contour_tracing_Abeer_George_Ghuneim/intro.html
12. Mathworks Image Grayscale- <https://www.mathworks.com/help/images/grayscale-images.html?fbclid=IwAR2Dz1gWFwEtsZa-JtXwS1nKq0KCxxtPQQd6-k460zgrT8aBV9M3b9JTWU>
13. Mathworks Image Thresholding. <https://in.mathworks.com/discovery/image-thresholding.html>
14. O. Sutton. Introduction to k nearest neighbour classification and condensed nearest neighbour data reduction (2012). http://www.math.le.ac.uk/people/ag153/homepage/KNN/OliverKNN_Talk.pdf
15. H. Yu, S. Kim. SVM tutorial: Classification, regression, and ranking. <https://pdfs.semanticscholar.org/cbc3/d8b04d37b2d4155f081cd423380220a91f13.pdf>

Using Convolutions and Image Processing Techniques to Segment Lungs from CT Data



Souvik Ghosh, Sayan Sil, Rohan Mark Gomes and Monalisa Dey

Abstract In this paper, we put forward a novel approach of lung segmentation using convolutional neural networks (CNNs). CNNs have outperformed traditional techniques in many visual recognition tasks over the years. In the traditional methods, the features of the images to be recognized and then segmented are often hand-coded. This might seem quite simple at first and easy to implement but then as the complexity of the problems increases, these methods start to crumble. Moreover, these methods do not handle edge cases accurately as the programs do not know what to do when the images deviate from the expected nature. This is where deep learning methods (CNNs in our case) come to the rescue. Instead of hard-coding, we have built a CNN which is fed with a dataset consisting of CT images and the segmented images of the lungs. CNN automatically detects the features, and the accuracy is increased by tuning the parameters as well as using other techniques. The model once trained is then tested using the test dataset. As evident from the results, this approach gives a much higher accuracy compared to traditional methods, ultimately paving the way to a better analysis of the lungs.

Keywords Convolutional neural networks · Lung segmentation · Deep learning

S. Ghosh · S. Sil (✉) · R. M. Gomes · M. Dey
Institute of Engineering and Management, Salt Lake Electronics Complex,
Gurukul, Y-12, Sector V, Kolkata 700091, India
e-mail: sayan.sil.mail@gmail.com

S. Ghosh
e-mail: culeshovi@hotmail.com

R. M. Gomes
e-mail: rohanmarkgomes@gmail.com

M. Dey
e-mail: monalisa.dey.21@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_13

1 Introduction

Computed tomography (abbreviated as CT) is a widely used technique in the field of medical imaging. Since most of the CT images are nowadays digital, many advanced image processing techniques can be used which makes it easier to substantially increase the quality of the images and thus extract much more information out of them. In a CT scan, X-rays are used to obtain cross-sectional images of the object being scanned from various angles after which digital geometry processing is used to generate a three-dimensional image of the volume of the object from a series of two-dimensional radiographic images.

The first step in the analysis of the CT image is the segmentation of the organ of interest, in our case the lung. This segmentation can be done either manually or automatically with the help of machines. Segmentation is a vital step in the treatment of the disease, and errors in this step can lead to loss of lives. With the rapid advancement of technology, in the twenty-first century we are in a position to automate this task with the help of machines and also perfecting its working over time. Over the years, many techniques have been employed like thresholding-based segmentation, region-based segmentation, neighboring anatomy-guided segmentation. In this paper, we have proposed a convolutional neural network (CNN)-based segmentation which gives quite a high accuracy compared to traditional methods [9].

2 Related Works

There have been various notable works done in the field of medical image segmentation. In V-Net, the authors propose an approach to 3D image segmentation based on a CNN which is trained end-to-end on MRI volumes depicting prostate and learns to predict segmentation for the whole volume at once [12]. A survey on deep learning in medical image analysis reviews the major deep learning concepts pertinent to medical image analysis [10]. In U-Net, the authors present a network and training strategy that relies on the strong use of data augmentation to use the available annotated samples more efficiently. The architecture consists of a contracting path to capture context and a symmetric expanding path that enables precise localization [14]. The authors of hierarchical 3D fully convolutional networks for multi-organ segmentation show that a multi-class 3D FCN trained on manually labeled CT scans of seven abdominal structures can achieve competitive segmentation results [15]. The authors of VoxResNet propose a novel voxelwise residual network with a set of effective training schemes [1]. Another paper studies the effectiveness of accomplishing high-level tasks with a minimum of manual annotation and good feature representations for medical images [16]. There is also a work on automatic lung segmentation for accurate quantitation of volumetric X-ray CT images [4]. The other works that use CNN to segment medical images are Hough-CNN [13], MRI abdomen images using cascaded fully convolutional neural networks (CFCNs) [2], and CNN-based segmentation of medical imaging data a uses CNN-based method [6].

3 Dataset and Preprocessing

The dataset used for this experiment is finding and measuring lungs in CT Data which is a reviewed dataset hosted at Kaggle [11]. In order to find diseases in the lungs, it is a necessary step to find and segment the lungs. This experiment goal is to help in processing and further create a path to find lesions in CT images of the lungs. The dataset consists of a total of 267 CT images of lungs as well as their masks. The resolution of them is 128×128 with channel size of 1. The input images as well as the output masks are normalized to train the network better. There was a train-validation split performed by a value of 0.1.

4 Proposed Model

The model architecture is described in Fig. 1. The model is divided into two parts consisting of the encoder and the decoder. Both of the encoder and the decoder use convolutions of 4×4 followed by a rectified linear unit (ReLU) [8]. The encoder uses convolutions with strides of 2 for downsampling, whereas decoder uses transposed convolutions with strides of 2 for upsampling. Transposed convolutions are often called as deconvolutions. The first layer of encoder has channel size of 64. The next layer of encoder has channel size of 128. Now, the network will start decoding, and the first decoding layer consists of channel size 64. After this, the final decoding operation gives an output of channel size of 1. Each convolution layer except the output layer is followed by batch normalization [5]. Batch normalization is used to speed up training time as well as regularize the network from over-fitting.

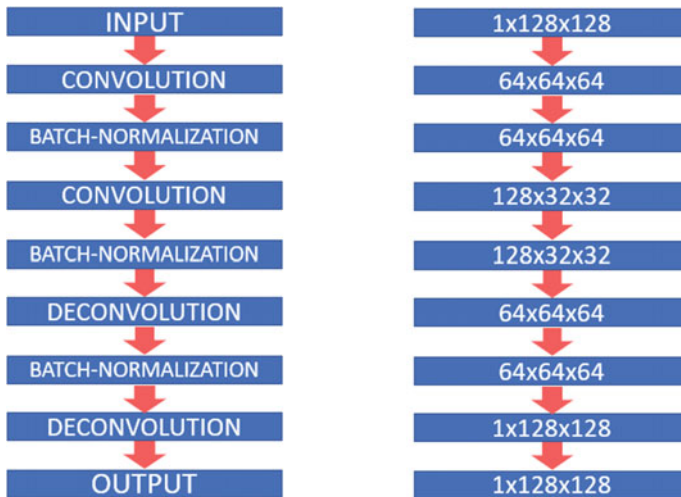


Fig. 1 Architecture of our model along with its dimensions

5 Results and Discussion

5.1 Model Experiment

Our model yielded a promising overall accuracy of 97.78% on the given dataset of CT scans, which can be considered exceptionally good going by the computationally less expensive model architecture and the limited set of training data for the scenario. The detailed analysis of the experimental results is mentioned below.



Fig. 2 Results of our techniques on test data. *CT scan* is the input, and *actual* is the true prediction. *Predicted* is the output of our network, and *masked* is the result of our image processing techniques

The network was trained for 150 epochs, and the model weights were saved when the validation loss decreased Fig.3. Learning rate decay was used when the validation loss did not improve. Binary cross-entropy was used as the loss function for our model. The network predicted output on test dataset is shown in Fig. 2.

We find that the model converges around epoch 100 with the final training loss of 0.0897 and a validation loss of 0.0540. The accuracy also shows no considerable improvement after epoch 100, with the final training accuracy being 0.9761 and validation accuracy being 0.9781 after 150 epochs.

On repeating the experiment for 100 epochs, expected results were seen as shown in Fig. 4.

The final training loss resulted to be 0.0823 and evaluation loss 0.0631 for 100 epochs. The overall training accuracy was found to be 0.9740, and evaluation accuracy was found 0.9778.

As one might notice, there were a few spikes observed on the loss and accuracy graphs given above. Such a behavior can be fundamentally explained due to the momentum in Adam [7]. While doing this experiment, there was a saddle point that was frequently encountered at a loss of 3.7865 [3]. The network was stuck at this point and did not converge. The network at this state only predicted a monotonous black output. To overcome this scenario, more random initialization of weights were

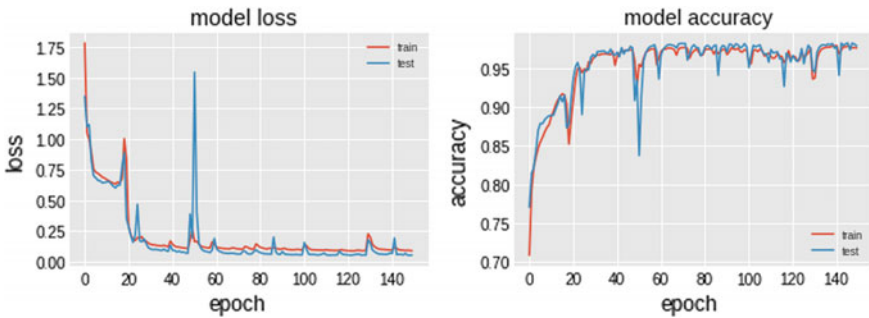


Fig. 3 Plot of model loss and accuracy on train and test data for 150 epochs

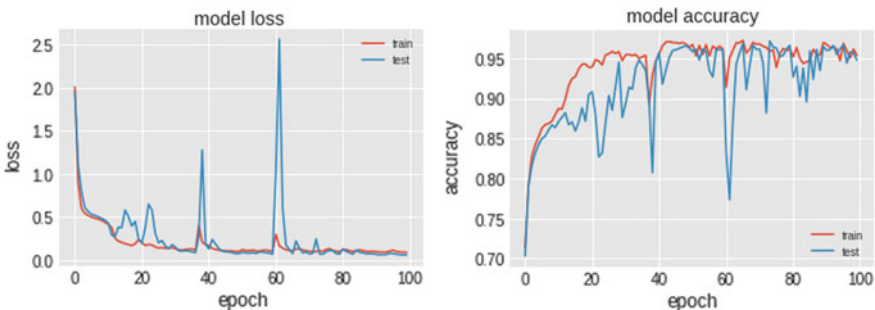


Fig. 4 Plot of model loss and accuracy on train and test data for 100 epochs

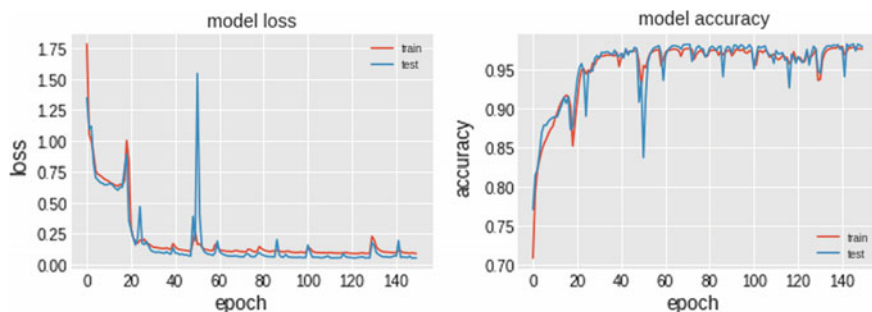


Fig. 5 Plot of the complex model loss and accuracy on train and test data for 150 epochs

performed and the aforementioned Adam was used as the optimizing algorithm. A more complex network was also used for evaluation, with two more convolutional and deconvolutional layers. However, the accuracy of model did not converge well.

As one might perceive from the above graph, the model began over-fitting around epoch 30 Fig. 5. The optimizer kept decreasing the loss; however, this made the model to perform poorly on the validation set, with an unsatisfactory accuracy of 0.4704 at epoch 150. Thus at the expense of extra computational time, we are not able to receive any increase in the overall accuracy.

The overall findings of the experiment have been tabulated below:

Network architecture	Epochs	Train loss	Valid loss	Train Acc	Valid Acc
Shallow CNN	100	0.0823	0.0631	0.9740	0.9778
Shallow CNN	150	0.0897	0.0540	0.9761	0.9781
Deep CNN	150	0.0284	0.0666	0.4158	0.4704

Thus, we are able to experimentally establish that our shallow CNN model is the best possible architecture for segmentation of lung images in CT scans among other similar possibilities, considering the trade-off between accuracy and computational time.

5.2 Post-processing of Predicted Output

Weighted Gaussian Blurring

A Gaussian kernel of size 5×5 is defined to remove Gaussian noise from the output image. Blurring in OpenCV uses a low-pass filter to remove high-frequency noise and sharp edges from image. The standard deviation in both the axes is set as zero for these images; thus, every element in the 5×5 kernel is the same as the collective mean. The Gaussian image is then added to the original image.

Thresholding

To form a clear and distinct image from the grayscale, we threshold each pixel value against a determined value to get a binary image of only pure white and pure black pixels. This is done to resemble more like the desired output format which can be further used in medical purposes, neatly resembling the shape and size of lungs.

Erosion and Island Removal

Islands of white pixels on the black backdrop are detected and correspondingly marked accordingly using breadth-first search approach. On detection of these islands, we find their area and correspondingly remove small islands from the image by resetting their pixel values to 0. After thresholding against area size, to remove any possibility of bigger unwanted islands which were not removed in the thresholding step, we take the top two biggest islands corresponding to two lungs in the resultant image and reset all other pixels to 0.

6 Conclusion

This study proposed a new spatial method of segmentation of medical images implementing the new concept of convolutional neural networks in medical segmentation, which was found to be reasonably simple and computationally efficient, to segment images of lungs from CT scans. On the one hand, like conventional manual segmentation, the proposed method has the ability to withstand against image noise. On the other hand, like color thresholding raw CT scans, it has the properties of working with shape/size invariant and grayscale invariant images also. The quality of the proposed approach was validated with multiple experiments to prove that this approach is robust to grayscale variation, rotation variation, and a variety of kinds of noise. This proposed CNN method gives better classification accuracy than traditional dense neural networks; however, it uses far less number of computational layers to achieve similar accuracy. Due to the computationally inexpensive nature of the model, the smaller computation time caused the model to not converge on specific circumstances which were dealt with using standard practices. It was experimentally proved that the optimal approach to reconstruction of the desired image for the proposed CNN model is by un-sharpening and de-noising, which are fundamental and more dominant over other approaches.

References

1. H. Chen, Q. Dou, L. Yu, J. Qin, P.A. Heng, Voxresnet: deep voxelwise residual networks for brain segmentation from 3D MR images. *NeuroImage* (2017)
2. P.F. Christ, F. Ettliger, F. Grün, M.E.A. Elshaera, J. Lipkova, S. Schlecht, F. Ahmaddy, S. Tatavarty, M. Bickel, P. Bilic, et al., Automatic liver and tumor segmentation of CT and

- MRI volumes using cascaded fully convolutional neural networks (2017). arXiv preprint [arXiv:170205970](https://arxiv.org/abs/170205970)
3. Y.N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, Y. Bengio, Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, in *Advances in Neural Information Processing Systems* (2014), pp. 2933–2941
 4. S. Hu, E.A. Hoffman, J.M. Reinhardt, Automatic lung segmentation for accurate quantitation of volumetric X-ray CT images. *IEEE Trans. Med. Imaging* **20**(6), 490–498 (2001)
 5. S. Ioffe, C. Szegedy, Batch normalization: accelerating deep network training by reducing internal covariate shift (2015). arXiv preprint [arXiv:150203167](https://arxiv.org/abs/150203167)
 6. B. Kayalibay, G. Jensen, P. van der Smagt, CNN-based segmentation of medical imaging data (2017). arXiv preprint [arXiv:170103056](https://arxiv.org/abs/170103056)
 7. D.P. Kingma, J. Ba, Adam: a method for stochastic optimization (2014). arXiv preprint [arXiv:14126980](https://arxiv.org/abs/1412.6980)
 8. A. Krizhevsky, I. Sutskever, G.E. Hinton, Imagenet classification with deep convolutional neural networks, in *Advances in Neural Information Processing Systems* (2012), pp. 1097–1105
 9. Y. LeCun, Y. Bengio et al., Convolutional networks for images, speech, and time series, in *The Handbook of Brain Theory and Neural Networks*, vol 3361(10) (1995)
 10. G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A. van der Laak, B. Van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis. *Med. Image Anal.* **42**, 60–88 (2017)
 11. K. Mader, Finding and measuring lungs in CT data (2017). <https://www.kaggle.com/kmader/finding-lungs-in-ct-data>
 12. F. Milletari, N. Navab, S.A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in *2016 Fourth International Conference on 3D Vision (3DV)* (IEEE, New York, 2016), pp. 565–571
 13. F. Milletari, S.A. Ahmadi, C. Kroll, A. Plate, V. Rozanski, J. Maiostre, J. Levin, O. Dietrich, B. Ertl-Wagner, K. Bötzel et al., Hough-CNN: deep learning for segmentation of deep brain regions in MRI and ultrasound. *Comput. Vis. Image Underst.* **164**, 92–102 (2017)
 14. O. Ronneberger, P. Fischer, T. Brox, U-net: convolutional networks for biomedical image segmentation, in *International Conference on Medical Image Computing and Computer-Assisted Intervention* (Springer, Berlin, 2015), pp. 234–241
 15. H.R. Roth, H. Oda, Y. Hayashi, M. Oda, N. Shimizu, M. Fujiwara, K. Misawa, K. Mori, Hierarchical 3D fully convolutional networks for multi-organ segmentation (2017). arXiv preprint [arXiv:170406382](https://arxiv.org/abs/170406382)
 16. Y. Xu, T. Mo, Q. Feng, P. Zhong, M. Lai, I. Eric, C. Chang, Deep learning of feature representation with multiple instance learning for medical image analysis, in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, New York, 2014), pp. 1626–1630

Analyzing Code-Switching Rules for English–Hindi Code-Mixed Text



Sainik Kumar Mahata, Sushnat Makhija,
Ayushi Agnihotri and Dipankar Das

Abstract In this work, we have proposed an efficient and less resource intensive strategy for parsing and analyzing switching points in code-mixed data. Specifically, we have explored the rules of code-switching in Hindi–English code-mixed data. The work involves code-mixed text extraction, translation of the extracted texts to its pure form, forming word pairs, annotation of these using of Parts-of-Speech tags and recognition of the rules that govern switching in code-mixed text. We have created three models, viz. baseline model, lexicon-based model, and machine learning-based model, and found out the individual accuracies of these models.

Keywords Code-switching · Code-mixing · Parts-of-Speech · Lexicon

1 Introduction

India has vast and linguistically diverse diaspora due to its long history of foreign acquaintances. English, one of those borrowed languages, became an integral part of the education system and hence gave rise to a population who are very comfortable using bilingualism when communicating with one another [13]. This kind of language diversity and dialects initiates frequent code-mixing in India. Further, due to the emergence of social media, the practice has become even more widespread. The phenomenon is so common that this is often considered as a different (emerging)

S. K. Mahata (✉) · D. Das
Jadavpur University, Kolkata, India
e-mail: sainik.mahata@gmail.com

D. Das
e-mail: dipankar.dipnil2005@gmail.com

S. Makhija · A. Agnihotri
Rajasthan Technical University, Kota, India
e-mail: sushant191296@gmail.com

A. Agnihotri
e-mail: ayushiagnihotri23@gmail.com

variety of the language, e.g., Benglish (English–Bengali), and Hinglish (English–Hindi) [10].

Code-switching is the practice of alternating between two or more languages or varieties of language in a conversation [12, 14]. It can be differentiated from code-mixing, as code-switching is juxtaposition within the same speech exchange of passages of speech belonging to two different grammatical systems or subsystems. On the other hand, code-mixing refers to the embedding of linguistic units such as phrases, words, and morphemes of one language into an utterance of another language [15].

Most Hindi (Hi) speakers tend to switch to English (En) because the code-switching English word is easier to use compared to its Hindi counterpart [6]. We also see that code-switching is also an important phenomenon while swearing [1]. Swearing is a rude or offensive language that someone uses, especially when they are angry [2].

This paper presents our latest investigations on code-mixed corpus representation. We have classified training corpus into four different word pair categories, namely Hindi–Hindi, Hindi–English, English–English, and English–Hindi, and also classified test data into three respective categories. We have focused basically on three models.

- **Baseline model**
- **Lexicon-based model**
- **Naïve Bayes model**

We have tried and researched whether syntactic rules, such as Parts-of-Speech (POS) tags, can determine code-switching or not. The focus was on the feasibility of a system predicting the language of the token pairs, when fed with only lexical information. As a by-product, we have also extracted lexicon-based rules pertaining to the above information.

The above-mentioned models work well on the test dataset and reflect the efficiency of work done.

The paper has been organized as follows: Section 2 describes a brief survey of previous work done in this domain. This will be followed by the methodology of preparation of the datasets, and creation and validation of the various models, in Sect. 3. This will be succeeded by the Conclusion and Future Work in Sect. 4.

2 Literature Survey

Code-mixing and code-switching are well-studied phenomena observed in bilingual societies [3, 4, 9].

Code-mixing and code-switching, though mostly observed in spoken language, are now becoming popular in text, due to the emergence of social media communication channels like Twitter, Facebook, and WhatsApp [7, 8].

Vyas et al. [15] reported that code-mixing is observed in social media, especially from multilingual users. The authors used POS tagging to determine rules. The authors implemented a pipelined approach in which the tasks—language identification, text normalization, and POS tagging—were performed sequentially.

Agarwal et al. [1] observed the code-switching and swearing patterns on social networks. In this survey, they have described emotions that can be described by swearing words. There have been several studies on the swearing behavior of multilingual users, most of which report the presence of a significant preference toward a particular language for swearing in multilingual speech communities.

Ahmed et al. [2] described back transliteration. There are a number of transliteration schemes that are used for generating Indian language text in Roman transliteration. Roman transliteration is widely used for inputting Indian languages in a number of domains.

Chopde [5] attempted to standardize the mapping between the Indian language script and the Roman alphabet, with ITRANS.

Patro et al. [14] presented a set of computational methods to identify the likeliness of a word being borrowed, based on the signals from social media. Their work further showed that it is not always true that only highly frequent words are borrowed; nonce words could also be borrowed along with the frequent words.

3 Methodology

3.1 Data Preprocessing

3.1.1 Data Extraction

We collected a code-mixed corpus from Sentiment Analysis of Indian Languages (SAIL2017) shared task.¹ The corpus contained code-mixed sentences, where the tokens were annotated with their respective language IDs, viz. EN (English), HN (Hindi), and UN (Unknown). We extracted bi-gram patterns of the tokens from the data and their respective language IDs. An example of the extracted data is given in Table 1.

3.1.2 Translation to Devanagari Script

The tokens that were classified as Hindi (HN) needed to be translated to its pure Devanagari script, contrary to which it would have become difficult to annotate it, with its respective POS tags. We took the help of ITRANS² for this specific task.

¹http://dasdipankar.com/ICON_NLP_Tool_Contest_2017/HI/discretionary-EN.json.

²<https://www.ashtangayoga.info/sanskrit/transliteration/transliteration-tool/>.

Table 1 Extracted bi-gram patterns and language ID from SAIL corpus

Word 1	Word 2	Language ID
Bhi	Agar	HN-HN
Agar	Next	HN-EN
Next	Time	EN-EN
Time	Kabhi	EN-HN

Fig. 1 Transliteration of Hindi tokens into Devanagari scripts

karna	↔	कर्न
parega	↔	परेग
lekin	↔	लेकिन्
sahi	↔	सहि

The transliterated Hindi tokens were fed to this translator, and in turn, it returned the Devanagari script of the Hindi token. This is shown in Fig. 1.

3.1.3 POS Tagging of the Tokens

POS tagging of the tokens was essential so as to investigate the occurrence of Hindi-to-English/Vice-Versa, code switching [11]. Later, we subjected the POS-tagged tokens to 2-gram/bi-gram model for creating the baseline system. For POS tagging of the tokens, we performed the following tasks:

1. The English words from Sect. 3.1.1 were annotated with their POS tags using NLTK³ Toolkit.
2. The Hindi words from Sect. 3.1.1 and subsequently their Devanagari scripts from Sect. 3.1.2 were annotated with their respective POS tags using a Hindi POS tagger (Table 2).⁴
3. A list with the POS tag bi-grams and its respective language IDs were created for use in Sect. 3.3.

3.2 Baseline Model

After our training set became ready in Sect. 3.1.1, we used the test data, obtained from SAIL shared task, to classify the pairs with the help of our training set. The steps followed are as discussed below:

³<http://www.nltk.org>.

⁴<http://sivareddy.in/downloads>.

Table 2 Example of the word pairs after POS tagging

Word 1	Word 2	POS 1	POS 2	Language ID
papa	जि	NN	RP	EH
wait	कर्	NN	VM	EH
papa	ग्	NN	SYM	EH
fan	हु	NN	VM	EH
in	the	IN	DT	EE
of	the	IN	DT	EE
for	the	IN	DT	EE
on	the	IN	DT	EE
कि	party	CC	NN	HE
सल्मन्	sir	NN	NN	HE
इ	have	SYM	VB	HE
भे	plz	RP	NN	HE
के	लिये	PSP	PSP	HH
क्	लिये	SYM	PSP	HH
रह	है	VM	VM	HH
हो	गय	VM	VAUX	HH

Table 3 Baseline accuracy of the language IDs

Total word pairs in test data	43,903			
Language ID	EE	EH	HH	HE
Word pairs found	16,694	3737	20,108	3364
Accuracy (%)	38	8	45	7
Average accuracy (%)	64.25			

1. The sentences in the test set were subjected to bi-gram modeling, and it extracted the word pairs in the format described in Table 1.
2. The word pairs of the test data were searched in the training data, prepared in Sect. 3.1.1, and were annotated with the searched language ID.

For a given language ID, the total number of pairs matching it divided by the total number of word pairs in the test data resulted in the accuracy of that particular language ID.

$$b = \frac{\text{Occurrence Of Words Of Specific Class}}{\text{Total Number Of Words In All Classes}} * 100$$

where b is baseline accuracy percentage.

The baseline accuracy of the language IDs is given in Table 3. The cumulative accuracy of the baseline model was reported as 64.25%.

3.3 *Lexicon-Based Model*

For creating the lexicon-based model, we took the help of the training data that was created in Sect. 3.1.3. The POS tag pairs were separated for each language ID as given in Table 4. Our main aim was to extract unique POS tag pairs pertaining to each language ID. This enabled us to formulate unique POS tag pair rules pertaining to different language IDs.

Since the POS tagset for English and Hindi was similar, there were repetitions in the POS tag rules when comparing English and Hindi languages. For example,

love | u | NN | NN
and
सल्मन् | sir | NN | NN

have the same POS tag rule. So, to find out the unique rule, the following steps were followed.

1. For a rule, its number of occurrences in each language ID class was calculated.
2. Also, for the same rule, its weighted average was calculated.

For example, suppose for Rule NN-NN, its number of occurrence in

Language ID EE is 4,
Language ID EH is 3,
Language ID HE is 2,
Language ID HH is 1,

The number of rules in language ID EE is 1200,
The number of rules in language ID EH is 600,
The number of rules in language ID HE is 500,
The number of rules in language ID HH is 1600,

Table 4 Lexicon-based model data template

POS 1	POS 2	Language ID
NN	RP	EH
NN	VM	EH
IN	DT	EE
IN	DT	EE
CC	NN	HE
NN	NN	HE
PSP	PSP	HH
SYM	PSP	HH

Table 5 Lexicon-based model statistics

Language ID	EE	EH	HE	HH
Number of rules	444	179	187	288
Number of word pairs pertaining	38,197	16,472	14,719	22,802
Confidence score (%)	41	17	15	24

So, the weighted average of the rule,
 For language ID EE is $4/1200 = 0.0033$.
 For language ID EH is $3/600 = 0.005$.
 For language ID HE is $2/500 = 0.004$.
 For language ID HH is $1/1600 = 0.0006$.

Therefore, Rule NN-NN was allotted to language ID *EH*.

The number of rules for each language ID and their cumulative *Confidence* score are given in Table 5.

The system was validated with a manually prepared test data that contained 100 word pairs (25 word pairs from each language ID). It yielded an accuracy score of 75%.

3.4 Naïve Bayes Model

To be more sure about the POS tag rules that govern language switching in code-switched data, we implemented the Naïve Bayes classifier on the training data prepared in Sect. 3.1.3.

Naïve Bayes is a high-bias, low-variance classifier, and it can build a good model even with a small dataset. It is simple to use and computationally inexpensive. Typical use cases involve text categorization, including spam detection, sentiment analysis, and recommender systems [11].

We have used the same POS tags as features for our model. We subjected the model to tenfold cross-validation mechanism, and the model reported an accuracy of 79.69%. Kappa (K) value was measured at 0.655. The detailed confusion matrix is given in Table 6.

4 Conclusion and Future Work

In the current work, we have tried to analyze the role of POS tag lexicons when governing language switching in code-switched data. We used the training data from SAIL shared task as gold standard. We created a baseline system from the test data

Table 6 Confusion matrix for model trained using Naïve Bayes classifier

	EE	HE	EH	HH	Kappa
EE	33,726	2459	2352	3931	0.655
HE	19	903	0	46	
EH	19	0	899	41	
HH	1087	3994	4780	37,971	
Precision	86.45%				
Recall	52.66%				
Accuracy	79.69%				

(from the same shared task) to benchmark our research. The creation of the lexicon-based model and the Naïve Bayes model and their outperforming the baseline model, direct light on the aspect that syntactic rules play an important rule when code-switching is concerned.

As a future work, we would like to deal with prediction of the next token as well as its language when the first token is given as input. We would also like to use ConceptNet⁵ as a feature vector when dealing with the above problem.

Acknowledgements This work is supported by Media Lab Asia, MeitY, Government of India, under the Visvesvaraya Ph.D. Scheme for Electronics & IT.

References

1. P. Agarwal, A. Sharma, J. Grover, M. Sikka, K. Rudra, M. Choudhury, I may talk in English but gaali toh hindi mein hi denge: a study of English-Hindi code-switching and swearing pattern on social networks, in *2017 9th International Conference on Communication Systems and Networks (COMSNETS)* (IEEE, New York, 2017), pp. 554–557
2. U.Z. Ahmed, K. Bali, M. Choudhury, V. Sowmya, Challenges in designing input method editors for Indian languages: the role of word-origin and context, in *Proceedings of the Workshop on Advances in Text Input Methods (WTIM 2011)* (2011), pp. 1–9
3. J.P. Blom, J.J. Gumperz et al., Social meaning in linguistic structure: code-switching in Norway. *The bilingualism reader* (2000), pp. 111–136
4. M.S. Cárdenas-Claros, N. Isharyanti, Code-switching and code-mixing in internet chatting: between ‘yes’, ‘ya’, and ‘si’-a case study. *Jalt Call J.* **5**(3), 67–78 (2009)
5. A. Chopde, Itrans-Indian language transliteration package (2006). <http://www.aczoom.com/itrans>
6. M. Choudhury, K. Bali, T. Dasgupta, A. Basu, Resource creation for training and testing of transliteration systems for Indian languages, in *LREC* (2010)
7. D. Crystal, *A Dictionary of Language* (University of Chicago Press, Chicago, 2001)
8. B. Danet, S.C. Herring, Introduction: the multilingual internet. *J. Comput.-Med. Commun.* **9**(1), JCMC9110 (2003)
9. B. Danet, S.C. Herring, Multilingualism on the internet, in *Language and Communication: Diversity and Change Handbook of Applied Linguistics* vol 9, (2007), pp. 553–592

⁵<http://conceptnet.io/>.

10. A. Dey, P. Fung, A Hindi-English code-switching corpus, in *LREC* (2014), pp. 2410–2413
11. S. Kalmegh, Analysis of WEKA data mining algorithm REPTree, simple cart and randomtree for classification of Indian news. *Int. J. Innov. Sci. Eng. Technol.* **2**(2), 438–446 (2015)
12. D.C. Li, Cantonese-English code-switching research in Hong Kong: a survey of recent research. *Hong Kong Engl.: Auton. Creat.* **1**, 79 (2002)
13. S. Mandal, D. Das, S.K. Mahata, Preparing Bengali-English code-mixed corpus for sentiment analysis of Indian languages, in *The 13th Workshop on Asian Language Resources* (2018), p. 57
14. J. Patro, B. Samanta, S. Singh, A. Basu, P. Mukherjee, M. Choudhury, A. Mukherjee, All that is English may be Hindi: enhancing language identification through automatic ranking of likeliness of word borrowing in social media (2017). arXiv preprint [arXiv:170708446](https://arxiv.org/abs/170708446)
15. Y. Vyas, S. Gella, J. Sharma, K. Bali, M. Choudhury, POS tagging of English-Hindi code-mixed social media content, in *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (2014), pp. 974–979

Detection of Pulsars Using an Artificial Neural Network



Rajarshi Lahiri, Souvik Dey, Soumit Roy and Soumyadip Nag

Abstract The research paper demonstrates how to devise an optimal machine learning classifier for the detection of pulsars and then analyzes the performance of various classification models. Pulsars are classified as zero (not a pulsar) or one (pulsar) using logistic regression, decision tree, random forests, KNN classifier, and an artificial neural network. The performance has been analyzed based on five parameters: accuracy, recall, precision, specificity, and prevalence. For this, emission data from the HTRU2 dataset which was collected from the High Time Resolution Universe Survey is being used. Before using any classification algorithms, the patterns in the data were analyzed to understand the correlation between the different characteristics. The classifier proposed in this paper was found to give 98.01% accuracy.

Keywords Supervised learning · Logistic regression · K-nearest neighbors · Decision trees · Neural networks

1 Introduction

Pulsars are rotating neutron stars that produce radio emission detectable here on Earth. They are significant in maintaining accurate time, work as interstellar probes, and provide important proofs in the area of general relativity. The rotation of pulsars

R. Lahiri (✉) · S. Dey · S. Roy · S. Nag
Institute of Engineering & Management, Salt Lake Electronics Complex,
Gurukul, Y-12, Sector V, Kolkata 700091, India
e-mail: rlahiri1008@outlook.com

S. Dey
e-mail: souvikd337@gmail.com

S. Roy
e-mail: roysoumit23@gmail.com

S. Nag
e-mail: soumyadipnag3@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_15

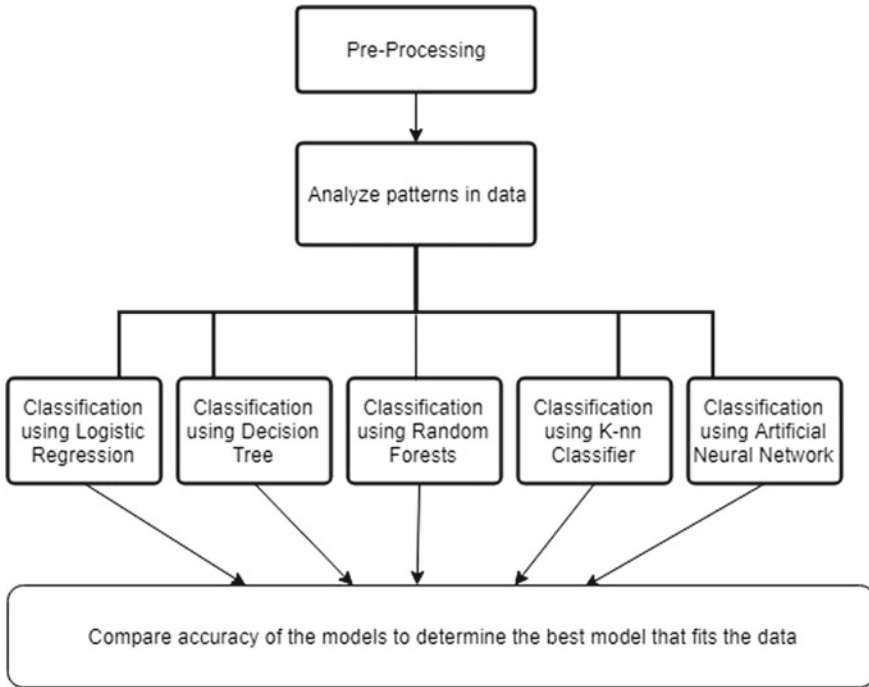


Fig. 1 Overview of the methodology and analysis

produces electromagnetic radiation across space. Human detection of these radiations is possible when they strike Earth’s atmosphere. We are concerned with finding periodic radio emissions produced by these pulsars. These radio emissions differ along with each rotation of the pulsar. We use a detection technique called “candidate,” which takes into account the number of rotations of the pulsar and length of the observation. Due to the limited amount of available information, most of these signals are caused by radio-frequency interference (RFI) or noise. This makes the problem of pulsar detection quite complicated. Machine learning algorithms solve this problem of pulsar detection by binary classification on a set of pulsar candidates. Figure 1 illustrates the approach taken by us in this research. We then compare the accuracy to determine the best classification model for the dataset.

2 Literature Survey

The algorithms follow the concept of supervised learning. This section highlights the various research papers for the detection of pulsars.

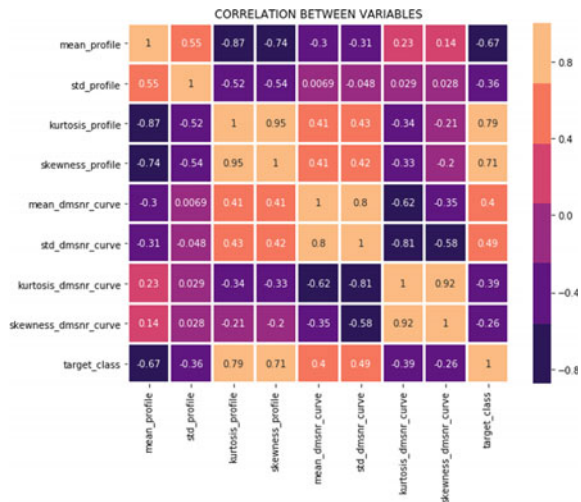
2.1 Analyzing the Dataset

HTRU2 [8] is a dataset, which describes a sample of pulsars collected during the High Time Resolution Universe Survey. Each pulsar candidate is described by eight features and a binary class variable. The first four are statistics obtained from the integrated pulse profile. This is a tuple of features that describe a longitude-resolved version of the signal that has been averaged in both time and frequency. The remaining four features are obtained from the delta modulation–signal-to-noise ratio (DM-SNR) curve. These features are as follows:

- Mean
- Standard deviation
- Excess kurtosis
- Skewness
- Mean
- Standard deviation
- Excess kurtosis
- Skewness
- Label (0 or 1)

It is very important to note that this dataset has a very big class imbalance. To circumvent the issue of class imbalance in a binary classification problem, we used the area under the ROC curve (AUC) [5] as a metric. This metric is insensitive to class imbalance as it measures the cumulative predictive power of the classifier under all trade-offs between sensitivity and specificity. The calculation of skewness and kurtosis is almost the same, and thus, skewness and kurtosis are highly correlated (Fig. 2).

Fig. 2 Correlation matrix



2.2 Related Work

Pulsar detection models have been previously made, namely the SPINN classifier [10]. C. M. Flynn et al. managed to provide a recall of 98%. SPINN's success is mainly attributed to the low complexity of the pulsar candidate. A. Punia et al. solved the problem of class imbalance by using synthetic minority oversampling technique (SMOTE) [11]. The SMOTE model was also successful in providing a recall of 98%. Bates et al. [3] demonstrated an ANN model that can detect 99% of the observations as noise and produce better classification results.

3 Common Supervised Learning Algorithms

3.1 Logistic Regression

Logistic regression is a powerful classification algorithm, where the probability of the output variable is calculated based on a given set of features. Here, the problem is a binomial logistic regression [4], where the response variable has two values 0 and 1. With binary classification, we assume x be some feature and y be the output which can be either 0 or 1. The probability that the output is 1 given its input can be represented as follows: $P(y = 1 | z)$

$$\log \left(\frac{p(X)}{1 - p(X)} \right) = \beta_0 + \beta_1 X \quad (1)$$

where the left-hand side is called the logit or log-odds function, and $p(X)/(1 - p(X))$ is called odds. The odds signifies the ratio of the probability of success to the probability of failure. We use maximum likelihood estimation [1]. There can be infinite sets of regression coefficients. The maximum likelihood estimate is that set of regression coefficients for which the probability of getting the data we have observed is maximum. If we have binary data, the probability of each outcome is simply π if it was a success, and $1 - \pi$ otherwise.

3.2 Decision Tree

Decision trees (DTs) [12] are simple supervised learning algorithms used for classification. Decision trees originate from the concept of conditional decision statements. The aim is to find the appropriate tree that produces the lowest cross-validation error on the given dataset. Decision trees are good when data is limited but performs bad on large-scale classification problems.

3.3 Random Forest Classifier

Random forests [7] are a learning method for classification that works by creating multiple decision trees when trained, and the class label is decided on the basis of the frequency of the outputs of these decision trees. Random forests are a good deterrent against decision trees' habit of overfitting to their training set.

3.4 K-Nearest Neighbor Classifier

KNN [2] is a popular classification algorithm. Its aim is to assign a class label to a data point based on the vote from its k-nearest neighbor, where k is an integer factor. The value of k is selected on the basis of the lowest cross-validation error.

3.5 Artificial Neural Network

To obtain control over the entire training process, a custom artificial neural network was created which is described in the Proposed Methodology section of the paper. In common artificial neural network (ANN) [9] implementations, the signal at a connection between artificial neurons is a real number, and the output of each artificial neuron is computed by some activation function of the sum of its inputs. Artificial neurons have a threshold such that the signal is only sent if the aggregate signal crosses that threshold. Different layers perform different kinds of computations on their inputs. The neural network consists of an input and output layer with some hidden layers. The layers are a set of neurons that are connected to the next layer by synapses.

4 Proposed Methodology

4.1 Preprocessing

Centering and scaling were done on each feature by calculating the necessary statistics on the samples in the training set (randomly split from the complete dataset). Standardization was done on the dataset, where each individual feature should look like standard normally distributed data (e.g., Gaussian with zero mean and unit variance). Our research mainly focuses on finding the best classifier for detecting pulsars. The selection of classifiers was done to maintain diversity from the simplification of decision trees to the complicated mathematics of neural networks and the practical application of logistic regression.

4.2 Building and Training the Neural Network

The classifier designed here consists of five dense layers. The input layer consists of ten neurons with an input dimension equal to 8 and ReLU [6] as the activation function. The second layer has six neurons with an input dimension equal to 8 and ReLU as the activation function. The third layer has four neurons with an input dimension equal to 4 and ReLU as the activation function. The fourth layer has one neuron with an input dimension as 4 and ReLU as the activation function. The output layer has one neuron with sigmoid [13] as the activation function as it gives a value between 0 to 1 that tells us the probability of the sample being a pulsar. The best model was obtained when it was trained for 200 epochs with a batch size of 100 and simple gradient descent as the optimizer and binary cross-entropy to calculate the loss. Cross-entropy loss measures the performance of a classification model whose output is a probability value between 0 and 1. Our classifier has a loss of 0.067. The neural network used in the classification of pulsars in this paper is depicted in Fig. 3.

The activation function used to connect the first four layers is ReLU (Rectified Linear Unit). The activation function used in the output layer is sigmoid. The main reason why we have used sigmoid function is that the output exists between 0 and 1. Therefore, it is especially used for models where we have to predict the probability as an output. Since the probability of anything exists only between the range of 0 and 1, sigmoid is the right choice. The sigmoid and the ReLU formulae used are as follows:

$$\text{Sigmoid: } \sigma(z) = \frac{1}{1 + e^{-z}} \quad (2)$$

$$\text{ReLU: } R(z) = \max(0, z) \quad (3)$$

5 Results and Analysis

The performance of our ANN model can be evaluated on the basis of metrics that are accuracy, specificity, and severity. The metrics used are false negatives (FN), true positives (TP), false positives (FP), and true negatives (TN).

True positives (TP): We predicted pulsars, when it is actually a pulsar.

True negatives (TN): We predicted not a pulsar, when it is actually not a pulsar.

False positives (FP): We predicted pulsar, but it is not a pulsar (also known as a “Type I error.”).

False negatives (FN): We predicted not a pulsar, but it is actually a pulsar (also known as a “Type II error.”).

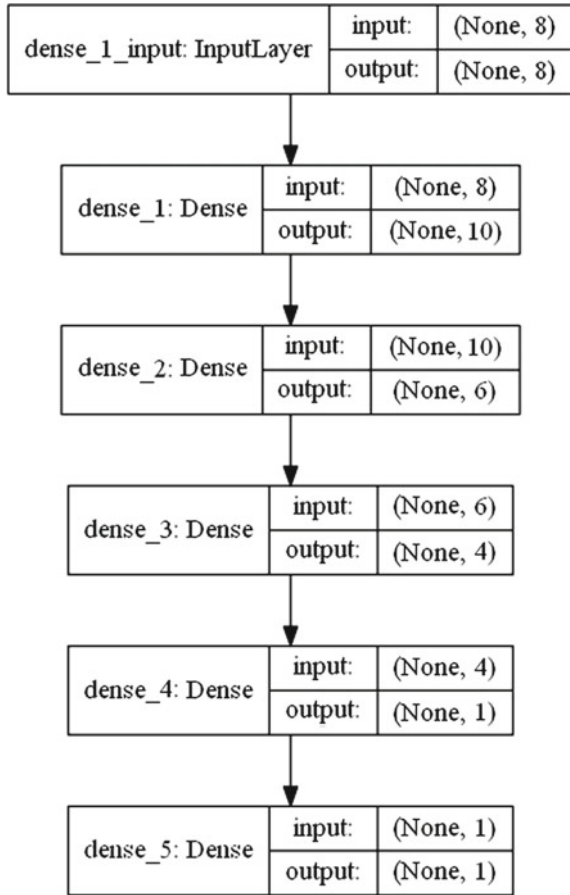
Total Positives (P): FN + TP

Total Negatives (N): FP + TN

Accuracy: The accuracy of a classifier on the given set of pulsars is the proportion of the total number of pulsars that are correctly identified.

Accuracy = (TP + TN) / (P + N)

Fig. 3 Neural network structure



Recall : Also referred to as sensitivity.

Recall = TP/P

Specificity : TN/N

A ROC curve is used to evaluate the performance of a binary classifier. It calculates the performance of a classifier over all possible thresholds. By plotting the true positive rate (y-axis) against the false positive rate (x-axis), we obtain the ROC curve. The area under the ROC curve measures the accuracy of a binary classifier. An area of 1 represents a perfect classifier; an area of 0.5 represents an inferior classifier. The desired area under the ROC curve for a classifier to be optimal should be greater than 0.8. Given below are the confusion matrices and the ROC curves along with the result of our performance metrics for the proposed algorithms.

Logistic regression

Accuracy: 0.976

Misclassification rate: 0.02

Recall: 0.815

Specificity: 0.993

Precision: 0.980

Prevalence: 0.095

According to our ROC metric, the classifier is graded as excellent (Fig. 4).

Decision tree

Accuracy: 0.963

Misclassification rate: 0.036

Recall: 0.814

Specificity: 0.980

Precision: 0.806

Prevalence: 0.095

According to our ROC metric, the classifier is graded as good (Fig. 5).

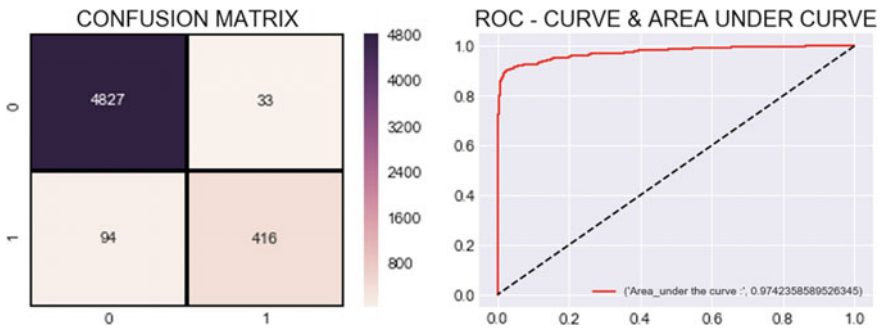


Fig. 4 Logistic regression

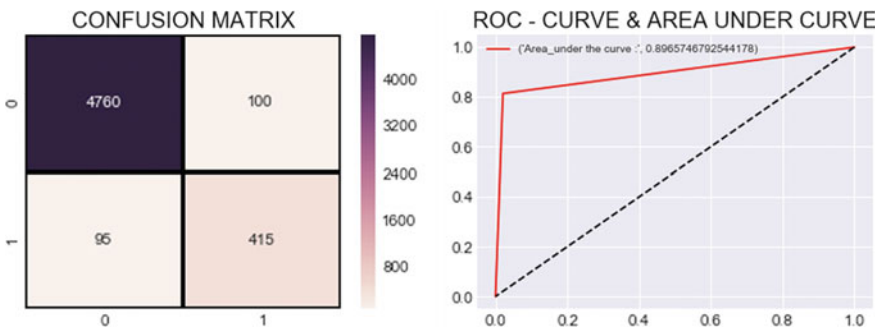


Fig. 5 Decision tree

Random forest

Accuracy: 0.977
Misclassification rate: 0.022
Recall: 0.839
Specificity: 0.992
Precision: 0.918
Prevalence: 0.095

According to our ROC metric, the classifier is graded as excellent (Fig. 6).

K-nearest neighbor

Accuracy: 0.969
Misclassification rate: 0.03
Recall: 0.776
Specificity: 0.989
Precision: 0.888
Prevalence: 0.095

According to our ROC metric, the classifier is graded as excellent (Fig. 7).

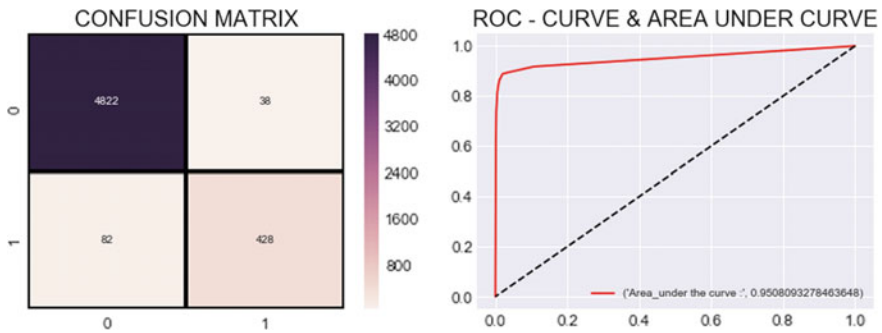


Fig. 6 Random forest

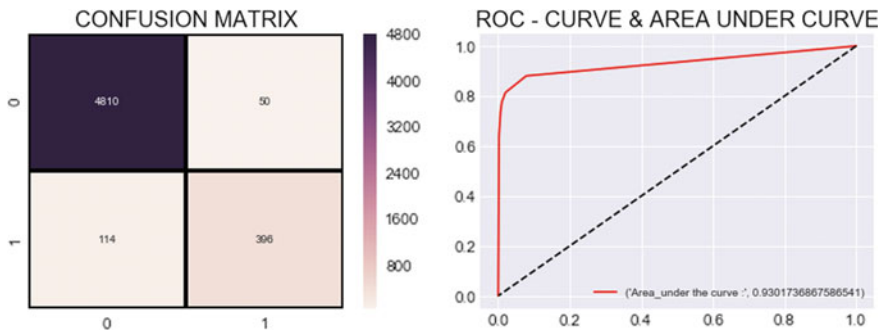
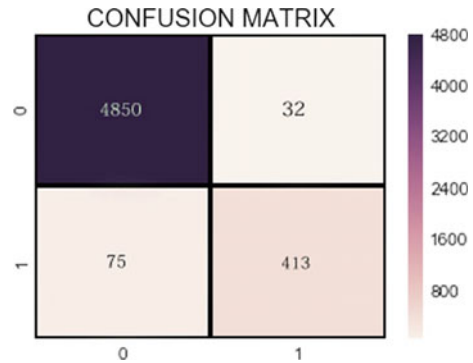


Fig. 7 K-nearest neighbor

Fig. 8 Neural network

Neural network

Accuracy: 0.980

Misclassification rate: 0.02

Recall: 0.984

Specificity: 0.928

Precision: 0.993

Prevalence: 0.095

According to our confusion matrix, the classifier is graded as excellent (Fig. 8).

From our calculations, the neural network performs best as it gives an accuracy of 98.01% on the test data with a loss of 0.0678. The neural network was trained for 200 epochs with a batch size of 100 using simple gradient descent as the optimizer. The next best algorithms that performed well are random forest and logistic regression at an almost identical accuracy of 97.7%. Also, the presence of high precision and low recall in these two algorithms suggests that those we predict as positive are indeed positive (low FP).

6 Conclusion and Future Work

The classification model proposed in this paper yielded an accuracy of 98.01% on the HTRU2 dataset. In future in place of the HTRU2 dataset a bigger dataset with more samples of pulsar candidates can be used to test the proposed classifier and compare accuracy.

References

1. J. Aldrich, R. A. Fisher and the making of maximum likelihood 1912–1922. *Statist. Sci.* **12**(3), 162–176 (1997). <https://doi.org/10.1214/ss/1030037906>
2. N.S. Altman, An introduction to kernel and nearest-neighbor nonparametric regression. *Am. Stat.* **46**(3), 175–185 (1992). <https://doi.org/10.1080/00031305.1992.10475879>
3. S.D. Bates, M. Bailes, B.R. Barsdell, N.D.R. Bhat, M. Burgay, S. Burke-Spolaor, D.J. Champion, P. Coster, N. D'Amico, A. Jameson, S. Johnston, M.J. Keith, M. Kramer, L. Levin, A. Lyne, S. Milia, C. Ng, C. Nietner, A. Possenti, B. Stappers, D. Thornton, W. van Straten, The high time resolution universe pulsar survey VI. An artificial neural network and timing of 75 pulsars. *Mon. Not. R. Astron. Soc.* **427**(2), 1052–1065 (2012). <https://doi.org/10.1111/j.1365-2966.2012.22042.x>
4. D.R. Cox, The regression analysis of binary sequences. *J. R. Stat. Soc. Ser. B (Methodol.)* **20**(2), 215–242 (1958). <http://www.jstor.org/stable/2983890>
5. T. Fawcett, An introduction to roc analysis. *Pattern Recognit. Lett.* **27**(8), 861–874 (2006). <https://doi.org/10.1016/j.patrec.2005.10.010>, <http://www.sciencedirect.com/science/article/pii/S016786550500303X> (ROC Analysis in Pattern Recognition)
6. R.H.R. Hahnloser, R. Sarpeshkar, M.A. Mahowald, R.J. Douglas, H.S. Seung, Digital selection and analogue amplification coexist in a cortex-inspired silicon circuit. *Nature* **405**, 947–951 (2000). <https://doi.org/10.1038/35016072>
7. T.K. Ho, Random decision forests, in *Proceedings of 3rd International Conference on Document Analysis and Recognition*, vol 1 (1995), pp. 278–282 <https://doi.org/10.1109/ICDAR.1995.598994>
8. R.J. Lyon, B.W. Stappers, S. Cooper, J.M. Brooke, J.D. Knowles, Fifty years of pulsar candidate selection: from simple filters to a new principled real-time classification approach. *Mon. Not. R. Astron. Soc.* **459**(1), 1104–1123 (2016). <https://doi.org/10.1093/mnras/stw656>
9. W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **5**(4), 115–133 (1943). <https://doi.org/10.1007/BF02478259>
10. V. Morello, E.D. Barr, M. Bailes, C.M. Flynn, E.F. Keane, W. van Straten, Spinn: a straightforward machine learning solution to the pulsar candidate selection problem. *Mon. Not. R. Astron. Soc.* **443**(2), 1651–1662 (2014). <https://doi.org/10.1093/mnras/stu1188>
11. A. Punia, A. Sardana, M. Subashini, Evaluating advanced machine learning techniques for pulsar detection from htru survey, in *2017 International Conference on Intelligent Sustainable Systems (ICISS)* (2017), pp. 470–474. <https://doi.org/10.1109/ISS1.2017.8389455>
12. J.R. Quinlan, Induction of decision trees. *Mach. Learn.* **1**(1), 81–106 (1986). <https://doi.org/10.1007/BF00116251>
13. P. Verhulst, Notice sur la loi que la population suit dans son accroissement. *Corresp. Math. Phys.* **10**, 113–126 (1838). <https://ci.nii.ac.jp/naid/10015246307/en/>

Modification of Existing Face Images Based on Textual Description Through Local Geometrical Transformation



Mrinmoyi Pal, Subhajit Ghosh and Rajnika Sarkar

Abstract The main objective of this paper is to modify the facial features, to change the existing image based on description of height and width of the face components and to generate a transformed image. There are many challenges in dealing with the modification of existing face images which are variations in facial expressions can be a cause for confusion, lack of large-scale supervised datasets, both speed and scalability of system are of prime importance as results are expected in real times. The scope of the study is to test 30 male and female face images from ORL database. The study has been done using MATLAB R2013a .

Keywords Extraction · Transformation · Interpolation-algorithm · Face-transformation · Face-components · Face-detection · Construction · Modification · Geometrical-transformation · Image · Image-warping · Scaling · Translation · Smoothing

1 Introduction

Over the last several decades, face recognition has received notable importance as one of the most effective applications of image processing and analysis. The reason for this trend is its application in person identification and the access to viable technologies. Although several biometric technologies exist (e.g., voice-waves, retinal-scan and iris-scan), these techniques have not been welcomed by the common people till now. Face recognition has an advantage over other types of biometric recognition because of its non-contact connection, long-distance recognition and user-

M. Pal (✉) · S. Ghosh · R. Sarkar
Kolkata, India
e-mail: mrinmoyi.pal.mp@gmail.com

S. Ghosh
e-mail: subhajit.ghosh022@gmail.com

R. Sarkar
e-mail: rajnikajhiliksarkar@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_16

friendliness. In particular, face detection has been applied to bio-information-based authentication technologies such as accurate face extraction for face recognition as well as a number of various areas for identifying criminals, automatic video surveillance and many other security systems. In recent years, increased importance of modification of existing face images based on textual description has been observed. Modification of existing face images is important because the database may not always contain the picture of the person to be identified.

We account a person in reference to the different important face components which are eyes, eyebrow, lips and nose. Integration of these features provides the recognition of face as a whole. It is essential to modify these features based on description given by an eyewitness. An alternative system is forensic sketch generation by interviewing a witness to gain textual descriptions of the suspect [1]. To generate more realistic forensic photo, we use existing images for modification based on textual description through local geometrical transformation. This approach needs an adequate face database which is used for local geometrical transformation. The face images generated after local geometrical transformation are more expressive than the photo sketches because the photo sketches may not be realistic. Difference between modalities is another problem for matching photos and sketches. Therefore, it is required to modify the facial features and generate the altered image.

The next sections contain as follows Literature Review, Methodology, Experimental Results, Conclusions and Future Scope.

2 Literature Review

An exhaustive survey is carried out on facial feature transformation. Various researches and studies on this field started several decades back. Relevant literature to this research work is presented in this part. Based on this important review papers, scope and objective of the present study are established.

This section gives an insight into the present state of face image modification-related issues by exploring the available literature.

In [1], a multiscale Markov Random Fields (MRF) model has been used for sketch synthesis from a given face photo and recognition. Only faces of front view and neutral expression have been considered. In [2], 2-D image warping has been used for synthesizing human facial expressions. Spatial displacement of face point has been obtained from test images. Methods of facial feature extraction have been explored in [3]. New face has been constructed based on textual description with these face components. In [4], face images have been transformed along perceived dimensions. Prototype images have been used to define axis of transformation. Non-linear image shape transformation has been proposed in [5]. Bilinear, quadratic, cubic, biquadratic and bicubic image transformation models have been considered in this paper. In [6], transformation between face images of the single person has been used. Content-based image retrieval has been utilized [6, 7]. Image feature rep-

resentation and extraction are the fundamental basis of content-based image retrieval. Text descriptions of the images have been handled in [7].

3 Methodology

At first, we consider the base image as the image which has the best similarity levels with the description. The modification will be based on:

(1) Left eye: Height, Width. (2) Right eye: Height, Width. (3) Left eyebrow: Height. (4) Right eyebrow: Height. (5) Nose: Height, Width. (6) Lip: Height, Width.

The assessment has been performed on 30 female and male face images from the ORL database. The modifications will be performed on these facial features, i.e., eyes, eyebrows, nose and lips. Mostly criminal's appearance and textual description are noted from any eyewitness's interview. Based on textual description, we perform local geometrical transformation on face images. We perform the following steps:

Preprocessing, extraction of facial components from a face image, store the facial components, perform local geometrical transformation based on textual description and reconstruct the face image with the transformed facial features.

One possible integrated system architecture, as shown in Fig. 1.

3.1 Extraction Methodology

To transform a facial image, extraction of face component from images is required. The different seven components are:

A. Eyebrows (Left and Right) B. Eyes (Left and Right) C. Nose D. Lip E. Face cutting.

The extraction of all the facial components with one algorithm is not possible. Therefore, we used separate algorithms for different facial component. These algorithms have been designed in [3].

In the first step of facial component detection, all the positions of the components are tracked. Figure 2 shows geometrical model of a face image for the prediction of the region of interest or the approx positions where the facial components may appear and then the facial components of the actual regions are extracted by applying the algorithms over the area at the predicted regions.

$a = 0.14 W, b = 0.15 W, c = 0.06, d = 0.05 W, e = 0.36 W, f = 0.32 W, g = 0.40 W, h = 0.55 W, i = 0.50 W, j = 0.25 W.$

Here, we considered two points (x_1, y_1) and (x_2, y_2) as the coordinates of the top left and bottom right corner of the predicted rectangular region for each facial component, respectively.

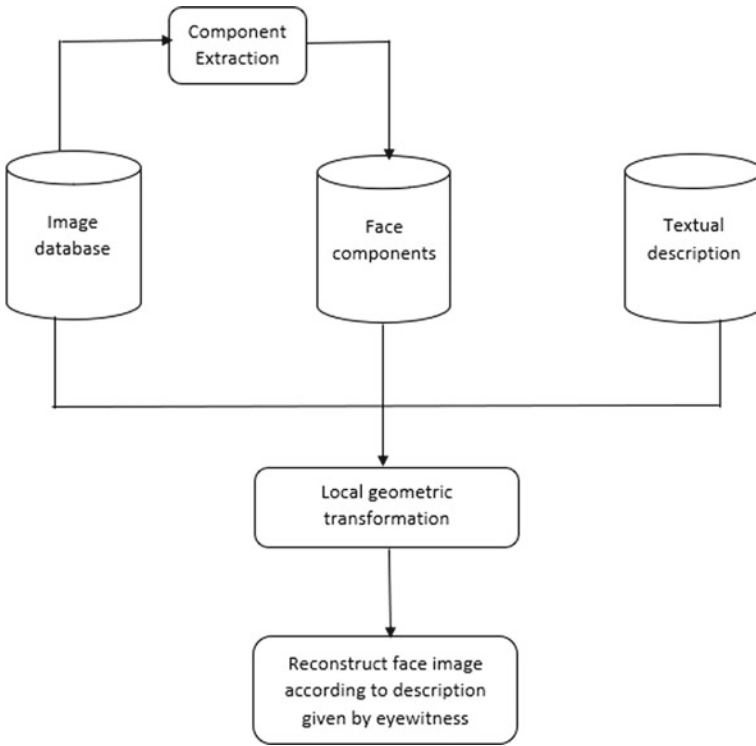


Fig. 1 System architecture

Fig. 2 Geometrical model of face

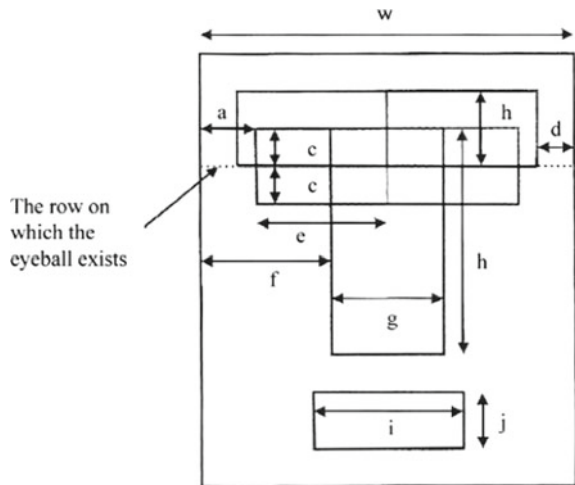
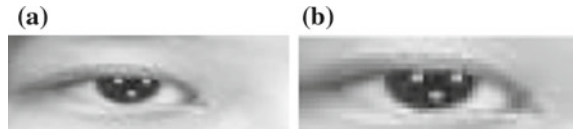


Fig. 3 **a** Anticipated right eye region. **b** Extracted region of right eye after applying the method



3.1.1 Right Eye Extraction

We are assuming that in the right half region of the face right eye is present and most changes of intensity values within two alongside pixels which happens to be in the mid-eye position. So, we acquire the row where the maximum changes happened. After that we can presume that particular area of eye. These following steps are:

1. Clip the right half region of the face.
2. Perceive the row where the most difference within two alongside intensities occurs. That row is called `Maximum_Index`.
3. Then clip the section of the image with $(x1, y1)$ and $(x2, y2)$ as per the model shown in Fig. 3 where

$$\begin{aligned} x1 &= \text{Maximum_Index} - (0.06 * W) \\ y1 &= 0.14 * W \\ x2 &= \text{Maximum_Index} + (0.06 * W) \\ y2 &= W/2 \end{aligned}$$

After that, apply median filter.

4. Normalize by adding up one integer value N ($N = 127$) with all the pixel intensity value and gain an updated matrix.
5. Detect the edges by Canny's algorithm of edge detection which yields a new matrix.
6. By applying algorithm of boundary detection, obtain the actual right eye region.

3.1.2 Face Cutting Extraction

To get empty face from a complete image of face, we needed to restore each and every facial component as per the skin tone.

Due to scarcity of face, we could not show the extraction for other parts of the face (lips, nose, eyebrows, etc.) (Fig. 4).

3.1.3 Face Component Modification

In this phase, the geometric transformation of the facial features is performed. Here, we analyze all facial components to find out the size of it as per its weight and height.

We have executed the modification in two ways:

1. Scaling and translation of the features.
2. Image warping using 2-D affine geometric transformation and then scaling.

3.2 *Scaling and Translation*

Image scaling changes the size of the face component by multiplying the coordinates of the points by scaling factors. It is the process of resizing an image. The scale operator performs a geometric transformation that can be used to reduce the image size. Image scaling can be also interpreted as image resampling or image reconstruction [8]. Scaling compresses or expands an image along the coordinate directions. An image can be zoomed either through pixel replication or interpolation. Bilinear interpolation algorithm has been used to perform scaling procedure. The face components are resized to a new width and height.

After scaling, translation of the facial components is performed. To translate a facial feature, we translate every point of the feature by the same amount to the same direction. Detailed algorithm could not be given due to lack of space [9].

An affine transformation is an important linear 2-D geometric class transformation which maps variables (e.g., pixel intensity values located at position (x_1, y_1) in an input image) into new ones (e.g., (x_2, y_2) in an output image) by applying a linear fusion of translation, rotation, scaling and/or shearing functions.

Affine transformation is a transformation which can be showed in the form of a matrix multiplication (linear transformation) sorted by a vector addition (translation) [10]. We can use this transformation to express: rotations (linear transformation), translations (vector addition) and scale operations (linear transformation).

We represent the affine transform by using a 3×3 matrix. If we want to transform a 2-D vector $[x \ y]$ by using A and B , we can do it with:

$$T = A \cdot \begin{bmatrix} x \\ y \end{bmatrix} + B$$

where A is a 2×2 matrix and B is a 2×1 matrix.

Fig. 4 **a** Complete image of face. **b** Empty face after restoring the facial components as per the skin tone

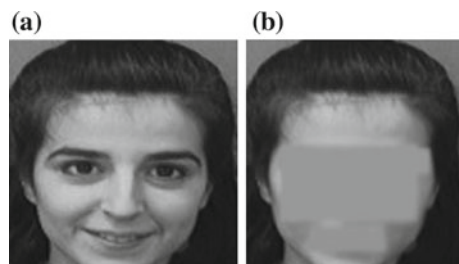


Fig. 5 Image warping implementation

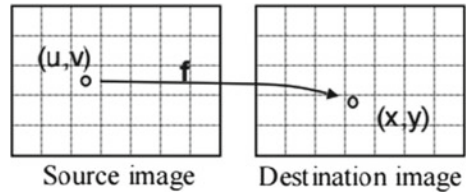


Image warping is an act of deforming a source image into a target or destination image according to the mapping between source space (u, v) and destination space (x, y) . The mapping is usually described by the functions $x(u, v)$ and $y(u, v)$. There are two basic components to an image warp: spatial transformation and resampling [11].

In homogeneous notation, 2-D points are represented by three vectors. For affine mappings, we denote points in source space by $P_s = (u, v, q)$ and points in the destination space by $P_d = (x, y, w)$.

Transformation matrices are denoted M_{ab} where a is the initial coordinate system and b is the final coordinate system. We have chosen $w = q = 1$ without loss of generality. The 3×3 matrix M_{sd} has six degrees of freedom. Geometrically, the vectors (a, d) and (b, e) are basic vectors of the destination space, and (c, f) is the origin.

Since an affine mapping has six degrees of freedom, it may be defined geometrically by specifying the source and destination coordinates of three points (Fig. 5).

3.3 Smoothing After Transformation

When the modified face components are overlapped on the face cutting, there may generate dissimilarities between overlapping pixels. To make the overall face natural to the user, smoothing should be used. We can apply image stitching which is a process of joining different images to form one image [12]. This process can be divided into three main components—image calibration, registration and blending [13]. Image calibration provides a pixel-to-real-distance conversion factor (i.e., the calibration factor, pixels/cm), that allows image scaling to metric units. Image registration involves matching features [6] in a set of images or using direct alignment methods to search for image alignments.

Image blending includes executing the adjustments found out in the calibration stage, fused with remapping of the images to an output projection.

There are two important approaches for image stitching.

1. Direct techniques: by shifting or warping the images relative to one other and to look, how much the pixels agree [14].

2. Feature-based techniques: by finding all corresponding feature points in an image pair by comparing every feature in one image with all features in the other using one of the local descriptors [15].

4 Experimental Results

In this section, images of faces after geometrical transformation of features are presented. The amount of changes in facial features is also projected in this section. In this study, a total of 30 images were considered. Among all the images, 23 images are of male and 7 are of female.

In this report, the following modifications of features are considered for a face image: (i) eye width change, (ii) eyebrow width change, (iii) nose height increase and width change, (iv) lip height and width change. The effects of glasses are not considered for this study. Amount of changes is shown in Fig. 6.

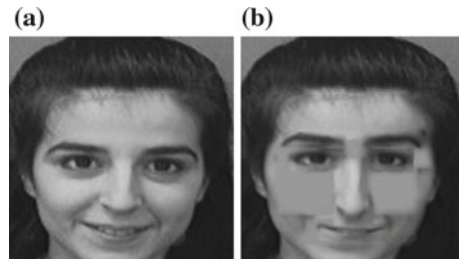
The width of eye has been decreased by 19%. The width of eyebrow has been increased by 50%. The height of nose has been increased by 28% and width has been decreased by 2%. The height of lip has been decreased by 3% and width has been increased by 30%.

Figure 7 shows the interface of modifications due to image warping using 2-D affine geometric transformation. There are six buttons depicting the changes, one button for viewing original image and one list box for selecting image. We have created a 2-D affine transformation object which is a 3×3 matrix. This matrix describes the 2-D forward affine transformation. The matrix used is as follows:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 20 & 20 & 1 \end{bmatrix}$$

After creating the geometric transformation object, we transform the image of face component according to that object. We generate six modified images.

Fig. 6 **a** Original face image. **b** Constructed face image after performing changes



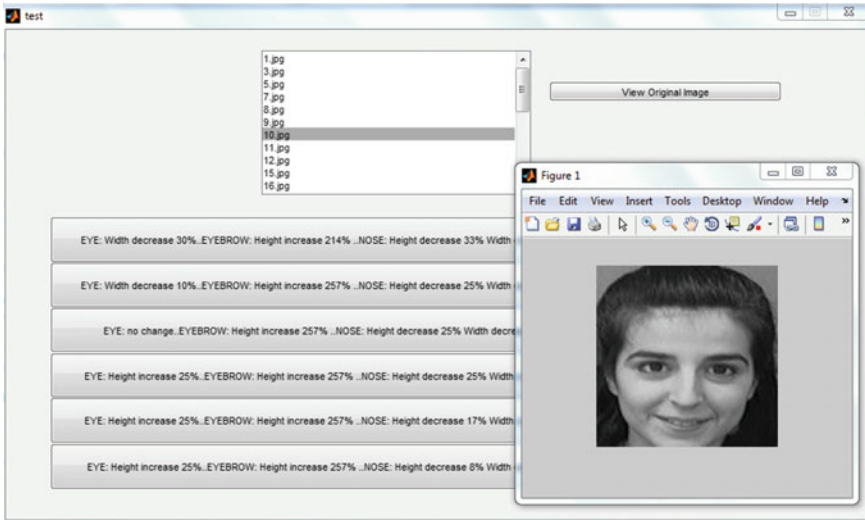


Fig. 7 User interface of modifications due to image warping using 2-D affine geometric transformation

In first modified image, the changes are, eye: Width decreases by 30%; eyebrow: Height increases by 214%; nose: Height decreases by 33%, Width decreases by 18%; lip: Height decreases by 35%, Width decreases by 47%.

In second modified image, the changes are, eye: Width decreases by 10%; eyebrow: Height increases by 257%; nose: Height decreases by 25%, Width decreases by 18%; lip: Height decreases by 35%, Width decreases by 43%.

In third modified image, the changes are, eye: no change; eyebrow: Height increases by 257%; nose: Height decreases by 25%, Width decreases by 10%; lip: Height decreases by 35%, Width decreases by 39%.

In fourth modified image, the changes are, eye: Height increases by 25%; eyebrow: Height increases by 257%; nose: Height decreases by 25%, Width decreases by 2%; lip: Height decreases by 35%, Width decreases by 45%.

In fifth modified image, the changes are, eye: Height increases by 25%; eyebrow: Height increases by 257%; nose: Height decreases by 17%, Width decreases by 2%; lip: Height decreases by 35%, Width decreases by 41%.

In sixth modified image, the changes are, eye: Height increases by 25%; eyebrow: Height increases by 257%; nose: Height decreases by 8%, Width decreases by 2%; lip: Height decreases by 35%, Width decreases by 37%.

Figure 8 presents the snapshot of the system when the button labeled as view original image is clicked.

In the below modified images, the upper left position of the face components is fixed. Only the size is varied. However, position of a face component may differ when the images are of different persons. But in different modified images of the

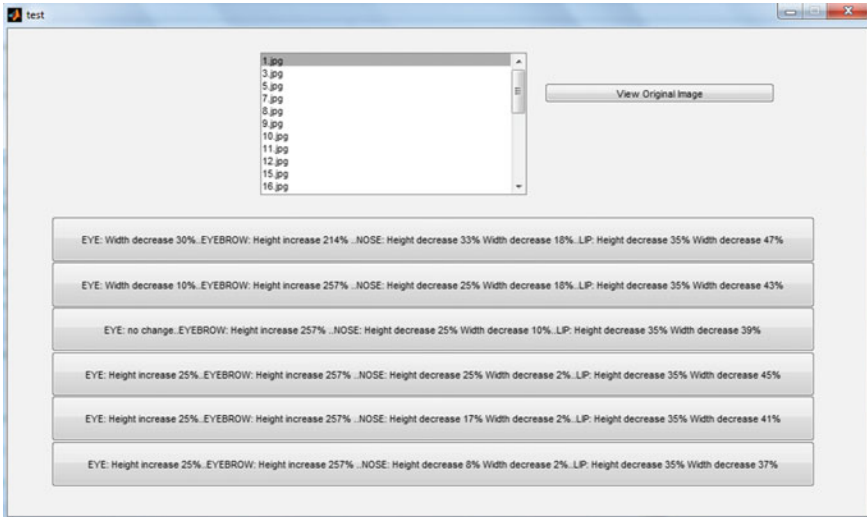


Fig. 8 Snapshot of the system when view original image button is clicked

same person, the upper left position of face components is always fixed. The position of face components is noted at the time of extracting those features.

Figure 9 shows the modified images after image warping.

5 Conclusions

A total of 30 images which include various aged male and female both from ORL database were tested but due to page limitation only 1 female image from ORL database has been shown here.

This system aims at constructing new faces based on textual representation with the geometrically transformed features from the same face. Image scaling, translation, 2-D affine transformation and image warping are used for local geometrical transformation of the facial features. For extraction of face components, we used a geometric model of a face [3]. This system is helpful when the intended image is unavailable in the database. Also, the user interface does not need any proficiency of the user. This system can also be used as a mechanism to generate database of different faces as it generates new modifies face images. This study can be very useful for criminal investigation where based on description given by eyewitness we have to generate a face of a criminal which is not available in the database.

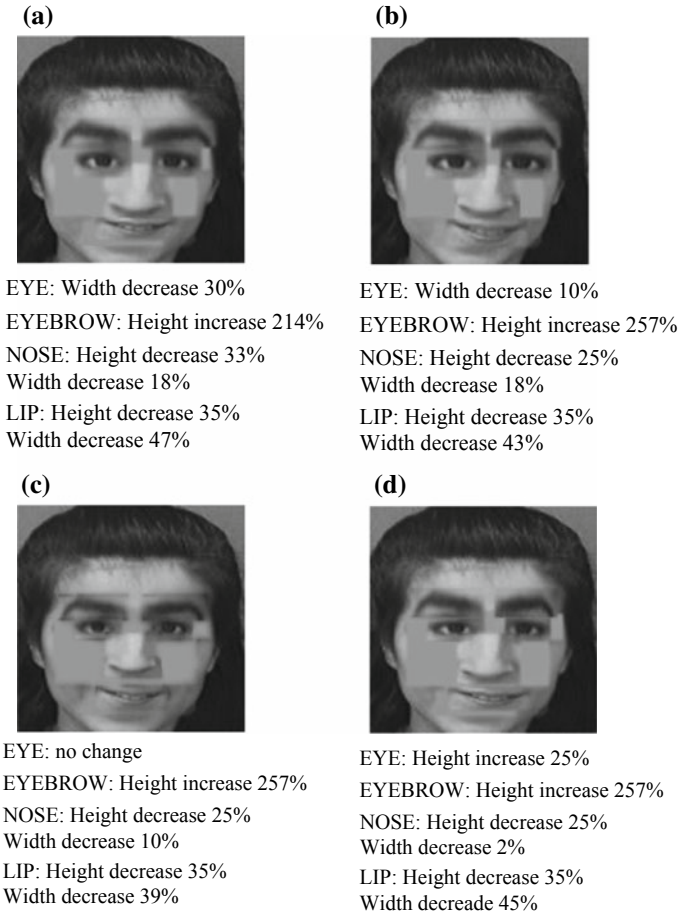


Fig. 9 a–d Images after modification

References

1. X. Wang, X. Tang, Face photo-sketch synthesis and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **31**(11), 1955–1967 (2009)
2. M. Leung, H.Y. Hui, I. King, Facial expression synthesis by radial basis function network and image warping, in *IEEE International Conference on Neural Networks*, vol. 3, pp. 1400–1405
3. D. Bhattacharjee, S. Halder, M. Nasipuri, D.K. Basu, M. Kundu, Construction of human faces from textual descriptions. *Soft Comput. Fusion Found. Methodologies Appl.* **15**(3), 429–447 (2009). (Springer)
4. Bernard Tiddeman, Michael Burt, David Perrett, Prototyping and transforming facial textures for perception research. *IEEE Comput. Graph. Appl.* **21**(5), 42–50 (2001)
5. Y.Y. Tang, C.Y. Suen, Image transformation approach to nonlinear shape restoration. *IEEE Trans. Syst. Man Cybern.* **23**(1), 155–172 (1993)

6. F. Perronnin, J.-L. Dugelay, K. Rose, A probabilistic model for face transformation with application to person identification. *EURASIP J. Appl. Sig. Process.* **4**, 510–521 (2004)
7. Y. Rui, T.S. Huang, Image retrieval: current techniques, promising directions, and open issues. *J. Vis. Commun. Image Represent.* **10**, 39–62 (1999)
8. G. Bebis, M. Georgiopoulos, N. da Vitoria Lobo, M. Shah, Recognition by learning affine transformations. *Pattern Recogn.* **32**(10), 1783–1799 (1999)
9. R.C. Gonzalez, P. Wintz, *Digital Image Processing*, 2nd edn. (Addison Wesley, Publisher, 1987)
10. OpenGL-ARB, OpenGL Reference Manual: The Official Reference Document to OpenGL, Version 1.1., 2nd edn. (Addison-Wesley, Reading, MA, 1997)
11. A. Watt, *3D Computer Graphics*, 3rd edn. (Addison-Wesley, Reading, MA, 1995)
12. P. Kale, K.R. Singh, A technical analysis of image stitching algorithm. *Int. J. Comput. Sci. Inf. Technol.* **6**(1), 284–288 (2015)
13. K. Shashank, N. SivaChaitanya, G. Manikanta, ChNV Balaji, V.V.S. Murthy, A survey and review over image alignment and stitching methods. *Int. J. Electron. Commun. Technol.* **5**, 50–52 (2014)
14. R. Szeliski, Image alignment and stitching. *Found. Trends Comput. Graph. Vis.* **2**(1), 1–104 (2006)
15. E. Adel, M. Elmogy, H. Elbakry, Image stitching based on feature extraction techniques: a survey. *Int. J. Comput. Appl.* **99**(6), 1–8 (2014)

A Neural Network Framework to Generate Caption from Images



Ayan Ghosh, Debarati Dutta and Tiyasa Moitra

Abstract With several applications in various different fields, computer vision can be found everywhere in our society. A fundamental problem of artificial intelligence (AI) is to give automatic description of the content in an image. In this paper, we focus on one of the visual recognition facets of computer vision, i.e., image captioning. To give the description for visual data has been studied for a long time but in the field of videos. In the recent few years, emphasis has been lead on still image description with natural text. Due to the recent advancements in the field of object detection, the task of scene description in an image has become easier. The aim of the paper was to train convolutional neural networks (CNN) with several hundreds of hyperparameters and apply it on a huge dataset of images (ImageNet) and combine the results of this image classifier with a recurrent neural network to create a caption for the classified picture. In this paper, we systematically analyze a deep neural networks-based image caption generation method. With an image as the input, the method can output an English sentence describing the content in the image. We analyze three components of the method: convolutional neural network (CNN), recurrent neural network (RNN) and sentence generation. By replacing the CNN part with three state-of-the-art architectures, we find the VGGNet which performs best according to the BLEU score. In this paper, we present the detailed architecture of the model used by us. We achieved a BLEU score of 56 on the Flickr8k dataset while the state-of-the-art results rest at 66 on the dataset.

Keywords Convolutional neural network · Recurrent neural network · BLEU score

A. Ghosh (✉) · D. Dutta · T. Moitra
Institute of Engineering and Management, Kolkata, India
e-mail: ayaghosh0627@gmail.com

D. Dutta
e-mail: debaratidutta39@gmail.com

T. Moitra
e-mail: tiyasamoitra1998@gmail.com

1 Introduction

Artificial intelligence (AI) is now at the heart of innovation economy. In the recent past, a field of AI namely deep learning has turned a lot of heads due to its impressive results in terms of accuracy when compared to the already existing machine learning algorithms. The task of being able to generate a meaningful sentence from a picture is not easy but can be very helpful.

Automatic caption generation is a demanding artificial intelligence problem where a textual interpretation must be created for a given image, which has great potential impact. Recently, deep learning methods have achieved state-of-the-art results on examples of this problem. For example, it could help visually impaired people better understand the content of images on the web. Also, it could provide more accurate and compact information of images/videos in scenarios such as image sharing in social network or video surveillance systems.

In the model proposed in the paper, we combine this into a single model consisting of a convolutional neural network (CNN) encoder which helps in creating image encodings. We use the VGG16 architecture proposed with some modifications. These encoded images are then passed to a LSTM network. The input to the network is an image which is first converted in a 224×224 dimension. We use the Flickr8k dataset to train the model. The model outputs a generated caption based on the dictionary it forms from the tokens of caption in the training set. The generated caption is compared with the human given caption via BLEU score measure.

Now comes the contribution of this paper.

2 Related Work

A relevant background on the previous works on this topic will be provided here. Recently, many have proposed different methods for creating image descriptions. Image to caption generation is a core part of scene understanding, which is important because of its use in a variety of applications such as image search, telling stories from albums, helping visually impaired people understand the web. Over the years, many different image captioning approaches have been developed.

The architectures used by the winners of ILSVRC have contributed a lot to this field. One such architecture used by us was the VGG16 proposed by He et al. in 2014 [1]. Their main contribution was a thorough evaluation of networks of increasing depth using an architecture with very small (3×3) convolution filters, which show that a significant improvement on the prior-art configurations can be achieved by pushing the depth to 16–19 weight layers. Apart from that the research in the tasks of machine translation has consistently helped in improving the state-of-the-art performance in language generation.

In 2015, researchers at Microsoft's AI Lab used a pipeline approach to image captioning [2]. They used a CNN to generate high-level features for each potential

object in the image. Then they used multiple instance learning (MIL) to figure out which region best matches each word. The approach yielded 21.9% BLEU score on MSCOCO. After the pipeline approach, researchers at Google came up with the first end-to-end trainable model. They were inspired by the RNN model used in machine translation.

Vinyt et al. [3] replaced this encoder RNN with CNN features of the image as the CNN features are widely used in all computer vision tasks. They called this model as neural image caption (NIC). Following this, two researchers at Stanford modified the NIC. They used an approach that leverages datasets of images and their sentence descriptions to learn about the inter-modal correspondences between language and visual data. Their alignment model was based on a novel combination of convolutional neural networks over image regions, bidirectional recurrent neural networks over sentences and a structured objective to align the two modalities through a multimodal embedding. They used the Flickr8K, Flickr30K and MSCOCO datasets and achieved state-of-the-art results in the same [4]. Their model was further modified by Jonathan et al. [5] in 2015 when they proposed a dense captioning task in which each region of an image was detected and a set of descriptions generated. Another model which used a deep convolutional neural network (CNN) and two separate LSTM networks were proposed by Wang et al. [6] in the year 2016.

3 Proposed Work

In this section of the paper we will discuss the proposed work in details.

3.1 *Applied Techniques*

The following techniques and ways are introduced and used to generate the caption from the image.

3.1.1 Convolutional Neural Networks

Convolutional neural networks (**ConvNets** or **CNNs**) are a category of artificial neural networks which have proven to be very effective in the field of image recognition and classification. First ConvNet was discovered in the year 1990 by Yann Lecun, and the architecture of the model was called as the LeNet architecture.

The entire architecture of a ConvNet can be explained using four main operations, namely

1. Convolution,
2. Linearity (ReLU),

3. Pooling or Sub Sampling,
4. Classification (Fully Connected Layer).

The above points are the main building blocks of a CNN, so to understand how all these will work is an important step to understand ConvNets.

Every image can be represented as a matrix of its pixel values. A digital camera image will have three channels—red, green and blue, and one can imagine those as three two-dimensional matrices stacked over each other, each having pixel values in the range of 0–255.

Convolution Operator

The purpose of convolution operation is to extract features from an image. We consider filters of size smaller than the dimensions of image. The entire operation of convolution can be understood with the example below.

Consider a small two-dimensional 5×5 image with binary pixel values. Consider another 3×3 matrix.

The 3×3 matrix is known as a filter or kernel or feature detector, and the matrix devised by gliding over the image and determining the dot product is called the convolved feature or activation map or the feature map. Stride is the number of pixels by which we slide the filter over the original image.

Introducing Nonlinearity

An additional operation is applied after every convolution operation. The most usually used non-linear function for images is the ReLU that stands for rectified linear unit. The ReLU operation is an element-wise operation which replaces the negative pixels in the image with a zero.

Although most of the operations in real life relate to nonlinear data, the output of convolution operation is linear because the operation applied is element-wise multiplication and addition. Some other commonly used nonlinearity functions are sigmoid and tanh.

Spatial Pooling

The pooling operation decreases the size of the image but preserves the important features in the image. The most common type of pooling technique used is max pooling. In max pooling, you slide a window of $n \times n$ where n is less than the side of the image and determine the maximum in that window and then shift the window with the given stride length. The complete process is specified by Fig. 1.

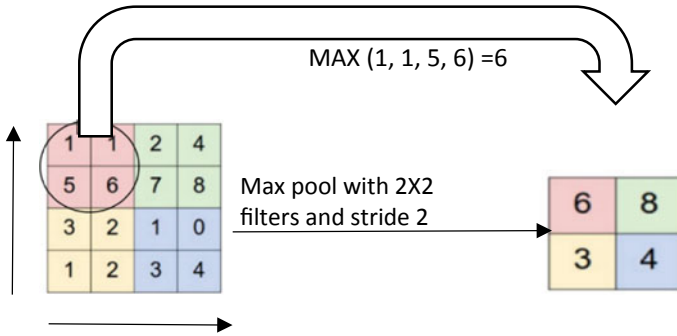


Fig. 1 Max pooling operation

Fully Connected Layer

The fully connected layer is the multiple layer perception that takes the help of SoftMax activation function in the output layer. The term “fully connected” refers to the fact that all the neurons.

In the previous layer are connected to all the neurons of the next layer. The convolution and pooling operation generate features of a picture. The task of this layer is to map all the feature vectors to the classes in the training data.

The task of image classification on cifar-10 has shown state-of-the-art results with the use of convnets. We use the Alex net architecture proposed by Alex krizhevsky with a few tweaks. Alex trained for images having 224*224 dimensions and hence need to be modified to be used for cifar-10 since the images in cifar-10 are 32*32. The model used by us has alternate layers of convolution and nonlinearity. We use a fully connected layer at the end which uses SoftMax activation to give the scores of the 10 classes present in the cifar-10 dataset.

The dataset on these convnets yields an accuracy of 85% within around 1.5 h of training on GPU’s. The plots of loss and accuracy on test and validation set are shown in the figures below.

The model is built using tensorflow. Tensorflow is an open source library developed by Google brain team for machine learning. Though being a python api, most of the code of tensorflow is written in C++ and CUDA which is nvidia’s programming language for GPU’s. This helps tensorflow in faster execution of code since python is slower than CPP. Also, the use of GPU enhances the performance of the code significantly.

3.1.2 Deep Cnn Architecture

The details of the CNN were discussed in Sect. 3.3. Convolutional neural network (CNN) has improved the task of image classification significantly. ImageNet Large Scale Visual Recognition Competition (ILSVRC) has provided various open source

deep learning frameworks like ZFnet, AlexNet, VGG16, ResNet, etc. have shown great potential in the field of image classification. For the task of image encoding in our model, we use VGG16 which is a 16-layered network proposed in ILSVRC 2014 [1]. VGG16 significantly decreased the top-5 error rate in the year 2014 to 7.3%. The image taken for classification needs to be a 224*224 image. The only preprocessing done is by subtracting the mean RGB values from each pixel determined from the training images.

The convolution layer consists of 3*3 filters and the stride length is fixed at 1. Max pooling is done using 2*2-pixel window with a stride length of 2. All the images need to be converted into 224*224-dimensional image. A rectified linear unit (ReLU) activation function follows every convolution layer.

The advantage of using a ReLU layer over sigmoid and tanh is that it accelerates the stochastic gradient descent. Also unlike the extensive operations (exponential, etc.), the ReLU operation can be easily implemented by thresholding a matrix of activations at zero. For our purpose, however, we need not classify the image and hence we remove the last 1*1*1000 classification layer.

The output of our CNN encoder would thus be a 1*1*4096 encoded which is then passed to the language generating RNN. There have been more successful CNN frameworks like ResNet, but they are computationally very expensive since the number of layers in ResNet was 152 as compared to VGG16 which is only a 16-layered network. A comparison between the layers versus top-5 error rate in the ILSVRC challenge is given below.

3.1.3 Recurrent Neural Net (RNN) Decoder Architecture

Recurrent neural nets are a type of artificial neural network in which connection between units forms a directed cycle. The advantage of using RNN over conventional feed forward net is that the RNN can process arbitrary set of inputs using its memory. RNNs were discovered in the year 1980 by John Hopfield who gave the famous Hopfield model. Recurrent neural nets in simple terms can be considered as networks with loops which allow the information to persist in the network.

A recurrent neural network can be considered as multiple copies of same network with each network passing the message to its successor.

Consider a machine that tries to generate sentences on its own. For instance, the sentence is “I grew up in England, I speak fluent English,” if the machine is trying to predict the last word in the sentence, i.e., *English*, the machine needs to know that the language name to be followed by fluent is dependent on the context of the word England. It is possible that the gap between the relevant information and the point where it is needed becomes very large in which case the conventional RNNs fail.

To overcome the above-mentioned problem of “long term dependencies,” Hochreiter and Schmidhuber proposed the Long Short-Term Memory (LSTM) networks in the year 1997. Since then, LSTM networks have revolutionized the fields of speech recognition, machine translation, etc. Like the conventional RNNs, LSTMs also have

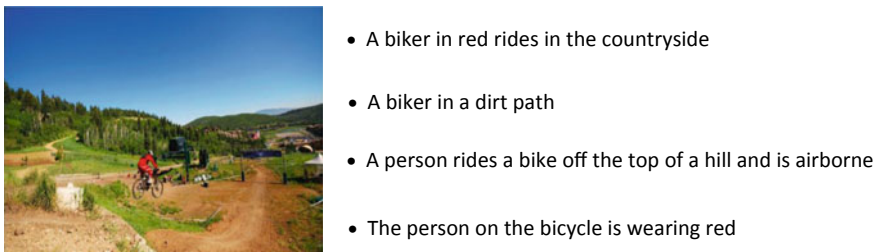


Fig. 2 Sample image and corresponding captions from the Flickr8k dataset

a chain like structure, but the repeating modules have a different structure in case of a LSTM network.

The key behind the LSTM network is the horizontal line running on the top which is known as the cell state. The cell state runs through all the repeating modules and is modified at every module with the help of gates. This causes the information in a LSTM network to persist. We use this LSTM network with a slight variation.

3.2 Dataset Preparation

For this paper, we use a Flickr8k dataset. We manually label the dataset for training purpose, and for test dataset, the given data are already labeled. This labeled test data are used for validation purpose. The dataset contains 8000 images with five captions per image. The dataset by default is split into image and text folders. Each image has a unique id, and the caption for each of these images is stored corresponding to the respective id. The dataset contains 6000 training images, 1000 development images and 1000 test images. A sample from the data is given in Fig. 2.

Other datasets like Flickr30k and MSCOCO for image captioning exist, but both these datasets have more than 30,000 images thus processing them becomes computationally very expensive. Captions generated using these datasets may prove to be better than the ones generated after training on Flickr8k because the dictionary of words used by RNN decoder would be larger in case of Flickr30k and MSCOCO.

3.3 Model Overview

The model proposed takes an image I as input and is trained to maximize the probability of $p(S|I)$ [3] where S is the sequence of words generated from the model and each word S_t is generated from a dictionary built from the training dataset. The input image I is fed into a deep vision convolutional neural network (CNN) which helps in detecting the objects present in the image. The image encodings are passed on to

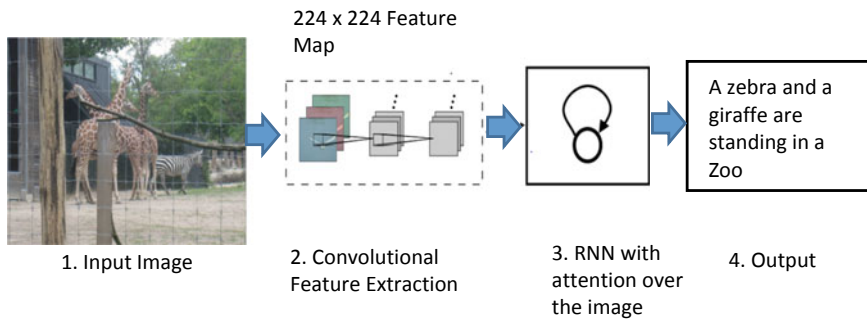


Fig. 3 Overview of the image captioning model

the language generating recurrent neural network (RNN) which helps in generating a meaningful sentence for the image. An analogy to the model can be given with a language translation RNN model where we try to maximize the $p(T|S)$ where T is the translation to the sentence S . However, in our model the encoder RNN which helps in transforming an input sentence to a fixed length vector is replaced by a CNN encoder. Recent research has shown that the CNN can easily transform an input image to a vector (Fig. 3).

For the task of image classification, we use a pre-trained model VGG16. The details of the models are discussed in the following section. A Long Short-Term Memory (LSTM) network follows the pre-trained VGG16 [1]. The LSTM network is used for language generation. LSTM differs from traditional neural networks as a current token is dependent on the previous tokens for a sentence to be meaningful and LSTM networks take this factor into account.

In the following sections, we discuss the components of the model, i.e., the CNN encoder and the language generating RNN in details.

4 Result Analysis

We define the accuracy of the model by BLEU score. Bilingual evaluation under-study (BLEU) is an algorithm that evaluates the standard of text which has been created by a machine. Blue score is always defined between 0 and 1, 0 being the machine translation is not at all related to the reference sentence. BLEU's evaluation system requires two inputs (a) a numerical translation closeness metric, which is then assigned and measured against, and (b) a corpus of human reference translations (Fig. 4).

The candidate in this example has all its words contained in the reference thus giving a unigram precision score of 1. A unigram precision score of 1 means the candidate and reference sentences are highly correlated. However, we can see that the two are very different from each other.

Fig. 4 Sample example

Candidate	the	the	the	the	the	the	the
Reference 1	the	cat	is	on	the	mat	
Reference 2	there	is	a	cat	on	the	mat

BLEU makes straightforward modification. In the candidate translation for each word, the maximum total count is taken by the algorithm, $mmax$. In the example above, the word “the” is appearing two times in reference 1. Thus, $mmax = 2$.

For the candidate translation, the count mw of each word is clipped to a maximum of $mmax$ for that word. In this case, “the” has $mw = 7$ and $mmax = 2$; thus, mw is clipped to 2. These clipped counts mw is then added over all distinct words in the candidate. This sum is then divided by the total number of words in the translation of the candidate. In the example above, the score of modified unigram precision would be: $2/7$.

For calculating BLEU score, we first generate captions for all the test images and then use these machine-generated captions as candidate sentences. We compare these candidate sentences with five of the captions given by humans and average the BLEU score of candidate corresponding to each of the references. Thus, for 1000 test images we calculate 1000 BLEU scores using Natural Language Toolkit (NLTK), a python package.

We averaged out these BLEU scores over the 1000 test images. The net BLEU score of the model after training for 70 epochs with a batch size of 512 was found to be 0.562 or 56.2% while the state of the art on Flickr8k is around 66%. On increasing the number of epochs, we may reach near state-of-the-art results but that would require higher computation. The net BLEU score can also be improved by decreasing the batch size.

The model was able to predict the meaningful captions for an image in most of the cases. However, in some of the cases it got confused due to lack of the tokens in dictionary to define an event. Also, it got confused in the cases where almost the entire image was covered with one color.

5 Conclusion and Future Scope

Our end-to-end system neural network system is capable of viewing an image and generating a reasonable description in English depending on the words in its dictionary generated on the basis of tokens in the captions of train images. The model has a convolutional neural network encoder and a LSTM decoder that helps in generation of sentences. The purpose of the model is to maximize the likelihood of the sentence given the image. Experimenting the model with Flickr8K dataset shows decent results. We evaluate the accuracy of the model on the basis of BLEU score. The accuracy can be increased if the same model is worked upon a bigger dataset.

Furthermore, it will be interesting to see how one can use unsupervised data, both from images alone and text alone, to improve image description approaches.

Since video processing generally involves processing key frames (images) from streaming video data, it is therefore possible to convert existing models and applications from images to video (e.g., automatic video summarization). The dataset given in this thesis can be further elaborated according to specific applications to make use of the news document by virtue of which the annotation keywords may be increased by identifying synonyms, even sentences that correspond to the image captions. A predictable extension would be considering spatial information when dealing with image representation. As of now, we take the concerned image regions or detected regions as bag-of-words, which might be extended to bigrams corresponding to their spatial relations. For example, experiment can be done with features concerning document structure such as titles and further exploit syntactic information in a more straightforward way.

References

1. K. Simonyan, A. Zisserman, in *Very deep convolutional networks for large-scale image recognition*. *arXiv preprint arXiv: 1409.1556* (2014)
2. H. Fang et al., From captions to visual concepts and back, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015)
3. O. Vinyals et al., Show and tell: a neural image caption generator, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015)
4. A. Karpathy, L. Fei-Fei, Deep visual-semantic alignments for generating image descriptions, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2015)
5. J. Johnson, A. Karpathy, L. Fei-Fei, Denscap: fully convolutional localization networks for dense captioning, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016)
6. C. Wang et al., Image captioning with deep bidirectional LSTMs, in *Proceedings of the 2016 ACM on Multimedia Conference* (ACM, 2016)

Remote Sensing and Advanced Encryption Standard Using 256-Bit Key



Sumiran Naman, Sayari Bhattacharyya and Tufan Saha

Abstract The most convenient way of data acquisition in recent times is undoubtedly through remote sensing. Remote sensing data has shown substantial growth in terms of pliability, reliability, cost-effectiveness, speed and efficiency. However, the growth is prone to data breaches and attacks. The commonly employed method for data security is encryption and decryption. This paper proposes the secure transmission of remote sensed data by using AES (Rijndael) algorithm with 256-bit key as the underlying encryption algorithm.

Keywords Remote sensing · Encryption · Decryption · AES-256

1 Introduction

National Remote Sensing Centre (NRSC) defines remote sensing as the most convenient way of data acquisition in recent times [1]. Remote sensing is acquiring information about an object without making any physical contact. Basically, it refers to satellite or aircraft-based sensor technologies for recognition, categorisation and capturing specific details of objects on Earth. Image processing technique is used to extract quantitative information from an image. Remote sensed image processing in recent days has been accounting to a larger space in the current research era. The data sources used in GIS are majorly comprised of remotely sensed data.

Data interception and breach attacks have shown an exponential increase in recent days. The data transmission occurring between the various communication channels

S. Naman · S. Bhattacharyya (✉) · T. Saha
Department of Computer Science and Engineering, Institute of Engineering
and Management, Kolkata, India
e-mail: iamsay253@gmail.com

S. Naman
e-mail: sumiran.naman.rs@gmail.com

T. Saha
e-mail: tufan.saha@outlook.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling
and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_18

are always vulnerable and the chances of illegal acquisition is directly proportional to the confidentiality attached with the data. The recent actions of satellite interceptions have shown alarming and catastrophic threats pertaining to the national security as well. Here is where encrypting of the data over communication channel becomes an inalienable data characteristic for secure and efficient transmission. The encryption algorithm used here is AES (Advanced Encryption Standard) also referred as Rijndael algorithm. AES algorithm was accepted by the US National Institute of Standards and Technology in the year 2001. Rijndael is a family of block ciphers with different block sizes as well as different key length.

This paper at first briefly describes history of AES and the research works that have been done in the past years in the field of satellite data security using different proposed algorithms including AES. This leads to the proposed work in Sect. 2 where each step of our proposed cryptosystem is described. Section 3 discusses about the need of AES-256 algorithm for the security of remote sensing data. And finally, Sect. 5 concludes the paper.

1.1 Advanced Encryption Standard (AES)

During the selection process, The National Security Agency (NSA) reviewed all the AES finalists including Rijndael and said they were secure enough to implement for the encryption of the unclassified documents. However after rigorous testing in May 2002, the Rijndael algorithm was selected as the most suitable of all and hence became effective as Federal Government standard. AES is also listed in ISO/IEC 18033-3 standards and is first and only of its kind to be publicly available block ciphers for top-level security information approved by NSA. Rijndael was accepted as AES with three variants of 128-, 192-, 256-bit key size with block size of 128 bits. The design and strength of all keys of AES algorithm (i.e. 128, 192, 256) are sufficient to protect classified information up to SECRET level. However, 192- or 256-bit key should be used for top secret-level information [2]. AES uses 10, 12, 14 rounds for 128-, 192- and 256-bit keys for encryption, respectively.

1.2 Related Works

AES algorithm, being a higher efficient algorithm proposed in data encryption, has been widely acceptable in the works implemented in the field of satellite data security mainly the image data security. Boukhatem Mohammed Belkaid, Lahdir Mourad and Cherifi Mehdi proposed a system to intensify the security of transmission of Meteosat images which was based on AES and RSA (Rivest–Shamir–Adleman) algorithm and the system creates a new password in every session of encryption [3]. Lijie Yin et al. introduced an encryption scheme that merges EZW and chaos theory for the encryption of remote sensing images that avoids transmission of multi-key [4]. In 2010,

Muhammad Usama et al. explained a system based on idea of chaotic cryptography using multiple chaotic maps and external secret key for encryption and decryption of satellite images [5]. Xiaoqiang Zhang et al. proposed an encryption algorithm in hybrid domain and experimented on remote sensing images using DWT and PWLCM system [6]. Faruk Ahmed and E. N. Ganesh implemented the processing of remote sensing images by using watermarking and encryption algorithm [7]. Naida H. Nazmudeen and Farsana F. J. proposed the idea of combining DWT-DCT watermarking algorithm and AES encryption algorithm for the improvement of satellite image security [8]. A secure image transformation algorithm was presented by Rafeeqe K. M. and S. Shiva Shankar and the transformed image was encrypted using remote sensing algorithm [9]. P. Gunavathy and A. Vinoth Kannan proposed the concept of enhancement and segmentation of a satellite image and encrypting that segmented image using RC4 algorithm (Stream Cipher) [10]. Pokhali Sayeda Roohi Banu presented a model of AES which is fault-tolerant based on Hamming error correction code for avoiding the corruption of satellite onboard data mainly due to the SEUs (single event upsets) [11]. Recently in 2018, NovelSat Satellite Technologies have launched ProtCASTER, a satellite broadcast solution that is a circuit implementation of AES-256 algorithm [12].

The increasing public and private data communication has also given rise to further space assets dependencies. Once data breach has happened, it becomes really notorious job for the respective country to deal with. The attacker or the adversaries can be anywhere in the range of the satellite beam coverage which is miles. Moreover, the attacked country has to face double jeopardy as the attack in outer space cannot surely be pointed out to a specific country. As it will not only pose threat to international political relations of the country but also the strategic advances could diminish. Many powerful countries have observed state-sponsored cyber attacks and the lack of any international law restricting attacks in space makes the space cybersecurity further vulnerable [13, 14]. The space cyber attacks in the present days cannot be legally penalised, so this leaves us with only option to safeguard our satellites with the best possible security. The need for securing the remote sensing data hence increases further as they carry more sensitive information as compared to others. So it is wise to establish a secure communication traffic to nullify any such malicious attacks posing threat to the national security jeopardising its place at international level. This has given rise to various methods of cryptography and data hiding techniques. The most refined and trusted algorithm is AES (Advanced Encryption Standard) with the variation of 256-bit key which is considered 'unbreakable'.

2 Proposed Work

In this work, a secure satellite communication system based on AES algorithm is proposed which involves a secure channel that uses a 256-bit key. The proposed system is designed to secure the entire data traffic for the transmission of remotely

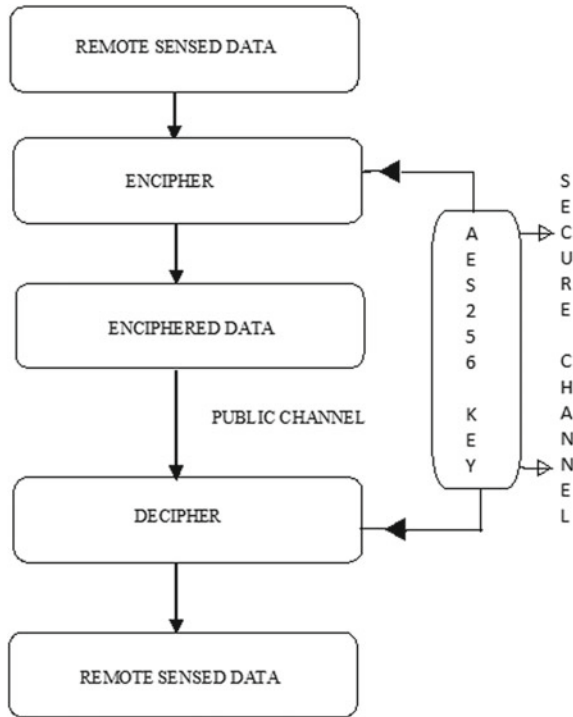


Fig. 1 Proposal framework

sensed data. The idea of implementing the scheme is laid out in the following flow chart (as shown in Fig. 1).

2.1 Remote Sensed Data

Remote sensing is the process of obtaining information about any object or surface phenomenon without establishing any physical contact with the object or its surface. The data acquisition for a particular object or phenomenon uses the principle of inverse problem. Sometimes, the parameters of the object or phenomenon in the particular area of interest cannot be measured directly. As a workaround to this, we acquire those parameters who have substantial relation with our area of interest, so that the required parameters can be estimated with maximum degree of certainty.

2.2 *Encipher*

During the data transmission to Earth stations, some losses are bound to happen due to channel noises and if the loss also incurs during the decryption, it would lead to an erroneous data keeping in mind the sensitivity of remote sensed data.

This would lead us to a trade-off between data security and data loss. Hence, the main idea is to use such algorithm that has the minimal chances of alterations being created in the image simultaneously ensuring that data security is also not jeopardised. This led us to AES-256 which successfully obliterates all the above concerns. In the proposed scheme, we intend to not only encrypt the message but instead the entire data traffic taking place between the satellite and the station.

The security of encryption can further be enhanced by adding several hashing methodology or coupling it with some competent encryption algorithm. However, this would depend on the design and demand of the remote sensing system as when and where it is implemented giving it a flexibility of adding or removing the security layers without much concern.

The encryption time for AES-256 is bit higher than its 128-bit counterpart due to more number of rounds involved in the encryption process. But it is still less than the other cryptographic modules. The security provided by AES-256 is much more higher than all of the existing modules and this has been justified in the later sections of the paper.

2.3 *Decipher*

This stage of the scheme will take place at the receiver end which can be earth station or any intermediate satellite. Decrypting follows the inverse algorithm of encryption (shown in Fig. 2). Since AES uses symmetric key, so we need to establish a secure channel to transfer the AES-256 key to the encipher stage of the scheme, making the keyless prone to attacks.

The transfer of key over a secured channel is to protect the key from any side-channel or known-key attacks. The key must be chosen wisely as side-channel or known-key attacks can some time bring down the brute-force attack time on any data communication channel. Let us assume if the attacker knows that the key bits are related to each other in some particular fashion then it becomes a relatively easy job as the brute force will have lesser domain of keys to search with. Therefore, if the keys are hashed and any related key is blocked by system in initial stage only, then it will almost completely nullify the chances of side-channel or known-key attacks. Moreover, the transfer of key over secured channel may be a subject to the data sensitivity of the information being transferred. The secured channel may or may not be a necessary subject for lesser sensitive or non-military grade data and may depend upon the demand and implementation of the project undertaken.

The following flow chart describes the AES-256 algorithm (shown in Fig. 2).

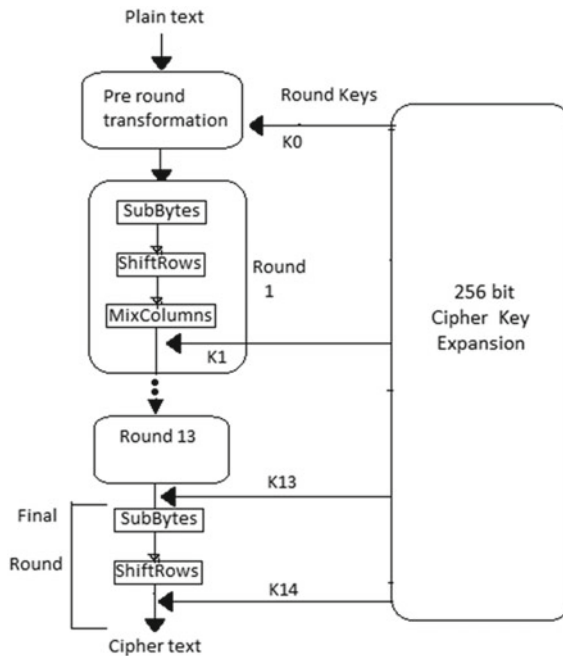


Fig. 2 Flow chart describing AES algorithm

3 Why AES-256 for Remote Sensing System

The proposed idea would affect the entire mechanism of remote sensing in a very significant way. As there are extra stages involved in the proposed mechanism, it is bound to create some overheads which may somewhat affect the efficiency of system, but addition of these extra security layers will weigh out the cons with it.

Many cryptographers and cryptanalysts term AES-256 bit as oversecure and rule out the possibility of 128-bit variant to be cracked and term it as safe for upcoming years. However, the term safe is relative and has no assurance of lifetime. Given the recent advancement if quantum computers come into existence, then it would not be an impossible task to carry out an attack on 128 bit but that would still take time even for 128-bit variant to be cracked. But this possibility is also ruled out because 256-bit variant will require more than twice the number of time taken to crack 128 bit. Remote sensing may have data security levels ranging from moderate to top secret-level. The proposed idea is robust enough to be implemented as a single module as well as holds the power to be coupled with other add-ons like hashing and hybrid cryptography to completely nullify the security vulnerabilities.

3.1 Implications on Remote Sensed Data

In the proposed work under the section of encipher, it is discussed that how the data sensitivity of remotely sensed data becomes inevitable when it comes to encryption and decryption and thus we chose AES to be best suited for the job. In comparison with AES-128, AES-256 stands out in terms of security but at the cost of speed and memory. Speed and memory overhead for AES-256 do not show a vast system requirement in terms of chip area and hardware installations. This trade-off between security and time and memory can easily be narrowed down with help of efficient designs and it is not of much concern given the recent technological advancements in chip fabrications. The implementation of design will take into account the demand of the remote sensing system in determining the security layers as mentioned in the proposed idea. For example, if the remotely sensed data has a higher military grade information, then it is wiser to add more security layers where security supersedes all the other factors, whereas this may not be the case for a topographic mapping satellite as it would require fewer layers of security.

3.2 Attacks and Vulnerability

AES has a comparatively simple algebraic framework. This property of AES was targeted in the XSL attack done by Nicolas Courtois and Josef Pieprzyk to expose the less complexity of the nonlinear components of the cipher [15]. However, the other follow-up papers have shown that the attack suggested by Courtois is impractical and hypothetical. The concern was raised by several other developers of the competing algorithm over the simplicity of nonlinear components of the Rijndael. The developer of competing algorithm Twofish, Bruce Schneier wrote that successful attacks on Rijndael would be developed someday but did not believe that anyone would ever discover the attack that could reveal the Rijndael traffic.

Some related-key attacks done by famous cryptographer Adi Shamir against AES-256 required 2^{39} time for key recovery in nine-round version, 2^{45} time for ten-round version [16] with a stronger related-key attack than the previous one. This becomes ineffective against AES-256 with 14 rounds and this can be completely nullified if full AES is implemented where the generation of the related key is constrained by proper software implementation. The key-recovery attacks and known-key distinguishing attacks on full AES have shown huge time and memory requirements to be utilised for the lower variants of the AES but that would still take billions of years to do brute-force attack and on the implementable hardware. These attacks are thus ruled out as these are impractical in nature and hence pose no threat to AES with 256-bit key [17].

Certain side-channel attacks have claimed to expose the AES-256 key but that requires the running of programme on the same node as the data. These attacks can

also be nullified by having built-in hardware instructions for AES that will prohibit and safeguard against any timing-based side-channel attacks on AES.

4 Implementation Results

In the implementation of the project, we have coupled AES-256 with SHA 256 to perform hashing of the 256-bit key. Hashing improves the key strength and minimises any related-key attacks. The key entered by the user is padded to 256-bit key by using the hashing algorithm. We have coupled AES with $(2, n)$ share cryptography. In this algorithm, we have divided the encrypted image into two shares (shown in Fig. 4). If either of the image share is lost, the remaining would be useless. The image shares would not reveal any data unless the other share is combined with it to produce the original encrypted image. This image sharing technique was suggested by Naor and Shamir [18]. The project was implemented to demonstrate the flexibility of the proposed algorithm as it can support hybrid cryptography and can be coupled with other algorithms also for better security layers in the communication channel.

The execution time for encryption using 256-bit key is more as compared to for the 128-bit variant due to more number of rounds involved in encryption of 256-bit variant. This leads to a trade-off between security and efficiency. However, the overheads involved in the process can be minimised by having upgraded hardware support which may incur some cost overheads leading to a triangular trade-off between cost, speed and security.

The proposed idea was implemented using python and python packages coupled with hashing (SHA 256) and visual cryptography by image sharing. The programme was implemented on Intel i3-380M CPU with 3M cache and clock rate of 2.53 GHz aided by 6 GB RAM. Figures 3, 4 and 5 are the results of our implementation of the proposed system. Figure 3 represents the input image which is encrypted using 256-bit key and image sharing technique represented as in Figs. 4 and 5 shows the decrypted image.

5 Conclusion

In this paper, we have tried to layout and implement a model for a secure data transmission in remote sensing by using Advanced Encryption Standard with 256-bit key size. We have tried to show why AES-256 is still now one of the safest cryptographic algorithms for the use of encryption and decryption. This moreover justifies the implementation of AES-256 in our proposed scheme to not only make the data encrypted, but we try to bring the entire data traffic happening between the various layers and nodes of remote sensing under the umbrella of AES-256 keeping in mind the data sensitivity and efficiency of the communication in accordance with the proposed idea for remote sensing.

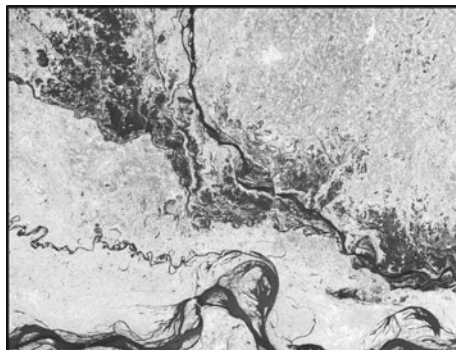


Fig. 3 Input image



Fig. 4 Encryption procedure with image sharing



Fig. 5 Decrypted image

References

1. Tea Area Development and Management using RS and GIS | National Remote Sensing Centre, https://nrsc.gov.in/Tea_area
2. L. Hathway, *National Policy on the Use of the Advanced Encryption Standard (AES) to Protect National Security Systems and National Security Information* (2003)
3. B. Mohammed, L. Mourad, C. Mehdi, Meteosat images encryption based on AES and RSA algorithms. *Int. J. Adv. Comput. Sci. Appl.* **6** (2015)
4. L. Yin, J. Zhao, Y. Duan, Encryption scheme for remote sensing images based on EZW and Chaos, in *2008 The 9th International Conference for Young Computer Scientists* (2008)
5. M. Usama, M. Khan, K. Alghathbar, C. Lee, Chaos-based secure satellite imagery cryptosystem. *Comput. Math Appl.* **60**, 326–337 (2010)
6. X. Zhang, G. Zhu, S. Ma, Remote-sensing image encryption in hybrid domains. *Opt. Commun.* **285**, 1736–1743 (2012)
7. F. Ahmed, D. Ganesh, Implementation of encryption and watermarking algorithm for remote sensing image. *Int. J. Eng. Comput. Sci.* (2016)
8. N. Nazmudeen, F.J. Farsana, Satellite image security improvement by combining DWT - DCT watermarking and AES encryption. *Int. J. Adv. Comput. Res.* **4** (2014)
9. R. Km, S. Shankar, Secure image transformation using remote sensing encryption algorithm. *Int. J. Sci. Eng. Res.* **5** (2014)
10. P. Gunavathy, A. Kannan, Segmentation and encryption of satellite images using stream cipher algorithm. *Int. J. Comput. Sci. Mob. Comput.* **5**, 743–750 (2016)
11. P.R. Banu, *Satellite On-Board Encryption* (2007)
12. NovelSat | NovelSat ProtCASTER, <http://novelsat.com/novelsat-protcaster/>
13. China-based hacking campaign is said to have breached satellite, defense companies, <https://www.cnbc.com/2018/06/19/china-based-hacking-breached-satellite-defense-companies-symantec.html>
14. E. Rosenfeld, US weather systems hacked by Chinese: Report, <https://www.cnbc.com/2014/11/12/chinese-hack-us-weather-systems-satellite-network-washington-post.html>
15. B. Schneier, Crypto-Gram. 15 Sept 2002 - Schneier on Security, <https://www.schneier.com/crypto-gram/archives/2002/0915.html>
16. A. Biryukov, O. Dunkelman, N. Keller, D. Khovratovich, A. Shamir, Key Recovery Attacks of Practical Complexity on AES Variants With up to 10 Rounds (2010)
17. J. Goldberg, AES Encryption isn't cracked, <https://blog.agilebits.com/2011/08/18/aes-encryption-isnt-cracked/> (2011)
18. M. Naor, A. Shamir, Visual cryptography, advances in cryptology, in *Eurocrypt '94 Proceeding LNCS*, 950. 1–12 (1995)

Grasp-Pose Prediction for Hand-Held Objects



Abhirup Das, Ayon Chattopadhyay, Firdosh Alia and Juhi Kumari

Abstract Robotic grasp prediction of small objects is a recent challenge, incorporating concepts such as depth perception, geometric matching, deep feature-engineering, etc. However, the entirety of the existing literature focuses on the usage of simple two-fingered robotic hands to grasp handheld objects. Our paper promotes the use of five-fingered humanoid hands for better grasping and proposes a novel hand grasp-pose prediction model that learns features present in small objects to predict appropriate poses that can be assumed by the fingers of the robotic hand to facilitate complex grasp prediction, as observed in humans. We make use of the coil-100 dataset that comprises images of various handheld objects and employ a supervised learning mechanism to perform multi-label classification over a set of eight selected grasp-poses.

Keywords Convolutional neural networks · Grasp estimation · Multi-label classification

1 Introduction

Robots today can be programmed to perform a multitude of tasks, ranging from assembling complex machinery, spinning a top on a sword's edge [1], to being neutrally controlled for aiding people suffering from paralysis [2]. This task of robotic grasping

A. Das (✉) · A. Chattopadhyay · F. Alia · J. Kumari
Institute of Engineering and Management, Salt Lake Electronics Complex,
Gurukul Y-12, Sector V, Kolkata 700091, India
e-mail: abhirupdas.iem@gmail.com

A. Chattopadhyay
e-mail: ayon01051998@gmail.com

F. Alia
e-mail: aliafrd8@gmail.com

J. Kumari
e-mail: namratajuhisingh@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_19

remains a problem even today, especially for previously unknown objects, involving several stages of planning, perception, and control. Most existing approaches make use of complex handcrafted feature-engineering techniques that are difficult to implement. Also, the existing literature is dominated by papers using two-fingered robotic arms. These allow no freedom in terms of grasp-pose, thus limiting the effectiveness of the grasping mechanism. Our paper promotes the use of multi-fingered humanoid arms and proposes an end-to-end model capable of feature extraction from 2-D images of objects and predicting appropriate poses for grasping handheld objects.

Much of the existing literature on robotic grasp planning consider 3-D model data a necessity for feature extraction, often stressing over the reconstruction of a complex three-dimensional model. A number of approaches, such as friction-cones [3] pre-store and restored primitives [4], and form force-closure technique [5] use the later mentioned approach. Practically, obtaining such 3-D data, or reconstruction of 3-D models is extremely difficult, cumbersome and time-consuming, especially when input modalities like RGB-D camera outputs are introduced. Features like texture, illumination, and occlusions are challenges most handcrafted feature engineering techniques struggle to overcome. We explore the use of deep-learning-based feature engineering mechanisms, which allow models to learn high-level representations of 2-D object data to estimate appropriate hand-poses for solving the robotic-grasp planning problem (Fig. 1).

Deep-learning approaches have become extremely popular, owing to their ability to learn useful features from data directly, without the need for any extra preprocessing. The earliest works by Hinton [8] have showed models being able to learn pen-stroke features from handwritten numerals. Later uses of convolutional neural networks were able to learn object parts when trained on natural images. In fact, these approaches are not restricted to the visual domain, but rather have shown to learn useful features for a wide range of domains, such as audio [9] and natural language data [10]. In all of these approaches, the use of CNNs has been notable. Contrary to handcrafted approaches like HOG, SIFT, etc. where a lot of expert knowledge is

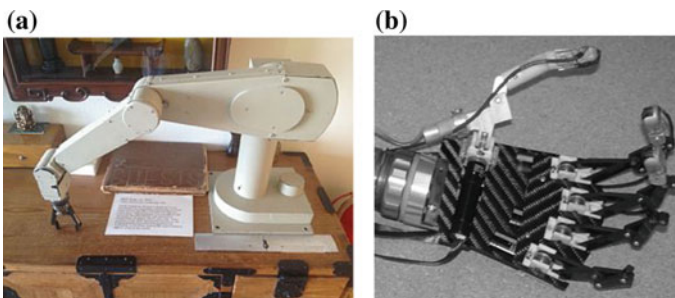


Fig. 1 a Two-fingered arm [6] and b five-fingered arm [7]

necessary, CNNs allow models to learn high-level representations of data and engineer features specifically suited to the task at hand. convolutional neural networks in word-spotting and handwriting recognition learn complex curvatures and spatial features, while those used in natural-image object detection, or action recognition learn to deal with illumination, textures as well as depth, thus automating the entire feature extraction process. Neural networks allow modelling of extremely complex features; CNNs with deep architectures and large number of layers model such features using millions of parameters that affect the learning process between individual layers. Deeper layers use the output (feature maps) of previous layers and model a better representation of those features, for example, using corners, lines, and curves to represent rectangles and squares, which in turn can be used to represent complex shapes like the letter p in the English alphabet. Some of the most popular of such CNN architectures include the VGGNet, ResNet, InceptionNet: winners of the ILSRVC ImageNet object-recognition challenge in consecutive years. Such architectures use Deep CNNs for feature extraction, after which the features are fed into feed-forward neural networks which use features modeled to predict labels (in our case, specific grasp-poses). Our model makes use of such a data-driven model like VGGNet as its basic architecture. There have been approaches such as [11] that focus on the grasp-planning for multi-fingered humanoid arm. These have the potential to grasp objects of various sizes under various situations, as well as execute a variety of grasp types: fingertip grasp, envelope grasp, etc. These grasping techniques have been modeled after human hands and are often referred to as precision grasping. Existing ideas estimate singular poses, contrary to human tendency to opt for a variety of possible gestures for single objects. In case of a can, one can grab it from the top or even envelope it with all fingers. We explore a similar pose-estimation strategy for grasp-execution by using supervised learning to perform a multi-label classification problem and ranking predicted labels, i.e., grasp-poses according to predicted scores.

2 Related Works

First, we discuss works comprising some of the perception and learning-based approaches on robotic grasping. Grasp is defined in most of the works as end-effort which achieves partial or complete closure of any object. Implementing this is a challenging work as there can be a lot of configurations in which an object can be gripped. Earlier works [12–14] were more focused about the testing of the grasping and remaking the closures and grasps which fulfill the condition of the object and can also be grasped efficiently. They did that by generating some quality scores. However, the recent works have refined the definition [15] and consider that they have proper knowledge of the object, shape, and other properties of the object in their concern.

The whole method of grasping can be quite a tedious task to do at first hand. The process of fast synthesis of 3-D models to determine the grasps that can facilitate to pick them up is still an active area of research [16–18] and the more modern methods use physical simulations to find the optimal grasping positions. However, there are some works like, [19] where spaces for graspable objects are defined, and then new objects are mapped on those spaces to discover the grasps which is an innovative approach to the problem in concern. Another constraint for grasping is that in real world there will not be complete information available about the 3-D object, hence, we mostly use half information from certain sensors like colour or depth cameras and tactile sensors which makes the problem more challenging [20] as the information source is not efficient enough, thereby the model must be made so that it can work on noisy data. Some implement the challenging task by estimating the pose of the known object [21, 22] and then apply it to the full-model algorithm on previous results, while others avoid this assumption. Such algorithms, however, make use of assumptions such as considering the total object to belong to a single primitive set of shapes [23, 24], or consider it to be planer [25] while there are some other works which provide better results and high accuracy for some particular cases, like grasping the corners of a towel [26]. Hence, the hand-coded grasping algorithms failed in many cases of real-world objects and were difficult and time consuming too.

Machine learning methods took the flaws of the previous algorithms and tried to solve it with more accuracy by replacing the hand-coded rules for grasping the real-world object with actual learning [8, 27–29]. Recent works have developed it further so that the grasping pose could be found out even if the object is partly occluded [30] or presented in an unknown pose [31].

Deep learning has showed its potential for learning useful features directly from the data for a wide variety of tasks. The work of Hinton et al. [8] showed that a deep network trained on the handwritten digits can learn the features correspond to the pen-strokes. Works using localized convolutional features by Lee et al. [28] showed that these types of deep networks can learn features related to object parts when given natural images as the only training data. This demonstrated that these features learned by these types of networks can adapt to the nature of the data given. These networks were not only limited in the fields of visual features but they showed prominent results in the field of audio [32, 33] and natural language data [34]. However, most of the works in deep learning was then done for improving the classification accuracy, and only few of the works involved solution to the detection problems [35, 36].

Multi-label classification has generally been an important problem for image recognition for a long time. The work of Duygulu et al. [37] focused on identifying object present in certain sub-sections of an image which involved multi-class classification of the segregation of the output tensor. The work by Wei et al. [38] showed that weights from pretrained networks with single-class classification had good success in multi-label classification and Wang et al. [39] also showed that by using recurrent neural networks to capture semantic label, and we can improve the multi-label classification accuracy. The paper by Daniel et al. [2] implemented a convolutional neural network model (CNN) to perform multi-label classification of Amazons satellite images. They implemented this on three different DCNNs (Deep

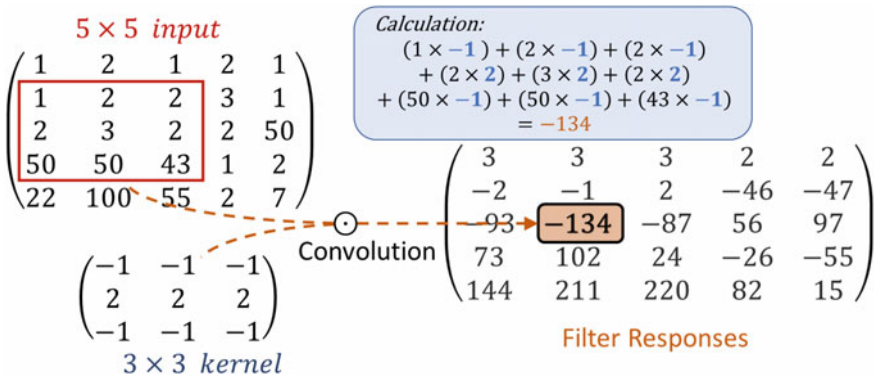


Fig. 2 The convolutional operation: feature extraction using a sliding 3 × 3 window on a binarized image. We use the example of a kernel suited for horizontal-edge detection. Observe the relative spikes in values in the fourth-row resulting from the edge (low-to-high junction) in the same row in the input image

CNNs) and achieved a 0.91 F-score. They did the multi-label classification by replacing the Softmax layer and adding a Sigmoid activation at the end so that each of the labels will have a score between 0 and 1, thereby showing the confidence of the network about the various possibilities that can occur. We, in this paper, take this concept and choose to generalize it further (Fig. 2).

3 Proposed Methodology

3.1 Requirements

We also discuss the VGG16 architecture and it’s use in transfer learning for the purpose of domain adaptation. At last we describe the eight hand-poses representing the eight classes our model is classified over.

3.1.1 Convolution Neural Networks

The feature descriptor we have used in our model are convolutional neural networks (CNNs) which output a condensed feature vector consisting of all the information present in the original image, and whose capacity can be controlled by varying the breadth and depth of the network. Convolutional neural networks (CNNs) constitute the class of the models which can compensate for the lack of information about the

object by having prior knowledge about other types of similar objects [1–5]. There have been several works using CNNs as their feature descriptor because of their way of preserving the prior knowledge about the objects and then applying their learned parameters for the classification, recognition or detection of other objects related to them. It showed exceedingly good results in the ImageNet challenge [6] by winning the top-five test error rate of 15.3% in ILSVRC-2012. Convolutional neural network in general have other some features because of which they are extensively used in all sorts of detection and recognition. In CNNs, various types of filters can be employed, namely pre-specified unstructured filters such as random filters [7, 8], and filters that are learned in a supervised [7, 9] or in an unsupervised manner [7, 10]. They also facilitate the use of a various types of non-linearities to be used in the network such as hyperbolic tangents [8, 9], rectified linear units (ReLU) [11, 12] and logistic sigmoid [13, 14] and the filters, non-linearities, and the pooling operators which can be applied on them can be changed for other networks.

3.1.2 Transfer Learning

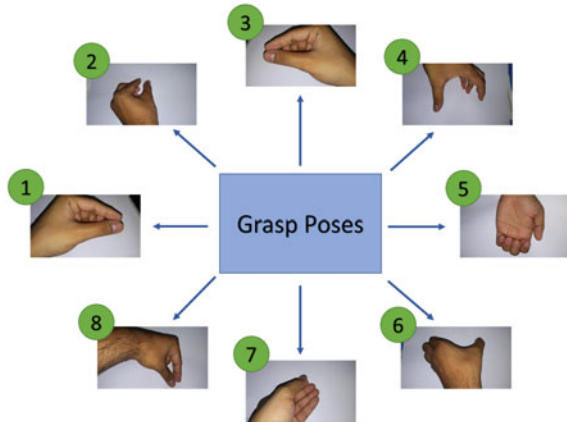
The VGG-Net inspired baseline architecture learns a high-level feature representation over a total of 10 convolutional layers, obtaining a feature map of 256 channels. Training a model to extract features over such deep-architectures turn out to be extremely time consuming as well as computationally expensive. For this reason, we opted for a transfer-learning approach for finetuning a trained model. A major assumption in machine-learning problems is that training and future model data need to belong to the same feature space and have similar data distribution. Transfer learning, however, allows a model pretrained on a source dataset \mathcal{D}_S (on which the model was initially trained) and learning task \mathcal{T}_S to be finetuned on a different target dataset \mathcal{D}_T and learning task \mathcal{T}_T to help learn the predictive function $f_t(\cdot)$ in \mathcal{D}_T using the knowledge in \mathcal{D}_S and \mathcal{T}_S , where $\mathcal{D}_S \neq \mathcal{D}_T$ and $\mathcal{T}_S \neq \mathcal{T}_T$. More specifically, our proposal makes use of inductive transfer learning, since the original pretrained VGG model focused on performing 1000 class single-label object classification, while our task \mathcal{T}_T is to perform eight-class multi-label classification. We use the pretrained model weights from the VGG-16 architecture used on the ImageNet dataset, classifying 1000 non-overlapping class categories for training and finetuning over the images in our coil-100 dataset, since both datasets deal with real-world images of objects, and thus, share some properties in their respective feature-spaces (Fig. 3).

3.1.3 Hand-Poses

Our pose-estimation model tries to predict the required optimal hand pose for the input handheld objects by classifying the poses w.r.t the objects. To implement this, we consider eight hand poses which are discussed as follows.

Pinching: We are considering pinch as the hand movement required to grab thin surfaces like the handle of a cup. We are further dividing the pinch movement into

Fig. 3 The eight grasp-poses chosen for prediction: in order, (1) left pinch, (2) middle pinch, (3) right pinch, (4) lift from top, (5) bucket hold, (6) grab from front, (7) flat pinch, and (8) pinch from top. *Source* Hand poses are of the Authors



three different ways, according to the orientation of the object in the training and testing images Left, Middle, and Right. If the objects handle (like the handle of the cup) is facing in the right-hand side then we use the right pinch, else if the handle is facing leftwards then we use the left pinch. Middle pinch is a complicated motion as the pose is not only meant for picking the cups with the handle occluded, but also for other objects which have similar thin body as a cups handle facing toward a camera. Another type is flat-pinch that focuses on objects such as a saucer or any other flat object.

Lifting: Lifting an object is also a very general movement that we do with most of the handheld object that we see around us in our everyday life. However, even though the object is handheld we cannot always lift the object from the top or the probability of picking an object from top may not be always valid. Few items like fish can and toy car can easily be moved/picked up using this motion. *Basket Hold:* Bucket-hold is limited but needed pose for many day-to-day handheld objects like baskets or a bag.

Grabbing: Many objects (especially those having cylindrical shape) can be picked up in this way, provided the objects are only handheld. *Flat Pinch:* This applies to objects like chocolate bars or saucers. *Pinch from top:* Pinch from top is viable pose, but it depends on both the size of the object and size of the hand. Hence, this pose is not perfectly reliable for all the cases, but for the surfaces which have a small surface area and can be held in hands, pinch from top can be used.

3.2 The Base Network

Our model is based on the VGG-16 architecture. The VGGNet is a deep network with 13 convolutional layers of filter size 3×3 and stride of 1. In the default architecture, 4 pooling operations are used to down-sample the size of input to a 16×16 -dimension

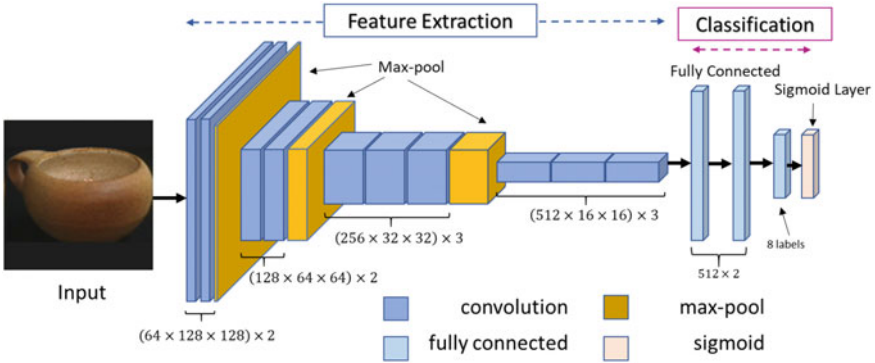


Fig. 4 Our base network, inspired by VGG-Net

feature map of 512 filters. Our proposed model applies 10 convolution layers and 3 similar max-pooling layers on input image of dimension $128 \times 128 \times 3$ (RGB image) to obtain a final feature map of size $16 \times 16 \times 512$. We exclude the final pooling layer as seen in VGG-16. The 512-filter feature map is then fed-forward into two 3 fully connected networks of size 512, 512, and 8 with ReLU activation in between (Fig. 4).

3.3 Multi-label Classification

Traditional object detection algorithms performing multi-class classification utilize the Softmax function to calculate losses. The output from the final layer is passed through it. However, the *Softmax* loss function is only suitable for single-label probability prediction in supervised learning problems.

$$\text{Softmax}(z_j) = \frac{e^{z_j}}{\sum_{k=1}^K e^{z_k}} \quad (1)$$

where z_j denotes individual prediction scores in the output prediction vector of length K . In our approach, we pass the output into a *Sigmoid* layer $\sigma(\cdot)$. The resulting values can then be used to predict the appropriate poses. Poses with scores above a certain threshold λ are considered appropriate outputs.

$$\sigma(z_j) = \frac{1}{1 + e^{-z_j}} \quad (2)$$

With the outputs L_p obtained from the σ layer, we train the classifier using our ground-truth labels L_g with a L2 loss regression function:

$$L2 = \frac{1}{\mathcal{N}} \sum_{i=1}^K (l_{g_i} - l_{p_i})^2 \quad (3)$$

where \mathcal{N} is the normalisation constant, and l_{g_i} and l_{p_i} are the i th instances of L_g and L_p , respectively.

In this paper, we have proposed a novel approach toward pose recommendation by using various handheld objects and labeling them according to the ways we can grasp the objects and then training them accordingly using the tools we have mentioned above.

4 Experiments

4.1 Dataset

We have labeled the database called Columbia Object Image Library [32] which was used in a real-time 100 object recognition system. However, we have used this object as a pose recommendation system by labeling the database according to the poses in which they can be picked up. COIL-100 consists of 7200 color images of 100 objects (72 images per object), all of which are handheld. The objects have a wide variety of shapes making the training more generalized and also have complex geometric characteristics. We have taken eight specific poses according to which objects can be held, which are: (1) Left Pinch (2) Middle Pinch (3) Right Pinch (4) Lift from Top (5) Bucket Hold (6) Grab from Front (7) Flat Pinch (8) Pinch from Top.

Lift from top is used for lifting most of the objects. Flat pinch is used for thin objects whereas grab from front is not used for thin objects, it is used when there is a huge surface in the front mainly cylindrical surface. Left pinch, middle pinch, and right pinch are mainly used for holding cup handles. These pinches are also used for holding object which is as thin as cup handle. When the object is not symmetric, we consider mainly lift from top. Hence, whichever poses are best for lifting an object, we select those poses by filling one in those columns in excel sheet of that particular object. Thus, we manufacture 7200 labels selected carefully by human experts, each label being a one-hot vector of length 8 that denotes 1 for all appropriate poses, and 0 for others.

4.2 Training

In the training, we have used pretrained VGG-16 model. VGG-16 has five stages of convolution, each consisting of convolution and pooling layers. First stage takes input of size 128×12 consisting of three channels, and then convolves over it using a kernel of size 3×3 increasing the number of channels to 64. Two such layers are

used and then pooling is done to give a resulting volume of 64×64 . Second stage also has two layers of pooling which increases the number of channels to 128 and after pooling the resultant volume is $32 \times 32 \times 128$. Then in the next stage, three convolution layers are used with the same kernel and then pooling is done which again reduces the image to half giving a volume of size $16 \times 16 \times 256$. In the fourth stage also, there are three levels which give a feature map of 8×8 as output with 512 channels and here we take the weights and flatten it to a vector of size $1 \times 1 \times 8$ and give a sigmoid activation to the flatten layer unlike the ReLU activation present in the other layers.

It is important to note that to get suitable results, it is important to adjust the learning rates of the entire deep network for proper training [40]. We set the learning rate of the deeper classification layers that outputs the predicted pose-labels to 0.01 while the learning rate for the feature-extraction unit is set to 0.002. This prevents the model from drastically changing the parameters in the shallower part of the network. As we observe, this slightly improves the overall training time. We choose the ADADELTA optimizer during training.

4.3 Results

After training the end-to-end framework for 500 epochs and batch size of 32 on a laptop-computer with an Intel i7 7th gen processor and a Nvidia GTX-1050 graphics card with optimizations for 6 h, we obtained an accuracy of 67.5%. However, the results varied according to conditions of the object, i.e., for the images with less complicated geometry like can it produced more accuracy then the other objects. The results can, however, be improved by using more labels and by using a bigger database. Here, we have labeled the data using human experts, but a larger verified database will improve the results. As observed, interestingly, in Fig. 5, the outputs of the cup with the broken handle shows low predicted outputs for ‘left-pinch’.

5 Conclusion

Even though work in the field of grasp recognition is extensive, a vast majority deals with 3-D model analysis, reconstruction, etc. We propose a simpler CNN architecture to predict a set of appropriate poses for grasp-planning process. We consider the problem of estimating poses and grasp planning for humanoid arms which are more diverse in their applications over only a few classes, encouraging the design of similar models classifying over a much larger and diverse range of classes, while realizing the power of convolutional neural networks as feature descriptors in standard deep neural network models. Our proposal would suitably be a part of an entire pipeline of methodologies involving object localization, depth perception, grasp estimation

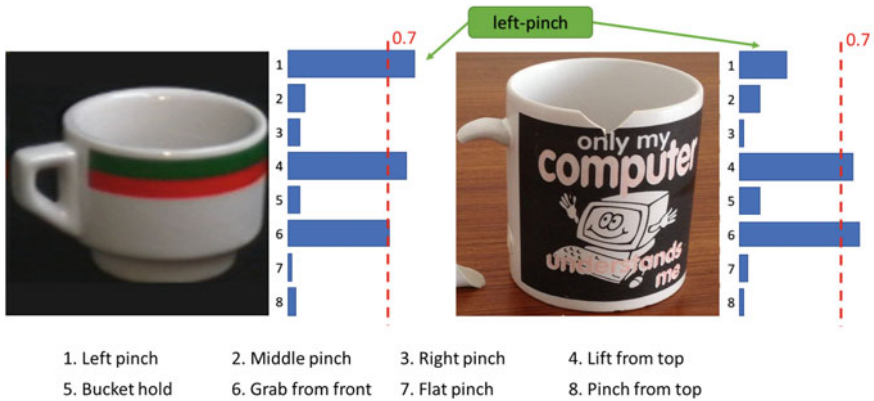


Fig. 5 The diagram portrays two pictures of cups from the COIL dataset (left) and a test image (right). The *left-pinch* label is observed to have a significant difference in prediction labels during inference. The red-line denotes the threshold; labels with *Sigmoid* outputs greater than the threshold are considered

and planning, and actual grasping that allows five-fingered robotic arm to grab small objects like cups, saucers, fruits, pencils, etc.

References

1. T. Shin-ichi, M. Satoshi, *Nipponia* **13**, 7 (2000)
2. L.R. Hochberg, D. Bacher, B. Jarosiewicz, N.Y. Masse, J.D. Simeral, J. Vogel, S. Haddadin, J. Liu, S.S. Cash, P. van der Smagt et al., *Nature* **485**(7398), 372 (2012)
3. M.T. Mason, J.K. Salisbury Jr, Compliance and force control for computer controlled manipulators. *IEEE Trans. Syst. Man, Cybern.* (1985)
4. A.T. Miller, S. Knoop, H.I. Christensen, P.K. Allen, in *IEEE International Conference on Robotics and Automation. Proceedings, ICRA'03*, vol. 2 (IEEE, 2003), pp. 1824–1829
5. A. Bicchi, V. Kumar, in *ICRA*, vol. 348 (Citeseer, 2000), p. 353
6. A. Reinhold, Category: victor scheinman (2013). https://commons.wikimedia.org/wiki/File:Scheinman_MIT_Arm.agr.jpg#filelinks
7. D. Reid, Southampton remedi hand beats hollywood (2005). <http://www.technovelgy.com/graphics/content05/southampton-remedi-hand2.jpg>
8. G.E. Hinton, R.R. Salakhutdinov, *Science* **313**(5786), 504 (2006)
9. S. Hershey, S. Chaudhuri, D.P. Ellis, J.F. Gemmeke, A. Jansen, R.C. Moore, M. Plakal, D. Platt, R.A. Saurous, B. Seybold, et al., in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2017), pp. 131–135
10. W. Yin, H. Schütze, B. Xiang, B. Zhou, arXiv preprint [arXiv:1512.05193](https://arxiv.org/abs/1512.05193) (2015)
11. V. Lippiello, F. Ruggiero, B. Siciliano, L. Villani, *IEEE/ASME Trans. Mechatron.* **18**(3), 1050 (2013)
12. K. Lakshiminarayana, in *ASME Paper*, 78-DET-32 (1978)
13. V.D. Nguyen, in *Proceedings of 1986 ACM Fall Joint Computer Conference* (IEEE Computer Society Press, 1986), pp. 129–137
14. J. Ponce, D. Stam, B. Faverjon, *Int. J. Robot. Res.* **12**(3), 263 (1993)
15. A. Rodriguez, M.T. Mason, S. Ferry, *Int. J. Robot. Res.* **31**(7), 886 (2012)

16. M. Dogar, K. Hsiao, M. Ciocarlie, S. Srinivasa, Manipulator motion planning using flexible obstacle avoidance based on model learning. *Int. J. Adv. Robot. Syst.* (SAGE Publications Sage UK: London, England) (2012)
17. C. Goldfeder, M. Ciocarlie, H. Dang, P.K. Allen, in *ICRA'09. IEEE International Conference on Robotics and Automation* (IEEE, 2009), pp. 1710–1716
18. J. Weisz, P.K. Allen, in *2012 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2012), pp. 557–562
19. F.T. Pokorny, Y. Bekiroglu, J. Exner, M. Björkman, D. Kragic, in *RSS 2014 Workshop: Information-Based Grasp and Manipulation Planning*, (2014)
20. J. Bohg, A. Morales, T. Asfour, D. Kragic, *IEEE Trans. Robot.* **30**(2), 289 (2014)
21. A. Collet, D. Berenson, S.S. Srinivasa, D. Ferguson, in *IEEE International Conference on Robotics and Automation, ICRA'09* (IEEE, 2009), pp. 48–55
22. C. Papazov, S. Haddadin, S. Parusel, K. Krieger, D. Burschka, *Int. J. Robot. Res.* **31**(4), 538 (2012)
23. J.H. Piater, Learning visual features to predict hand orientations (2002)
24. D.L. Bowers, R. Lumia, *IEEE Trans. Fuzzy Syst.* **11**(3), 320 (2003)
25. A. Morales, P.J. Sanz, A.P. Del Pobil, in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2 (IEEE, 2002), pp. 1711–1716
26. J. Maitin-Shepard, M. Cusumano-Towner, J. Lei, P. Abbeel, in *2010 IEEE International Conference on Robotics and Automation (ICRA)* (IEEE, 2010), pp. 2308–2315
27. P. Viola, M. Jones, in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001*, vol. 1 (IEEE, 2001), p. I
28. H. Lee, R. Grosse, R. Ranganath, A.Y. Ng, in *Proceedings of the 26th Annual International Conference on Machine Learning (ACM, 2009)*, pp. 609–616
29. R. Socher, B. Huval, B. Bath, C.D. Manning, A.Y. Ng, in *Advances in Neural Information Processing Systems*, (2012), pp. 656–664
30. J. Glover, D. Rus, N. Roy, in *Proceedings of Robotics: Science and Systems IV* (Zurich, Switzerland, 2008), pp. 278–285
31. R. Detry, E. Baseski, M. Popovic, Y. Touati, N. Kruger, O. Kroemer, J. Peters, J. Piater, in *IEEE 8th International Conference on Development and Learning, ICDL 2009* (IEEE, 2009), pp. 1–7
32. H. Lee, P. Pham, Y. Largman, A.Y. Ng, in *Advances in Neural Information Processing Systems*, (2009), pp. 1096–1104
33. A.R. Mohamed, G.E. Dahl, G. Hinton et al., *IEEE Trans. Audio Speech Lang. Process.* **20**(1), 14 (2012)
34. R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, P. Kuksa, *J. Mach. Learn. Res.* **12**(Aug), 2493 (2011)
35. M. Osadchy, Y.L. Cun, M.L. Miller, *J. Mach. Learn. Res.* **8**(May), 1197 (2007)
36. Y. LeCun, F.J. Huang, L. Bottou, in *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2004*, vol. 2 (IEEE, 2004), pp. II–104
37. P. Duygulu, K. Barnard, J.F. de Freitas, D.A. Forsyth, in *European Conference on Computer Vision* (Springer, 2002), pp. 97–112
38. Y. Wei, W. Xia, J. Huang, B. Ni, J. Dong, Y. Zhao, S. Yan, arXiv preprint [arXiv:1406.5726](https://arxiv.org/abs/1406.5726) (2014)
39. J. Wang, Y. Yang, J. Mao, Z. Huang, C. Huang, W. Xu, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2285–2294 (2016)
40. A. Sharma, et al., in *2015 13th International Conference on Document Analysis and Recognition (ICDAR)* (IEEE, 2015), pp. 986–990

A Multi-level Polygonal Approximation-Based Shape Encoding Framework for Automated Shape Retrieval



Sourav Saha, Soumi Bhunia, Laboni Nayak,
Rebeka Bhattacharyya and Priya Ranjan Sinha Mahapatra

Abstract Due to exponential growth of image data, textual annotation of every image object has failed to provide practical and efficient solution to the problem of image mining. Computer vision-based methods for image retrieval have been drawing significant attention with the progress in achieving fast computational execution power. One of the fundamental characteristics of an image object is its shape which plays a vital role to recognize the object at a primitive level. We have studied the scope of a shape descriptive framework based on multi-level polygonal approximations for generating features of the shape. Such a framework explores the contour of an object at multiple approximation stages and captures shape features of varying significance at each approximation stage. The proposed algorithm determines polygonal approximations of a shape starting from coarse-level to more refined-level representation by varying number of polygon sides. We have presented a shape encoding scheme based on multi-level polygonal approximation which allows us to use the popular distance metrics to compute shape dissimilarity score between two objects. The proposed framework when deployed for similar shape-retrieval task demonstrates fairly good performance in comparison with other popular shape-retrieval algorithms.

Keywords Shape retrieval · Shape descriptor · Shape analysis · Shape recognition

S. Saha (✉) · P. R. S. Mahapatra
Kalyani University, Kalyani, India
e-mail: souravsaha1977@gmail.com

P. R. S. Mahapatra
e-mail: priya_cs@yahoo.co.in

S. Bhunia · L. Nayak · R. Bhattacharyya
Institute of Engineering and Management, Kolkata, India
e-mail: soumi.mecheda@gmail.com

L. Nayak
e-mail: nayak.laboni@gmail.com

R. Bhattacharyya
e-mail: rebeka.bhattacharyya@gmail.com

1 Introduction

With the progress in digital imaging technologies, the ever-growing volume of image data has propelled the demand of an automated system for retrieving images of interest from a repository which are relevant to a user-specific query. Searching images using their textual annotations is a subjective process and is not practical for large databases. Shape is regarded as one of the most potential criteria to identify an image object. In general, shape is an intuitive characteristic of an object for which all kinds of mathematical frameworks fall short to find a proper crisp description [1]. It is easy to perceive but difficult to find a precise descriptor for any shape because of its almost infinite variations while preserving the underlying identity. Shape has always been an important visual information attracting the attention of researchers over the past few decades specifically dealing with the problems of computer vision. With the ever-growing demand of automated emulation of a human vision system, research interest among today's scientists in computer vision domain has been gradually driven toward obtaining maturity in automated shape analysis from both theoretical and practical points of view. Based on a user survey [2], 71% of the users were found to be interested in recognizing objects by shape. The problem of shape-based similarity retrieval demands a solution for retrieving an ordered list of shapes from a repository which is similar to a given query shape, and the list is preferably ranked based on the degree of similarity with the query shape. The overall motivation of this research work is to concentrate on exploring contour of an image object at multiple approximation stages for shape-based object retrieval.

1.1 Related Work

Shape-based image retrieval has been receiving substantial attention of researchers since the advent of advanced digital image processing. One of the earliest efforts made in solving this task is on the basis of global features, such as invariant moments [3], geometric features [4], Zernike moments [5, 6], fusion of invariant moments and histogram of edge directions [7], and the angular radial transform descriptor [8]. In [9], a shape is partitioned into a set of closed contours and each contour is interpreted as a chain code for facilitating shape matching. Another successful approach for shape retrieval relies on the construction of attributed relational graph (ARG) [10, 11] by decomposing shape. An ARG of a shape is a fully connected graph, where a node denotes an interior region and edges manifest spatial relationships among the interior regions. For 2D shape matching, many approaches have been reported in the literature [2, 12]. Adamek and O'Connor proposed an effective shape descriptor that explores the degree of both concavity and convexity of the contour points [13]. They called it as multi-scale convexity-concavity (MCC) descriptor. Multiple scales are rendered using Gaussian kernels of varied widths for smoothing the contour. The curvature of a contour point is estimated by measuring the relative displacement with

reference to its location in the previous scale level. Later, the matching is performed by a dynamic programming approach. Belongie et al. [14] proposed the idea of *Shape Context* as a shape descriptor. In shape context, a histogram is linked to every contour point for marking the relative distribution of the remaining contour points. The matching operation can be executed by finding the correspondence between every pair of contour points. One of the most widely accepted shape descriptors for object retrieval is based on the curvature scale space (CSS) method proposed by Mokhtarian et al. [15, 16]. MPEG-7 community has recommended it as the standard for boundary-based shape description [17]. In this method, the boundary is gradually smoothed using a set of Gaussian kernels resulting in approximated curves (CSS contours) until the boundary becomes totally convex. For matching, only the maximal inflection points of CSS contours are explored. The repeated application of Gaussian smoothing filters leads to the understanding of the contour evolution from coarse-level representation to fine-grained representation. Motivated by this idea, we have proposed a multi-level polygonal approximation of a contour by varying the number of sides of the shape approximating polygon at each level.

1.2 Objective of the Proposed Work

In this paper, we have tried to develop a multi-level polygonal approximation framework based on suboptimal thick-edged polygonal approximation method as described in [18]. The multi-level approximation strategy represents the contour with different degrees of approximation achieved by varying the number of sides of a shape approximating polygon at each level (Fig. 1). To facilitate shape-based matching, a new shape encoding strategy along with an effective shape matching algorithm has been developed to quantitatively measure the shape similarity between two objects. The effectiveness of the proposed framework is shown by deploying it for the task of retrieving similar shapes with reference to a query object.

Rest of the paper is organized as follows. In Sect. 2, the proposed technique is detailed focusing on a special thick-sided polygonal approximation, shape descriptor feature vector generation, and shape similarity score evaluation strategy. The subsequent section to the detailed framework description has aimed to produce a report on how efficiently our framework has performed on shape-retrieval task by substantiating results. The concluding section summarizes the proposed work.

2 Proposed Framework

This section discusses our proposed framework in detail. The proposed model for shape retrieval initially extracts the contour of an image object using Moors neighborhood boundary tracing algorithm. In the proposed framework, a shape is approximated at multiple stages by varying the number of sides of the approximating

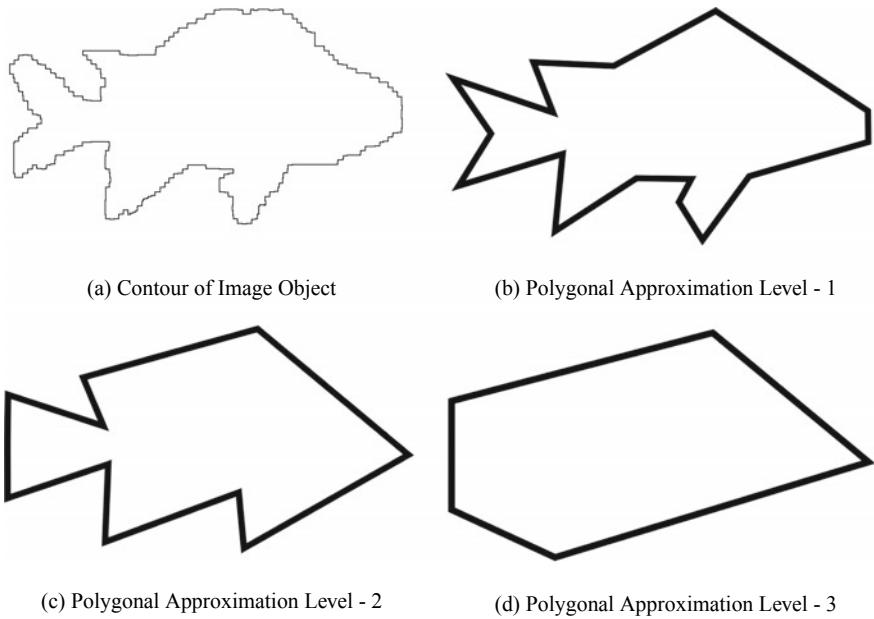


Fig. 1 Multi-level polygonal approximation

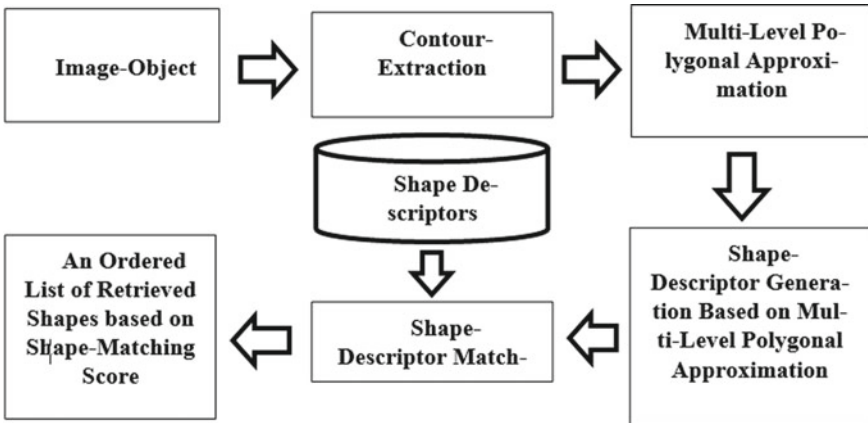


Fig. 2 Overall flow of the proposed scheme

polygon. For a polygonal approximation of 2D planar shape, we have applied suboptimal thick-edged polygonal approximation method based on the principle described in [18]. Figure 2 intuitively demonstrates the overall flow of the proposed strategy. The description of each step of the overall flow is detailed next.

2.1 *Thick-Poly-Line Approximation of Contour*

In discrete geometry, a poly-line is a connected series of line segments. The poly-line is extensively used for approximating a discrete curve with a polygon in order to represent its shape [19]. Given the specification of number of poly-line-segments for a digital curve, we have generated a suitable linear approximation of a curve using special thick-poly-line, wherein every line segment is suboptimally thick to approximate a curve segment based on the idea proposed in [18]. The proposed solution in [18] for polygonal approximation of a closed contour explores a unique idea of piece-wise thick-line-fitting to a curve based on digital geometry and uses greedy best-first heuristic strategy to repetitively split a curve so that the vertices of the approximating polygon can be obtained. Figure 3 shows the outcome of the thick-edged polygonal approximation method proposed in [18].

2.2 *Multi-level Polygonal Approximation*

One of the interesting advantages of our poly-line approximation method is that we can apply it for a closed discrete curve which leads to the generation of a shape

Fig. 3 Thick-edged polygonal approximation



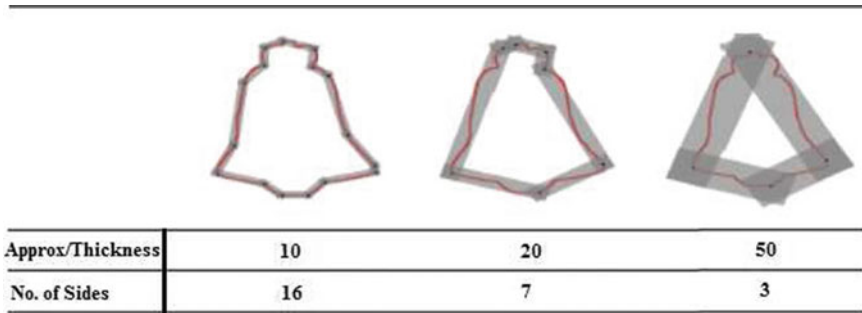


Fig. 4 Multi-level polygonal approximation

approximating polygon. In our proposed framework, users can flexibly specify the number of edges as an input parameter to obtain a thick-edged shape approximating polygon. The accuracy of the approximation increases if users specify more number of edges for approximating polygon. We assume that the degree of approximation varies inversely with the accuracy of shape approximation. This implies that the degree of approximation also varies inversely with the increase in the number of edges as evident in Fig. 5. We have utilized this characteristic to approximate a shape at various approximation levels by varying number of edges of approximating polygon. It helps us to explore the evolutionary nature of the approximating polygons in capturing shape-detailing characteristics at different levels (Fig. 4). The average thickness of the edge depends on the specified number of edges of the polygon. As we increase the number of edges, the average thickness of the edges starts decreasing (Fig. 4). The average thickness of an edge of the shape approximating polygon corresponds to the degree of approximation.

2.3 Shape Descriptor for Multi-level Polygonal Approximation

As discussed earlier, each approximation level corresponds to a specific number of edges of the approximating polygon. Multi-level polygonal approximations are carried out repeatedly by specifying more number of polygon edges at each iteration till the approximation error (i.e., average thickness of the edge) reaches a threshold value. At every stage of multi-level polygonal approximation, a shape descriptor feature vector is formed with respect to the corresponding shape approximating polygon. A special polygon-shape descriptive feature is formed based on its internal angles as illustrated below. While comparing two shapes at a specific level, the following sequence of steps is applied for forming shape descriptive features and these features are used for shape comparison. At each level, a shape dissimilarity score is computed based on Manhattan distance between two shape descriptive features.

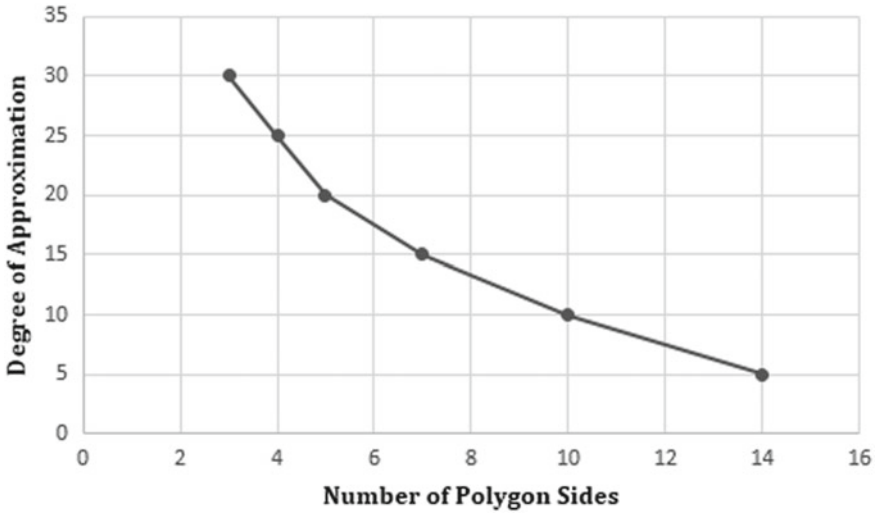
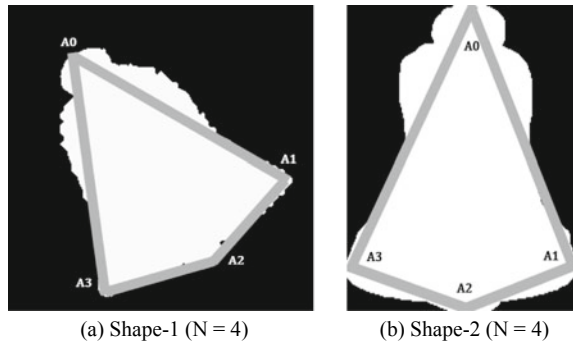


Fig. 5 Degree of approximation versus number of polygon sides

Fig. 6 Shape approximating polygon with sides = 4



These steps are illustrated below with reference to Fig. 6a, b. Furthermore, Fig. 7 and Table 1 illustrate the shape descriptor generation strategy and matching scheme with reference to a few more shapes.

2.3.1 Basic Principle of the Shape-Descriptor Formation and Comparison Strategy

- Step 1: The vertices of the shape approximating polygon are indexed clockwise.
- Step 2: Internal angle at each vertex (v_c) of the shape approximating polygon is computed based on the angle needed to rotate edge (v_c, v_{c+1}) counterclockwise to align with edge (v_c, v_{c-1}) keeping vertex v_c fixed. Equation 1 is used to obtain internal angle θ_c° at vertex (v_c) with respect to adjoining vertices








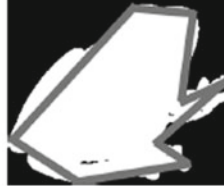
N	Image-1	Image-2	Dissimilarity Score
4			$(B_{41} - B_{42} / 4) + (B_{81} - B_{82} / 8) = 13.5$ $(F_{41} - F_{42} / 4) + (F_{81} - F_{82} / 8) = 9.8$
	$B_{41} = [31, 99, 102, 126]$	$B_{42} = [42, 88, 97, 131]$	
8			
	$B_{81} = [22, 27, 31, 49, 53, 93, 104, 157]$	$B_{82} = [23, 30, 42, 53, 55, 80, 105, 148]$	
4			
	$F_{41} = [63, 90, 99, 107]$	$F_{42} = [58, 92, 101, 107]$	
8			
	$F_{81} = [40, 51, 77, 81, 88, 99, 127, 153]$	$F_{82} = [35, 68, 70, 73, 78, 107, 132, 154]$	

Fig. 7 Feature vector formation at multiple stages

v_{c-1} and v_{c+1} .

$$\theta_{i,c} = \tan^{-1} \left(\frac{y_i - y_c}{x_i - x_c} \right) \times \frac{180^\circ}{\pi}, \quad i = c - 1, c + 1$$

$$f(\theta_{i,c}) = \begin{cases} 360^\circ + \theta_{i,c} & \text{if } \theta_{i,c} < 360^\circ \\ \theta_{i,c} & \text{otherwise} \end{cases} \quad (1)$$

$$\theta_c = (360^\circ + f(\theta_{c+1,c}) - f(\theta_{c-1,c})) \pmod{360^\circ}$$

Step 3: Equation 2 is applied on internal angles of a shape approximating polygon to compute a new set of angles. In case of Fig. 6a, the set will be $\{A0 = 31^\circ, A1 = 99^\circ, A2 = 126^\circ, A3 = 102^\circ\}$.

$$f(\theta_i) = \begin{cases} \theta_i^\circ & \text{if } \theta_i < 180^\circ \\ \theta_i^\circ - 180^\circ & \text{otherwise} \end{cases} \quad (2)$$

Step 4: The new set of computed angles at the previous step are sorted in ascending order. In case of Fig. 6a, the ordered sequence will be $\{A0 = 31^\circ, A1 = 99^\circ, A3 = 102^\circ, A2 = 126^\circ\}$. In case of Fig. 6b, the ordered sequence will be $\{A0 = 42^\circ, A1 = 88^\circ, A3 = 97^\circ, A2 = 131^\circ\}$. These sequences are used as feature vectors at a specific level determined by the number of sides of the shape approximating polygon.

Step 5: Manhattan distance between two shape representative feature vectors is computed and the distance is divided by the number of edges to obtain shape dissimilarity score at the respective level. The shape dissimilarity score for Fig. 6a, b is computed as 8.0, while the number of polygon edges is 4.

Step 6: The final shape dissimilarity score between two shapes is determined by adding dissimilarity scores computed at different levels, i.e., with varying edges (N) being 4, 8, etc. The final shape dissimilarity score for Fig. 6a, b is computed as 13.5 as shown in Fig. 7. Mathematically, the shape dissimilarity score between two shapes— S_1 and S_2 —can be computed based on Eq. 3 where $f_{j,k}^i$ denotes j th feature element of S_i at k th approximation level.

$$\text{Dissimilarity}(S_1, S_2) = \sum_{k=4}^N \left(\frac{1}{k}\right) \sum_{j=1}^k |f_{j,k}^1 - f_{j,k}^2| \quad (3)$$

3 Experimental Results and Discussion

For evaluating the performance of the proposed framework, experiments have been carried out on the publicly available MPEG-7 database [17]. The dataset comprises 1400 shapes grouped into 70 classes, and every class contains 20 similar shapes. Some shapes of the same class have gone through various view transformations in addition to scaling, rotation, and shearing processes. Figure 8 presents the top three retrieval results of the proposed framework based on a set of sample images from MPEG-7 test database. It is evident from the observation that the proposed framework is able to retrieve most relevant shapes against a query image within the topmost three retrievals depending on the illustrated similarity score ranking applied on respective shape descriptive feature codes. Table 1 presents the shape dissimilarity score matrix for two classes where each class contains two image objects. The results listed in Table 1 show that the dissimilarity score between any two objects belonging to the same class is smaller than the scores reported for the objects belonging to different classes.









































Query Image	Top Three Retrieved Images			Query Image	Top Three Retrieved Images		
 1				 6			
 2				 7			
 3				 8			
 4				 9			
 5				 10			

Fig. 8 Retrieval result

Table 1 Shape similarity score matrix

Figure	Bell-1	Bell-2	Frog-1	Frog-2
Bell-1	0.0	13.5	39.5	36.8
Bell-2	13.5	0.0	34.7	34.6
Frog-1	39.2	34.7	0.0	9.8
Frog-2	36.8	34.6	9.8	0.0

Figure 9 presents the execution time profile of the proposed framework generated based on varying sides of a 32-sided star polygon image object.

3.1 Metric for Evaluating Performance

In content-based image retrieval, the performance evaluation is difficult because of the varying subjectiveness of the manual assessment. Ideally, the merit of a shape-retrieval system relies on the relevance of the retrieved images which are ranked based

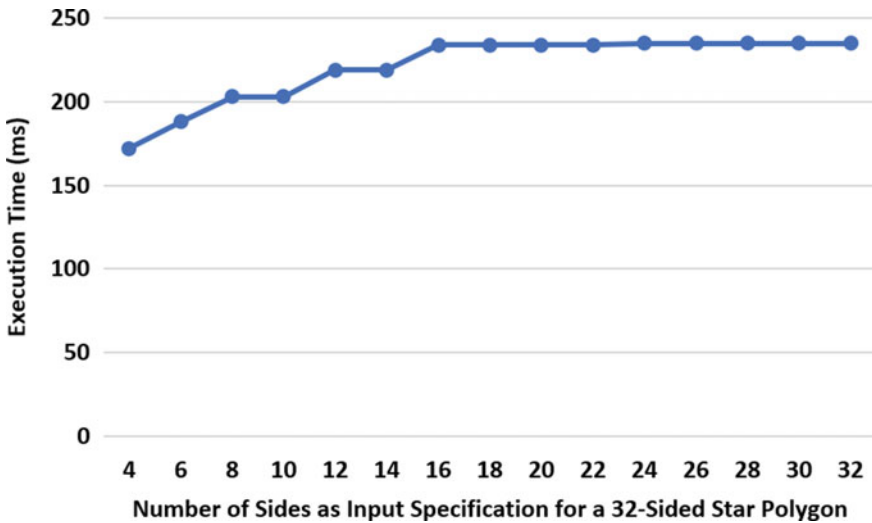


Fig. 9 Execution time profile

on some similarity score. However, many schemes for estimating the performance of such models have been attempted for making a comparative study of solutions. Probably, the widely used criteria for judging retrieval effectiveness have been *Bull's Eye Test* [17]. *Bull's Eye Test* provides us with a platform for comparing the effectiveness of an approach in comparison with other shape-retrieval approaches. Each shape of the dataset is compared with all other shapes, and among the top 40 retrieved shapes, we only consider the number of shapes retrieved by an applied model from original class of the query shape. The retrieval rate based on *Bull's Eye Test* against a query shape is computed as the ratio of the number of retrieved shapes from the class of the query shape to the total number of shapes originally included in the query shape's class. The overall *Bull's Eye Percentage (BEP)* score can be computed by averaging *BEP* scores obtained for different query shapes. Table 2 reports *BEP* scores of the proposed model in comparison with popular CSS scheme [15] with respect to the sample dataset listed in Fig. 8. Each class of image dataset has 20 members, and relevant retrievals correspond to the retrieved shapes of the query shape's class. The overall accuracy of the proposed framework is estimated as 85% (Table 2) according to *Bull's Eye Test*, which is comparable to the widely referred algorithm.

4 Conclusion

An effective shape descriptive framework called multi-level polygonal approximation is proposed for similar shape retrieval which explores various polygonal approximations of an object's contour while varying the number of edges of the shape

Table 2 Result: Bull's Eye Test

Sample no.	CSS algorithm [15]		Proposed	
	Retrievals	BEP score (%)	Retrievals	BEP score (%)
1	17	85	17	85
2	18	90	17	85
3	18	90	17	85
4	16	80	16	80
5	19	95	18	90
6	15	75	18	90
7	18	90	18	90
8	16	80	15	75
9	18	90	15	75
10	19	95	19	95

approximating polygon, and a special shape encoding scheme has also been developed to measure shape similarity between two objects. The performance of the proposed framework is experimentally found to be fair and comparable with widely referred algorithm. However, we need to further analyze the robustness of the proposed framework under many other plausible deformations of closed contours.

References

1. S. Marshall, Review of shape coding techniques. *Image Vis. Comput.* **7**(4), 281–294 (1989)
2. D. Zhang, G. Lu, Review of shape representation and description techniques. *Pattern Recogn.* **37**(1), 1–19 (2004)
3. M.K. Hu, Visual pattern recognition by moment invariants. *IRE Trans. Inf. Theor.* **8**, 179–197 (1962)
4. C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, W. Equitz, Efficient and effective querying by image content. *J. Intell. Inf. Syst.* **3**(3), 231–262 (1994)
5. A. Khotanzad, Y.H. Hong, Invariant image recognition by zernike moments. *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(5), 489–497 (1990)
6. Y.S. Kim, W.Y. Kim, Content-based trademark retrieval system using a visually salient feature. *Image Vis. Comput.* **16**(12), 931–939 (1998)
7. A.K. Jain, A. Vailaya, Shape-based retrieval: a case study with trademark image databases. *Pattern Recogn.* **31**(9), 1369–1390 (1998)
8. W.Y. Kim, Y.S. Kim, A new region-based shape descriptor. Technical Report ISO/IEC MPEG99/M5472, 1999
9. H.L. Peng, S.Y. Chen, Trademark shape recognition using closed contours. *Pattern Recogn. Lett.* **18**(8), 791–803 (1997)
10. E.G.M. Petrakis, Design and evaluation of spatial similarity approaches for image retrieval. *Image Vis. Comput.* **20**(1), 59–76 (2002)
11. E.G.M. Petrakis, C. Faloutsos, K. Lin, Imagemap: an image indexing method based on spatial similarity. *IEEE Trans. Knowl. Data Eng.* **14**(5), 979–987 (2002)
12. S. Loncaric, A survey of shape analysis techniques. *Pattern Recogn.* **31**(8), 983–1001 (1998)

13. T. Adamek, N.E.O. Connor, A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE Trans. Circ. Syst. Video Technol.* **14**(5), 742–753 (2004)
14. S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(4), 509–522 (2002)
15. F. Mokhtarian, M. Bober, *Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization*, vol. 25 (Springer Science & Business Media, Berlin, 2013)
16. F. Mokhtarian, A. Mackworth, Scale-based description and recognition of planar curves and two-dimensional shapes. *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(1), 34–43 (1986)
17. R. Ralph, Mpeg-7 core experiment. (1999). <http://www.dabi.temple.edu/~shape/MPEG7/dataset.html>
18. S. Saha, S. Goswami, P.R.S. Mahapatra, A heuristic strategy for sub-optimal thick-edged polygonal approximation of 2-d planar shape. *Int. J. Image Graph. Sig. Process. (IJIGSP)* **10**(4), 48–58 (2018)
19. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 3rd edn. (Prentice-Hall Inc, Upper Saddle River, NJ, USA, 2006)

A Hand Gesture Recognition Model Using Fuzzy Directional Encoding of Polygonal Approximation



Sourav Saha, Soma Das, Shubham Debnath and Sayantan Banik

Abstract *Sign Language* is the basic communication medium for interacting with deaf and dumb people. Hand gesture is usually regarded as one of the principal components of *Sign Language*. Each gesture has a unique meaning assigned to it. This paper proposes a framework for hand gesture-based *Sign Language* recognition from a video stream using computer vision approaches. The proposed framework segments hand-palm region in the video stream using background subtraction method and skin color models. Thereafter, a fuzzy directional encoding strategy is employed on the polygonal approximation of a gesture to form features which are used to train an ANN classifier for developing a gesture prediction model. This proposed model considers digits as a significant part of *Sign Language* and the model seems working with acceptable accuracy in predicting hand gestures representing various digits in *Sign Language*.

Keywords Gesture recognition · Sign Language recognition · Human computer interaction

1 Introduction

As per sociolinguistic survey, deaf and dumb of older generations used to be neglected in our society. In the modern society, the scenario has changed a lot and today's people have started realizing their potential and treating them as intelligent

S. Saha (✉)
Kalyani University, Kalyani, India
e-mail: souravsaha1977@gmail.com

S. Das · S. Debnath · S. Banik
Institute of Engineering and Management, Kolkata, India
e-mail: soma.das@iemcal.com

S. Debnath
e-mail: shubhamagain7@gmail.com

S. Banik
e-mail: baniksayantana19@gmail.com

as any other normal human being. But their inability of formal verbal communication is posing the main obstacle to lead a normal social life. Instead of using formal language of verbal communication, the deaf and dumb people need Sign Language for communicating their expressions. Sign Language is a visual form of communication which is normally defined as an organized collection of standardized gestures. Among various kinds of gestures, hand gestures are extensively used being the simplest constituents of Sign Language, and almost all the basic requirements for the communication through Sign Language can be fulfilled using hand gestures. Many researchers have attempted to recognize hand gestures for different languages through various techniques. In this paper, a computer vision-based framework is proposed to interpret hand gestures in a video stream representing various digits as symbols of Sign Language. Gesture recognition systems can be broadly categorized into two different approaches [1]. One of the categories uses gloves or external sensors attached to a user for capturing gestures, while other approaches capture gestures using a video camera and process them with computer vision techniques. The first category captures the gesture more precisely by recording signals emitted from sensors, but the system may not be comfortable, as most of the times the user needs to wear a complicated sensor framework. The second category has recently been widely explored due to easy availability of smartphones equipped with a high-resolution camera. A number of works can be reported on the development of hand gesture-based Sign Language recognition models from computer vision perspective. Zaki and Shaheen [2] introduce an ASL (American Sign Language) recognition system that uses principal component analysis (PCA), Kurtosis position, and motion chain codes as descriptors for gestures. They claim PCA's role in retaining configuration and orientation of hands, while dynamic gestures are dealt with Kurtosis position and motion chain codes. A recognition system for single-handed gestures of Persian Sign Language (PSL) is proposed by Karami et al. [3]. It employs 2D discrete wavelet transform with 9 levels of decomposition for each gesture image. The coefficient matrices obtained from wavelet decomposition tree are used as descriptors, and a multi-layer perceptron neural network (MLP-NN) is trained with these descriptors for classification. In an ASL alphabet recognition work by Munib et al. [4], Canny's edge detection algorithm is applied on gesture images to detect sharp intensity changes. The descriptors are prepared by applying Hough transform on edges which are further customized to recognize arbitrary gestures using a three-layer feed forward back propagation neural network. Rahaman et al. [5] proposed a technique for Bengali Sign Language (BSL) recognition that extracts the hand gesture from video stream based on hue and saturation value from HSI color model, and subsequently, a K-nearest neighbors (KNN) classifier is employed for recognition. Jasim et al. [6] used LDA and LBP to obtain the features from hand gesture which are later identified by the KNN algorithm. In [7], Yasir et al. also used LDA and ANN to recognize two-handed hand signs for BSL. Ayshee et al. [8] proposed fuzzy rules to develop a hand gesture recognition system for Bengali characters.

A desirable characteristic of a gesture recognition system is its ability to recognize gestures in real time. A real-time gesture recognition system needs to predict the gesture from video stream within the reasonable response time. Although there have

been many studies which are using computer vision for gesture recognition, a problem like significant computational time along with the accuracy issue in the presence of complex background has always been challenging for the algorithms. To use gesture recognition techniques in a real-time environment, it is necessary to reduce their computational time. Such reduction can be achieved by using simple feature vectors for gestures with simple prediction technique.

1.1 Objective of the Proposed Work

Our study presents a computer vision-based gesture recognition model that uses a special gesture encoding scheme based on fuzzy theory and applies a simple ANN classifier for Sign Language-specific gesture recognition. The proposed method is able to classify hand gestures corresponding to symbols of Sign Language from a video stream. The proposed framework extracts the foreground of the input video stream by subtracting background. A skin color model in combination with a few geometric criteria is used to detect the hand-palm region in the foreground image. The contour of the detected hand-palm area is approximated by a polygon, and subsequently, a fuzzy-directional encoding strategy is employed on polygon edges for generating a feature vector with respect to the gesture. These gesture representative feature vectors are used to train an ANN model to build a gesture classifier. During prediction, the ANN classifier is applied to a gesture representative feature vector for recognizing symbols of the Sign Language.

Rest of the paper is organized as follows. In Sect. 2, the proposed technique is detailed focusing on the technique for hand region extraction, fuzzy gesture descriptor for feature vector generation based on polygonal approximation, and ANN-based gesture classification strategy. The subsequent section to the detailed framework description has been aimed to produce a report on how efficiently our framework has performed on gesture recognition task. The concluding section summarizes the proposed work highlighting the pros and cons of the proposed model.

2 Proposed Model

The proposed framework is discussed elaborately in this section. The overall flow of the proposed model is depicted in Fig. 1. Following steps are required to identify a gesture in a video stream.

- Step 1: **Hand-palm Region Extraction:** The objective of this step is to segment the hand-palm region from a video frame.
- Step 2: **Polygonal Approximation:** The objective of this step is to obtain a polygonal approximation of the segmented hand-palm region.

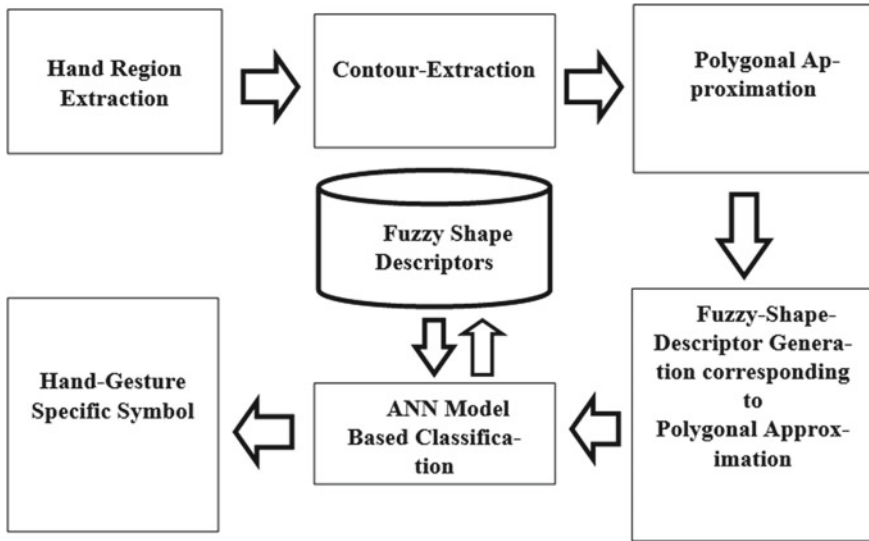


Fig. 1 Overall flow of the proposed model

- Step 3: **Fuzzy Rule-based Polygon Encoding**: The objective of this step is to form a feature code corresponding to the polygonal approximation of the segmented hand-palm region based on a special fuzzy rule. These feature codes are used for training an ANN model.
- Step 4: **ANN Classification**: The objective of this step is to classify a gesture representative feature code and predict its respective gesture class.

2.1 Hand-Palm Region Segmentation

Each frame of the input video stream is examined, and foreground objects are identified by subtracting the background. A foreground image frame comprising foreground objects is constructed, and each foreground object is further analyzed based on a skin color model along with a few geometric criteria for segmenting hand-palm region.

2.1.1 Background Subtraction

A background subtraction scheme is used to isolate foreground parts in a video stream. The foreground pixels at a time t are isolated in the current frame $C(t)$ by comparing with the background frame $B(t)$. The foreground frame $F(t)$ is obtained

by comparing each pixel $P[C(t), x, y]$ at location (x, y) of $C(t)$ with the spatially corresponding pixel $P[B(t), x, y]$ of $B(t)$. In terms of mathematical formulation, it is expressed as $P[F(t), x, y] = |P[C(t), x, y] - P[B(t), x, y]|$ where $F(t)$ represents pixel difference matrix at time t . A foreground pixel in $F(t)$ at (x, y) is identified as prominent pixel if $P[F(t), x, y]$ is found to be larger than a threshold. The threshold is computed through the widely used OTSU method [9]. The background frame is updated to adapt with a slow but gradual change in background. The following equations are used for obtaining updated background frame $B(t + 1)$ and prominent foreground region $M(t)$ at time t .

$$P[M(t), x, y] = \begin{cases} 1, & \text{if } P[F(t), x, y] > \text{threshold} \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

$$P[B(t + 1), x, y] = \begin{cases} 0.9 \times P[B(t), x, y] + 0.1 \times P[C(t), x, y], & \text{if } P[M(t), x, y] = 0 \\ P[B(t), x, y], & \text{otherwise} \end{cases} \quad (2)$$

2.1.2 Hand Region Localization

The foreground frame is processed to identify probable hand region using skin color-based filtering. In our proposed framework, a green wristband is used to isolate hand palm from the forearm. We have applied skin color-based filtering on the foreground image which results in three disconnected probable regions—(a) facial region, (b) hand-palm region, and (c) forearm regions. Thereafter, facial and forearm regions are eliminated based on (a) the aspect ratio of the bounding box, (b) the proportion of area coverage relative to entire frame area, and (c) location of individual centers of skin regions with reference to the center of the combination of three disconnected skin regions. The overall process is illustrated in Fig. 2.

Many skin color detection algorithms are available on the basis of the characterization of color spaces like RGB, HSV, YUV [10]. The effectiveness of each color space depends on its robustness to the variation of lighting and the ability to distinguish skin color pixels in images from a complex background. For an effective skin pixel detection, we use multiple threshold-based approaches considering a combination of RGB, HSV, and YUV color spaces to benefit from the characteristic of each color model. Prior to the deployment of the model, the threshold values are primarily computed with the help of the color histogram by exploring the boundary of manually extracted skin regions. These threshold values are applied to the foreground image to extract skin regions.

The *RGB color model* represents the colors that are in the red, green, and blue planes and does not isolate the luminance from the chrominance components, thereby making it a poor choice for color-based pixel detection. However, following threshold values [11] are used in RGB model to detect skin region $S(t)$.

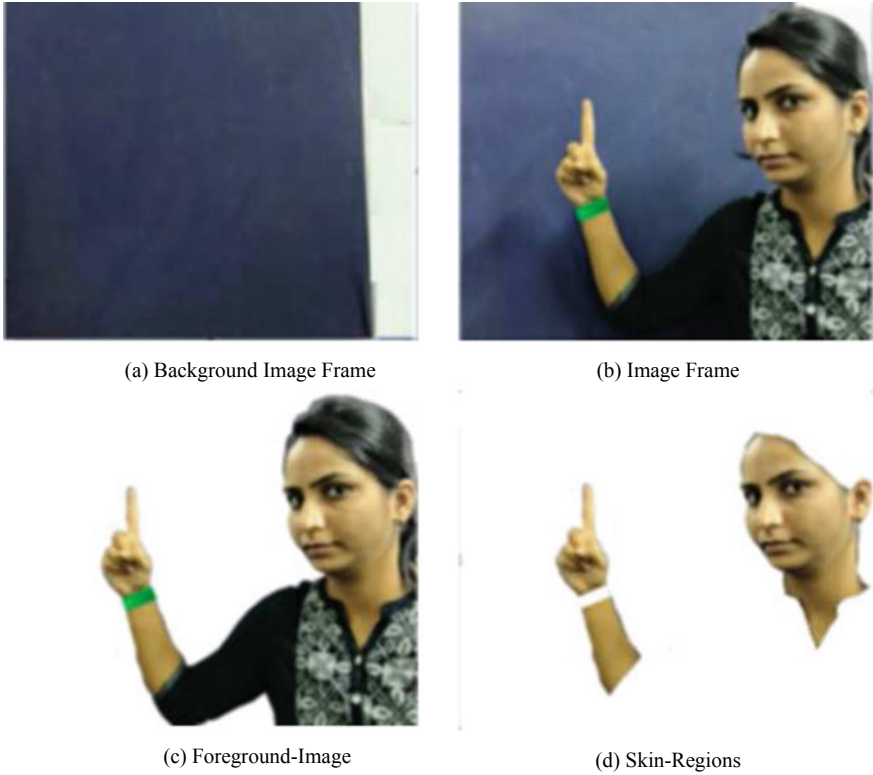


Fig. 2 Hand-palm region extraction

$$P[S(t), x, y] = 1 \text{ if } \begin{cases} R[F(t), x, y] > 95, G[F(t), x, y] > 40, B[F(t), x, y] > 20 \\ \max(R, G, B) - \min(R, G, B) > 15 \\ |R - G| > 15, R > G, R > B \end{cases} \quad (3)$$

The *HSV color model* is described by hue, saturation, and value. The hue plane indicates the color itself, whereas the saturation plane represents the brightness as a combination of the illumination and hue. Sobottka and Pitas [12] proposed threshold values as stated below to identify the skin regions.

$$P[S(t), x, y] = 1 \text{ if } \begin{cases} 0.23 < S < 0.68 \\ 0 < H < 50 \end{cases} \quad (4)$$

The *YUV color model* is used in some TV systems, such as Phase Alternation Line (PAL) and National Television System Committee (NTSC). The luminance plane (Y) is isolated from the chromatic planes (U,V) for reducing the effect of illumination changes. The following equation shows the range for detecting skin colors [13].

$$P[S(t), x, y] = 1 \text{ if } \begin{cases} 65 < Y < 170 \\ 85 < U < 140 \\ 85 < V < 160 \end{cases} \quad (5)$$

2.2 Polygonal Approximation

The approximation of arbitrary two-dimensional curves by polygonal figures is an imperative technique in digital image processing. A digital curve can be effectively simplified by polyline without loss of its visual property. Techniques for polyline approximation of digital curve have been driving interest among the researchers for decades [14]. We have approximated the shape of hand-palm region based on the thick-sided polygon approximation method as proposed in [15]. The contour of the hand-palm region is extracted using Moor's neighborhood principle [9] before applying the approximation method and Fig. 3 shows the result of the polygon approximation of a simple gesture. Subsequently, the gesture-approximating polygon is encoded using a special fuzzy rule characterizing orientations of its sides.

2.3 Fuzzy Rule-based Polygon Encoding

In the proposed framework, we have applied an overlapping triangular fuzzification rule for encoding the orientation of a polygon edge with respect to a reference frame (Fig. 4). All the vertices of a gesture-approximating polygon are indexed in a clockwise manner starting from the rightmost vertex of the bottom-most polygon edge as shown in Fig. 4. An i th polygon edge is represented as edge (v_i, v_{i+1}) where v_i acts as its tail vertex. An edge is indexed based on the index of its tail vertex. This implies that i th edge has i th vertex as its tail. The orientation of each edge is determined in terms of its deviation relative to its previous edge as illustrated below.



Fig. 3 Polygonal approximation

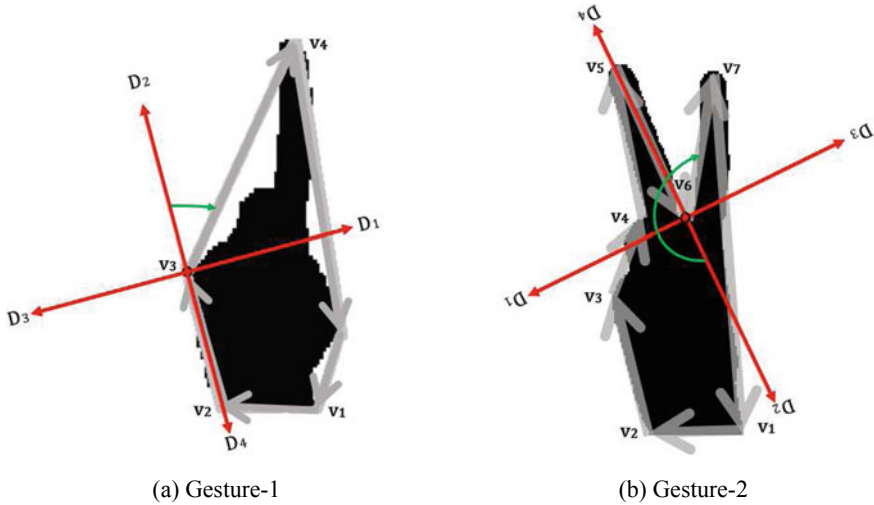


Fig. 4 Determining orientation of a polygon edge

Basic Principle of Determining Orientation of a Polygon Edge:

- Step 1: Generate a reference frame having four quadrants based on four directional vectors $\{D_1, D_2, D_3, D_4\}$ as shown in Fig. 5. The reference frame is placed at each polygon vertex in such a way that if index finger of the left hand is kept facing toward direction D_2 , then left thumb points to direction D_1 .
- Step 2: Polygon edges are assumed to be vectors arranged in a clockwise direction as shown in Fig. 4 to enclose the respective gesture.
- Step 3: In order to obtain orientation of edge (v_i, v_{i+1}) , the reference frame's origin is placed at v_i in such a way that direction D_2 gets aligned with previous edge (v_{i-1}, v_i) facing toward the direction of edge (v_{i-1}, v_i) and also ensures that direction D_1 is clockwise oriented toward the right of direction D_2 .
- Step 4: The orientation of edge (v_i, v_{i+1}) is measured in terms of the clockwise turning angle of direction D_2 to make it aligned with the direction of edge (v_i, v_{i+1}) . This implies that the clockwise angular deviation of direction D_2 toward edge (v_i, v_{i+1}) gives us the orientation of edge (v_i, v_{i+1}) relative to its previous edge. In Fig. 4, the orientation of edge (v_3, v_4) is 45° in case of Gesture-1, whereas the orientation of edge (v_6, v_7) is 225° in case of Gesture-2.

Fuzzification of Orientation of a Polygon Edge: We have considered four triangular fuzzy membership functions with respect to four directions, namely D_1, D_2, D_3, D_4 . These directions are used as the basis of our triangular fuzzy membership functions (Fig. 5). The triangular fuzzy membership function is a function of a vector, x , and depends on scalar parameters a, b , and m , as given by Eq. 6. Table 1 depicts the indi-

Table 1 Fuzzy membership functions

	a	m	b
μ_{D_1}	-90	0	90
μ_{D_2}	0	90	180
μ_{D_3}	90	180	270
μ_{D_4}	180	270	360

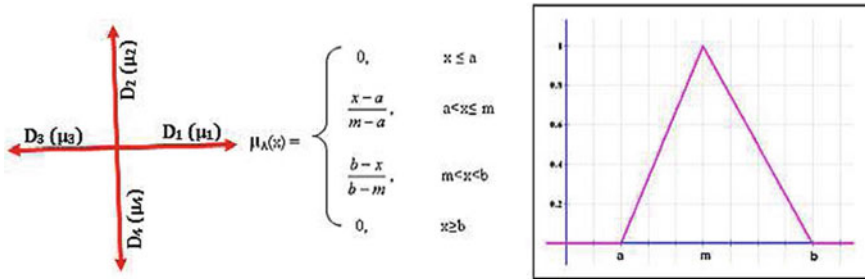


Fig. 5 Triangular fuzzy membership function

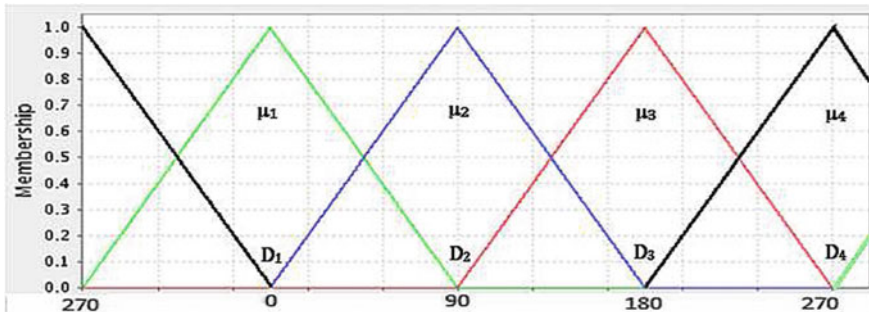


Fig. 6 Overlapped fuzzy membership functions for four different orientations

vidual parameters for each membership function, whereas Fig. 6 diagrammatically shows the overlapped nature of membership functions with respect to the four directions in our reference frame. We generate a special fuzzy descriptor for a polygon edge based on its orientation. The orientation of a polygon edge (say x) is fuzzified using four fuzzy membership functions, and the sequence of values— $\{\mu_{D_1}(x), \mu_{D_2}(x), \mu_{D_3}(x), \mu_{D_4}(x)\}$ —forms a fuzzy descriptor for the respective polygon edge. For example, the fuzzy descriptor for edge (v_3, v_4) of Gesture-1 in Fig. 4 with orientation 45° is $\{0.5, 0.5, 0.0, 0.0\}$.

$$\mu_{D_i}(x) = \begin{cases} 0 & x \leq a \\ \frac{x-a}{m-a} & a < x \leq m \\ \frac{b-x}{b-m} & m < x \leq b \\ 0 & x \geq b \end{cases} \quad (6)$$

Gesture-Approximating Polygon Encoding and Feature Formation for ANN Model: The fuzzy descriptor of an edge of a gesture-approximating polygon is assigned an index as same as the index of the edge. Thereafter, a chain of fuzzy descriptors is formed by ordering them from the first index to the last index. This chain of fuzzy descriptors is used as a final gesture descriptor feature code for encoding the corresponding gesture. Figure 7 illustrates the feature formation strategy for a gesture. An ANN model is trained with these gesture descriptor features, and later, we classify a gesture on applying the trained ANN model to a gesture descriptor feature code.

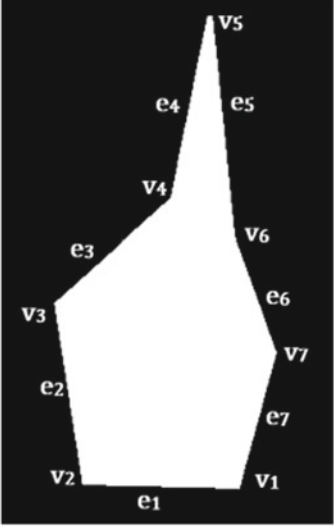
Gesture-Polygon	Edge	Fuzzy-Descriptor
	1	{0.16, 0.84, 0.00, 0.00}
	2	{0.12, 0.88, 0.00, 0.00}
	3	{0.38, 0.62, 0.00, 0.00}
	4	{0.60, 0.00, 0.00, 0.40}
	5	{0.00, 0.22, 0.78, 0.00}
	6	{0.86, 0.00, 0.00, 0.14}
	7	{0.61, 0.39, 0.00, 0.00}
Feature Code: [0.16, 0.84, 0.00, 0.00, 0.12, 0.88, 0.00, 0.00, 0.38, 0.62, 0.00, 0.00, 0.60, 0.00, 0.00, 0.40, 0.00, 0.22, 0.78, 0.00, 0.86, 0.00, 0.00, 0.14, 0.61, 0.39, 0.00, 0.00]		

Fig. 7 Fuzzy gesture descriptor feature code

3 Experimental Results and Discussion

In order to evaluate the performance of the proposed gesture recognition model, experiments have been conducted on a synthetic dataset. The dataset consists of video clips of 100 gestures grouped into 10 classes. Each class represents a specific digit and contains video clips of 10 similar gestures. Some of the gestures are shown in Fig. 8 along with their polygonal approximations. Table 2 lists the performance of the proposed model in terms of recognition accuracy of the ANN model for ten different gesture categories. In Table 2, (i, j) th cell shows what percentage of gesture- i is predicted as gesture- j by the ANN classifier. The values listed along the diagonal of the confusion matrix denote the true prediction ability of the classifier. The overall accuracy of the proposed model is observed as 96% which is reasonably acceptable as compared to existing models. The experiment is run on an Intel Core i7 3.0GHz processor with 8 GB RAM. The average running time is observed as 5 frames (320×240) per second. We have also tested our model under dynamic gesture transition, wherein the classifier decision is taken on two-third majority voting per 15 frames. In case of two-third majority voting, we examine responses of the classifier for 15 consecutive frames, and if any response appears more than 9 times as output, we consider the response as the predicted gesture. The overall accuracy of our model while applied on dynamic gesture recognition from real-time video streaming is found to be 93% with response time being nearly 3 s.

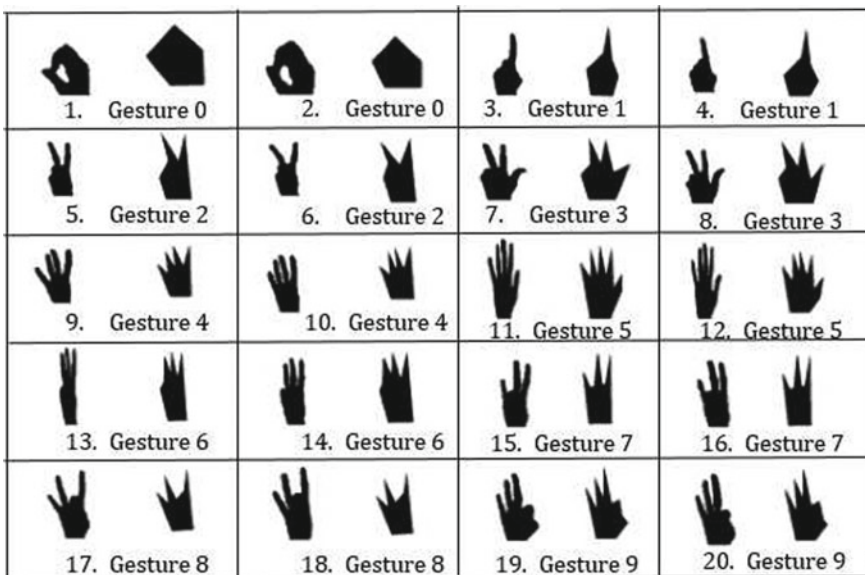


Fig. 8 Gesture dataset and respective polygonal approximation

Table 2 Gesture recognition accuracy (%)

Gesture	0	1	2	3	4	5	6	7	8	9
0	98.00	0.00	0.00	0.00	0.00	2.00	0.00	0.00	0.00	0.00
1	2.00	96.00	2.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
2	0.00	3.00	97.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
3	0.00	0.00	2.00	95.00	3.00	0.00	0.00	0.00	0.00	0.00
4	0.00	0.00	1.00	3.00	96.00	0.00	0.00	0.00	0.00	0.00
5	1.00	0.00	0.00	0.00	0.00	99.00	0.00	0.00	0.00	0.00
6	0.00	0.00	0.00	0.00	0.00	0.00	96.00	04.00	0.00	0.00
7	0.00	0.00	0.00	0.00	0.00	0.00	04.00	96.00	0.00	0.00
8	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	97.00	03.00
9	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	03.00	97.00

4 Conclusion

This paper presents a computer vision-based framework for hand gesture recognition from a video stream. The hand-palm region in a video stream is primarily segmented based on a background subtraction and a skin color model. A polygonal approximation of hand gesture in combination with a fuzzy encoding scheme is used to generate gesture representative feature set for developing an ANN model-based gesture classifier. The proposed framework focuses on a few gestures of Sign Language representing digits, and its ability to interpret those hand gestures accurately in a video stream is fairly acceptable. Nevertheless, there are challenges which are to be addressed to make the proposed model effectively applicable for dynamic sign language recognition in real time, considering all kinds of symbols even in the presence of complex background under unconstrained natural condition.

Acknowledgements We hereby give consent to the use of all the photographs which are presented in the article revealing the identity of a person for publication.

References

1. S.S. Rautaray, A. Agrawal, Vision based hand gesture recognition for human computer interaction: a survey. *Artif. Intell. Rev.* **43**(1), 1–54 (2015)
2. M.M. Zaki, S.I. Shaheen, Sign language recognition using a combination of new vision based features. *Pattern Recogn. Lett.* **32**, 572–577 (2011)
3. A. Karami, B. Zanj, A.K. Sarkaleh, Persian sign language (psl) recognition using wavelet transform and neural networks. *Expert Syst. Appl.* **38**, 2661–2667 (2011)
4. Q. Munib, M. Habeeb, B. Takruri, H.A. Al-Malik, American sign language (asl) recognition based on hough transform and neural networks. *Expert Syst. Appl.* **32**(1), 24–37 (2007)

5. M.A. Rahaman, M. Jasim, M.H. Ali, M. Hasanuzzaman, Real-time computer vision-based Bengali sign language recognition, in *2014 17th International Conference on Computer and Information Technology (ICCIT)*, IEEE (2014), pp. 192–197
6. M. Jasim, M. Hasanuzzaman, Sign language interpretation using linear discriminant analysis and local binary patterns, in *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*, IEEE (2014), pp. 1–5
7. R. Yasir, R.A. Khan, Two-handed hand gesture recognition for Bangla sign language using LDA and ANN, in *2014 8th International Conference on Software, Knowledge, Information Management and Applications (SKIMA)*, IEEE (2014) pp. 1–5
8. T.F. Ayshee, S.A. Raka, Q.R. Hasib, R.M. Rahman, M. Hossain, Sign language recognition for Bengali characters. *Int. J. Fuzzy Syst. Appl.* **4**, 1–14 (2015)
9. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 3rd edn. (Prentice-Hall Inc, Upper Saddle River, NJ, USA, 2006)
10. A.-T.Z. Hamid, R.W.O. Rahmat, M.I. Saripan, P.S. Sulaiman, Skin segmentation using YUV and RGB color spaces. *JIPS* **10**(2), 283–299 (2014)
11. J. Kovac, P. Peer, F. Solina, Human skin color clustering for face detection, in *The IEEE Region 8 EUROCON 2003. Computer as a Tool*, vol 2 (Sept 2003), pp. 144–148
12. K. Sobottka, I. Pitas, A novel method for automatic face segmentation, facial feature extraction and tracking. *Sig. Process. Image Commun.* **12**(3), 263–281 (1998)
13. C. Prema, D. Manimegalai, Survey on skin tone detection using color spaces. *Int. J. Appl. Inf. Syst.* **2**, 18–26 (2012) (Published by Foundation of Computer Science, New York, USA)
14. D. Zhang, G. Lu, Review of shape representation and description techniques. *Pattern Recogn.* **37**(1), 1–19 (2004)
15. S. Saha, S. Goswami, P.R.S. Mahapatra, A heuristic strategy for sub-optimal thick-edged polygonal approximation of 2-D planar shape. *Int. J. Image Graph. Sig. Process. (IJIGSP)* **10**(4), 48–58 (2018)

Signature-Based Data Reporting in Wireless Sensor Networks



Monika Bhalla and Ajay Sharma

Abstract Authenticated data transmission is a vital concern in wireless sensor networks (WSNs). Clustering approach in WSNs assists in improving the performance of resource-constrained sensor networks. In this paper, we propose an authenticated variation of LEACH (Sig-LEACH) for cluster-based WSNs (CWSNs). The cluster head node is chosen on a random rotation basis. The proposed protocol uses certificateless digital signature for node authentication. The security of the signature relies on elliptic curve discrete logarithm (ECDLP) hardness assumption. Sig-LEACH is. The simulation results illustrate the efficiency of the proposed protocol. In terms of energy consumption and security overhead, the results calculated from the propose protocols illustrated better performance than other existing protocols to secure CWSNs.

Keywords Certificateless digital signature · Cryptography · Cluster-based WSNs · LEACH

1 Introduction

The recent advancement and modification in low-power sensor motes have guided various innovative applications in the field of wireless sensor networks (WSNs) [1]. The major applications involve monitoring of temperature, pressure for physical and environmental conditions, monitoring the marine environment, and traffic control management which means it is significant to physical, commercial, and industrial applications. Secure data transmission is a fundamental need in WSNs. Cryptography is used in achieving this goal [2].

M. Bhalla
Amity University, Noida, Uttar Pradesh, India
e-mail: mbhalla2@amity.edu

A. Sharma (✉)
BlackRock Gurugram, Gurgaon, India
e-mail: sharma1_ajay@yahoo.com

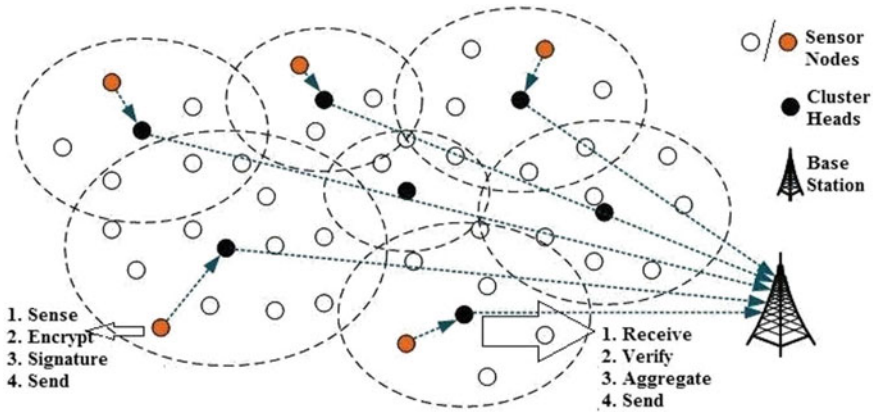


Fig. 1 Overall structure of entire communication from sensor node to base station

Unlike ad hoc networks, WSNs are more challenging to secure because they are more restrained in terms of resources. Most of the resources that are constraints are memory, processing power, and bandwidth. As the communication is wireless in nature and deployment is unattended, it also attracts additional vulnerabilities, and hence, breaking the security of the network is a common event. Secure data transmission in cluster-based WSNs has been widely explored by researchers [3]. To reduce bandwidth usage and extend node lifetime, cluster-based network architecture is more favored. The cluster head (CH) acts as a leader in this architecture and collects the data collected by the leaf nodes. This aggregated data is then forwarded to the base station (BS). The overall communication structure is better explained in Fig. 1.

2 Related Work

Heinzelman et al. [4] presented an effective energy-balanced protocol, low-energy adaptive clustering hierarchy (LEACH) for CWSNs by randomly rotating the responsibility of cluster head. This approach has major advantage of improving network lifetime. APTEEN [5] is another variation of LEACH protocol with a different approach of assigning the responsibility of cluster head. Secure data transmission in these CWSNs is a challenging task and is widely studied in last decade. The major challenge is to deal with dynamic allocation of cluster heads randomly. Therefore, various LEACH similar protocols have been presented recently. Some of the examples are GS-LEACH [6], Sec-LEACH [7], and R-LEACH [8]. The common approach used for security in these protocols is symmetric key management. In this approach, all the nodes keep a collection of keys and they communicate with the neighboring sensor nodes on the basis of common key. A major suffering of this key management system is orphan node problem [9]. Also, to permanently store the key pool, an addi-

tional memory overhead is considered. As the network size grows, this complexity is unacceptable. Sometimes, in such cases, due to lack of non-sharing of any key with neighboring nodes, a node is not able to join the cluster head and therefore has to elect itself as a CH. The increasing number of cluster heads degrades the overall performance of entire sensor network.

This orphan node problem allowed the researchers to check the feasibility of the asymmetric key cryptography. However, the computational complexity of asymmetric systems is considered high, but some concrete results make it acceptable in WSNs [10]. Digital signature is a key feature of public key cryptography which assures authenticity, integrity, and nonrepudiation in the network. Shamir [11] introduced an identity (ID)-based public key cryptography that does not need any certificate of bonding between private and public key pair. In ID-based cryptography, user selects its public key, for instance, their phone number, address, driving license, or any other unique identities and third party, say private key generator (PKG), generates the private key of user. This concept does not consider the dishonesty of PKG and leads to key escrow problem. In key escrow problem, a malicious PKG can easily send corrupt messages on behalf of any user.

Al Riyami and Peterson [12] suggested a solution of key escrow problem inherited from ID-based cryptography by presenting a novel concept of certificateless signature scheme, in which a third party, say key generation center (KGC), generates the partial private key of user instead of private key. Private key is generated by the user with the help of partial private key. In this case, the KGC does not know this private key directly.

3 Our Contribution

The Sig-LEACH protocol is a LEACH-based protocol which uses certificateless signature scheme motivated by Tsai et al. [13]. The embedding of this low-cost signature in LEACH will ensure the security in an energy-efficient way. The computational complexity of this signature is one elliptic curve scalar point multiplication in signature part and four point multiplications in verification part. The best part of this proposal is to consider the use of certificateless signature. In identity-based signature schemes (SET-IBOOS or SET-IBS), key management mechanism for authentication security in wireless sensor networks is key escrow affected [14].

The subsequent sections of this paper will cover in the following manner: Section 2 describes the preliminaries and security model of certificateless signature scheme. Section 3 introduces the proposed Sig-LEACH protocol for clustered WSNs. Section 6 analyzes and evaluates the performance of Sig-LEACH. The last section concludes this work.

4 Proposed Approach

The protocol proposed here is named as Sig-LEACH. The proposed protocol operates in two phases. The phases are setup phase and steady phase. Like LEACH protocol, time has been divided into successive time intervals. The notations for time stamp in the following algorithm are T_{BS} (BS-sensor node communication) and t_s (node to cluster head communication). The purpose of using time stamp is to maintain the freshness of data. The preliminary computations for setup and key extraction part have been already stored (preloaded) in the sensor nodes.

5 Protocol Setup Phase

Let the sensor nodes have the identities as $\{ID_1, ID_2, ID_3, \dots, ID_n\}$. The base station broadcasts a time stamp and nonce to all sensor nodes. The sensor nodes (including cluster heads) are aware of the base station's identity. The volunteer cluster head is selected and now

- **Encryption:** To make sure the confidentiality of sensed data m , we exploit a lightweight symmetric encryption algorithm f to encrypt the data and forward it to the cluster head. The ciphertext is computed as $m' = m + f_\tau(ID||i)$ where τ is the symmetric secret key for encryption.
- **Key Management:**
 - i. **Setup:** Let E be an elliptic curve over a prime field F_p defined by the generalized Weierstrass equation $y^2 = x^3 + ax + ba$ and G be an additive cyclic group. The base station chooses its master secret key $x \in Z_q$ and computes public key as $P_{pub} = xP$. The base station chooses a one-way hash function $H: \{0, 1\}^* \times G \rightarrow Z_q$ and declares the public parameters as $\{E/F_p, G, P, P_{pub}, H\}$.
 - ii. **Set Secret Value:** The sensor node chooses a secret random value $x_{ID} \in Z_q$ and computes its public value as $P_{ID} = x_{ID} P$.
 - iii. **Extraction of Private/Public Keys:** The base station chooses $r_{ID} \in Z_q$ and computes $R_{ID} = r_{ID} P$ and $h_{ID} = H(ID, R_{ID}, P_{ID})$ for each sensor node with his identity ID . The base station then computes $s_{ID} = r_{ID} + h_{ID} x \pmod n$ and sends (s_{ID}, R_{ID}) to the sensor node via a secure channel. The private and public key pair for sensor node is (x_{ID}, s_{ID}) and (P_{ID}, R_{ID}) , respectively.
 - iv. **Sign:** The sensor node signs the message m in the following steps:
 1. The sensor node computes $R = tP$.
The signer computes $k_1 = H(m || R || t_s || P_{ID} || R_{ID})$ and $k_2 = H(m || R || P_{ID} || t_s || R_{ID} || P_{pub})$.
 2. Repeat step 1 if $\gcd(t + h, n) = 1$.
 3. Now, the node computes the signature $s = (t + k_1)^{-1}(k_2 x_{ID} + s_{ID}) \pmod n$ and sends (ID_j, m', R, t_s, s) to the cluster head.

- v. **Verification:** The cluster head node confirms the received signature validity with the following equation:

$$s(R + k_1 P) = k_2 P_{ID} + R_{ID} + h_{ID} P_{pub}$$

Steady State Phase: In steady state phase, firstly, the sensor node transmits the sensed data to cluster head, and secondly, the cluster head aggregates and sends that particular data to base station.

As the protocol runs in rounds (similar to LEACH), each round includes a setup phase and then steady state phase. The basic assumption here is to ensure that all the nodes are synchronized, i.e., all the sensor nodes are aware of the starting and ending time of each round. The overall transmission time has been divided into time slots using TDMA. For cluster head selection, remaining energy in the node is considered as a primary factor. Other than this, the received signal strength will play a major role.

Setup Phase:

Step 1: BS $\Rightarrow G_S: \langle ID_{BS}, T_{BS}, nonce \rangle$

Step 2: CH_{*i*} $\Rightarrow G_S: \langle ID_i, t_s, adv, R_i, s_i \rangle$

Step 3: S_{*j*} $\Rightarrow CH_i: \langle ID_i, ID_j, t_s, join, R_i, s_i \rangle$

Step 4: CH_{*i*} $\Rightarrow G_S: \langle ID_i, t_s, sched (\dots ID_j/t_j, \dots), R_i, s_i \rangle$

Steady Phase:

Step 5: S_{*j*} $\Rightarrow CH_i: \langle ID_i, ID_j, t_s, m', R_j, s_j \rangle$

Step 6: CH_{*i*} $\Rightarrow BS: \langle ID_{BS}, ID_i, t_s, Agg(m'_j), R_i, s_i \rangle$.

6 Implementation and Results

There are usually two types of costs associated with sensor networks: computational cost and communication cost. The computational cost at the end of sensor nodes includes the encryption of sensed information and signature on that message. At cluster head node, the received signature is verified and all the received messages are aggregated to be forwarded to base station. The cost of encryption in Sig-LEACH, SET-IBS, and SET-IBOOS is almost same, but signature verification process in Sig-LEACH is comparatively cheaper than others. The following computational table describes the costs associated with related protocols.

Here, PM represents scalar point multiplication, BP refers bilinear pairing, and EXP is used for exponentiation operation. According to [15], the timings for one point multiplication and one bilinear pairing are 0.295 and 1.90 s, respectively. The time for one exponentiation is approximately equal to one point multiplication. In SET-IBOOS protocol, the offline computation is also being performed by sensor

Table 1 Computational cost

Scheme	Sign		Verify	Signature size (bytes)
SET-IBS	1PM		4PM	65 ^a
SET-IBS	1BP + 1PM		2BP	
SET-IBOOS	4EXP (offline)	1EXP (online)	3EXP	48

^aOnly one elliptic curve coordinate is enough to transmit a point. The other coordinates can be calculated easily

Table 2 Energy consumption

Scheme	Sign (time in s)	Energy consumption (mJ)	Verify (time in s)	Energy consumption	Total energy consumption (mJ)
Sig-LEACH	0.295	7.07	1.14	28.36	35.36
SET-IBS	2.195	52.60	3.76	91.24	143.68
SET-IBOOS	1.475	35.36	0.865	21.20	56.52

node itself. There is no third party to compute this offline part, especially in case of homogeneous sensor network. Also, the signature size is comparable in Sig-LEACH. Therefore, from Table 1, it is clear that Sig-LEACH is more economical than others in terms of computation cost. To compute the energy consumption, RELIC-Toolkit cryptographic library MICAz and MICAz sensor mote with TinyOS is used. To support the performance of Sig-LEACH, Table 2 shows the calculated energy consumption.

To better analyze the performance of sensor network, three parameters have been focused. The parameters taken into consideration are data packets received, number of dead nodes, and sum of energy. As all the protocols are based on LEACH, the performance is also analyzed in rounds. Figure 2a–c demonstrates the performance of SET-IBS, SET-IBOOS, and Sig-LEACH, respectively. When the number of round increases on one axis, the performance of other parameters can be captured on other axes. The results clearly indicated that the performance of certificateless Sig-LEACH is better than other two identity-based protocols.

7 Conclusion

When there is a need for transmitting the sensed data to cluster head as well as to base station, authentication in wireless sensor networks is very crucial. To avoid key escrow problem, certificateless signature scheme is used. Comparison with SET-IBOOS as well as SET-IBS is extrapolated with the proposed Sig-LEACH protocol. The comparative analysis clearly describes that Sig-LEACH outperforms the compared protocols.

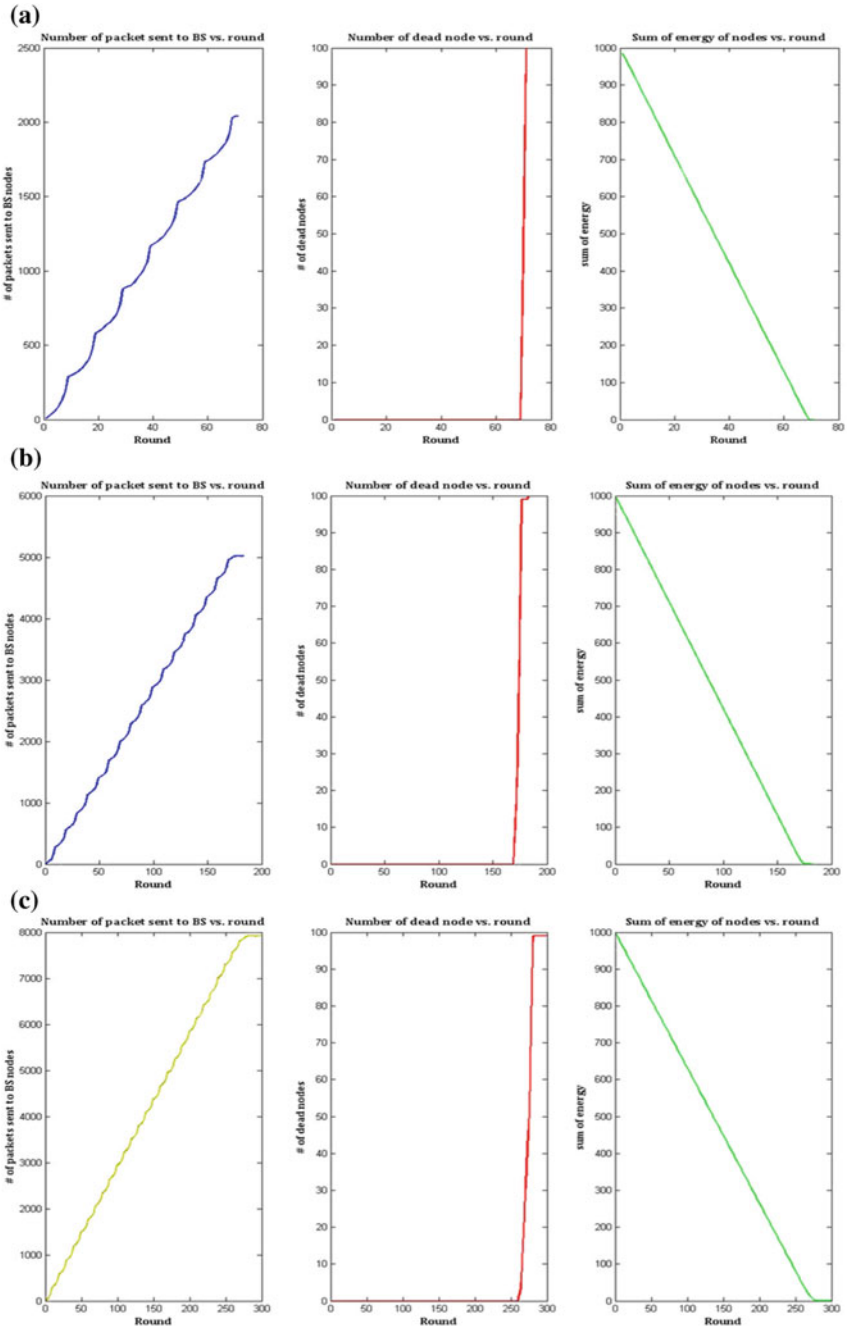


Fig. 2 Performance comparison a IBS b IBOOS c Sig-LEACH

References

1. I.F. Akyildiz, W. Su, Y. Sankara Subramaniam, E. Cayirci, Wireless sensor networks: a survey. *Comput. Netw.* **38**(4), 393–422 (2002)
2. J.P. Walters, Z. Liang, W. Shi, V. Chaudhary, Wireless sensor network security: a survey. *Secur. Distrib. Grid Mob. Pervasive Comput.* **1**, 367 (2007)
3. A.A. Abbasi, M. Younis, A survey on clustering algorithms for wireless sensor networks. *Comput. Commun.* **30**(14), 2826–2841 (2007)
4. W.R. Heinzelman, A. Chandrakasan, H. Balakrishnan, Energy-efficient communication protocol for wireless microsensor networks, in *Proceedings of the 33rd Annual Hawaii International Conference on System Sciences, 2000, IEEE* (Jan 2000), p. 10
5. A. Manjeshwar, Q.A. Zeng, D.P. Agrawal, An analytical model for information retrieval in wireless sensor networks using enhanced APTEEN protocol. *IEEE Trans. Parallel Distrib. Syst.* **13**(12), 1290–1302 (2002)
6. P. Banerjee, D. Jacobson, S.N. Lahiri, Security and performance analysis of a secure clustering protocol for sensor networks, in *Sixth IEEE International Symposium on Network Computing and Applications, IEEE (NCA 2007)* (July 2007), pp. 145–152
7. L.B. Oliveira, A. Ferreira, M.A. Vilaça, H.C. Wong, M. Bern, R. Dahab, A.A. Loureiro, SecLEACH—on the security of clustered sensor networks. *Sig. Process.* **87**(12), 2882–2895 (2007)
8. K. Zhang, C. Wang, C. Wang, A secure routing protocol for cluster-based wireless sensor networks using group key management, in *2008 4th International Conference on Wireless Communications, Networking and Mobile Computing, IEEE* (Oct 2008), pp. 1–5
9. S. Sharma, S.K. Jena, A survey on secure hierarchical routing protocols in wireless sensor networks, in *Proceedings of the 2011 International Conference on Communication, Computing & Security, ACM* (Feb 2011), pp. 146–151
10. A. Liu, P. Ning, TinyECC: A configurable library for elliptic curve cryptography in wireless sensor networks, in *International Conference on Information Processing in Sensor Networks, 2008. IPSN'08, IEEE* (Apr 2008), pp. 245–256
11. A. Shamir, Identity-based cryptosystems and signature schemes, in *Proceedings of CRYPTO'84 on Advances in Cryptology* (1984), pp. 47–53
12. S.S. Al-Riyami, K.G. Paterson, Certificateless public key cryptography, in *International Conference on the Theory and Application of Cryptology and Information Security* (Springer, Berlin Heidelberg, Nov 2003), pp. 452–473
13. J.L. Tsai, N.W. Lo, T.C. Wu, Weaknesses and improvements of an efficient certificateless signature scheme without using bilinear pairings. *Int. J. Commun. Syst.* **27**(7), 1083–1090 (2014)
14. Y. Wang, X. Peng, J. Bian, Key management mechanism for authentication security in wireless sensor network. *Appl. Math.* **9**(2), 711–719 (2015)
15. R. Yasmin, E. Ritter, W.A.N.G. Guilin, An authentication framework for wireless sensor networks using identity-based signatures: implementation and evaluation. *IEICE Trans. Inf. Syst.* **95**(1), 126–133 (2012)
16. H. Lu, J. Li, M. Guizani, Secure and efficient data transmission for cluster-based wireless sensor networks. *IEEE Trans. Parallel Distrib. Syst.* **25**(3), 750–761 (2014)

Wine Quality Analysis Using Machine Learning



Bipul Shaw, Ankur Kumar Suman and Biswarup Chakraborty

Abstract Almost from the beginning of mankind, there has been the existence of different kinds of wine. It has also become very important for us to know the quality of the wine, before consuming it. In the last few decades, the food industry has grown enormously and so are the food quality analysis and its “rating” process. We often come across cases in which a consumer falls sick because of consuming low-quality food, so it has become a necessary evil for us to have a quality analysis of a product before selling the product, “evil” because it adds up extra cost to the production of the final product. Similarly, it is also necessary to do a quality analysis of wine and there have been different methods used to determine the quality of the wine, but we often get confused regarding which method to rely on! This paper focuses on the comparative study over different classification algorithms for wine quality analysis which are: SVM, random forest and multilayer perceptron and to know which of the above-mentioned classification algorithms give more accurate result.

Keywords Machine learning · Support vector machine · Random forest · Multilayer perceptron · Wine quality

1 Introduction

As we all know, that heavy research is being done in the field of machine learning, but surprisingly it is not a new topic! It was started back in 1950 by Alan Turing. He proposed a “learning machine” that could learn and become artificially intelligent,

B. Shaw (✉) · A. K. Suman · B. Chakraborty
Department of Computer Science and Engineering, Institute of Engineering
and Management, Saltlake, India
e-mail: bipulshaw2014@gmail.com

A. K. Suman
e-mail: aksankuraks@gmail.com

B. Chakraborty
e-mail: biswarup.98.kkr@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling
and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_23

which marked the starting of the era of machine learning! But even though it was proposed so early, it has gained an enormous interest growth in last few years. The main four reasons for this sudden change are: rise in computing power, heavy amount of data generation, growth of deep learning and the rise of digital era. We can broadly classify machine learning into three subtopics: supervised learning, unsupervised learning and reinforcement learning. In supervised learning, we train the algorithm with the training data sets which includes both inputs and labels (targets or outputs). After training the algorithm with a lot of data sets, we try to create the logic for predicting the labels for the new data. There are two types of supervised learning: classification and regression. Then comes unsupervised learning in which training data sets does not include targets; here, we do not tell the system where to go, but the system has to understand itself from the data we give because here training data is not structured. It also has two types of method, namely clustering and anomaly detection. Reinforcement learning differs in the aspect of input/output pairs which may need not be presented and suboptimal actions need not be explicitly corrected. Instead of that, the focus is given upon performance, which involves balance between exploration and exploitation.

Our food industry has also become a large-scale user of these techniques, and in that we have a small section of wine quality analysis in which machine learning is used extensively. The reason for the sudden growth in wine quality check is because it has a direct relation with our health, and moreover it helps us to check the variation in the condition of heart. Another valid reason for this can be the increase in the amount of wine consumption and forcing their respective companies to have assessment on wine quality and grading certification to maintain their name and survive in the corporate world. And hence, the above-stated reasons motivated us to write a paper on wine quality analysis and to compare the result of different algorithms.

Similarly, we can also state that since the quality of a wine does not depend on a single factor but on multiple attributes so it becomes comparatively easier to check the quality of a wine by machine learning rather than the human tasters! Moreover, it also helps us to know that which physical/chemical attribute is affecting the quality of the wine in which way (either positive or negative). In this paper, we have implemented and compared the results of three techniques, namely support vector machine, random forest and multilayer perceptron to check the quality of wine.

2 Literature Survey

There have been many researches done on topics like “wine quality analysis”, “price prediction of wines” and “conditions in favour of preparing wine”. Some of the earlier works on these topics are as follows: in [1], the authors associated wine drinking with increase in heart rate variability in women along with coronary heart disease. In [2], the authors tell the wine applications with the help of electronic noses. In [3], the authors used Gaussian regression process and multitask learning to predict the price of wine. They used past data of wine price to predict the price of wine in

future. They concluded that advanced machine learning technique has the potential to predict the price of wine. In [4], the authors mentioned that the price and quality of wine depend on the weather in which the grapes were cultivated. He derived a price equation using several factors. In [5], the authors predicted wine verification using data mining tools. In [6], the authors analysed the quality of wine using a decision tree and other tools. In [7], the authors model the wine preferences using data mining from some physiochemical properties. In [8], the authors introduce us to the elimination of recursive features with random features for PTR-MS analysis of agro-industrial products. In [9], the authors extract rules from multilayer perceptron in some classification problems. It is a clustering-based approach. In [10], the authors compare multivariate analytic techniques. They also listed out the pros and cons of recursive partitioning analysis. In [11], the author shows us an efficient algorithm to generate classification rules. In [12], the author has done the wine quality analysis using several classification approaches with different feature sets like principal component analysis, recursive feature elimination and nonlinear decision tree. In [13], the author has also classified wine quality with imbalanced data using synthetic minority over-sampling technique (SMOTE), decision tree, adaptive boosting and random forest. And hence after being encouraged by the previous works as stated above, we applied more than one ML algorithms to predict the quality of the wine, namely SVM, random forest and multilayer perceptron.

3 Methodologies

See Fig. 1.

3.1 Data Preparation

The data set contains 4898 instances of red wine from the UCI machine learning repository. The physical properties which are in the data set are: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulphur dioxide, total sulphur dioxide, density, pH, sulphates, alcohol and finally quality. There are physical as well as chemical variables that influence the quality of wines. Tartaric acid, citric acid and malic acid are present in the wine, whereas ascorbic and sulphurous acids are added during the winemaking. Residual sugar determines the sweetness of the wine. Although it is not the only factor which determines the sweetness of the wine, it still plays the major role in determining the taste of the wine. And due to yeast metabolism, there is generation of alcohol by which wine gets its alcoholic properties.

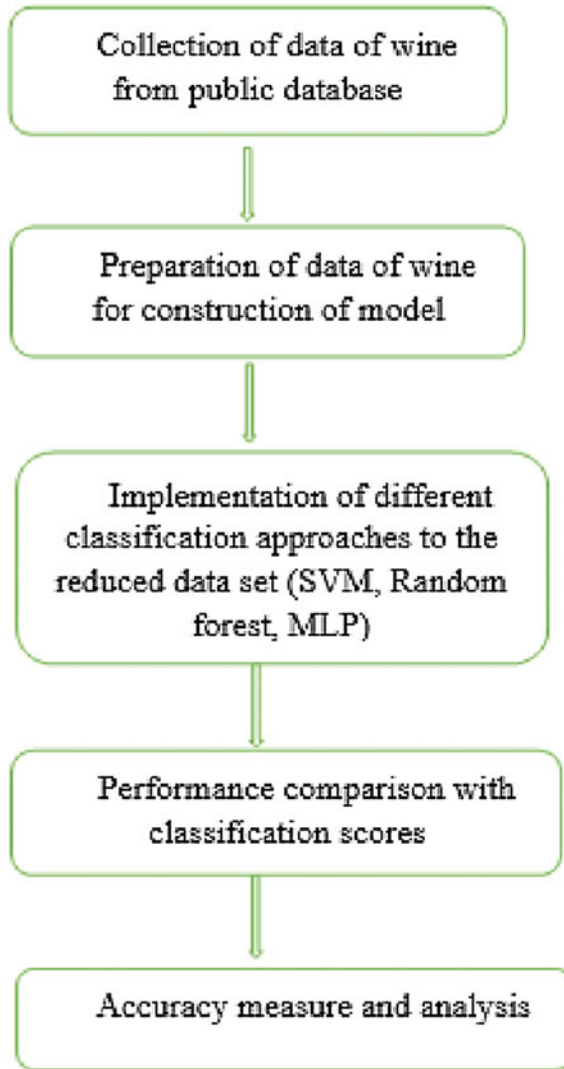


Fig. 1 Flow chart showcasing our working structure in this paper

3.2 Selection of Algorithms

In this paper, our motive was to analyse the quality of wine using supervised learning method and to be more specific, by different classification techniques. So in order to remain on our motive, we decided to take three algorithms in supervised learning method, but each one of its kind. So, we took SVM as first algorithm, random forest as second algorithm and multilayer perceptron as our third algorithm. So basically,

we will be comparing among a simple classifier such as SVM, an algorithm which works on the principle of decision tree such as random forest and a classifier which uses artificial neural network like multilayer perceptron to see which type of classifier gives best result in our stated problem.

3.3 Comparative Study

In this paper, we have used three different kinds of techniques, namely: SVM, random forest and multilayer perceptron, to find and predict the quality of the given wine data with the help of created logic and by training the data sets. First of all, we will implement the SVM algorithm to the test data set and calculate its result, after this we will implement the random forest algorithm to the test data set and calculate its result, and finally we will implement the multilayer perceptron algorithm and calculate its result as well. After calculating the results, we will make a bar graph for better distinction between the performances of the three algorithms used in this paper.

Support Vector Machine

Support vector machine or SVM is a supervised learning model with associated learning algorithms that analyses data used for classification and regression analysis. Its basic approach is to separate the positive and negative class with the largest margin. It is based on Vapnik statistical learning theory [14]. It has many good properties like margin maximization, high fitting accuracy, small number of tuneable parameters and kernel technology adopted in high-dimensional feature space. We can also say that it is a non-probabilistic binary linear classifier.

Random Forest

Random forest is a supervised learning method which consists of random decision tree which works in coordination with classification and regression. They operate by constructing multiple layers of decision tree at the time of training and output the classes or mean prediction of the individual trees. Now by taking the probability of all the decision trees into account, the overall probability is calculated. It is basically used for the data sets having high dimensionality where the individual variables are non-stationary and highly noisy.

Multilayer Perceptron

Multilayer perceptron is a class of feed forward artificial neural network. An MLP consists of at least three layers of nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and nonlinear activation distinguish MLP from a linear perceptron. It can even distinguish data that are not linearly separable. For now, MLP classifier supports only cross-entropy loss function, which allows probability estimates by running predict_proba method.

3.4 Performance Measure Metrics

The parameters used to determine the quality of the wine by the above-mentioned algorithms are: fixed acidity, volatile acidity, citric acid, residual sugar, chlorides, free sulphur dioxide, total sulphur dioxide, density, pH, sulphates and alcohol content. In this paper, the parameter used to compare the performance and validations of the classifiers is accuracy score, whereas according to the definition, accuracy in machine learning is the number of correct predictions made, divided by the total number of prediction.

4 Result

There are several solutions for the wine quality analysis as well as several ways to do it. The above-stated problem has been solved by various algorithms but which one gives the best result? We have considered three basic algorithms which were used to determine the quality of the wine and studied that which algorithm gives the best possible result. In this paper after dividing the data set into two groups, namely training data set and test data set, we trained each classifier based on the training data set and tested their (classifier’s) efficiency on the test data set. So, each classifier is able to show the performance metrics, i.e. accuracy based on the test data set. We made a bar graph plot for better understanding of the comparative study of the classifiers based on the accuracy parameter. And hence, we were able to see that random forest algorithm gave the result with best accuracy and SVM with the worst among the three. The classification scores of SVM are given in Table 1, classification scores of random forest are given in Table 2, and classification scores of multilayer perceptron are given in Table 3. And a bar graph is shown in Fig. 2.

The following given graph is based on the comparative study among the three algorithms we used in this paper, namely SVM, random forest and multilayer perceptron by parameterizing their accuracies.

Table 1 Classification score for SVM

Quality	Precision	Recall	F1-score	Support
3	0.00	0.00	0.00	9
4	0.00	0.00	0.00	76
5	0.60	0.65	0.63	669
6	0.56	0.75	0.64	960
7	0.57	0.21	0.31	349
8	0.00	0.00	0.00	82
Average/total	0.53	0.57	0.53	2145

Accuracy score: 0.5729603729603729

Table 2 Classification score for random forest

Quality	Precision	Recall	F1-score	Support
4	0.98	0.46	0.62	140
5	0.84	0.86	0.85	1469
6	0.77	0.90	0.83	1846
7	0.91	0.67	0.77	730
8	1.00	0.39	0.56	111
9	1.00	0.20	0.33	5
Average/total	0.83	0.82	0.81	4352

Accuracy score: 0.8196231617647058

Table 3 Classification score for multilayer perceptron

Quality	Precision	Recall	F1-score	Support
4	0.79	0.56	0.65	140
5	0.85	0.79	0.82	1469
6	0.75	0.87	0.80	1846
7	0.78	0.68	0.72	730
8	0.88	0.40	0.55	111
9	1.00	0.20	0.33	5
Average/total	0.83	0.82	0.81	4352

Accuracy score: 0.787812548529

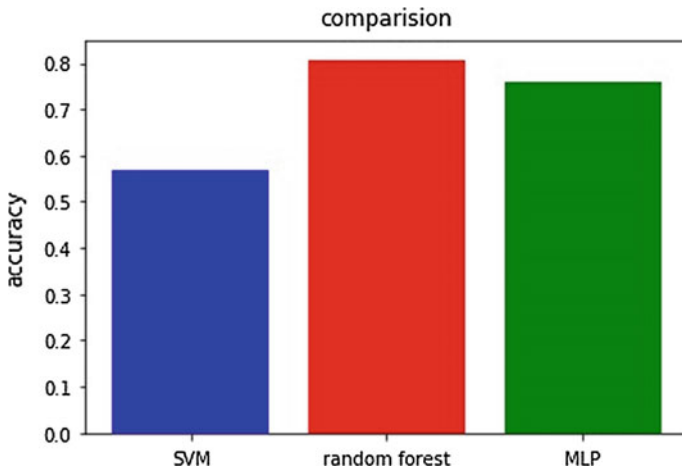


Fig. 2 Bar graph showing the comparison among the used algorithms, on the basis of accuracy

5 Conclusion

In this paper, our basic motive was to see that which of the three algorithms, SVM, random forest and MLP, give the best accurate result in the analysis of wine quality. During our comparative study between those algorithms, we came to know that random forest algorithm gives the best result with an accuracy percentage of 81.96, then comes multilayer perceptron algorithm with an accuracy percentage 78.78 and lastly comes the support vector machine algorithm with an accuracy percentage of 57.29. The reason why random forest algorithm gives the best result is because when a node is split during the formation of tree, the split that is chosen is no longer the best split among all the features. Instead, the split chosen is the best and due to this randomness, the prejudice of the system increases and hence yields an overall better model. Finally, we can conclude our paper with a result that random forest algorithm is the best among the above-stated algorithms and a question that, “Is there any other classification algorithm which can give better accuracy result than random forest algorithm?”.

References

1. I. Janszky, M. Ericson, M. Blom, A. Georgiades, J.O. Magnusson, H. Alinagizadeh, S. Ahnve, Wine drinking is associated with increased heart rate variability in women with coronary heart disease. *Heart* **91**(3), 314–318 (2005)
2. V. Preedy, M.L.R. Mendez, Wine applications with electronic noses, in *Electronic Noses and Tongues in Food Science* (Academic Press, Cambridge, MA, USA, 2016) pp. 137–151
3. M. Yeo, T. Fletcher, J. Shawe-Taylor, Machine learning in fine wine price prediction. *J. Wine Econ.* **10**(2), 151–172 (2015)
4. O. Ashenfelter, Predicting the quality and prices of Bordeaux wine. *J. Wine Econ.* **5**(1), 40–52 (2010)
5. J. Ribeiro, J. Neves, J. Sanchez, M. Delgado, J. Machado, P. Novais, Wine vinification prediction using data mining tools, in *Conference Proceedings, Computing and Computational Intelligence* (Tbilisi, Republic of Georgia, 2009, June) pp. 78–85
6. S. Lee, J. Park, K. Kang, Assessing wine quality using a decision tree, in *IEEE International Symposium on Systems and Engineering (ISSE)* (2015, Sept) pp. 176–178
7. P. Cortez, A. Cerdeira, F. Almeida, T. Matos, J. Reis, Modeling wine preferences by data mining from physicochemical properties. *Decis. Support Syst., Elsevier* **47**(4), 547–553 (2009)
8. P.M. Granitto, C.L. Furlanello, F. Biasioli, F. Gasperi, Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products. *Chemometr. Intell. Lab. Syst.* **83**(2), 83–90 (2006). <https://doi.org/10.1016/j.chemolab.2006.01.007>
9. E.R. Hruschka, N.F. Ebecken, Extracting rules from multilayer perceptrons in classification problems: a clustering-based approach. *Neurocomputing* **70**(2), 384–397 (2006)
10. E.F. Cook, L. Goldman, Empiric comparison of multivariate analytic techniques: advantages and disadvantages of recursive partitioning analysis. *J. Chronic Dis.* **37**(9–10), 721–731 (1984)
11. S. Vijayarani, M. Divya, An efficient algorithm for generating classification rules. *Int. J. Comput. Sci. Technol.* **2**(4), 512–515 (2011)
12. S. Aich, A.A. Al-Absi, K.L. Hui, J.T. Lee, M. Sain, A classification approach with different feature sets to predict the quality of different types of wine using machine learning techniques, in *International Conference on Advanced Communication Technology* (2018), pp. 139–143

13. G. Hu, T. Xi, F. Mohammed, H. Miao, Classification of wine quality with imbalanced data, in *IEEE International Conference on Industrial Technology (ICIT)* (2016), pp. 1712–1717
14. V.N. Vapnik, *The Nature of Statistical Learning Theory*. New York: Berlin: Springer (1995)

Generalized Smart Traffic Regulation Framework with Dynamic Adaptation and Prediction Logic Using Computer Vision



Vishal Narnolia, Uddipto Jana, Soham Chattopadhyay and Shramana Roy

Abstract Traffic signals are essential to ensure safe driving at road intersections and to maintain a constant flow of vehicles in a convenient manner. However, sometimes inefficient traffic control themselves restrict the constant flow creating commotions and delays. Therefore, in this work, we are introducing a generalized smart traffic regulation algorithm (G-STRA) which is going to consider the real-time traffic density using image processing at each lane, to reduce the wait time and improve the total throughput. The system will be smart enough to identify vehicles of importance (VoIs), to give them extra clearance from the regular commuting vehicles. The G-STRA is called “smart” as it will adapt itself with the phases of the day, days of importance (DoIs), the condition of road, and weather and the type of commotion to predict the type of flow it should maintain learning from its past experiences.

Keywords Self-learning · Vehicle detection and classification · Real time · Vehicle of importance · Generalized · Traffic regulation

1 Introduction

Before the advent of traffic lights, people were employed as traffic police to regulate the flow of traffic. The earliest documented example was in London Bridge in 1722. On December 9, 1868, the first non-electric traffic lighting system was introduced.

V. Narnolia (✉) · U. Jana · S. Chattopadhyay · S. Roy
Department of Computer Science and Engineering, Institute of Engineering and Management,
Salt Lake, Kolkata, India
e-mail: vishalnarnolia747@gmail.com

U. Jana
e-mail: ujana10@gmail.com

S. Chattopadhyay
e-mail: scha99@gmail.com

S. Roy
e-mail: shramanaroy23@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_24

People used different colored lamps to drive the traffic, each of conveying a different meaning. It was not until 1912 that electrically controlled traffic lights were introduced and later in the 1920s instead of people manipulating the lights humanity saw the lights being directed automatically by timers [1].

Traffic signaling has improved a lot since then. To improve the fluency of movement, many researchers have developed intelligent algorithms to optimize the scheduling of traffic light timings. Many of them have considered monitoring the real-time flows of traffic which could be done when human intervention played a huge part in the traffic signaling system. The traffic control at each traffic light is presented by the variable sequence phasing cycle that represents the traffic flow that intends to pass the road intersection at that point of time. A lot of things might be considered as parameters for determining the schedule: the number of vehicles, traffic speed, and traffic volume at each flow to mention a few.

It is seen that the smartest of the working model present today also fails to incorporate many aspects. Human intervention is required to sort those aspects. There is a lack of a self-learning and adapting system which learns from the past experiences and predicts the solution themselves. This does not imply that human forces are not at all required, but they are required only for some unique situations which the system had not faced yet. Once a problem is known to the system and it is solved, the system will be able to solve such problems thereafter. If such a model comes into action, the need of involving manpower is reduced.

Since the time vehicles like ambulances and fire engines are levied off the normal traffic rules to serve their purpose efficiently, it is difficult to incorporate the right these prioritized vehicles are given into the regulation system. No previous work has been done to account for the priority of these vehicles, as predictive models have not been utilized before. On the flip side, the lack of predictive models also makes them indecisive about what to do if, say, there is a political procession is passing through a junction. Currently, road intersections are often blocked intentionally to let the rally pass because of the lack of any appropriate solution to tackle it, adding to the woes of the commuters.

The problems that the researchers have faced until now are to incorporate a mechanism that is robust to every condition of traffic. The researchers have tried to improve the regulation by leveraging on algorithmic improvements. However, it seems that algorithmic enhancements to improve the throughput has already reached its saturating points as recent advances in traffic regulation algorithms has shown very meager changes in performances. Monitoring real-time flows using image data has added versatility to the judgment of parameters which resulted in considerable improvements in throughput. Yet it still lacks the capability to be applied to roads in a generalized manner. Now, here is the ambitious part. This paper intends to introduction of a machine learning model that is going to learn from historical data, the variations in the density pattern with respect to various aspects such as the time of the day, part of the year, weather and road conditions and is going to allocate appropriate green time to the competing flows at any N -way intersection such that they come into conflict as less as possible.

2 Related Works

Khekare, G. S., Sakhare A. V. proposed in [2] the development of Vehicular ad hoc networks (VANETs), which are the quintessential of the new types of networks emerging in the wireless technologies. VANET is to facilitate the vehicles of a city with a wireless communication between vehicles themselves (V2V) and between vehicles and traffic control consoles (V2I). The system tells the intelligent console the present status, destination, and future movement of the vehicle at the next intersection. Meanwhile, it also connects with other nearby vehicles to inform them the same. This enables the roadside units to calculate and predict the next imposing signal while, to the other vehicles their next relative movement. The system also enables the vehicles to know the traffic conditions from inside and thereby, reducing the chances of undesirable outcomes.

The paper [3], Naeem Abbas, Muhammad Tayyab, M. Tahir Quadri proposed a smart traffic control algorithm by measuring traffic density. It counts density by computing the real-time frame live video by the reference image and searching vehicles on the road.

In the works of Anurag Kanungo, Chetan Singla, and Ayush Sharma regarding traffic control using video processing [4], they proposed an intelligent algorithm for allotting time to each road based on density to avoid more congestion, prioritize the vehicles, and predicting the flow of traffic.

A traffic control algorithm that is embedded within Simulation of Urban Mobility (SUMO) has been discussed in [5]. This algorithm measures the length of the traffic at each incoming lane and attempts to reduce the waiting time. This algorithm checks the lanes and determines the lane with the longest length of traffic and clears the allowance of that lane.

Again, another intelligent algorithm has been introduced by Bento et al. [6] which is based on V2V and V2I communications, which allow the exchange of information between vehicles and the intelligent traffic management device. A traffic simulator was developed to study the traffic management techniques, evaluate and make a comparison of their performance for mainly two types of road intersections, namely roundabout and crossroads. The paper claims that the proposed algorithm works very well for autonomous vehicles but also gives a reasonable performance for human drivers, provided they follow the speed profile proposed by their model.

In the paper [7], Webster proposed a real-time traffic signaling algorithm that makes use of VANETs to gather information about the instantaneous speed and position of individual vehicles to optimally calculate the signals that needed to be activated at a specific intersection at that time. The paper considers the traffic regulation as a job scheduling problem, where a job is the group of vehicles waiting at a respective incoming lane for their share of the green signal. An online algorithm was developed to minimize the delay across the intersection by scheduling the optimal sequence of several phases at each traffic light. The First-Come-First-Served principle was applied to schedule the competing jobs at each flow. Mathematical

analysis and simulation have been used to prove the correctness and advantages of the introduced algorithm.

3 Our Proposed Idea

The proposed logic to create smart traffic solution, by letting the system, knows few details like:

1. Number of roads in the intersection;
2. Nature of each road connected;
3. Densities of each incoming and outgoing lanes;
4. The speed of traffic flow in the intersection;
5. Status of each road at present;
6. Current part of day and other requisite environmental conditions.

All these parameters will be classified and defined by our self-learning and predicting unit.

CCTV cameras (each overhead of the signal facing each lane) will take the frames of the present scenario and process it to get the density of each road at regular intervals. Using these data, we will allot both time for a specific lane to remain open and the order in which it is done. Meanwhile, we will also set an offset for each lane to ensure that each road is given a green at least once in a cycle. The proposed system will also use object tracking systems to ensure prioritized traffic flow and will also help in road safety.

4 Design and Methodology

4.1 *Image Enhancement and Processing Unit*

Robust and reliable vehicle detection and density determination are critical to implement our proposed framework efficiently. This is done by implementing few of the pre-proposed and optimized computer vision modules.

The steps this framework incorporates are:

1. At the time of installation, a still image of the empty road is preprocessed into the system for reference.
2. Extracting images from the overhead cameras in real time.
3. Enhancing the images using Edge Canny and dilation filters [8].
4. Comparing the white spaces with the reference image to get the density of traffic in real time (in %) [3] (Figs. 1 and 2).



Fig. 1 Image enhancement of empty road (here density in right lane is taken 0%)



Fig. 2 Image enhancement of busy road (here density in right lane is given 47%)

4.2 The Traffic Regulation Unit

Variable Description:

1. N : No. of roads (odd numbers of roads are roads going into the junction, even ones are outgoing).
2. Speed $[N]$: This has the exact speed of flow of each lane.
3. Density Variables:
 - a. $D_{\text{odd}}[N/2]$: This will contain 1 if the density of the corresponding odd lane falls below 10%.
 - b. $D_{\text{even}}[N/2]$: This will contain 1 if the density of the corresponding even lane falls below 10%.
 - c. $D_{0-100}[N/2]$: This will contain the density of the odd lanes in percentage.
4. Flags:
 - a. Signal $[N/2][N/2]$: The i th row represents the $(2i + 1)$ th lane (incoming), the i th column represents the $(2i + 2)$ th lane (outgoing). The matrix represents all possible lane-to-lane connections. A cell contains 1 then it represents that connection or signal is enabled currently and vice versa.
 - b. Status $[N/2]$: This indicates whether the corresponding odd lane has been opened before in the current cycle or not.
5. Time Variables:
 - a. C : Counter for counting the T -state.
 - b. Time $[N/2]$: It contains the time allotted to each incoming lane.
 - c. T_E : This contains the time elapsed.
 - d. P : This will contain 1 if all the values of the status array are 1 and $T_E < T_{\text{max}}$.

Pseudocodes:

Algorithm initialize(N) // This algorithm initializes the Variables and Flags for a new Cycle

```
{
  D0-100[0 to (N/2-1)] = Deven[0 to (N/2-1)] = Dodd[0 to (N/2-1)] = 0;
  Status[0 to (N/2-1)] = 0;
  C = TE = P = Time[0 to (N/2-1)] = 0;
  if N > 10 then
    Tmax = 30 * N;
  else
    Tmax = 300;

  if i != j then
    Signal[i][j] = 0;
  else
    Signal[i][j] = -1;
}
```

Algorithm regulate() // This algorithm regulates the traffic flow, T = Equivalent Time of 1 T-State

```
{
  temp = max(D0-100);
  if Status[temp] == 1 then
    D0-100[temp] = 0;
    regulate();
  else
    Status[temp] = 1;
    Signal[temp][0 to (N/2-1)] = 1;
    if P == 1 || C == Tmax/T then
      initialize();
    else
      optimize(temp);
}
```

Algorithm optimize(x) // This algorithm optimizes the traffic flow by allowing other inactive lane's traffic to flow into the lane to which traffic of the currently active road is not moving into.

```
{
  for i = 0 to (N/2-1)
    if Deven[i] == 1 then
      Signal[x][i] = 0;
      checkPossibilities(Signal, I, D0-100);
    if D0-100[x] == 1 || TE == Time[x] then
      C++;
      TE = TE + T;
      regulate();
}
```

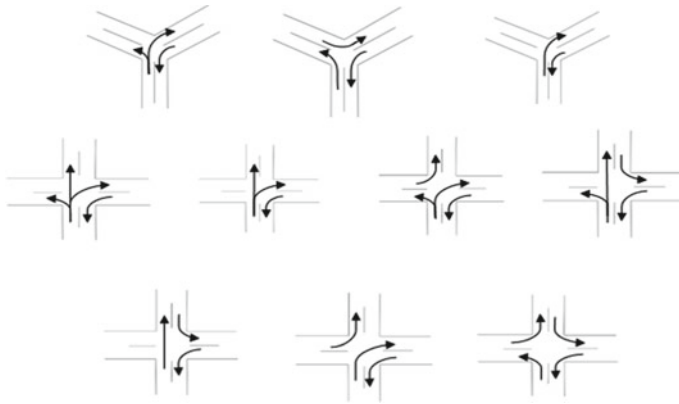



Fig. 3 These figures are to demonstrate all the possible combinations taking a single road in as the prime road. Since everything is generalized so are these combinations

Algorithm checkPossibilities (Signal, I , D_{0-100} []): This is a self-learning algorithm which has been pre-trained for all the possible combinations of traffic flow for smaller number of intersections, with the only constraint that “traffic cannot converge at a point but it may diverge from a point.” This will expand itself for a N number of intersections and the density will enable the unit to predict which possible combination will have maximum throughput.

G-STRA ideally is based on some assumptions. Those are:

1. The ideal flow of traffic is constant over a time with a speed of 20 m per 5 s.
2. The maximum or 100% of density means 200 m.
3. All the roads are of the same type.

All these assumptions will change its definitions based on our predictive and learning unit (Fig. 3).

4.3 Self-learning and Predictive Unit

Now, as mentioned in the title of our paper that G-STRA will be Self-Learning Predictive Algorithm. Based on the classifications (as shown in Fig. 4) and learning from the past experiences, the system will itself classify the various parameters it needs to decide. This makes the system self-learning and predictive. Moreover, since the parameters are assigned values real time, so the overall process makes the system dynamic and logical.

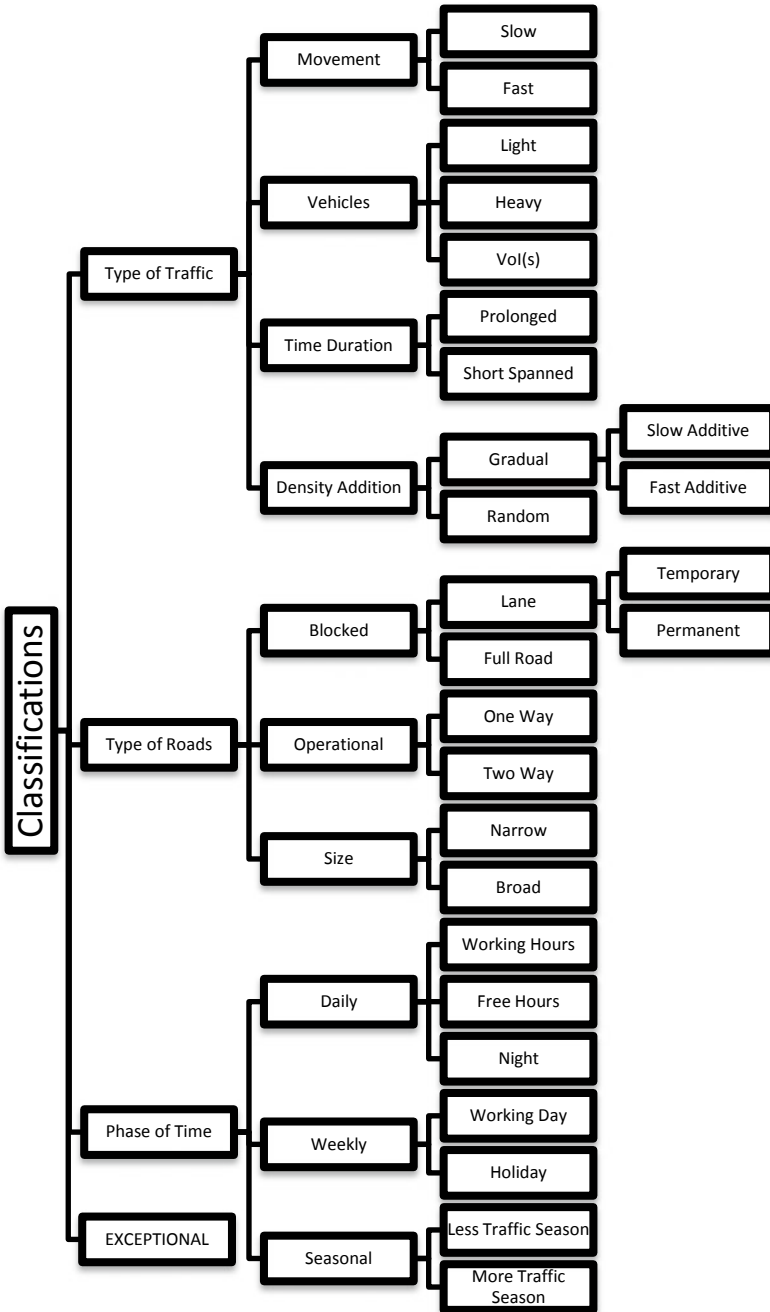


Fig. 4 Different types of classifications for the self-learning and predicting unit

4.4 *The Vehicle(s) of Importance Unit (VOI(s) Unit)*

Vehicles like ambulances, fire engines, and police vans (sometimes) need more attention and thus this unit is here to provide priority-based clearance. To make that thing possible, we are using multiple instance learning [9] which will classify the vehicles in lightweight, heavyweight, and VOI(s).

The system is going to use video tracking algorithm to track the VOI(s) to ensure that the vehicle has been given clearance and the purpose is served. Now, in case of several VOI(s) we can sort them and put them into a queue structure which our main algorithm will use to produce interrupts and give a green signal to the road the VOI is in. This again may cause disturbance in our traffic regulation so the sorting algorithm is also based on the density of the roads. The other factors taken into consideration are distance of the VOI from the intersection (farthest VOI will be given highest priority factor and vice versa) and destination of the VOI (if VANET is available).

The factors are evaluated and then given into a SVM model which will assign the total priority value of each VOI. This will make the system more robust and efficient.

5 Comparison

5.1 *Traditional or Hard-Coded Algorithm [10]*

This algorithm is the one used in today's day-to-day traffic control. This algorithm uses regular time intervals to move the traffic from the lanes.

1. An interval of time ' t ' is set within the lights to switch the signals.
2. After the first ' t ' time, one of the lanes is given a green signal, whereas the rest of the lanes are stopped.
3. The above step is repeated in a cyclic manner, i.e., the other lanes are allowed individually after the simultaneous ' t ' intervals.
4. The process continues until the traffic lowers down; the monitoring is done manually.

5.2 *Hard Coded Versus G-STRA*

Let us assume a situation of a five-road junction/intersection, which follows the following points:

1. Each lane has two-way flow.
2. The traffic densities for each lane are as follows:
 - a. Lane 1 \rightarrow 80; Lane 2 \rightarrow 55; Lane 3 \rightarrow 43; Lane 4 \rightarrow 77; Lane 5 \rightarrow 20.

3. Flow of the traffic is uniform in every lane, i.e., 2% of the traffic in the lane.
4. The check is going to stop when the density in every lane becomes zero.
5. Clock cycle is the unit time that the framework will require to calculate the densities, and this is the minimum time for the signals to remain unaffected.

Case 1: Ideal, the flow of density is equal for every lane. We derive the observation tables using (Tables 1 and 2):

Comparing the clock cycles:

Hard-Coded → 270 Clock Cycles **G-STRA** → **100 Clock Cycles.**

Case 2: Practical, the flow of density varies from lane to lane. We derive the observations, Table 3:

Comparing the clock cycles:

Hard-Coded → 270 Clock Cycles **G-STRA** → **100 Clock Cycles.**

Table 1 Hard-coded algorithm

Clock pulse	Signals (green only)	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	
0	Lane 1	80	55	43	77	20	
30	Lane 2	20	55	43	77	20	
60	Lane 3	20	0	43	77	20	
90	Lane 4	20	0	0	77	20	
120	Lane 5	20	0	0	17	20	
150	Lane 1	20	0	0	17	0	
180	Lane 2	0	0	0	17	0	
210	Lane 3	0	0	0	17	0	Stray
240	Lane 4	0	0	0	17	0	Stray
270	Complete	0	0	0	0	0	

Table 2 G-STRA

Clock pulse	Signals (green only)	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5
0	Lane 1 → All, Lane 5 → 1	80	55	43	77	20
40	Lane 4 → All, Lane 3 → 4	0	55	43	77	15
80	Lane 2 → All, Lane 1 → 2	0	55	32	0	15
110	Lane 3 → all, Lane 2 → 3	0	0	32	0	15
135	Lane 5 → all, Lane 4 → 5	0	0	0	0	15
145	Complete	0	0	0	0	0

Table 3 Calculated observations for variable load

Time	Signals	Lane 1					Lane 2					Lane 3				
<i>Hard-coded algorithm</i>																
0	Lane 1	1 → 2	1 → 3	1 → 4	1 → 5	Total	2 → 1	2 → 3	2 → 4	2 → 5	Total	3 → 1	3 → 2	3 → 4	3 → 5	Total
30	Lane 2	30	10	28	12	80	14	20	12	10	56	8	4	0	4	16
60	Lane 3	15	0	13	0	28	14	20	12	10	56	8	4	0	4	16
90	Lane 4	15	0	13	0	28	0	5	0	0	5	8	4	0	4	16
120	Lane 5	15	0	13	0	28	0	5	0	0	5	0	0	0	0	0
150	Lane 1	15	0	13	0	28	0	5	0	0	5	0	0	0	0	0
180	Lane 2	0	0	0	0	0	0	5	0	0	5	0	0	0	0	0
210	Lane 3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
240	Lane 4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
270	Complete	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
<i>G-STRA</i>																
0	4 → All; 3 → 4	1 → 2	1 → 3	1 → 4	1 → 5	Total	2 → 1	2 → 3	2 → 4	2 → 5	Total	3 → 1	3 → 2	3 → 4	3 → 5	Total
5	4 → All; 3 → 4; 1 → 2	30	10	28	12	80	14	20	12	10	56	8	4	0	4	16
		30	10	28	12	80	14	20	12	10	56	8	4	0	4	16

(continued)

Table 3 (continued)

Time	Signals	Lane 1					Lane 2					Lane 3				
		20	10	28	12	70	14	20	12	10	56	8	4	0	4	16
15	4 → 1, 3, 5; 3 → 4; 1 → 2															
40	4 → 1; 2 → 3, 4; 1 → 2	10	10	28	12	60	14	20	12	10	56	8	4	0	4	16
45	1 → All; 5 → 1	0	10	28	12	50	14	20	12	10	56	8	4	0	4	16
60	1 → 4, 5; 3 → 2; 5 → 1	0	10	28	12	50	14	15	7	10	46	8	4	0	4	16
65	1 → 4; 3 → 2; 4 → 1, 5	0	0	18	2	20	14	0	0	10	24	8	4	0	4	16
70	2 → All; 1 → 2	0	0	10	0	10	14	0	0	10	24	8	0	0	4	12
80	2 → 1; 1 → 2; 5 → 4, 3	0	0	0	0	0	14	0	0	10	24	8	0	0	4	12
85	5 → All; 4 → 5	0	0	0	0	0	4	0	0	0	4	8	0	0	4	12
90	5 → 2; 3 → 4, 5	0	0	0	0	0	0	0	0	0	0	8	0	0	4	12
95	3 → All; 2 → 3	0	0	0	0	0	0	0	0	0	0	8	0	0	0	8
100	Complete	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Time	Signals	Lane 4					Lane 5									

Hard-coded algorithm

Time	Signals	Lane 4					Lane 5					Total
		5 → 1	5 → 2	5 → 3	5 → 4	Total	4 → 1	4 → 2	4 → 3	4 → 5		
0	Lane 1	44	8	28	12	92	8	12	16	0	36	
30	Lane 2	44	8	28	12	92	8	12	16	0	36	
60	Lane 3	44	8	28	12	92	8	12	16	0	36	
90	Lane 4	44	8	28	12	92	8	12	16	0	36	
120	Lane 5	29	0	13	0	42	8	12	16	0	36	

(continued)

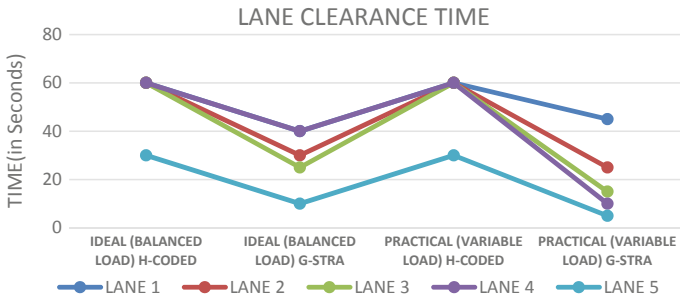


Fig. 5 Graph comparing time taken by lanes in different cases

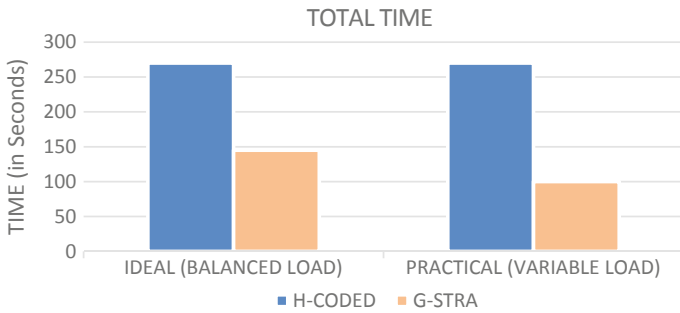


Fig. 6 Graph comparing time taken by the algorithm in different cases

This shows us that the proposed algorithm, i.e., G-STRA would perform even better for the cases in which the flow of densities is irregular for every pair of lanes (Figs. 5 and 6).

Such a huge difference has occurred due to the dataset selection, some of which already had '0' density flow from the very beginning. For more generalized datasets and the reduced traffic flow, the clock cycles may increase but the overall time would be much less when compared to hard-coded algorithm.

6 Conclusion

Summarizing all the points discussed in the previous sections of this paper, we have introduced a generalized smart traffic regulation algorithm that makes use of real-time image data of the concerned lanes to extract instantaneous traffic density characteristics and machine learning to learn patterns in density variations with respect to time of the day and year, weather and road conditions. One of the key inferences worth highlighting is that our algorithm performs increasingly better than the algorithm that is implemented currently in the traffic regulation systems, as the number of roads

at the intersection increases. On top of that, the algorithm works for any number of roads. However, the biggest selling point of this algorithm is that even though our algorithm requires a lot of set-up in comparison with the existing ones, in terms of taking frames from the CCTV cameras at the junction and extracting features from them to training our machine learning model with a good dataset, we just must set it up once. Our algorithm will learn from experience then onwards, and its predictions will get more accurate overtime. Moreover, it is adaptive to traffic conditions which seldom occur, thus making it very robust.

References

1. Wikipedia.org: https://en.wikipedia.org/wiki/Traffic_light
2. G.S. Khekare, A.V. Sakhare, Intelligent traffic system for VANET: a survey. *Int. J. Adv. Comput. Res.*, **2**(4, 6), 99
3. N. Abbas, M. Tayyab, M.T. Qadri, Real time traffic density count using imageprocessing. *Int. J. Comput. Appl.* **83**(9), 16–19 (2013)
4. A. Kanungo, C. Singla, A. Sharma, Smart traffic light switching and traffic density calculation using video processing, in *Proceedings of 2014 RAECS UIET Panjab University*, Chandigarh, Mar 2014
5. M. Behrisch, L. Bieker, J. Erdmann, D. Krajzewicz, SUMO simulation of urban mobility: an overview, *SIMUL 2011*, in *The Third International Conference on Advances in System Simulation* (Barcelona, Spain, Oct 2011), pp. 63–68
6. L.C. Bento, R. Parafita, U. Nunes, Intelligent traffic management at intersections supported by V2V and V2I communications, in *15th International IEEE Conference on Intelligent Transportation Systems Anchorage* (Alaska, USA, Sept 2012)
7. K. Pandit, D. Ghosal, H.M. Zhang, C. Chen-Nee, Adaptive traffic signal control with vehicular ad hoc networks. *IEEE Trans. Veh. Technol.* **62**(64), 1459–1471 (2013)
8. S. Van der Walt, J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, N. Yager, E. Gouillart, T. Yu, scikit-image: image processing in Python. *PeerJ* **2**, e453 (2014)
9. B. Babenko, M.H. Yang, S. Belongie, Robust object tracking with online multiple instance learning. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(8), 1619–1632 (2011)
10. P. Jadhav, P. Kelkar, K. Patil, S. Thorat, Smart traffic control system using image processing. *Int. Res. J. Eng. Technol.* **3**(3), 2278 (2016)

A Short Review on Applications of Big Data Analytics



Ranjit Roy, Ankur Paul, Priya Bhimjyani, Nibhash Dey,
Debankan Ganguly, Amit Kumar Das and Suman Saha

Abstract Big Data is a revolution in the twenty-first century which aspires to uplift the working, living and thought process by improving our decision-making capabilities. It is quite evident that the outreach of big data has spread across all scientific and engineering domains. With such a huge growth of data, utilizing it to its maximum potential requires modern approaches, novel learning techniques so that challenges related to such massive data sets can be addressed. We have tried to present a survey of how machine learning techniques are inter-related with big data analytics. We have also focused on some promising areas where big data has a key role like health care, banking and finance, security, aviation, astronomy, agriculture and natural calamities and disasters.

Keywords Big data analytics · Review · Machine learning

R. Roy · A. Paul (✉) · P. Bhimjyani · N. Dey · D. Ganguly · A. K. Das
Institute of Engineering and Management, Salt Lake, Kolkata, India
e-mail: ankurpaul2@gmail.com

R. Roy
e-mail: royranajit1@gmail.com

P. Bhimjyani
e-mail: priya.27bhimjyani@gmail.com

N. Dey
e-mail: nicks.dey@gmail.com

D. Ganguly
e-mail: debankanganguly18@gmail.com

A. K. Das
e-mail: amitkumar.das@iemcal.com

S. Saha
Cognizant Solutions Technology, Salt Lake, Kolkata, India
e-mail: suman7028@yahoo.com

1 Introduction

Big Data basically refers to the volume of data. The quantity of data is not only the sole factor we are concerned about, rather what the data generating organizations do with that data is important. It is evident that the era we are living in needs a push to overcome this imminent problem. Enormous amount of data is being continually generated at an extraordinary rate, as a product of advancement in Web technologies, social media, cell phones and sensing devices. Gathering and maintaining large collections of data might look quite easy, but deriving appropriate patterns or information from these assemblages is quite demanding and challenging. The traditional way of interpreting small and structured data was by trial-and-error analysis, but in the present scenario, quite a plethora of tools are being developed to cope with the huge unstructured data sets. Trial-and-error analysis becomes impossible when data sets are large and heterogeneous. For example, Twitter generates around 8 terabytes of data daily by processing 80 million tweets per day, and Walmart generates about 2.5 petabyte data per hour [1]. These large data sets, or in other words these Big Data, possesses astounding possibilities as a business model predictor in various fields such as health care, biology, transportation, online advertising, online marketing platforms, banking and financial services.

Recent studies reveal that majority of the world's data comes in the form of unstructured data. Unstructured data comprises of e-mails, word processing documents, presentations, photographs, audio files and videos. Big data may be explained in terms of five traits volume, veracity, variety, velocity and value. Volume depicts the size or the exponential growth of data (PB, EB and TB), velocity refers to the frequency or pace at which data is being created, stored and analysed. Variety is defined by the increased diversity of data, i.e. whether the data is structured or unstructured, veracity refers to the quality of the data or in other words the reliability or accuracy of the data because not all data which are gathered from various sources are reliable. The last 'V' or the value is the most significant V's of Big Data because Big Data will have importance if and only if the data being generated can be put to use by the organizations.

Machine learning is an area of research to develop various efficient algorithms to accurately predict variables on feeding data much like humans. This includes statistics, mathematics, artificial intelligence, etc. to better understand and predict qualified results [2, 3]. It is a technique by which machine learns by itself through some given data sets. This process begins with observations or data which are provided by users. These data sets are then analysed in an efficient [4, 5] way in search of new patterns in the data [6–8] and improve to make more accurate predictions in the future. Machine learning has an important role in Big data analytics [9, 10]. The basic goal is to allow computers or machines to learn without any human intervention. It can be classified into three types.

1.1 Supervised Learning

Supervised learning is a way in which classification of objects is done by comparing the data which is fed into the system. Data consists of several attributes such as colour, size, features and situations. These attributes are then compared on the mathematical and statistical basis to precisely classify the objects with specific features. Some of the algorithms are k-nearest neighbour, decision trees, Naïve Bayes, regression and many more.

1.2 Unsupervised Learning

Unsupervised Learning is a section of machine learning where the algorithm detects hidden patterns in data and segregates them into groups or clusters. This kind of algorithm is heavily used in data mining and data science world where some unknown conclusions can be achieved with unlabelled data. Some notable algorithms are k-means, hierarchical clustering and hidden Markov rule.

1.3 Reinforcement Learning

In this domain of machine learning, an object is kept in a spontaneous environment with which it maintains interaction and tries to get a maximum output for a result. This section of machine learning is called reinforcement learning. It is often used in gaming, navigation and robotics. Q-Learning, Sarsa are notable algorithms in this paradigm. Machine learning contributes a plethora of ways to make an approximate data model for specific data sets for precise predictions and analysis using probability and statistics.

Some of the tools of Big Data ecosystem are HDFS, YARN, MapReduce, HBase, Spark, Mahout, Hive, Pig, Zookeeper and many more.

Earlier, when data generation was less, relational databases could cope with those data sets, but with technological advancements, the rate data generation increases exponentially. This led to the emergence of Hadoop Distributed File System or HDFS. HBase is column-oriented NoSQL database which runs abreast HDFS. It eliminates the limitations of HDFS by allowing concurrent writes. For data insertion in HDFS, Sqoop and Flume are used, the latter for unstructured data such as logs files, event data, etc. and other former for relational databases. Hadoop Yarn provides flexible job scheduling and acts as resource manager in the Hadoop Distributed Processing Framework. Some of the higher-level programming models include Pig and Hive which have a similar goal—they are tools that help in writing complex Java MapReduce programs quite easily using scripting and SQL queries, respectively. Mahout is used to implement machine learning algorithms in Big data. And Zookeeper is a cen-

tralized management system for providing synchronization, maintaining distributed configuration service and naming registry for distributed systems.

MapReduce is a programming model that supports all the requirements of big data modelling like processing of big data, splitting complications into different parallel tasks, etc., i.e. it simplifies parallel computing. Google uses MapReduce for indexing of websites

These components serve as great tools for the control and maintenance of Big data and its analysis using efficient machine learning techniques to overcome the challenges [3, 11].

2 Related Works

During the last two decades, the overwhelming growth of data has led to the rise of Big Data in almost every aspect we come across in our lives. To cope with this huge amount of data, researchers have come up with new machine learning algorithms, incorporated ideas about image processing, artificial intelligence, bio-informatics, data analytics, etc.

Few of the domains which have come to the limelight in recent years have been discussed below.

2.1 Health Care

Big Data has brought about a revolution in the health care sector across many countries around the world. The industry pursues to become a system, focused on value i.e. providing better quality care at a reasonable cost. In health care domain, Big data has stimulated the process to identify and treat illness, enhance standard of life and avoid preventable deaths [12, 13]. With recent advancements of genomic research, machine learning on genomic data can provide insights over a patients health using real-time medicines [14–16].

With increased life spans and an outburst of population, this industry faces challenges in providing proper and personalized treatment to the patients. In the present scenario, our goal is to study a patients history and then raise necessary warnings if there is any probability of serious illness that he might face in the near future and henceforth do the treatment. Few devices have been developed which shall take care of patient monitoring to a new level. For example, Asthmapolis has fabricated a GPS-enabled tracker that keeps a record of inhaler usage of asthma patients. This valuable information is then directed to a central database management system which then analyses and brings in new insights from the data to meet the necessities of that individual [17]. The Centers for Disease Control and Prevention information use

this information to identify the places with high pollutant levels like volcanic fog in Hawaii or high counts of pollen in the north-east.

Another company named Ginger.io has devised a mobile application where people allow themselves to be tracked through their mobile phones and in turn get assisted with behavioural health therapies [18]. This app keeps a record of the calls, texts, current location and physical movements. Moreover, patients also respond to the regular-based surveys conducted by the company over their smartphones.

New strategies are being developed for predicting mortality over provided symptoms of pneumonia [19]. Big data and machine learning can predict post-operative symptoms of a patient [20] and can be even used in radiation oncology [21]. The patient data is then integrated with the research made on public health by the National Institutes of Health and other sources. The insights obtained may reveal, for instance, a lack of movement, or improper exercise patterns or other activity which could lead to a person feeling physically unwell. There might also be irregular sleeping patterns which are revealed through late-night calls or texts causing an anxiety attack and many more psychological illnesses [22] (Fig. 1).

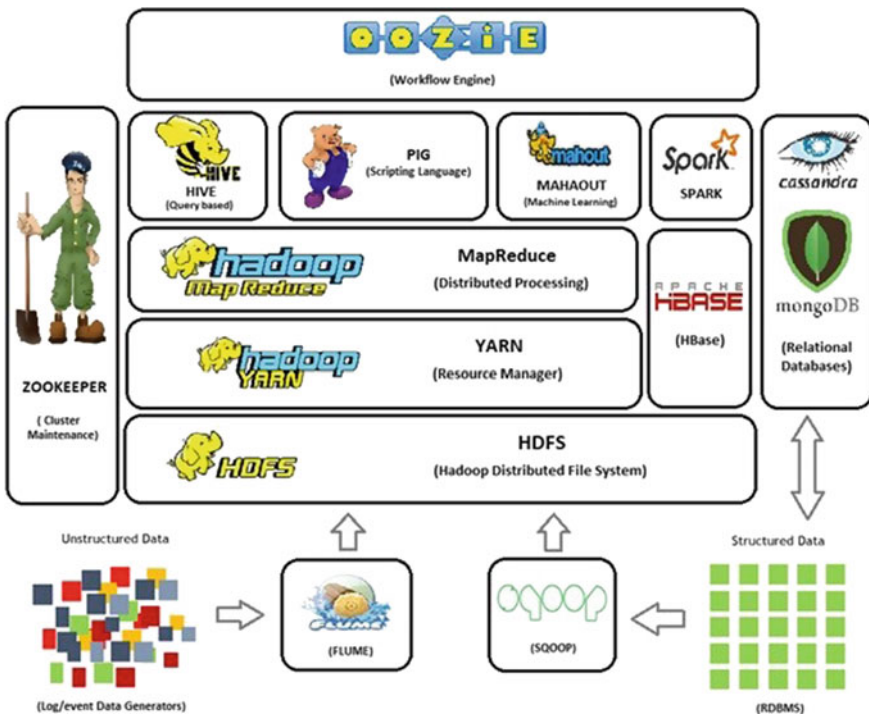


Fig. 1 Hadoop ecosystem

2.2 *Banking, Finance and Security*

Big Data is a great source of information for risk analysts. In order to evaluate risk and to get efficient business planning, a complete market insight is necessary. A more updated data is necessary to assess business models and for accurate analysis. But maintaining and getting such large updated data is itself a challenge [23]. For both banks and consumers, credit risk is an important hurdle, due to which reliability of the bank gets affected. Machine learning algorithms with Hadoop ecosystem can provide banks and different financial institutions, an opportunity to derive questionable insights in loan data to check for reliability and performance of that lending institutions [24].

Processing of data is a challenging task. It can be highly paralyzed but processing such a large amount of data is computationally intensive work. Data pre-processing is a must as some parts of the data might be missing. For this, many financial institutions are leaning towards huge data centres to allocate their data banks for processing. Processing this data can lead to new market insights and viable investments. But reserving a part of the revenue for these institutions for this emerging technology is a difficult decision. Investment banks have been more gradual in their grasping of big data than consumer retailers. Goldman Sachs has taken the lead with a 15 million dollar investment in big data analytics start-up Kensho [25].

One of the most debated problems is network intrusion. This breach can cause a leak in classified data which can a huge problem. Shan Suthaharan provided an extensive analysis of the problem using Big data and machine learning techniques for network classification [26]. Use of SVM and other machine learning algorithm can provide a secure environment for general computing over Big data [27, 28]. This distributed system can have a data breach which leads to privacy concerns. Machine learning provides several approaches for finding solutions for such problems. The Apache Mahout supports mainly two machine learning paradigms which are supervised learning and unsupervised learning. Supervised learning is a way in which objects are classified into different categories with the help of different attributes in a data set. Classifying e-mail messages between spam and safe, different recommendation systems like for music, movies, ads, Web blogs, etc. classification of patient data are well-used examples of supervised learning. Unsupervised Learning is a section of machine learning paradigm where the algorithm detects hidden patterns in data and used for segregating different objects into groups or clusters. This paradigm has a useful technique to reduce the complexity supervised learning by eliminating unwanted attributes from a data set. Mahout uses these paradigms to implement machine learning in big data analytics [29].

2.3 *Agriculture*

Smart farming is a breakthrough in technological advancement in the field of agriculture. Technologies like the IOT, cloud computing, machine learning and Big Data are incorporated into Smart Farming. Farmers have a better vision of the consequences of their actions which facilitate them to implement improved farming practices [30]. Big data combined with various machine learning techniques plays a major role in providing predictive insights in farming operations and introducing a overhaul to the processes involving farming for modern business models.

For instance, SatSure, a company founded by Abhishek Raju, works on using Big Data, IOT, machine learning to improve the lives of the Indian farmers. In India, agriculture is quite disorganized and cash oriented with the absence of electronic transaction, everything remains unrecorded. The company aims to provide insights about farm productivity so that farmers will have an idea as to when to sow, or harvest or even which part of the land can be used for farming. Provisions, like satellite-based monitoring, wind direction predictor, embedded sensors on crops and fields, GPS-enabled tractors, fertilizer requirement notifications, are rich sources of data which are used for improving agricultural practices in India. Apart from these Big Data analytics monitors, the growth rate and nutrient requirement for individual plants. Information on soil health, water availability and rainfall pattern enables farmers to decide which crop to plant for their next harvest [31].

2.4 *Aviation*

Traditionally where aircrafts captured only 125+ flight parameters, the Boeing 787 captures more than 1000 parameters thus adding up to half a terabyte of data every flight. These figures clearly mention the outbreak of data in aviation domain [32] and productive analysis should be done with this data [33]. Today, there are about thousands of airline companies in the world competing each other to get a firm grip in the aviation industry and business. Each of these industries generates revenues in the order of billions of dollars. Some of them, worthy of notice, are Delta Airlines, American Airlines and Qatar Airways and so on. But their success comes upon proper usage and application of appropriate technology on airline systems such as in planes and in air traffic control bases. The equipment used must be well equipped with all the necessary technologies to ensure the best and optimum performance [34]. Machine learning algorithms have played a key role in optimizing the performance of airlines and hence gave the industries a whole different level to be on. Machine learning in the field of artificial intelligence has always a key role in almost every technology that is currently under mankind's realisation—be its voice recognition, credit card fraud detection systems, etc. But even in a field like aviation, where human intervention is nearly inevitable, machine learning, still has managed to play a very important and dominant role in improving the performance of aviation industries.

Some of the questions that could arise with regard to this are:

1. How are these industries using machine learning to better their performance?
 2. What are some of the tangible results of incorporating machine learning in aviation?
 3. What is the possible future of using machine learning in aviation, if any?
- **AI Assistants:** Responding to customer inquiries through voice recognition for domestic flight information and ticket availability and this is achieved through natural language processing.
 - **Smart Logistics:** Machine learning algorithms are now applied to automate airline operations.
 - **Facial Recognition:** Machine learning algorithms are now being used for facial recognition for customer identity verification and matching the luggage to the respective customers.

Machine learning has been used to calculate the trajectory of flights in the air. The exact trajectory calculation for flights is very important with regard to safe air traffic. In order to cater the problem of heavy air traffic, plans have been developed for future purposes by Single European Sky ATM Research (SESAR) and Next Gen that depend on 4D trajectory calculations.

2.5 *Social Media and Advertisements*

Due to the emergence of technologies and smartphones, we have come to a point where we see everyone of us are connected to any one of the social media platforms it may be Facebook, Twitter, Instagram, Web blogs, etc. Social media is one domain where everyday petabytes of data are generated.

Public opinions can be gathered from this social media [36]. Data collected from online sources contains various formats like text, images, videos, logs, etc. Among these the most commonly used format is text. With these text data, many social and political predictions can be made, as these contain elements like Facebook status, tweets from Twitter, different reviews from product or online media. To process such a huge pile of data, a large unstructured database of words is maintained called text corpus. This provides precise prediction from a large amount of text data. Different machine learning algorithms can be used to detect spam in these data sets [37] (Fig. 2).

Online advertisement industry is a prime example for a company to have data analytics as a huge part of it. Irrelevant advertisements do not leave an impact on the user. So different machine learning algorithms are deployed for real-time analysis of the content that a user is watching and recommending relevant ads towards the user [38]. YouTube uses such framework to recommend ads to its user base. Another side to these companies is to predict the ad revenue with relevancy of an ad using machine learning [39]. This creates an estimation of how much audience an advertisement will reach.

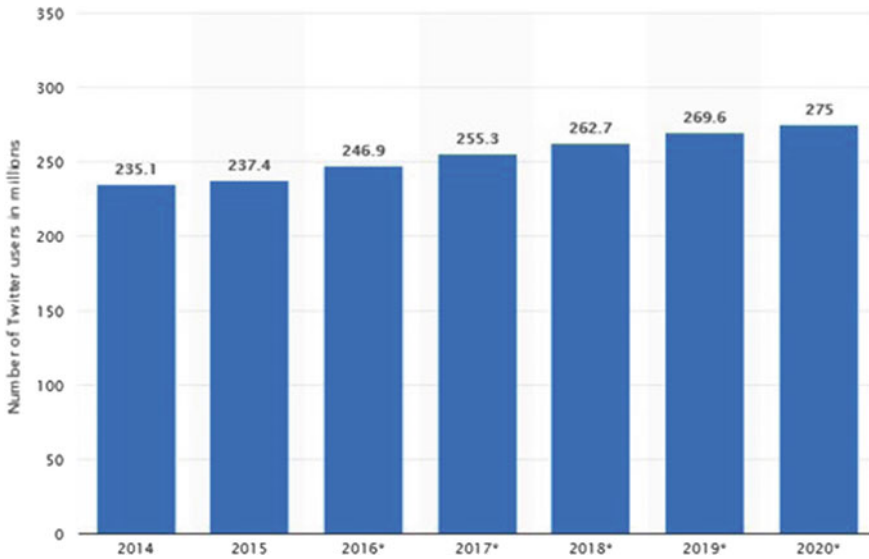


Fig. 2 Increment Twitter users over the years [35]

2.6 Astronomy

In the present scenario, due to the evolution of a wide range of surveys in the ground as well as space, astronomy has entered into the era of big data. SDSS or The Sloan Digital Sky Survey is known to have done one of the largest astronomical surveys till date the SDSS telescope every night produces about 210 gigabytes of data, and moreover, till now a million field images have been captured, about 250 million galaxies and several stars have been detected [40]. Proposed ML algorithms such as specially long short-term memory (LSTM), recurrent neural network (RNN) model can help processing the data in real time which can be used for activity forecasting [41].

- **Astrostatistics:** It is a study of complex data sets which are generated by automated scan of cosmos. This study includes data mining and statistical analysis to find linkage between astrophysics and generated data. Multivariate classification and multivariate regression, time series analysis, nonparametrics and Bayesian inference are included in the statistical analysis [42].
- **Astroinformatics:** This domain of study contains astronomy, astrophysics and informatics which ensures the solution of upcoming big data challenges. Astroinformatics also includes a set of naturally related specialities like data organization, data description, astronomical classification taxonomies, etc. [43].

SKA or The Square Kilometre Array is a modern and highly technologically advanced global radio telescope that is on the verge of completion; the telescope

Table 1 Various algorithms in the field of astronomy

Machine learning techniques	Approach	Applications in astronomy
1. Regression	<ul style="list-style-type: none"> – K-nearest neighbour – Support vector regression (SVR) – Artificial neural networks (ANN) – Random forest – Decision trees 	<ul style="list-style-type: none"> – Stellar physical parameter measurement – Photometric redshifts (galaxies, quasars)
2. Clustering	<ul style="list-style-type: none"> – K-means – Principal component analysis (PCA) – Cobweb – OPTICS 	<ul style="list-style-type: none"> – Special object detection
3. Feature selection	<ul style="list-style-type: none"> – Greedy stepwise – Exhaustive search – Best first – Race search – Rank search – Random search 	<ul style="list-style-type: none"> – Reducing dimension – Choose effective features

dish array is expected to reach South Africa by end 2020. This telescope has been designed and developed by 11 countries in the world. Several thousand radio antennas will be installed in a square kilometre land in Southern Africa once the telescope is launched. The telescope will cater to questions related to the origin and evolution of the universe. It is going to produce images with 50 times more resolution than famous Hubble Telescope. The telescope itself can produce upto 0.3–3 TB of data per second which can lead up to 1 exabyte data every year [44]. Two of the machine learning algorithms which came into limelight for processing data in real time are recurrent neural network (RNN) and long short-term memory (LSTM) model which can be used for activity forecasting [41].

Table 1 shows all the machine learning techniques that is implemented in different fields of astronomical research [45].

2.7 Natural Disasters

Every year, thousands of people die due to frequently occurring regional natural disasters. From 1995 to 2014, according to statistics although having been victims of 26% of the storms, around 89% of the fatalities due to storm have occurred in the developing or underdeveloped countries. With changing climate and increasing

extremities of weather, the number of victims will increase substantially [46]. A new project by DataKind and the World Bank's Global Facility for Disaster Reduction and Recovery (GFDRR) intends to help build resilience for communities that are menaced by natural disaster every year by using data appropriate advanced technology to operate on that data.

Satellite imagery with machine learning can help determine people at risk. With these data in hand, description of structure and buildings and their location can be identified during a natural disaster. By knowing the building location, the current technology goes a step further by knowing the type of the building. This information, when made available to GFDRR, helps it to distribute required resources to the distraught victims.

Whenever we want to know or track the forthcoming events, we go for predictive analysis—a characteristic of machine learning. The task that machine learning does is that it searches for patterns and similarities and accordingly analyses information to know about the possibilities of the same occurring in future, and moreover, it can also help provide details about the specific domain—centric areas. For example, in case of any natural calamities, predictive analytics can provide information about any particular place which may be less affected—thus in such situations, rescue teams may be sent to these places. Accordingly, the disaster management team can also be notified about the occurrence of such events as, and to how much extent the area has been affected. There are numerous organizations that allow authorities to scan through areas affected by tragedies. One such organization is Optima Predict. Machine learning along with image processing helps in providing a heat map of the affected area, thus helping the disaster management team in providing immediate response and supplies [47]. What we discussed here is just one of the many tasks which machine learning can do. Government authorities should look for ways through which they can use this technology.

3 Summary Table

Big data has touched almost every domain one could think of. Here goes a quick summary to enlist a few domains and how big data is applied there (Table 2)

4 Conclusion

The outreach of Big Data has spread across all science and engineering domains. Major changes are anticipated by acquiring knowledge and deriving new insights from these data sets. These high-volume, high-velocity, high-variety data cannot be processed by the prevailing ML algorithms. Hence, machine learning algorithms need to reinvent itself to tackle the shortcomings of such large data sets. This area of research is now interlinked with many other fields such as image processing, genetic

Table 2 Summary table

Domain	Applications
1. Health care	<ul style="list-style-type: none"> – GPS-enabled tracker for tracking usage of inhalers for asthma patients [17] – Real-time alerting to monitor health conditions of patients using various apps [18]
2. Banking and finance	<ul style="list-style-type: none"> – Risk management [23–25] – Fraud detection and security [26–28]
3. Agriculture	<ul style="list-style-type: none"> – Smart Farming [30] – SatSure – Satellite-based monitoring – Wind direction predictor – GPS-enabled tractor [31]
4. Astronomy	<ul style="list-style-type: none"> – Special/rare object detection [45] – Photometric redshifts (galaxies, quasars) [45] – Astrostatistics and astroinformatics [16, 42] – Square Kilometre Array (SKA) [41, 44]
5. Aviation	<ul style="list-style-type: none"> – Facial recognition – AI assistants – Smart logistics automation of airlines [34]
6. Social media	<ul style="list-style-type: none"> – Spam detection [37] – Microtargeting via social media advertising [38] – Analysing unstructured data [36] – Prediction of advertisement revenues [39]
7. Natural calamities and disasters	<ul style="list-style-type: none"> – Predictive analytics – Tracking impacts of disasters and monitoring recovery efforts [46] – Determining the exposure of human societies to disaster risk [47]

algorithms, deep learning and AI. With the advent of newer and modern technologies, the rate at which data is being generated is going on increasing with every passing day, so the need for progress in this field is on the rise. The overall objective of this survey is to analyse how big data is playing a major role in various domains and also portrays some of the optimized solutions which are currently being used to handle these data sets.

References

1. B. Marr. Really big data at walmart: real-time insights from their 40+ petabyte data cloud. *Forbes*, Jan 23, 2017, <https://www.forbes.com/sites/bernardmarr/2017/01/23/really-big-data-at-walmart-real-time-insights-from-their-40-petabyte-data-cloud/>. Accessed 2 July 2018
2. L. Zhou, S. Pan, J. Wang, A.V. Vasilakos, *Neurocomputing* **237**, 350 (2017)
3. S. Athmaja, M. Hanumanthappa, V. Kavitha, in *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*, *IEEE* (2017), pp. 1–4
4. O.Y. Al-Jarrah, P.D. Yoo, S. Muhaidat, G.K. Karagiannidis, K. Taha, *Big Data Res.* **2**(3), 87 (2015)
5. F. Sun, G.B. Huang, Q.J. Wu, S. Song, D.C. Wunsch II, *IEEE Trans. Syst. Man. Cybern. Syst.* **47**(10), 2625 (2017)
6. J.L. Torrecilla, J. Romo, *Stat. Probab. Lett.* **136**, 15 (2018)
7. R. Elshawi, S. Sakr, D. Talia, P. Trunfio, *Big Data Res.* (2018)
8. J.L. Berral-García, in *2016 18th International Conference on Transparent Optical Networks (ICTON)*, *IEEE* (2016), pp. 1–4
9. R. Swathi, R. Seshadri, in *2017 International Conference on Intelligent Computing and Control Systems (ICICCS)*, *IEEE* (2017), pp. 204–209
10. J. Lakshmi, A. Sheshasaayee, in *2015 International Conference on Green Computing and Internet of Things (ICGCIoT)*, *IEEE* (2015), pp. 480–484
11. A. Rathor, M. Gyanchandani, in *2017 International Conference on Electrical, Electronics, Communication, Computer, and Optimization Techniques (ICEECCOT)*, *IEEE* (2017), pp. 1–7
12. M. Chen, Y. Hao, K. Hwang, L. Wang, L. Wang, *IEEE Access* **5**, 8869 (2017)
13. W. Raghupathi, V. Raghupathi, *Health Inf. Sci. Syst.* **2**(1), 3 (2014)
14. F.S. Collins, E.D. Green, A.E. Guttmacher, M.S. Guyer, *Nature* **422**(6934), 835 (2003)
15. K. He, D. Ge, M. He, *Int. J. Mol. Sci.* **18**(2), 412 (2017)
16. J. Kim, P.W. Groeneveld, in *Big data, health informatics, and the future of cardiovascular medicine* (2017)
17. Z. Obermeyer, E.J. Emanuel, *N. Engl. J. med.* **375**(13), 1216 (2016)
18. *Ginger.io*, <https://www.ginger.io/>. Access 4 July 2018
19. G.F. Cooper, C.F. Aliferis, R. Ambrosino, J. Aronis, B.G. Buchanan, R. Caruana, M.J. Fine, C. Glymour, G. Gordon, B.H. Hanusa et al., *Artif. Int. Med.* **9**(2), 107 (1997)
20. S. Kobashi, B. Hossain, M. Nii, S. Kambara, T. Morooka, M. Okuno, S. Yoshiya, in *2016 International Conference on Machine Learning and Cybernetics (ICMLC)*, *IEEE*, vol 1 (2016), pp. 195–200
21. J.E. Bibault, P. Giraud, A. Burgun, *Cancer Lett.* **382**(1), 110 (2016)
22. R. Iniesta, D. Stahl, P. McGuffin, *Psychol. Med.* **46**(12), 2455 (2016)
23. A. Srivastava, S.K. Singh, S. Tanwar, S. Tyagi, in *2017 3rd International Conference on Advances in Computing, Communication & Automation (ICACCA)(Fall)*, *IEEE* (2017), pp. 1–6
24. S. Yadav, S. Thakur, in *2017 2nd International Conference on Telecommunication and Networks (TEL-NET)*, *IEEE* (2017), pp. 1–8
25. A. Munar, E. Chiner, I. Sales, in *2014 International Conference on Future Internet of Things and Cloud*, *IEEE* (2014), pp. 385–388
26. S. Suthaharan, *ACM SIGMETRICS Perform. Eval. Rev.* **41**(4), 70 (2014)
27. K. Xu, H. Yue, L. Guo, Y. Guo, Y. Fang, in *2015 IEEE 35th International Conference on Distributed Computing Systems*, *IEEE* (2015), pp. 318–327
28. M. Mayhew, M. Atighetchi, A. Adler, R. Greenstadt, in *MILCOM 2015-2015 IEEE Military Communications Conference*, *IEEE* (2015), pp. 915–922
29. J. Singh, in *2014 Conference on IT in Business, Industry and Government (CSIBIG)*, *IEEE* (2014), pp. 1–4
30. M.S. Moran, P. Heilman, D.P. Peters, C. Holifield Collins, *Ecosphere* **7**(10), e01493 (2016)
31. R.H. Ip, L.M. Ang, K.P. Seng, J. Broster, J. Pratley, *Comput. Electron. Agric.* **151**, 376 (2018)
32. A.M. Chandramohan, D. Mylaraswamy, B. Xu, P. Dietrich, in *2014 IEEE International Conference on Cloud Computing in Emerging Markets (CCEM)*, *IEEE* (2014), pp. 1–6

33. S. Ayhan, J. Pesce, P. Comitz, D. Sweet, S. Bliesner, G. Gerberick, in *2013 Integrated Communications, Navigation and Surveillance Conference (ICNS)*, IEEE (2013), pp. 1–13
34. T. Armes, M. Refern, in *2013 IEEE AUTOTESTCON*, IEEE (2013), pp. 1–5
35. Number of twitter users worldwide from 2014 to 2020 in millions, <https://www.statista.com/statistics/303681/twitter-users-worldwide/>. Access 14 July 2018
36. F. Shaikh, F. Rangrez, A. Khan, U. Shaikh, in *2017 International Conference on Intelligent Computing and Control (I2C2)*, IEEE (2017), pp. 1–6
37. S. Sharmin, Z. Zaman, in *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, IEEE (2017), pp. 137–142
38. R.V. Kaushik, R. Raghu, L.M. Reddy, A. Prasad, S. Prasanna, in *2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS)*. IEEE (2017), pp. 2434–2437
39. İ. İşlek, E. Karamatlı, A.T. Cemgil, in *2018 26th Signal Processing and Communications Applications Conference (SIU)*, IEEE (2018), pp. 1–4
40. J. Kremer, K. Stensbo-Smidt, F. Gieseke, K.S. Pedersen, C. Igel, IEEE Intell. Syst. **32**(2), 16 (2017)
41. L.X.Y. Yan, in *2017 IEEE Visual Communications and Image Processing (VCIP)*, IEEE (2017), pp. 1–4
42. Carnegie mellon university: Statistics and data science, <http://www.stat.cmu.edu/research/cross-disciplinary-research/262>. Access 14 July 2018
43. K.D. Borne, arXiv preprint [arXiv:0909.3892](https://arxiv.org/abs/0909.3892) (2009)
44. R. Spencer, in *IET Seminar on Data Analytics 2013: Deriving Intelligence and Value from Big Data*, IET (2013), pp. 1–26
45. Y. Zhang, Y. Zhao, Data Sci. J. **14**, (2015)
46. Global facility for disaster reduction and recovery (gfdrr), <https://www.gfdrr.org/en>. Access 15 July 2018
47. M. Arslan, A.M. Roxin, C. Cruz, D. Gin hac, in *2017 13th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, IEEE (2017), pp. 370–375

A Survey of Music Recommendation Systems with a Proposed Music Recommendation System



Dip Paul and Subhradeep Kundu

Abstract With the advent of digital music and music-streaming platforms, the amount of music available for selection is now greater than ever. Sorting through all this music is impossible for anyone. Music recommendation systems reduce human effort by automatically recommending music based on genre, artist, instrument, and user reviews. Although music recommendation systems are widely used commercially, there does not exist any perfect recommendation system that can provide best music recommendation to the user with the minimal user effort. In this paper, we reviewed the various recommendation systems that are currently in use including content-based, collaborative, emotion-based, and other techniques. We have also explored the strengths and weaknesses of each recommendation technique and at the end, we have provided an overview of a music recommendation system that may solve many of the challenges that existing recommendation systems face through an improved hybrid recommendation system.

Keywords Recommendation system · Metadata · Collaborative filtering · Content-based filtering

1 Introduction

The access to large amounts of music online has made it increasingly difficult to properly enjoy so much content. Music recommendation systems help users to find songs that they may like. Music recommendation systems aim to provide real-time recommendations of both new and old songs to the users. Music-streaming platforms like Spotify and Apple Music use recommendation systems heavily.

D. Paul (✉) · S. Kundu
Institute of Engineering and Management, Kolkata, West Bengal, India
e-mail: Pauldip04@gmail.com

S. Kundu
e-mail: subhradeep311@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_26

Music is an integral part of human life. People responds to music better than any other form of media, so recommending songs properly and efficiently has become important. The challenge of music recommendation systems is that human beings are unpredictable, and they may like one music today but may like an entirely different music tomorrow. It is impossible to identify users real-time reactions [1]. All recommender systems, at their basic form, rely on user behavior by recording activities of users who interact with a service or system, or by simply asking users for their preference [2].

Popular recommendation models that are currently in use are content-based filtering, context-based filtering, and metadata-based model. We are also seeing music being recommended based on mood, emotions, and social media interactions of the user. In this paper, we first explained popular music recommendation models that are in use and their challenges, and in the later part of the paper, we proposed a recommendation model that aims to overcome some of the challenges that current recommendation models face.

2 Methodology

We first searched online about various music recommendation systems and how they use user data to recommend songs. Then, we studied articles that provided us with different music recommendation systems. We studied the methods they described in the articles for writing this paper. We focused on the existing recommendation systems and their shortcomings and studied articles that can help us to give an idea of a recommendation system that can solve many of the shortcomings of the existing recommendation systems.

3 Music Recommendation Techniques

Commercially, the music-streaming sites like Spotify, Apple Music, and Pandora use different recommendation techniques to recommend music to its users. Broadly speaking, they use hybrid recommendation model or a recommendation model that is a combination of one or more recommendation technique.

3.1 Collaborative Filtering

Collaborative filtering is a method of generating automatic responses or predictions about the interests of a user by collecting preferences or taste information from many users. This filtering technique uses user ratings for its recommendation. Collaborative

systems are built on the assumption that users who rate items similarly in the past will continue to rate them similarly in the future [3].

This filtering technique uses k -nearest neighbor algorithm to provide recommendations. Ratings can be divided into two categories—explicit or implicit. Examples of explicit ratings are one-to-five-star rating systems that e-commerce sites use. These ratings are explicitly provided by the users. Implicit ratings can be obtained by interpreting user behavior. Play counts can be used for implicit rating. A song which is being played multiple times will automatically get higher implicit rating.

The biggest drawback of this system is that at early stages it provides poor recommendation. Especially, for items with very few ratings, recommendations performed in a fashion as outlined are not very reliable [4]. This is known as the cold-start problem. When a new user enters the system, the system cannot give effective recommendation as the user has not rated anything yet, so the system does not know what to recommend. Human effort is another challenge for this system. The more effort it will take to generate a recommendation, the less the users will be willing to rate.

3.2 *Content-Based Filtering*

In content-based filtering technique, songs are recommended based on the comparison done by the system between the content of the items and a user profile.

Several issues must be considered when implementing a content-based filtering system. First, terms can either be assigned automatically or manually. When terms are assigned automatically, a method must be chosen that can extract these terms from items. Second, the terms must be represented such that both the user profile and the items can be compared in a meaningful way. Third, a learning algorithm must be chosen that is able to learn the user profile based on seen items and can make recommendations based on this user profile.

The content of each item is represented as a set of descriptors or terms, typically the words that occur in a document. Acoustic features of the song like loudness, tempo, rhythm, and timbre are analyzed to recommend songs. Most common methods to compute similarity are: K-means clustering [5] and expectation-maximization with Monte Carlo sampling. This technique solves the cold-start problem as it can recommend songs based on very few data.

The major drawback of content-based model is that it relies on the correctness of the item model [6]. It also faces glass-ceiling effect. Another major drawback is that this technique fails to differentiate important differences between otherwise similar songs.

3.3 Metadata-Based Filtering

Metadata-based filtering uses metadata of a song like artist name, genre, and album name. This system uses metadata to recommend songs to the users. It is the most basic and traditional form of filtering technique.

The recommendation results are relatively poor, since it can only recommend music based on editorial metadata, and none of the user's information has been considered [6].

3.4 Emotion-Based Filtering

Music and human emotions are closely connected, so the recommendation model that considers human emotions is emotion-based filtering. Both in commercial and academic sectors, huge research is ongoing about music and its impact on human emotions.

Different acoustic features of the song are used to determine emotions that a song may trigger. Research has also shown that user's mood also plays a key role in selecting the songs [7]. Music-streaming sites create playlists based on human emotions and moods to better suit an emotion that a listener might feel. Recommendation system based on emotion can provide highest satisfaction to the listeners.

The biggest advantage of emotion-based filtering model is also its biggest disadvantage as this model requires huge data collection, huge number of datasets require a lot of human effort [8]. Another drawback is that one song may create different feelings to different persons, and this results in ambiguity of the datasets.

3.5 Context-Based Model

Context-based model uses public perception of a song in its recommendation. It uses social media sites like Facebook, twitter, and reddit and video platforms like YouTube to gather information about public perception of a song and recommend them accordingly to the listeners. It uses users' listening history to gather information about the user and recommends similar songs based on the engagement the songs are seeing in the social media sites. This model can behave efficiently with small amount of data. Platforms like Apple Music and Spotify use top charts or similar methods, where songs that are listened most by its entire user-base is reflected as well as the songs that see most social media engagements are recommended on that list. Context-based model can create a For You section for the user based on users listening history and social media engagement of different songs. Another method of context-based model uses location of the user to recommend songs. Listeners of

the same region may tend to like similar songs and through this method the system recommends songs.

Research has suggested that this model performs well due to the collection of social information [7]. Context-based model can recommend better than other recommendation models with few data as it uses social media sites to gather the songs that are currently popular around the globe.

4 Commercially Used Models

Commercially, a combination of these models along with several other parameters is used in recommendation systems.

4.1 Hybrid System

A hybrid model recommendation system uses a combination of previously mentioned recommendation models. It can recommend songs far more efficiently than a system that uses only a single recommendation technique.

An example of a hybrid system is a system that combines content and collaborative techniques. These models can mitigate the shortcomings of each recommendation models.

4.2 Listening History

Users' listening history is another important parameter that the music platforms use to recommend songs. A user is most likely to listen to similar songs and maybe even the same song, so in the commercial sector, listening history plays a critical role in recommendation models.

Different techniques are available to extract the information from the users' listening history, and one such approach is breaking down the entire listening history in sessions [9]. This can provide information on the songs that the user listens in succession. Breaking down the histories also gives an idea to the recommendation model about the difference between the user's short-term and long-term preference [10].

4.3 *User-Centric Experience*

Not a single person's music preference is identical to another person, but traditionally music recommendation models measured performance of new systems through comparing new datasets to existing datasets [11]. So, the modern music-streaming sites are using recommendation techniques that create a user-specific experience by understanding the music-listening habits of the users.

5 Proposed Model

Commercial recommendation models use a variation of hybrid recommendation systems. When a user first signs up with their email address, recommendation system works by asking the user to input some artists that he/she may like than with this user selection and location data the songs are recommended to the user. Then over time, as the user listens to more and more songs, the recommendation system uses the user's listening history to recommend songs.

We propose a slightly different recommendation model. We propose a recommendation model where initially during the sign-up process the system will ask the user for information like age, gender, location, and then music preference like which language of music user likes, what genre they prefer, and then the artists the user likes. Using this data, the system will initially be recommended songs in the following way to solve cold-start problem. System will categorize the user by their age groups. For example, if the user falls under 18–35 age demographic, chances are that the user will like to listen to currently popular songs. So, the system will initially offer currently popular songs by analyzing music charts like Billboard or equivalent charts and using social media sites. Through these combinations of metadata filtering, context-based model, and content-based model, the system will solve the cold-start problem. Then over time, as the system learns about the listening habits and listening patterns and creates an emotion-based model of the user than accordingly, the system will try to recommend a user-specific experience.

Several issues remain with this type of recommendation system. Firstly, human beings are unpredictable. A person in one age demographic may prefer a song that the system thinks the person will not like; so, this recommendation model cannot account for the unpredictability factor.

6 Conclusion

Through the course of working on this paper, we studied different music recommendation systems and their shortcomings and came to realize that there does not exist any perfect recommendation technique. Several issues remain in each recommenda-

tion model. Even, the hybrid recommendation technique that we proposed will not be accurate in predicting songs correctly all the time. One thing we came to realize is that creating a personalized experience for the user though difficult is the best way to recommend songs. User-centric models are far more effective in predicting accurately than other models. In the future, we hope for a hybrid recommendation model that will be more accurate than existing models with minimum human effort.

References

1. J. Lee, et al., Music recommendation system based on genre distance and user preference classification. *J. Theor. Appl. Inf. Technol.* **96**(5) (2018)
2. M. Millicamp, et al., Controlling spotify recommendations: effects of personal characteristics on music recommender user interfaces, in *Proceedings of the 26th Conference on User Modeling, Adaptation and Personalization, ACM* (2018)
3. J. O'Bryant, *A Survey of Music Recommendation and Possible Improvements.* (2017)
4. P. Knees, S. Markus, A survey of music similarity and recommendation from music context data. *ACM Trans. Multimedia Comput. Commun. Appl. (TOMM)* **10**(1), 2 (2013)
5. S. Ferretti, Clustering of musical pieces through complex networks: an Assessment over Guitar Solos, in *IEEE MultiMedia* (2018)
6. Y Song, D. Simon, P. Marcus, A survey of music recommendation systems and future perspectives, in *9th International Symposium on Computer Music Modeling and Retrieval*, vol 4 (2012)
7. P. Deshmukh, K. Geetanjali, *A Survey of Music Recommendation System* (2018)
8. J. Skowronek, M.F. McKinney, S. Van De Par, Ground truth for automatic music mood classification, in *ISMIR* (2006)
9. S.E. Park, S. Lee, S. Lee, Session-based collaborative filtering for predicting the next song, in *2011 First ACIS/JNU International Conference on Computers, Networks, Systems and Industrial Engineering (CNSI)*, IEEE (2011)
10. I. Kamehkhosh, D. Jannach, L. Lerche, Personalized next-track music recommendation with multi-dimensional long-term preference signals, in *UMAP (Extended Proceedings)* (2016)
11. M. Schedl, A. Flexer, J. Urbano, The neglected user in music information retrieval research. *J. Intell. Inf. Syst.* **41**(3), 523–539 (2013)

Interactive Systems for Fashion Clothing Recommendation



Himani Sachdeva and Shreelekha Pandey

Abstract An interactive system for fashion clothing recommendation is upcoming and exploring area that deals with image retrieval. Basically, it aims to have efficient online shopping systems that take clothes image as an input and automatically retrieve similar clothing images from a massive collection of clothing image dataset. In addition, such systems are trained to generate relevant style tags or annotations for the query image. Existence of factors like heterogeneous style attributes, body poses and appearances, and background generates several challenges for both the tasks. Past few years have identified part-based representations as useful tool in this domain. A survey of such systems is presented in this manuscript drawing attention toward their contribution. An attempt is also made to summarize must have features of fashion clothing recommendation systems thus presenting the future directions in this domain.

Keywords Clothing retrieval system · Style attributes · Fashion clothing

1 Introduction

Increasing popularity of e-commerce Web sites and advances in mobile computing not only explodes number of Web images but also opens several challenges related to the application of image processing techniques in the domain of online shopping. Shop anywhere anytime convenience associated with online shopping particularly clothes has attracted a large crowd thus making it a huge market. Online clothing shopping model continues to gain popularity in the present era due to easily available products and attractive deals. Most of the shopping Web sites, like Amazon.com and shopstyle.com, facilitates keyword-based search to find favorite clothing easily [2,

H. Sachdeva · S. Pandey (✉)
TIET Patiala, Punjab, India
e-mail: shreelekha.pandey@thapar.edu

H. Sachdeva
e-mail: himani.sachdeva93@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_27

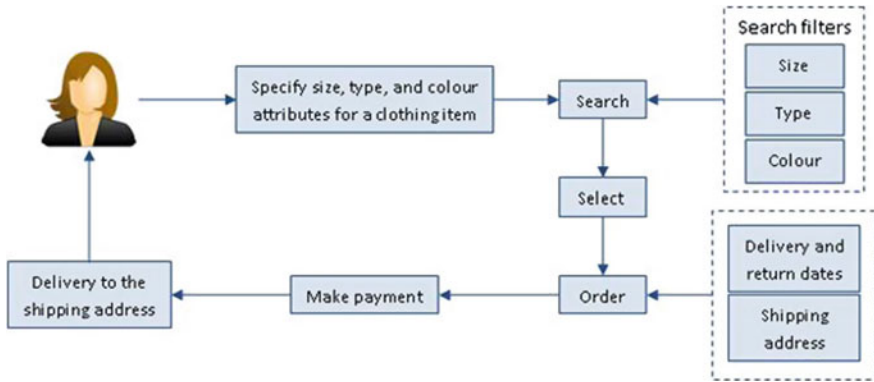


Fig. 1 Basic process flow of a keyword-based online clothing search system



Fig. 2 Top 10 images retrieved from www.myntra.com for text query ‘Black top’

7]. The basic process flow of such systems is shown in Fig. 1 and an example from www.myntra.com is shown in Fig. 2. These keywords typically belong to an attribute vocabulary (sometimes erroneous as well as incomplete) whose size is too small and is thus unable to characterize existing visual diversity in clothing properly.

As a result, past few years have observed an effective search and annotation of clothing images using image processing concepts an active research area [7]. Style is identified as an important clothing description in addition to colors and patterns. Also defining a style is easier using similar looking product images than descriptive words. Hence, researchers tried to develop recommender systems that work on combination of image and text processing principles. Such systems follow some basic sequence of steps: firstly in addition to basic preprocessing different regions in clothing images are

localized (mostly using convolutional neural network—CNN), visual features like histogram of oriented gradients—HOG, local binary patterns—LBP, color histogram, etc. are extracted; training is performed using style attributes and extracted visual features. During retrieval, the obtained images with fixed set of attributes, e.g. color, length, and material, are presented to the users and a few provide flexibility to modify these attributes for refinement of results. In this manuscript, several such systems are analyzed highlighting their major contributions toward this domain.

The rest of the manuscript is organized into three more sections. Section 2 presents the related studies in the domain of fashion clothing recommendation system. The observations made during the survey along with the discussions are listed in Sect. 3. Finally, the manuscript is concluded along with the discussion of a few research directions in the fashion clothing using computer vision techniques in Sect. 4.

2 Literature Survey

Image processing approaches are mainly used to retrieve and annotate similar clothing images automatically in the domain of clothing images. The annotation task is difficult when it comes to clothing images due to numerous style attributes covering appearances and body poses. Background variations complicate this task further. Thus, studies have explored part-based clothing and found it to be a valuable and meaningful task. However, factors like huge collection of clothing items, varying viewpoints, and lighting conditions create several challenges [3].

The problem of attribute classification for people is considered to be important but challenging too as it has many practical applications [1]. A simple and efficient system utilizing histogram of oriented gradients (HOG) features from image patches is trained with support vector machine (SVM) classifier. It is robust against camera views, occlusions, and communication. The system is capable to gain cues at any scale from body parts without any inference-related pose or alignment. Results are presented using a dataset consisting of nine annotation attributes for 8035 people. The system is flexible and can easily be extended to more attributes and visual categories. Another work uses SVM and develops ‘magic closet’ which is an occasion oriented and practical clothing pairing system [8]. It works on wearing properly as well as aesthetically principles. Rules matching visual features with clothing attributes are described to suggest the most apposite clothing provided a specific occasion automatically. The system is quite accurate but unable to detect a few clothing in the photograph album. Although with the maturity in state-of-the-art methods for feature detection systems, performance can be improved.

A scalable, fast, and fully automatic clothing detection and suggestion system working on Fashionista dataset with million’s of products is developed [6]. It distinguishes variations and styles within clothing class using attributes and considers co-occurrences among classes during training for better detection and lesser misclassifications. It outperforms state-of-the-art approaches. Model can further be generalized by including attributes and gender data for clothing class recognition along

with the proposed suggestion stages. A generalized approach employing transfer learning applicable to a range of fashion style categories and datasets with different domains is presented [2]. It utilizes aspect classifiers to allow searching based on attribute/stylistic element as well as to automatically generate tags or descriptions based on detected attributes. The work provides Women's Fashion Coat dataset with fine-grained clothing classes and also presents study of distinctive classifiers for every style attribute associated visual characteristics. In the future, multiple binary visual aspects help in category-level classification and development of other methods.

One more work attempts to approximate influence of social factors and content illustrations based on vision to approximate fashion photos quantitatively [10]. It utilizes information, namely user identity, number of comments, and user expertise. Statistical evidences as well as subjective measures (aesthetics from social content) are effectively evaluated and analyzed that prove the significance of system in comparison with other state-of-the-art systems. Focusing on cross-domain feature, a practical approach suitable for clothing retrieval in real-world application based on ranking network of dual attributes is presented [5]. Its characteristic learning phase encloses semantic features and visual correspondence force simultaneously to effectively model irregularities. Results show significant improvement over the baseline approaches. In addition, the work provides a distinct large-sized clothing dataset that can be used in person re-identification applications.

Another unique retrieval system, 'Particular Street to Shop' based on convolution neural network (CNN) and selective search method is presented [4]. It introduces a new dataset too and evaluates three techniques to learn a correspondence between street and shop domains. Retrieval results are exceptionally good and the presented system is an attempt to allow precise retrieval of clothing stuff from online owners. Performance can further be improved by developing techniques for better alignment of street and shop objects. In an attempt to improve clothing annotation and retrieval accuracy, a model using concepts of pose detection and part-based features is developed [9]. The study establishes the importance of part-based representative tags; also, it presents an analysis to link images and tags based on intra- and inter-clusters clothing key parts. Effective results are obtained for millions of clothing images with fashion models from Taobao. In the future, model can be investigated for all kinds of clothing photographs and performance improvements are expected using deep learning methods.

A two-level system aiming for better personalized interaction of customers is presented based on image retrieval technology [11]. It allows modification of original query descriptors, particularly color, texture, shape, and attribute, to handle semantic gap. Clothing descriptors are converted to bag-of-words (BOW) formats to ease descriptors modifications and effective use of topic model for semantic analysis. Later, a probabilistic network-based hybrid topic model combines multi-channel descriptors to depict each image. Results show improvement by 24% and are found to be promising. Recently, an out of the way study focuses on the development of CNN-based system that can learn effectively even in the presence of erroneous or incomplete clothes labels as well as category imbalances [7]. The presented multi-weight deep convolutional neural network utilizing the concepts of multi-label and

multitask outperforms and shows enhanced retrieval results. Working with limited training dataset, a multitask curriculum transfer (MTCT) network for fine-grained clothing attributes is formulated [3]. It optimally utilizes the concepts of transfer learning in specific scenario, i.e. orderly source domain with large labeled information and a noisy target domain. It consists of two main components, multitask deep learning and curriculum transfer deep learning, which helps in dealing all the issues related to small training dataset. Experiments are performed on a dataset collected from a wide source ranging from images via web retailers to in-the-wild street images. Effective attribute recognition results as well as accurate image retrieval results are showcased for unconstrained images from the street views.

3 Discussions and Observations

Table 1 list the main objectives along with the solution proposed by researchers for all the studies considered in this manuscript. The following are some observation for the literature survey. Most of the studies in this domain focus on human psychology while designing a clothing retrieval system to appropriately satisfy their demands. Architectures like support vector machines and different versions of neural networks are noticed to outperform any other architecture. Inclusion of semantic attributes during learning in any of the proposed work seems to improve the retrieval performance. Scalability and speed are two important properties of such systems. Recently, Dual Attribute-aware Ranking Network (DARN) is developed which is found to solve several issues related to cross-domain image retrieval, like street to shop image retrieval system. Majority of related works have observed that the utilization of concepts, like person detection and pose estimation, greatly helps in achieving quite accurate results.

Although the researchers are on their way to success and have achieved a few milestones in this domain, there are some fundamental challenges associated with clothing images. Diversity of style, in terms of sleeves, collar, length, texture, cuts, etc., can easily confuse a person and thus any computer system too. Thus, knowledge about style annotations is extensively used along with low-level features for enhanced recognition, retrieval, and classification of clothing images. Similarly, it is common to have dissimilarity in human postures and attaining a clear frontal view of clothes always is critically difficult. Background variations are quite common due to easy availability of cameras these days as people click images at any place. In fact, having clear background scenario is rare and objects in clutter are general. Lack of standard datasets is another point of concern. Very recently DeepFashion dataset is introduced; it's large-scale dataset with rich annotations.

Table 1 Summarization of objectives focused and the solutions proposed in existing studies

References	Objectives	Solution proposed
[1]	Recognize gender and other attributes like hairstyle, style of clothes. Effective recognition and segmentation	Extracts HOG features and trains SVM classifier using Bourdev method and combines together in discriminative model
[8]	Recommend suitable clothing for occasions by retrieving the most matched ones with the reference clothing image from online shops	Uses latent SVM model to aesthetically pair reference clothing that is suitable for specified occasion
[6]	Develop a finer clothing detection approach using image retrieval techniques to detect and retrieve visually similar products belonging to each class present in query image	Develops a fast and scalable clothing product suggestion framework based on detected classes and visual appearances. Works for hundreds of product classes and millions of product images. Performance is 50 times better
[2]	Recognize and retrieve similar fashion imagery as given in a query like suit, dress, and sweater. Return a ranked list of related items that have same visual attributes	Applies individual classifier for each style attribute and trains attribute classifiers on fined grained clothing. Defines a set of style-related visual attributes and provides attribute-oriented image retrieval. Accepts either a query image or target attribute group specification
[10]	Analyze the effects of visual, textual, and social factors on popularity in real-world networks focusing fashion. Utilize social factor for research involving social network photographs	Presents a new feature for style based on clothing parsing. Does a large-scale empirical study on how social versus content influences popularity on a real-world uncontrolled fashion network
[5]	Retrieve similar clothing items from a large-scale gallery of professional online shopping images. Create a dataset containing tens of thousands of online and offline clothing pairs obtained from user review pages	Develops a Dual Attribute-aware Ranking Network (DARN) for retrieval and feature learning. It simultaneously embeds semantic attribute information and visual similarity constraints into the feature learning stage
[4]	Match real-world garment item to the same item in an online shop. Collect large dataset by pairing of exactly matching items worn in a street	Achieves similarity measures between the street and shop domains that enable more accurate retrieval of clothing items from online shops
[9]	Address part-based clothing image annotation	Proposes effective and efficient clothing image retrieval system based on pose detection and part-based feature alignment
[11]	Allow users to specify query inputs as image and keywords too. Also multi-dimensional requirements are flexible and can be refined by editing visual features and attributes	Proposes a Hybrid Topic (HT) model, a probabilistic network integrating multi-channel descriptors into a unified framework. It uses bag of words (BOW) format

(continued)

Table 1 (continued)

References	Objectives	Solution proposed
[7]	Use deep convolutional neural networks for imbalance learning as images labeled by shop retailers from Web pages are largely erroneous or incomplete	Proposes multi-weight deep convolutional neural networks for coping with the noisy and imbalanced clothing image data in real world
[3]	Recognize clothing characteristics in detail, termed as fine-grained attributes in real-world cluttered images	Develops a novel deep learning-based multitask curriculum transfer architecture that utilizes multi-labeled annotations from different Web sources with fine-grained attributes

4 Conclusion and Future Scope

Most of the existing systems are based on retrieval of similar clothing. A very few of them has incorporated the concept of recommending a matching pair corresponding to a given query. In the past few years, researchers have tried to develop a flexible system to enhance user performance, where performance is measured how well retrieved images can match a fixed set of attributes. The system is chosen so that efficient results can be achieved in less time.

As it is very dynamic domain, a system would be much more accurate, user experience would be increased, and efficient results would be achieved in less period of time. The most important thing is that such systems should be much more close to human psychology. Development of the system is still on their way. New techniques are introduced everyday whose incorporation may lead to better results.

References

1. L. Bourdev, S. Maji, J. Malik, Describing people: a poselet-based approach to attribute classification, in *Proceedings of IEEE International Conference on Computer Vision (ICCV)* (2011), pp. 1543–1550
2. W. Di, C. Wah, A. Bhardwaj, R. Piramuthu, N. Sundaresan, Style finder: fine-grained clothing style detection and retrieval, in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (2013), pp. 8–13
3. Q. Dong, S. Gong, X. Zhu, Multi-task curriculum transfer deep learning of clothing attributes, in *Proceedings of IEEE Winter Conference on Applications of Computer Vision (WACV)* (2017), pp. 520–529
4. M. Hadi Kiapour, X. Han, S. Lazebnik, A.C. Berg, T.L. Berg, Where to buy it: matching street clothing photos in online shops, in *Proceedings of IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 3343–3351
5. J. Huang, R.S. Feris, Q. Chen, S. Yan, Cross-domain image retrieval with a dual attribute-aware ranking network, in *Proceedings of IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 1062–1070

6. Y. Kalantidis, L. Kennedy, L.J. Li, Getting the look: clothing recognition and segmentation for automatic product suggestions in everyday photos, in *Proceedings of 3rd ACM International Conference on Multimedia Retrieval* (2013), pp. 105–112
7. R. Li, F. Feng, I. Ahmad, X. Wang, Retrieving real world clothing images via multi-weight deep convolutional neural networks, in *Cluster Computing* (2017), pp. 1–2
8. S. Liu, J. Feng, Z. Song, T. Zhang, H. Lu, C. Xu, S. Yan, Hi, magic closet, tell me what to wear!, in *Proceedings of 20th ACM International Conference on Multimedia* (2012), pp. 619–628
9. G.L. Sun, X. Wu, Q. Peng, Part-based clothing image annotation by visual neighbor retrieval. *Neurocomputing* **213**, 115–124 (2016)
10. K. Yamaguchi, T.L. Berg, L.E. Ortiz, Chic or social: visual popularity analysis in online fashion networks, in *Proceedings of 22nd ACM International Conference on Multimedia* (2014), pp. 773–776
11. Z. Zhou, Y. Xu, J. Zhou, L. Zhang, Interactive image search for clothing recommendation, in *Proceedings of ACM Conference on Multimedia* (2016), pp. 754–756

Embedded Implementation of Early Started Hybrid Denoising Technique for Medical Images with Optimized Loop



Khakon Das, Mausumi Maitra, Minakshi Banerjee and Punit Sharma

Abstract This paper represents the architecture of an embedded system for Early Started Hybrid Denoising Technology for Medical Images (ESHDT) using a very popular embedded processor, ATmega processor, which is inexpensive in terms of computation. Embedded implementation of the ESHDT algorithm is chosen because hardware presents a good scope of parallelism and pipelining over software. The Proposed System can work efficiently in a noisy environment. The proposed embedded system architecture has less space complexity because the number of memories is used here. The system starts performing the algorithmic task as soon as two-pixel values are received, and parallelism is achieved by the sixth level of software pipelining and two levels of dedicated hardware pipelining. Due to the design simplicity of our algorithm, it used very few memory resources. Here we used the 8-bit AVR processor which run on low power, and we optimized SD card access so we can say the power consumption by our system is low. Instead of the set of filters, simple predict and update mechanism is introduced. ESHDT provides very good-quality denoised output image with respect to PSNR, UIQI, and MSE.

Keywords Embedded · Image fusion · Denoising · Wavelet transform · In-place calculation · Lifting

K. Das (✉) · M. Banerjee
RCC Institute of Information Technology, Kolkata 700015, India
e-mail: khokon.phd@gmail.com

M. Banerjee
e-mail: mbanerjee23@gmail.com

M. Maitra
Government College of Engineering and Ceramic Technology, Kolkata 700010, India
e-mail: mou1232005@yahoo.com

P. Sharma
Apollo Gleneagles Hospitals, Kolkata 700054, India
e-mail: dr_punitsharma@yahoo.com

1 Introduction

Noise is a basic characteristic that is present in all types of images. Noise decreases the visibility level of some structures and entities, especially those that have quite low contrast. Operations on images are done at the lowest level of abstraction to enhance the image information without changing the basic image information content [1]. The main objective is the noise reduction up to an acceptable level for medical images. In PET (denoted as positron emission tomography) imaging systems, the high-energy photons are emitted due to the annihilation of positrons and electrons, which provides useful information about the object. During the signal collection period, noise is also generated and added to the corresponding images [2]. Improvement of the image quality is insured by denoising techniques for PET, CT, and MRI images. Therefore, the type of the noise of the signal should be recognized and factual properties ought to be concentrated to outline ideal denoising technique [3–5]. But details of image information like edge, curve, etc., are lost during denoising, which may be retrieved by using image edge preservation [6], image contrast enhancement [7, 8], and fusion of heterogeneous images like PET-CT [9–11]. Discrete wavelet transform (DWT) was first given by Mallat [12]. Now DWT is broadly used for medical images to make denoised. Several modifications of DWT have been made to amend its performance. Hardware implementation of DWT was made to reduce processing cost, time, and power as it involves sizably voluminous computations [13–15]. An incipient concept of lifting scheme [16] was given by Sweldens in 1996 which is plerarily predicated on in-place calculation [17] and thereby eliminates the requirements for auxiliary memory. Depending on particular application, mother wavelet is selected. We have taken Haar's wavelet [18, 19] as the mother wavelet because it is simple and widely used in the medical field of image processing. K. Das et al. (2017) designed a hybrid algorithm to denoise images using a fusion of two images (CT and PET) and modified Haar's wavelet [20]. Here PET image provides very good information about soft tissue, and CT image gives us the information about structures. Early Started refers that the algorithm starts performing the wavelet transform as soon as information of the two pixels is received and algorithm stops at just after processing the last pixels' information. Waiting for receiving full image information at a time is not needed. We proposed modified form of the basic Haar's wavelet transform algorithm using in-place, fusion [9–11, 21–24], and lifting techniques [25, 26]. Here we use lifting techniques with an in-place calculation to perform modified wavelet transform, it applies on PET and CT images of the patients, and then both denoised images are fused to obtaining a better result with respect to various denoising parameters like MSE, PSNR, and UIQI. Due to simple calculation and fewer memory uses, the algorithm is suitable for embedded implementation.

2 Methodology

We have already developed a denoising algorithm for medical images using PET-CT image fusion and modified Haar's wavelet. Here PET image contains quality information about soft tissue, and CT image provides structural information. Early Started refers that the algorithm starts performing the wavelet transform as soon as information of the two pixels is received and algorithm stops at just after processing the last pixels' information, so no need to wait for full image information at the time of algorithm begins. Now we implement ESHDT algorithm on an embedded system. Embedded system implementation of the ESHDT algorithm is chosen because hardware presents a good scope of parallelism and pipelining over software. ESHDT can work well in the high level of noise. A basic study has been done based on the above method. The proposed algorithm uses few amount of on-chip-memories so the space complexity is reduced. Due to the uncomplicated design, ESHDT consumes less power and using less resource. We reconstruct ESHDT with an optimized loop. Instead of a set of filters, simple predict and update mechanism is introduced. ESHDT provides very good-quality output image with respect to PSNR, UIQI, and MSE. ESHDT algorithm is a modified form of Haar's wavelet transform using lifting, in-place, and fusion techniques [20] which has been described below.

2.1 Proposed Algorithm

Step 1: Divide all incoming pixels into even and odd sets based on pixel index, S_{Even} and S_{Odd} , where n is a level of denoising.

$$S_n = S_{n(\text{Even})} U S_{n(\text{Odd})} \quad (1)$$

Step 2: Calculation of detailed coefficient of Haar's wavelet transform is incorporated with predict step of lifting method. In-place calculation is used [i refers to pixel index and ' n ' is a denoise Level].

$$S_{n,i} = S_{n,i+1} - S_{n,i} \quad (2)$$

Step 3: To calculate Haar's approximate part in update step of lifting technique, in-place calculation is used.

$$S_{n,i} = s_{n,i} + \left(\frac{S_{n,i+1} - S_{n,i}}{2} \right) \quad (3)$$

Step 4: Lifting with fusion

4.1: PET image resize by the factor K [Depending on PET-CT scanner, the value

of $K = 1.5]$

4.2: Resized PET image starts growing from the seed point of CT image; both images must belong to the same subjects and same image slice number [$S_{n,PET-CT}$ is the fused image]

$$S_{n,PET-CT}(\text{index}) = \left(\frac{S_{n,PET}(\text{index}) + S_{n,CT}(\text{index})}{2} \right) \quad (4)$$

Step 5: End.

Now we design a dedicated embedded hardware system to take the advantages of hardware system like resource, time, and cost of customized hardware system [27–30].

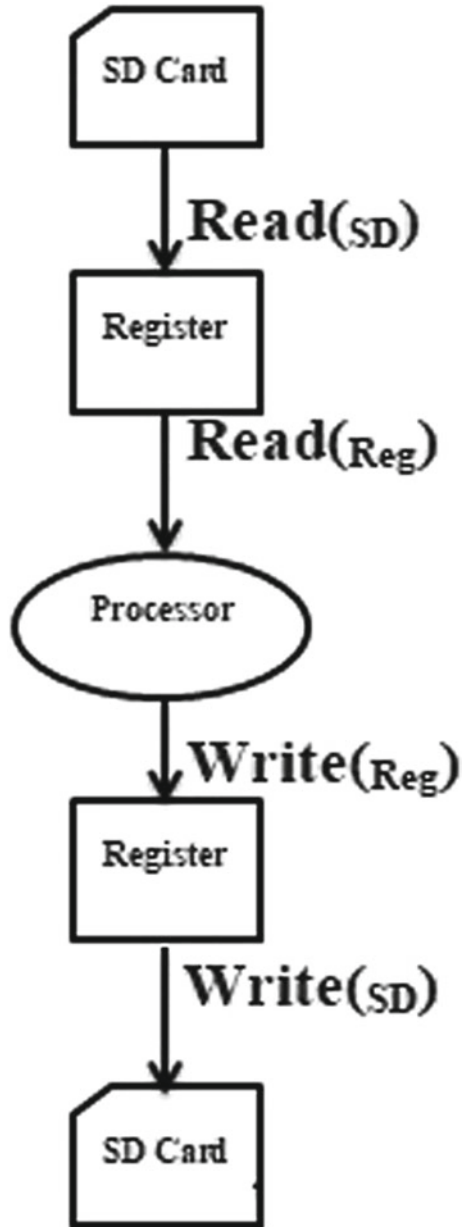
2.2 Applying Pipeline Concept

2.3 Software Pipelining

In Fig. 1, we have shown the data flow diagram for a single-pixel denoising, i.e., a single-cycle format of our proposed algorithm. In this figure, the storage device (SD card) is used for storing both input and output data after completion of each process. At the very first stage, the input data or rather the pixel is read from the SD card to the internal register after obtaining the input data processor performs the denoising operation on that and returns the output data to the register from where the output data is written back to the SD card. Through those steps, we complete a single cycle of our algorithm and the operation can repeatedly be done for multiple pixels of image, whereas the Fig. 2 describes the method for multiple pixels and multiple-cycle format in our proposed algorithm. In this method, we have considered an image of n pixels where the processor needs n times SD card read operations to read every pixel and after denoising write back n times to SD card. In our proposed algorithm in every level of denoising, the pixels are denoised but we found after a certain number of denoising level significant data lose may arise.

We have examined and found that in sixth denoising level, we acquired a clinically acceptable denoised image. For 6th level denoising, the system needs $6*N$ times SD read and write operation this process takes a long time because of the too many reads and write operation on the sd card. We have solved the issue using software pipelining as shown in Fig. 3. In Fig. 4, we have described the software pipelining through a required resource and availability graph where we have to consider the resource availability and number of denoising level to complete. For denoising, first pixel is read from the SD card and stored in the register and after getting the next pixel the processor have got the required resources, i.e., pixel to complete the first level of denoising and store it into the register for further use as a resource. While

Fig. 1 Single-cycle format of proposed algorithm



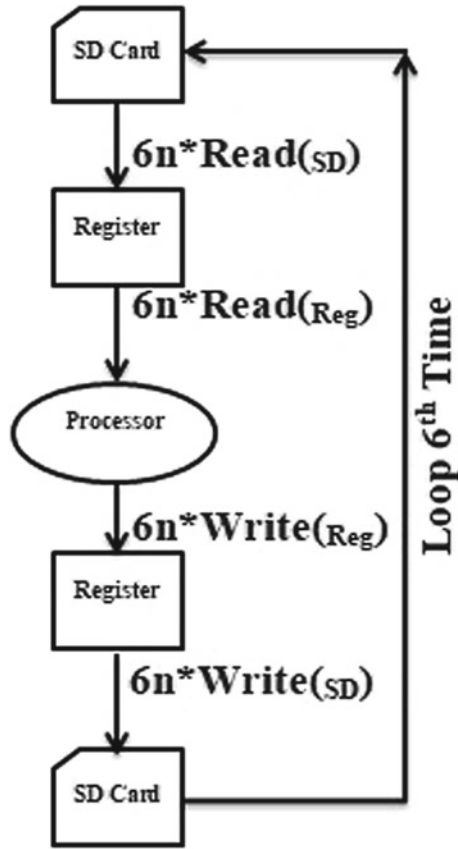


Fig. 2 Multiple-cycle format of proposed algorithm

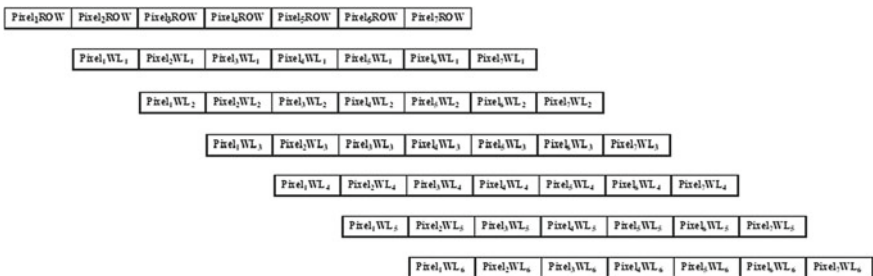


Fig. 3 Software pipelining of level six

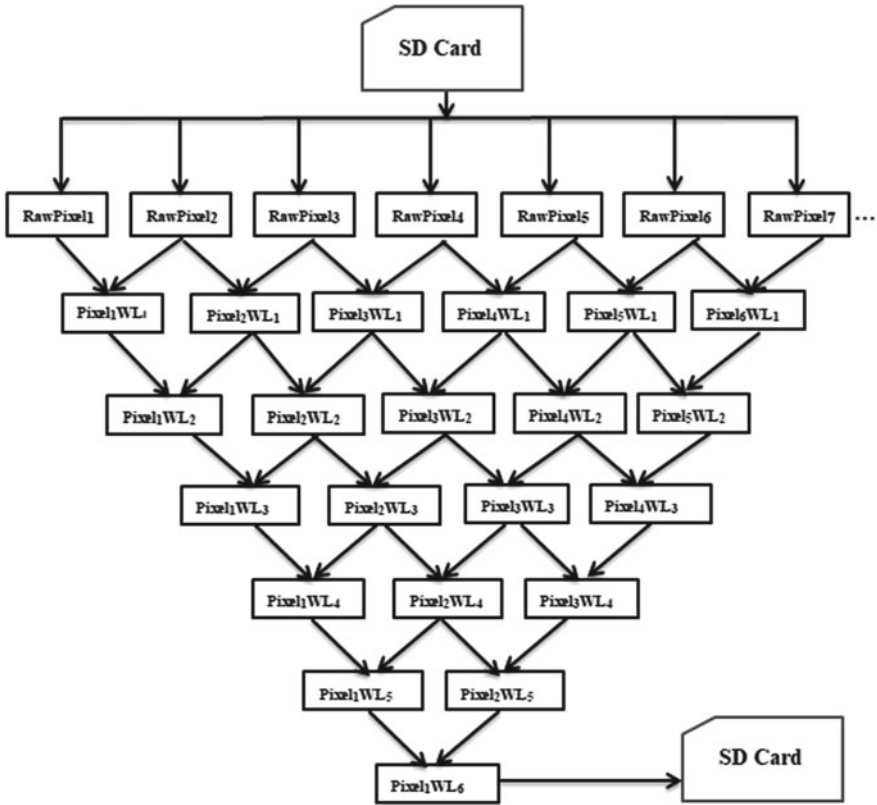
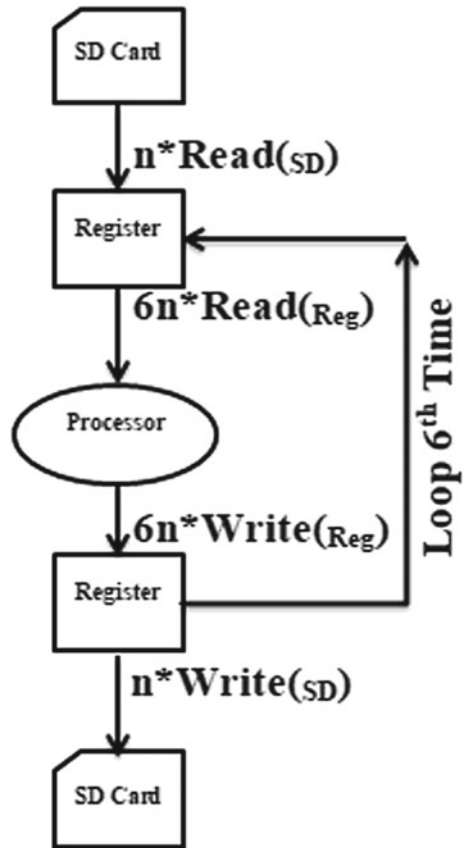


Fig. 4 Required resource and availability graph

the third pixel is read from the SD card, the processor has the resources available to complete the first denoising level as well as to complete the second denoising level. In this way, if we wish to get a sixth-level denoised single pixel, we must have at least seven pixels at a time. In this way for an input image, all denoised pixels of sixth denoising level stored in the register and written back to the SD card one by one as shown in Fig. 5 where we achieved the optimized loop. With the flow of the process, comparing Fig. 5 with Fig. 3 we can conclude that in Fig. 5, the time required to fetch data from SD card and write back the output data to SD card is much less than that of Fig. 3.

Fig. 5 Multicycle execution of proposed algorithm with optimization loop



2.4 Instruction Execution Timing

Concepts of the general access time of instruction execution are described in this section. The clock (clkCPU) is generated directly from the clock source, and it is used to drive the AVR CPU. No internal clock division is used here. In this picture below shows that the fast-access Register File concept and the Harvard architecture are good enough to fetches and executions instruction parallel way. To obtain 1 MIPS par MHz, the basic pipelining concepts are used. Figure 6 shows the timing concept (internal) for the Register File. Stored back the result to destination register after calculation, this is performed by only one clock cycle of an ALU operation using two register operands [31].

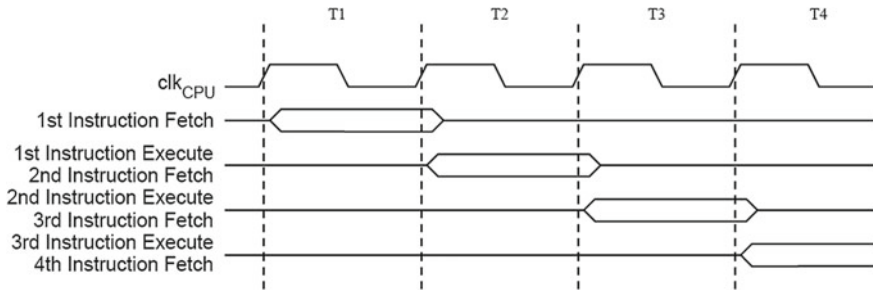
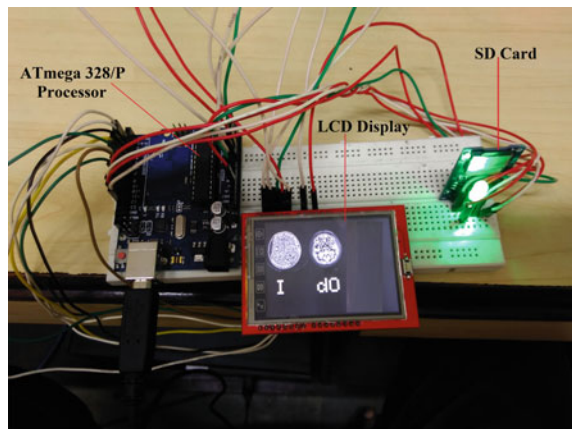


Fig. 6 The parallel instruction fetches and instruction executions

3 Embedded System Design

Our system received four pixels as input at a time two from each CT image and from PET image and performs wavelet transform separately after that fused both pixels and getting our desire output pixel as we called PET-CT fused pixel(s) (Image). Specifically, two pixels of CT and PET images, respectively, denoted as odd and even index pixels. Each of CT and PET images by this steps we perform split operation of lifting scheme. In predict step, we subtract odd pixel from even pixel. In update step, we add half amount of predict value with odd pixel, in this way we update lifting scheme procedure. This half amount of predict value contains the high-frequency part of input image which means the edge information of given image and some noise also, but amount of noise is very less. As we know that average and below of average of any pixel intensity value represent the image part of that pixel (image). In this way, it is defined that the high-intensity part of the given image contains the edge and noise part of the input image, and average intensity part of the high-intensity part (L(H)-Part) represents the image part of that image (H-Part), means the edge part [Image = L + L(H)]. We apply above process into PET and CT images separately

Fig. 7 Working model of embedded Early Started Hybrid Denoising Technique for Medical Images with optimized loop



and fused them to make a PET-CT fused image [23, 32–34] (In Fig. 7, display the working model of our proposed system).

4 Result and Output

See Tables 1 and 2.

Figure 8a shows the input image of Patient 1 (PET image), and Fig. 8b shows the hardware output of denoised PET image using the proposed technique.

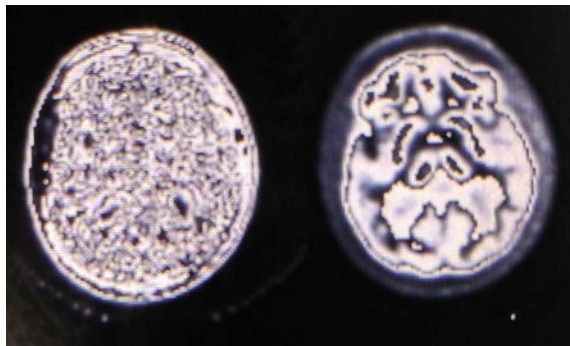
Table 1 UIQI, PSNR and MSE using our proposed algorithm for Patients 1 and 2

Patient 1						
	CT image			PET image		
	MSE	PSNR (dB)	UIQI	MSE	PSNR (dB)	UIQI
Input image	68105.80980	±0.33408	0.38399	537148.72821	±18.27229	0.49209
Level 6	0.08292	+112.50151	0.90017	0.36621	+102.88651	0.72842
Patient 2						
	CT image			PET image		
	MSE	PSNR (dB)	UIQI	MSE	PSNR (dB)	UIQI
Input image	72501.99099	±0.79892	0.35959	202041.73016	±9.01512	0.48919
Level 6	0.07289	+118.22391	0.91052	0.25130	+108.07501	0.60512

Table 2 UIQI, MSE, PSNR, and MI of fused images using our proposed algorithm for Patients 1 and 2

PET-CT fused image					
Patient	Level	MSE	PSNR (dB)	MI	UIQI
Patient 1	Level 6	0.05800	+119.06052	4.27176	0.91931
Patient 2	Level 6	0.05779	+119.98212	4.28070	0.92109

Fig. 8 Input image (8a—left side) and output image (8b—right side) of our proposed algorithm (PET image of Patient 1)



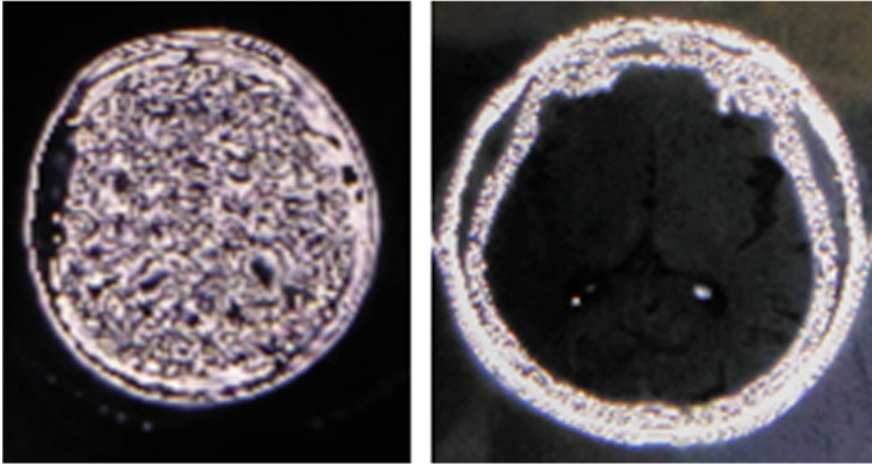
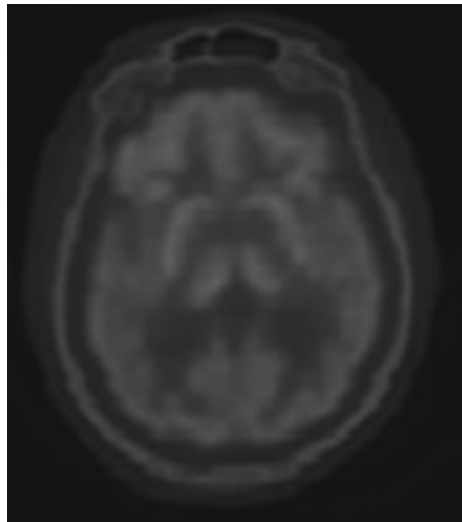


Fig. 9 PET and CT input images of Patient 1

Fig. 10 Output images
(PET-CT fused) of Patient 1



5 Conclusions

Our primary focus to design an embedded system for medical image denoising which is based on heterogeneous image fusion and modified Haar's wavelet transform. It is a hybrid technique. PET and CT brain images are considered for two patients for our studies. Based on various noise check parameters like PSNR, MSE, MI, UIQI, we

concluded that the performance of this method is very good, with respect to software. Comparing Fig. 5 with Fig. 3, we can also conclude that in Fig. 5, the time required to fetch data from SD card and store the output data to SD card is much less than that of Fig. 3 so in this way loop optimization is done through software pipelining. Due to the simple calculation, less memories are required so we can also conclude that the proposed hardware cost is low (Figs. 9 and 10).

Acknowledgements We thank to BRNS, DAE, Govt. of India, for sponsoring fellowship to Mr. Khakon Das, SRF in The Project Embedding Segmentation of PET Images for Medical Applications in Brain Disorders, Sanction No. “34/14/13/2016-BRNS with ATC, BRNS.”

We have obtained all ethical approvals from the Ethics Committee of Apollo Gleneagles Hospitals, Kolkata, India. The consents of the participants have also been taken to use the data in this research. Neither the publisher nor the editors will be responsible for any misinterpretation or any misuse of the data.

References

1. V. Hlavac, M. Sonka, R. Boyle, Image pre-processing, in *Analysis and Machine Vision* (1993), pp. 56–111
2. H.H. Muhammed, Y. Sicong, Noise type evaluation in positron emission tomography images, in *2016 1st International Conference on Biomedical Engineering (IBIOMED)* (IEEE, 2016), Yogyakarta, Indonesia. 10.1109/IBIOMED.2016.7869828
3. A. Aldroubi, M. Unser, A review of wavelets in biomedical applications. *Proc. IEEE* **84**(4), 626–638 (1996). <https://doi.org/10.1109/5.488704>
4. Robert D. Nowak, Wavelet-based rician noise removal for magnetic resonance imaging. *IEEE Trans. Image Process.* **8**(10), 1408–1419 (1999). <https://doi.org/10.1109/83.791966>
5. Martin J. Wainwright, Javier Portilla, Vasily Strela, Eero P. Simoncelli, Image denoising using scale mixtures of gaussians in the wavelet domain. *IEEE Trans. Image Process.* **12**(11), 1338–1351 (2003). <https://doi.org/10.1109/TIP.2003.818640>
6. B.K. Singh, K. Verma, A.S. Thoke, An enhancement in adaptive median filter for edge preservation. *Procedia Comput. Sci.* **48**, 29–36 (2015). <https://doi.org/10.1016/j.procs.2015.04.106>
7. P. Pal, S. Singh, Contrast enhancement of medical images: a review. *IJRDO J. Health Sci. Nurs.* **1**(4) (2016)
8. E.S. Kaur, A.R. Rivera, *Image Enhancement Techniques: A Selected Review* (2013)
9. N. Kumar, M. Chandana, S. Amutha, A hybrid multi-focus medical image fusion based on wavelet transform. *Int. J. Res. Rev. Comput. Sci.* **2**, 1187–1192 (2011)
10. S.S. Bedi, J. Agarwal, Implementation of hybrid image fusion technique for feature enhancement in medical diagnosis. *Human-centric Comput. Inf. Sci.* **5**(3), 1–17 (2015). <https://doi.org/10.1186/s13673-014-0020-z>
11. F. Ghante, L. Tawade, A.B. Aboobacker, Image fusion based on wavelet transforms. *Int. J. Bio-Sci. Bio-Technol.* **6**(3), 149–162 (2014). <https://doi.org/10.14257/ijbsbt.2014.6.3.18>
12. S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 674–693 (1989). <https://doi.org/10.1109/34.192463>
13. T. Lindblad, J. Chilo, Hardware implementation of 1D wavelet transform on an FPGA for infrasound signal classification. *IEEE Trans. Nucl. Sci.* **55**(1), 9–13 (2008). <https://doi.org/10.1109/TNS.2007.914322>

14. C. Souani, Optimized VLSI design of wavelet transform architecture, in *Proceedings of the 16th International Conference on Microelectronics, 2004. ICM 2004* (IEEE, 2014), Tunis, Tunisia, pp. 558–563. DOIurl10.1109/ICM.2004.1434724
15. I. Chakrabarti, A.S. Motra, P.K. Bora, An efficient hardware implementation of DWT and IDWT, in *TENCON 2003. Conference on Convergent Technologies for Asia-Pacific Region* (IEEE, 2003), Bangalore, India, pp. 95–99. <https://doi.org/10.1109/TENCON.2003.1273240>
16. W. Sweldens, I. Daubechies, Factoring wavelet transforms into lifting steps. *J. Fourier Anal. Appl.* **4**(3), 247–269 (1998). <https://doi.org/10.1007/BF02476026>
17. Koichi Kuzume, Koichi Nijjima, Shigeru Takano, Fpga-based lifting wavelet processor for real-time signal detection. *Sign. Process.* **84**(10), 1931–1940 (2004). <https://doi.org/10.1016/j.sigpro.2004.06.020>
18. R.S. Radomir, B.J. Falkowski, The haar wavelet transform: its status and achievements. *Comput. Electr. Eng.* **29**(1), 25–44 (2003). [https://doi.org/10.1016/S0045-7906\(01\)00011-8](https://doi.org/10.1016/S0045-7906(01)00011-8)
19. Haar Alfred, Zur theorie der orthogonalen funktionensysteme. *Math. Ann.* **69**(3), 331–371 (1910). <https://doi.org/10.1007/BF01456326>
20. P. Sharma, M. Banerjee, K. Das, M. Maitra, Early started hybrid denoising technique for medical images, in H. Bhaumik, S. Das, K. Yoshida, S. Bhattacharyya, A. Mukherjee (ed.), *Recent Trends in Signal and Image Processing* (Springer Singapore, 2019), pp. 131–140. ISBN 978-981-10-8863-6. https://doi.org/10.1007/978-981-10-8863-6_14
21. M.B. Haghghat, A. Aghagolzadeh, H. Seyedarabi, Multi-focus image fusion for visual sensor networks in DCT domain. *Comput. Electr. Eng.* **37**(5), 789–797 (2011). <https://doi.org/10.1016/j.compeleceng.2011.04.016>
22. M.B.A. Haghghat, A. Aghagolzadeh, H. Seyedarabi, A non-reference image fusion metric based on mutual information of image features. *Comput. Electr. Eng.* **37**(5), 744–756 (2011). <https://doi.org/10.1016/j.compeleceng.2011.07.012>
23. A.P. James, B.V. Dasarathy, Medical image fusion: a survey of the state of the art. *Inf. Fusion* **19**, 4–19 (2014). <https://doi.org/10.1016/j.inffus.2013.12.002>
24. M.J. Gooding, K. Rajpoot, S. Mitchell, P. Chamberlain, S.H. Kennedy, J.A. Noble, Investigation into the fusion of multiple 4-d fetal echocardiography images to improve image quality. *Ultrasound Med. Biol.* **36**(6), 957–966 (2010). <https://doi.org/10.1016/j.ultrasmedbio.2010.03.017>
25. Wim Sweldens, The lifting scheme: a construction of second generation wavelets. *SIAM J. Math. Anal.* **29**(2), 511–546 (1989). <https://doi.org/10.1137/S0036141095289051>
26. Wim Sweldens, Wavelets and the lifting scheme: A 5 minute tour. *Z. Angew. Math. Mech* **76**, 41–44 (1996)
27. B.A. Draper, J.R. Beveridge, A.P.W. Bohm, C. Ross, M. Chawathe, Accelerated image processing on FPGAS. *IEEE Trans. Image Process.* **12**(12), 1543–1551 (2003). <https://doi.org/10.1109/TIP.2003.819226>
28. J. Chilo, T. Lindblad, Hardware implementation of 1D wavelet transform on an FPGA for infrasound signal classification. *IEEE Trans. Nuclear Sci.* **55**(1), 9–13 (2008). <https://doi.org/10.1109/TNS.2007.914322>
29. R.M. Jiang, D. Crookes, FPGA implementation of 3D discrete wavelet transform for real-time medical imaging, in *2007 18th European Conference on Circuit Theory and Design* (IEEE, 2007), pp. 519–522. <https://doi.org/10.1109/ECCTD.2007.4529647>
30. A. Jarrah, M.M. Jamali, Optimized FPGA based implementation of discrete wavelet transform, in *48th Asilomar Conference on Signals, Systems and Computers, ACSSC 2014*, Pacific Grove, CA, USA, 2–5 Nov 2014, pp. 1839–1842 (2014). <https://doi.org/10.1109/ACSSC.2014.7094786>
31. Embedded microcontroller (ATMega328p) datasheet, http://ww1.microchip.com/downloads/en/DeviceDoc/Atmel-42735-8-bit-AVR-Microcontroller-ATmega328-328P_Datasheet.pdf. Accessed: 20 July 2018
32. Shutao Li, Xudong Kang, Leyuan Fang, Hu Jianwen, Haitao Yin, Pixel-level image fusion: a survey of the state of the art. *Inf. Fusion* **33**, 100–112 (2017). <https://doi.org/10.1016/j.inffus.2016.05.004>

33. A. Polo, Image fusion techniques in permanent seed implantation. *J. Contemp. Brachytherapy* **2**(3), 98–106 (2010). <https://doi.org/10.5114/jcb.2010.16920>
34. D. Besiris, V. Tsagaris, N. Fragoulis, C. Theoharatos, An fpga-based hardware implementation of configurable pixel-level color image fusion. *IEEE Trans. Geosci. Remote Sens.* **50**(2), 362–373 (2012). <https://doi.org/10.1109/TGRS.2011.2163723>

Stress Profile Analysis in n-FinFET Devices



T. P. Dash, S. Das, S. Dey, J. Jena and C. K. Maiti

Abstract In this paper, a TCAD framework has been used to study the effects of stress due to the nitride contact etch stop layers (CESL) on the performance improvement of FinFET devices. Two-dimensional stress distribution in different segments of the device has been presented. The stress level generated in the channel depends on the process parameters as well as the physical dimensions of the CESL layer. Also, a 3D view of the stress profile in the channel and the CESL layers has been shown. In addition, improvement in device electrical characteristics due to the stress effects has been investigated. Increment in drain current by ~60% has been observed when a CESL stressor is employed.

Keywords Stress · FinFETs · CESL · 3D stress analysis

1 Introduction

Stress engineering is being used in all Si CMOS-based devices. There are three commonly used methods for stress engineering in nano-scaled transistors, e.g., stress memorization technique (SMT), Si₃N₄ contact etch stop layers (CESL) stressors, and epitaxy of SiGe on S/D regions [1]. The stressor techniques induce uniaxial

T. P. Dash (✉) · S. Das · S. Dey · J. Jena · C. K. Maiti
Department of Electronics and Communication Engineering, Siksha 'O'
Anusandhan (Deemed to be University), Bhubaneswar 751030, Odisha, India
e-mail: taradash@soa.ac.in

S. Das
e-mail: sanghamitradas@soa.ac.in

S. Dey
e-mail: supravadey@soa.ac.in

J. Jena
e-mail: jhansiranijena@soa.ac.in

C. K. Maiti
e-mail: ckmaiti@soa.ac.in

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_29

strain and used for carrier mobility boost up which has been implemented in conventional MOSFETs by Ghani et al. [2]. The contact etch stop layer technology for inducing strain has been one of the possible solutions among all other feasible stress engineering techniques (e.g., strained spacer, silicide, SiGe, strained gate) [3]. The CESL consists of a nitride layer used to stop the etching of the metallic contact. The CESL technology exploits the intrinsic strain of the nitride contact etch liners.

CESL employs highly stressed Si_3N_4 films to exert stress through the active area and gate electrode which were first integrated into CMOS process below 90 nm technology [2, 4–6]. Huang et al. [7] have shown the performance improvement with CESL as one of the local strain techniques for nanoscale MOSFETs. The performance enhancement of compressively strained 90-nm-gate NMOSFET using trench-based structure has been shown by Di et al. [8]. Even more, the effect of CESL with SiGe channel has been shown for p-type MOSFETs [9]. CESLs can produce up to 3GPa of tensile or compressive stress, depending on the deposition conditions, which makes it an effective local stress source for both types of devices [10]. Though a lot of research has been carried out, the effect of CESL layer thickness and its optimization for better stress transfer efficiency has not been explicitly reported yet.

Although the implementation of strain techniques has been developed for the planar transistor such as strain induced by CESL is yet to be explored for non-planar devices such as FinFETs and nanowires. The effects of CESL are considered with simple mechanical simulations in planar devices, whereas the understanding of stress distribution becomes more complex when it is considered for three-dimensional (3D) devices such as dual gate (DG) and trigate (TG) FinFETs. Introducing uniaxial strain to a non-planar trigate FinFET and analyzing stress inside the device and the strain contribution to carrier mobility improvement in channel are challenging [11–13]. This work focuses on the TCAD prediction of stress effects in n-type FinFETs.

In this paper, a TCAD simulation framework is used to explain how the CESL transmits stress to the Si channel in FinFETs. The stress level created in the channel by CESL films depends on process parameters as well as its physical dimensions. A detail insight into the generation and distribution of stress in the device has been presented with 2D and 1D view of representation. We have also included 3D analysis of stress profile in FinFETs due to CESL layer. The improvement in the device performance due to stress has been presented in the electrical characteristics.

2 Stress Profile Analysis

In this section, we have discussed the stress profiles of the n-FinFET shown in Fig. 1. We have followed gate first process with tensile stress n-FinFET. We have used crystal orientations (100) for silicon wafer and the fin oriented is in the $\langle 100 \rangle$ direction. The FinFET with 50×50 nm (width \times height) and $1 \mu\text{m}$ in length fin has been virtually fabricated. The fin was deposited on a SiO_2 substrate layer, and a 2-nm-gate isolation layer separated it from the 50 nm polysilicon gate crossing it at right

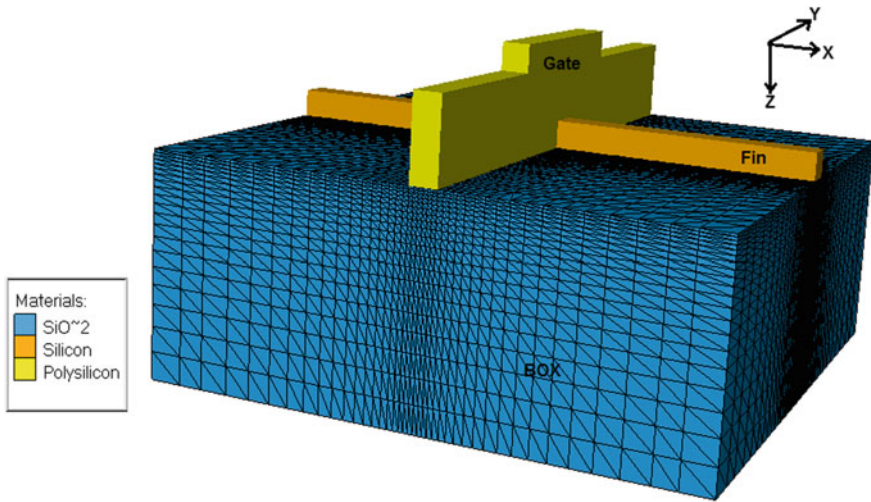


Fig. 1 A FinFET with a 50×50 nm fin on SiO_2 BOX substrate

Table 1 Material parameters used in simulation

Materials used	Young's modulus (Dyne/cm ²)	Poisson's ratio	Thermal Expansion (m/mK)
Silicon nitride	3.10×10^{12}	0.27	3.300×10^{-6}
Oxide	4.50×10^{11}	0.20	1.206×10^{-7}
Silicon	1.30×10^{12}	0.28	2.600×10^{-6}
Polysilicon	1.87×10^{12}	0.28	3.052×10^{-6}

angles. A 100 nm thick Si_3N_4 capping layer was deposited on top of the polysilicon. In order to investigate the stress profile due to CESL layer, finite element (FE) simulations have been performed. The material properties included in the simulation have been listed in Table 1. We have obtained the 3-D stress profiles in the FinFET induced by a tensile CESL capping layer. A tensile CESL cap layer simultaneously introduces tensile parallel stresses (Stress XX) and compressive vertical stress (Stress ZZ) which enhances electron mobility. We discuss the parallel and vertical stress in the following. The stress profile for the whole device is shown in Fig. 2. If we focus only on the stress in nitride layer, the stress in parallel direction varies from 140 MPa to 1.2 GPa as shown in Fig. 3 which induces stress in fin.

To have quantitative determination of stress in nitride layer, we have plotted Fig. 4 by taking 1D cut line parallel to y-axis within the nitride layer. It shows quite uniform parallel tensile stress profile of magnitude 500 MPa and vertical stress of -300 MPa. This stress in nitride layer will induce stress in the fin. Besides the nitride layer, the stress from the polysilicon also contributes to fin stress.

To evaluate the stress profile in the fin, we have to analyze the stress profile in polysilicon layer. The 1D stress profile in the polysilicon layer has been shown in

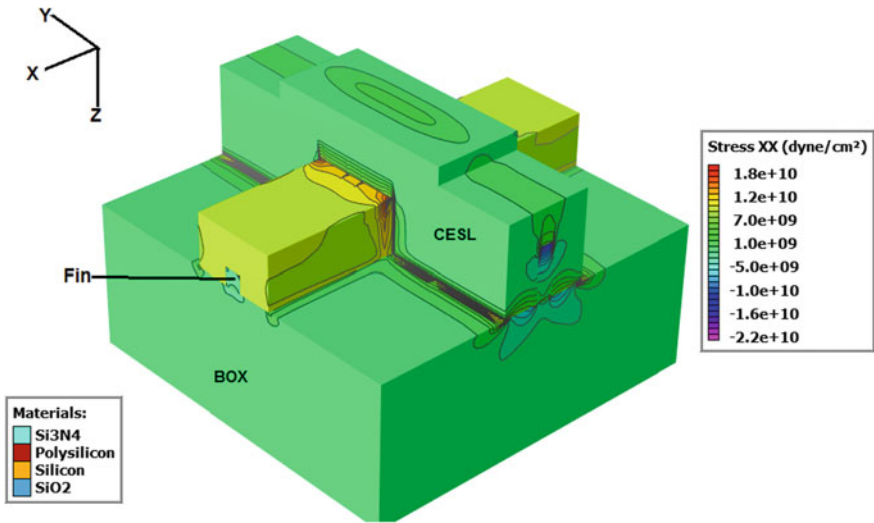


Fig. 2 3D stress profile in n-type FinFET structure obtained from stress simulations

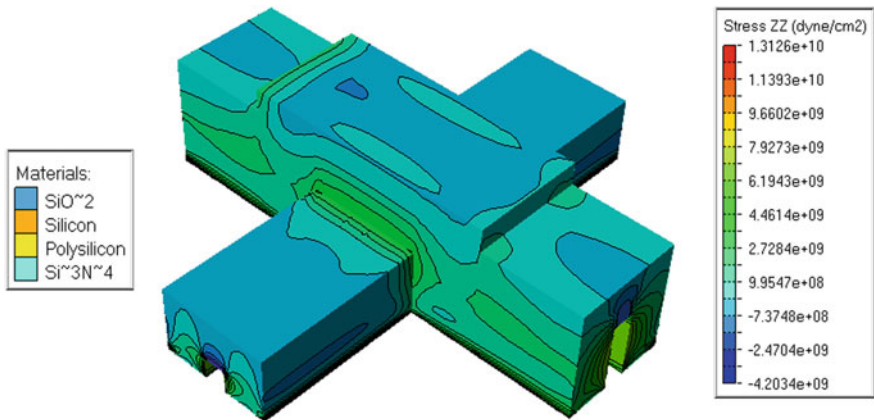


Fig. 3 Vertical direction stress field in the capping nitride layer

Fig. 5. The vertical stress is compressive and significantly high as -2.2 GPa whereas the stress in the parallel direction is insignificant.

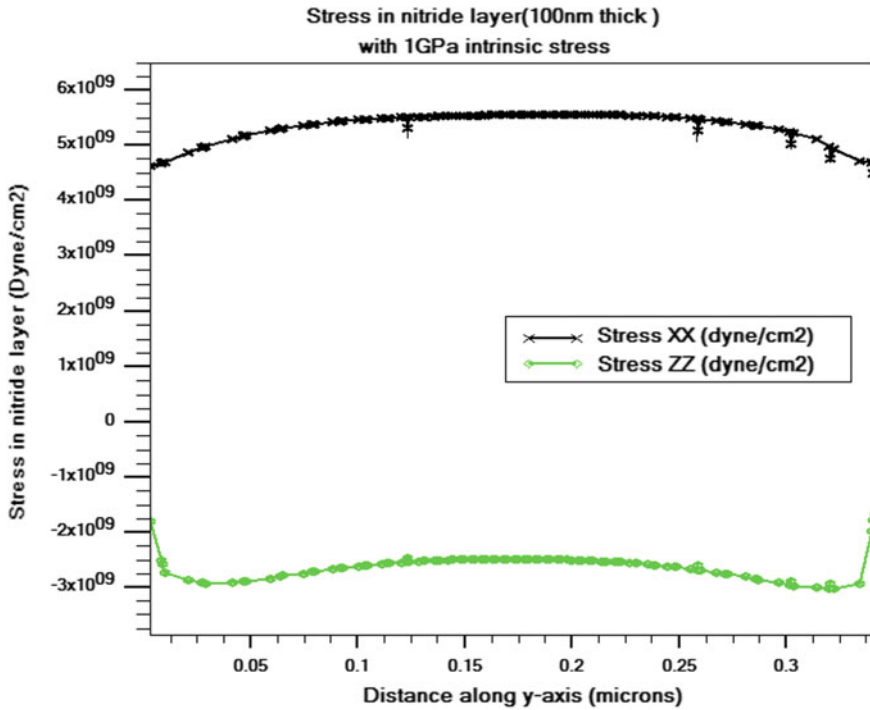


Fig. 4 1D stress profile inside the nitride layer with intrinsic stress +1 GPa

As our focus is to obtain the stress performance in the channel or fin, we have calculated the stress in the fin. A 3D stress profile is shown in Fig. 6 which shows that the parallel stress is nonuniform. To know both parallel and vertical stress field distribution in the fin, we took 2D YZ plane (cut line shown in Fig. 5) and mapped the stress profile in channel part of the fin. In Fig. 6, it can be clearly observed that the parallel stress becomes compressive to tensile from top of fin height to bottom and varies from -220 MPa to $+360$ MPa. The vertical stress is expected to be compressive which varies from -1.5 GPa to -750 MPa. Stress induced in the channel is dependent on the distance between the fin and the capping layer, and it increases toward the bottom of the fin due to the presence of the gate electrode. This results in a nonuniform stress distribution along the fin height. Table 2 summarizes the average stress along each fin surface induced by a 1 GPa tensile nitride layer. The tensile CESL cap generates tensile parallel stress and vertical compressive stress which enhances the electron mobility in n-type FinFET, is reported in Table 2 (Fig. 7).

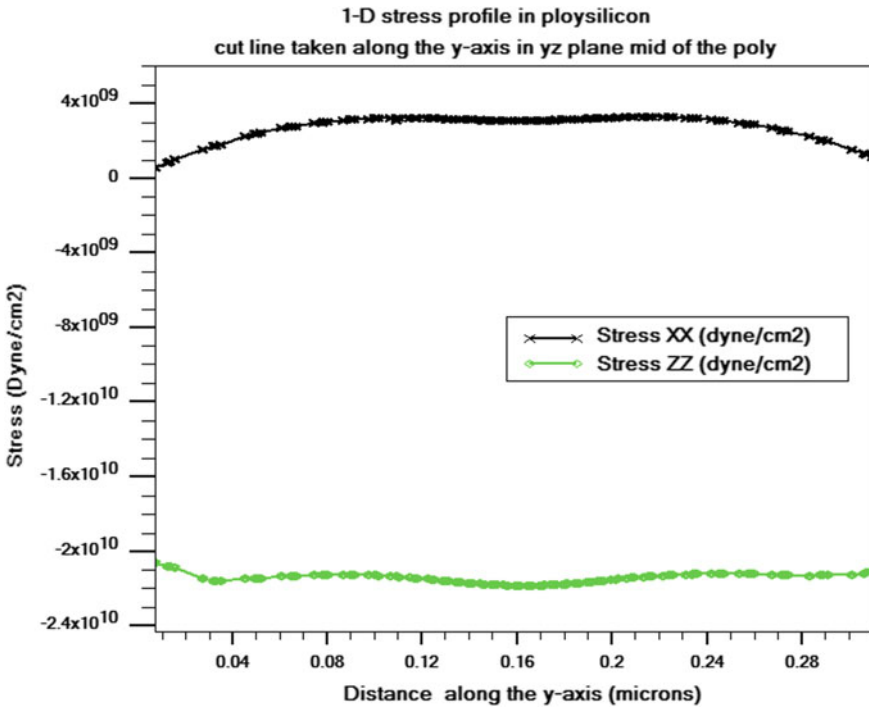


Fig. 5 1D stress distribution in the poly gate

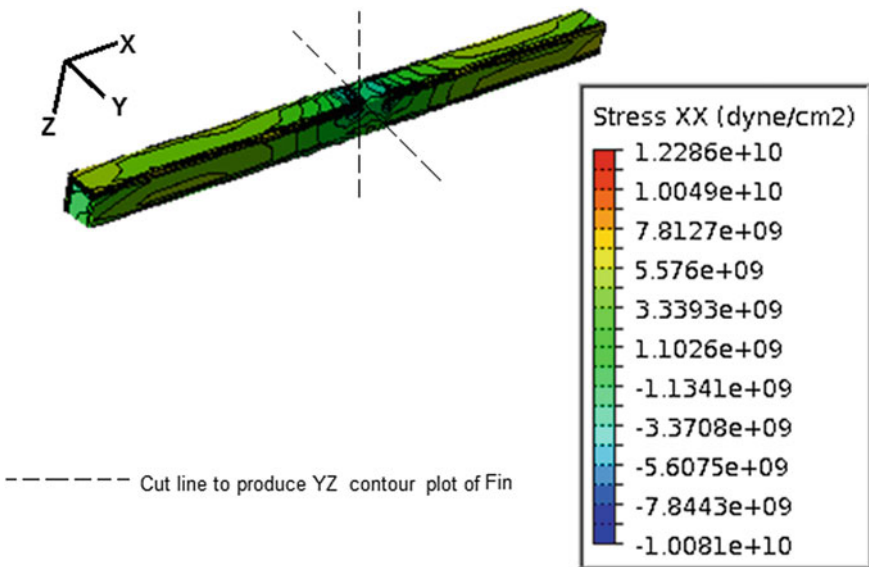


Fig. 6 3D stress Sxx (parallel) profile in the fin

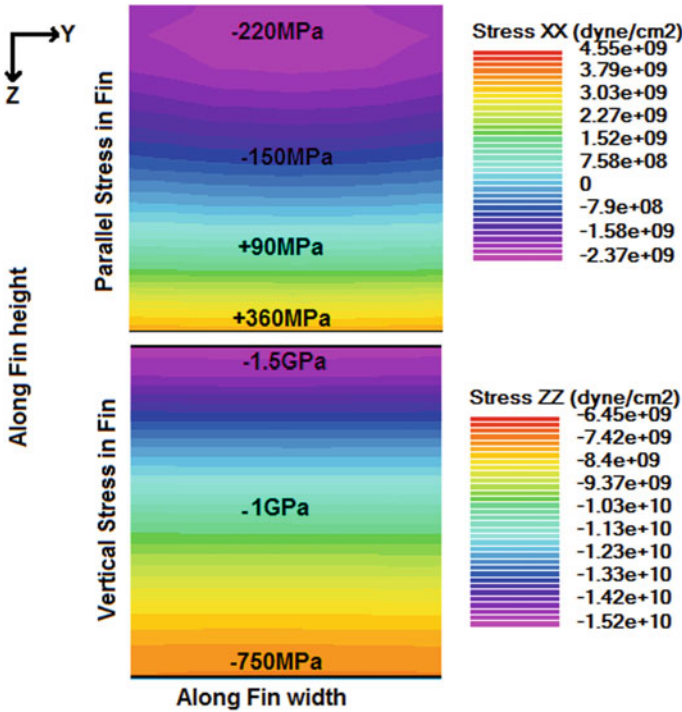


Fig. 7 2D mapping of stress in parallel and vertical direction across the fin

Table 2 Average stress in each direction

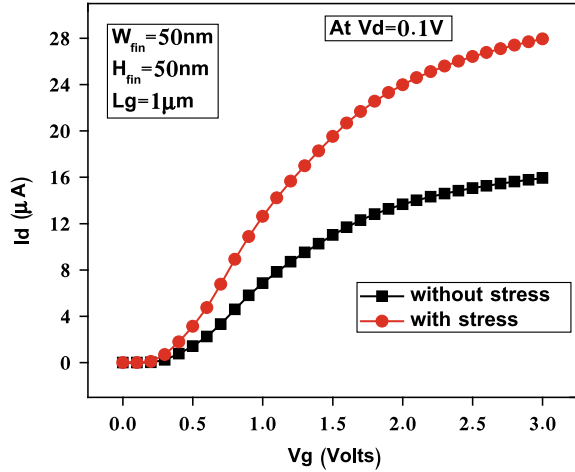
Stress components	Avg. stress in the channel (GPa)
σ_{xx}	0.272
σ_{yy}	-0.206
σ_{zz}	-0.215
Enhancement in μ_n	1.38815

3 Device Simulation

For the device characteristics simulation, silicon models such as Lombardi CVT mobility model, Augur Recombination model, and band gap narrowing model and concentration-dependent recombination model have been used. Besides this, the strain-dependent mobility enhancement models have been used to compute the mobility improvement. The evaluation of mobility enhancement factors along the fin is based on a full 3D piezoresistive model.

The piezoresistance mobility model is used to study the stress impact on transistor performance. The piezoresistance model provides accurate stress-dependent mobility values (within about 20%) at stress levels below 1 GPa [14, 15]. At higher stress levels,

Fig. 8 $I_d - V_g$ characteristics



holes exhibit linear mobility gain with increasing stress, whereas electron mobility gain with stress becomes nonlinear and eventually saturates [16, 17]. The following relation is used for the calculation of n-mobility enhancement factors in silicon with the coefficients of piezoresistivity discussed in [18].

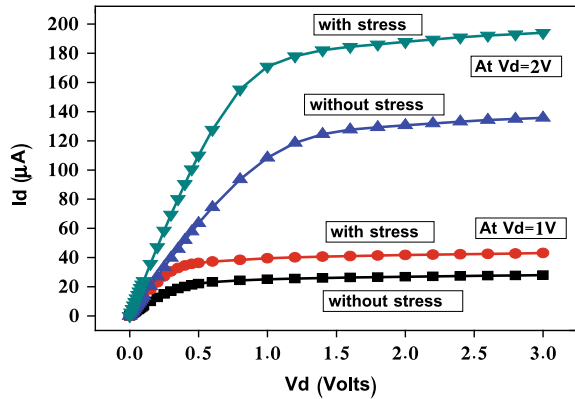
$$\mu_n = 1 - (-1.022\sigma_{xx} + 0.534\sigma_{yy} + 0.534\sigma_{zz})$$

In the equation, the stresses are in units of GPa and coefficients of piezoresistivity are in 1/GPa. All the above transport models along with the mobility improvement factor have been incorporated in the simulation to obtain the electrical performance using device simulator [19].

4 Enhancement in Device Performance

Once stress analysis simulation is complete, the three-dimensional structure is saved and used for further analysis. The aim of introducing stress in device is to obtain the improved electrical performance. The physical models discussed in the last section have been followed to investigate the effect of stress in transfer and output characteristics. Figure 8 shows the variation of drain current with the variation of gate voltage at $V_d = 0.1 V$. Figure 9 shows the output characteristics ($I_d - V_d$) at $V_g = 1$ and 2 V. We have observed 60% enhancement in drain current with stress compared to no stress condition. The enhancement in drain current with stress is due to enhancement in mobility following the piezoresistance mobility model discussed in the last section.

Fig. 9 $I_d - V_d$ characteristics of the n-FinFET



5 Conclusion

A 50×50 nm fin and $1 \mu\text{m}$ long channel FinFET has been virtually fabricated with height and width of 50 nm and deposited with 100 nm thick nitride layer. 3-D stress field distribution has been shown and analyzed in the nitride, polysilicon and within the fin. The contribution of nitride cap layer and polysilicon layer in the stress distribution in the channel part of the fin has been investigated. The amount of induced stress in the channel is dependent on the distance between the capping layer and the fin, which increases toward the bottom of the fin due to the nonzero thickness of the gate electrode. Then, the electrical performance assessment has been carried out considering the piezoresistance effect into account. The mobility enhancement factor has been observed to be 1.38 with the consideration of average stress in all direction. The drain current shows $\sim 60\%$ enhancement with nitride CESL in comparison with that without the CESL layer.

References

1. H.H. Radamson et al., The challenges of advanced CMOS process from 2D to 3D. *Appl. Sci.* **7**, 1047 (2017)
2. T. Ghani et al., A 90 nm high volume manufacturing logic technology featuring novel 45 nm gate length strained silicon CMOS transistors, in *2003 IEEE International Electron Devices Meeting* (2003), pp. 11.6.1–11.6.3
3. G. Eneman et al., Stress simulations for optimal mobility group IV p- and nMOS FinFETs for the 14 nm node and beyond, in *2012 International Electron Devices Meeting* (2012), pp. 6.5.1–6.5.4
4. H.S. Yang et al., Dual stress liner for high performance sub-45 nm gate length SOI CMOS manufacturing, in *IEDM Technical Digest. 2004 International Electron Devices Meeting* (2004), pp. 1075–1077
5. T. Sugii et al., High-performance low operation power transistor for 45 nm node universal applications, in *2006 Symposium on VLSI Technology Digest* (2006). pp. 156–157

6. K.-L. Cheng et al., A highly scaled, high performance 45 nm bulk logic CMOS technology with $0.242 \mu\text{m}^2$ SRAM cell, in *2007 IEEE International Electron Devices Meeting (2007)*, pp. 243–246
7. H.L. Huang, J.-K. Chen, M.P. Houng, Nanoscale CMOSFET performance improvement and reliability study for local strain techniques. *Solid-State Electron.* **79**, 31–36 (2013)
8. Z. Di, L. Qian, W. Xiangzhan, Y. Qi, C. Wei, T. Kaizhou, Performance enhancement of c-CESL-strained 95-nm-gate NMOSFET using trench-based structure. *J. Semiconductors* vol. 36, pp. 014010-1 -4 (2015)
9. H.-W. Hsu, C.C. Lee, Stress analysis of n-channel MOSFET with SiGe channel for different dummy poly gate number and pitch, in *2016 13th IEEE International Conference on Solid-State and Integrated Circuit Technology (ICSICT)* (2016), pp. 1242–1244
10. C.K. Maiti, T.K. Maiti, *Strain-Engineered MOSFETs* (CRC Press, Boca Raton, FL, 2012)
11. D. Guo et al., FINFET technology featuring high mobility SiGe channel for 10 nm and beyond, in *2016 IEEE Symposium on VLSI Technology* (2016), pp. 1–2
12. S. Chatterjee, S. Chattopadhyay, Fraction of insertion of the channel fin as performance booster in strain-engineered p-FinFET devices with insulator-on-silicon substrate. *IEEE Trans. Electron Devices* **65**, 411–418 (2018)
13. D. Bae et al., A novel tensile Si (n) and compressive SiGe (p) dual-channel CMOS FinFET co-integration scheme for 5 nm logic applications and beyond, in *2016 IEEE International Electron Devices Meeting (IEDM)* (2016), pp. 28.1.1–28.1.4
14. S.E. Thompson et al., A logic nanotechnology featuring strained-silicon. *IEEE Electron Device Lett.* **25**, 191–193 (2004)
15. F. Nouri et al., A systematic study of trade-offs in engineering a locally strained pMOS-FET, in *IEDM Technical Digest, 2004 IEEE International Electron Devices Meeting* (2004), pp. 1055–1058
16. Y. Kanda, Graphical representation of the piezoresistance coefficients in silicon. *IEEE Trans. Electron Devices* **29**(1), 64–70 (1982)
17. C.S. Smith, Piezoresistance effect in germanium and silicon. *Phys. Revolution* **94**, 42–49 (1954)
18. M. Yang et al., Hybrid-orientation technology (HOT): opportunities and challenges. *IEEE Trans. Electron Devices* **53**, 965–978 (2006)
19. Silvaco Inc, *Victory Device User's Manual* (2018)

An Annotation System to Annotate Healthcare Information from Tweets



Nixon Dutta, Anupam Mondal and Pritam Paul

Abstract This paper presents a unique idea to utilize social media data for the betterment of healthcare service, provided by the doctors and related industries. The researchers have observed that every day, on average, around 5 million tweets are tweeted on Twitter and around 2 billion tweets per year related to health care. This huge data source can be used and analyzed to obtain better knowledge about recent trends and discoveries in a particular field. Hence, we are motivated to develop a structured corpus from scratch, identify concepts and categories using the machine learning approach on the extracted unstructured and semi-structured corpora. In order to build the system, we have employed two well-known classifiers, namely multinomial Naive Bayes and support vector machine on the top of our prepared experimental dataset. The training and test datasets are part of the experimental dataset and have been used to build the module and validate them, respectively. The proposed module is able to assign healthcare concepts and their categories for tweets. Finally, the validation offers F-measure 0.67 and 0.57 for concept identification as well as categorization system individually. This annotation system may help to design various applications such as summarization and recommendation system in health care for assisting medical practitioners.

Keywords Twitter · Health care · Diseases · Treatment · Machine learning

N. Dutta (✉) · A. Mondal · P. Paul
Department of Computer Science & Engineering, Institute of Engineering and Management,
Y-12, Salt Lake Electronics Complex, Sector-V, Kolkata 700091, India
e-mail: nixondutta402@gmail.com

A. Mondal
e-mail: anupam.mondal@iemcal.com

P. Paul
e-mail: pritampaul9007@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_30

1 Introduction

The recent trends in healthcare services focus on extracting the knowledge from the daily produced electronics data using natural language processing (NLP). Primarily, the extracted data are presented as medical concepts and their categories. Additionally, when it comes to healthcare knowledge and management, Google Health and Microsoft HealthVault are making people feel more powerful by managing and keeping better track of our health. Presently, healthcare domain uses electronic health records as its standard. EHR system primarily presents how we can have immediate access to patient's laboratories test results, allergies, and diagnoses that help to make better decisions. We need better and more reliable access to information in order to expand the current situation with the EHR system. Currently, Medline is the fastest and most used data source for information and consists of articles on life science [1, 2].

Nowadays, many healthcare-related information has been discussed and forwarded through social media. This information provides a knowledge-based support to doctors and medical practitioners to know the recent discoveries in their field. They also need to have an idea of the trends and highly discussed topics in social media in their relevant fields. Being notified of this information will help the healthcare providers to provide better service. Reading the latest life science topics will also help the people who want to be in charge of their health.

In this paper, we are motivated to develop a platform for converting a structured corpus from the unstructured tweets. The structured corpus provides the healthcare concepts and categories to extract the knowledge-based information from tweets. Hence, we have designed an annotation system in health care to convert structured corpus from largely available unstructured corpora such as extracted data from social media. Our purpose is to build a platform, which will take search query as input, and in return give a highly discussed relevant set of information in the structured form as output. This paper deals with identifying informative sentences related to medical fields which will be extracted from Twitter and later stored in the database. It will also discuss assigning categories and concept to the tweets. To obtain the most accurate result, machine learning and NLP techniques will be used.

2 Related Work

Oana Frunza et al. have tried to identify disease–treatment relations in short texts by a machine learning approach [3, 4]. The system is able to build an application, which identifies and disseminates healthcare information. The system identifies semantic relations between diseases and treatments and extracts sentences from Medline that mention diseases and treatments. They complete two tasks: firstly, they identify between informative and non-informative sentence from the text, and secondly, they identify three semantic relations: cure, prevent, and side effect. They propose a unique

way to do both tasks in the same framework at once by putting the output of the first task as the input of the second task. This unique approach helped to increase the accuracy as the second task only has to deal with the informative sentence collected from the first task, which reduces lots of error. Three representation techniques, namely bag-of-word representation, NLP and biomedical concept representation, and medical concept representation, are used to obtain the treatment relation from the short text.

Aurangzeb Khan et al. proposed a review of machine learning algorithms [5]. This paper highlights the important techniques and methodologies, employed in text classification, and attempts to create awareness of some unsolved challenges, focused mainly on text representation and machine learning techniques. This paper discusses support vector machine (SVM), genetic algorithm, fuzzy correlation, artificial neural network, decision rules classification, Naïve Bayes algorithm, K-nearest neighbor (k-NN), Rocchio's algorithm, and hybrid techniques.

Abdur Rehman et al. developed a feature extraction algorithm for classifying text documents [6]. They have suggested various features like words, term frequency (TF) and inverse document frequency (IDF) to produce good classification results [7]. Unfortunately, these features have not provided adequate accuracy in every case.

Bharti E. Nerkar et al. proposed a method in which using machine learning approach, he tried to identify best treatment for diseases in relation to short text [8]. The paper discusses two tasks, (I) to identify and extract of informative sentences on diseases and related treatment topics and (II) to classify of these sentences with respect to semantic relations that exist between diseases and treatments. It takes symptoms as input and gives summary of disease as outputs. It also identifies the sentence as informative or non-informative and labels them as such. First, it downloads a pdf. Then it pre-processes the data, extracts relevant data, and stores them in the database. Next, it classifies symptoms using SVM. It uses three semantic relations to identify the exact diseases. After identification, the disease is sent to the medical database to extract the articles pertaining to it. Tokenization, stemming, and stop words removal is done in pre-processing steps. Combination of the bow, NLP, and biomedical concepts are put together to identify semantic relations. Keyword searching algorithm is used to extract pertinent information.

3 Proposed Method

3.1 Task and Datasets

The main task undertaken here helps to design the desired framework that identifies the relevant tweets and presents it to the end users. The first task identifies and extracts all the related tweets from Twitter according to the query. The second task recognizes different feature from data and performs classification of these texts according to the features. The final product can be a browser plug-in or an application which will

find and extract the most recent and highly discussed medical-related issues and will present them to the user. From an e-commerce point of view, the value of the product stands out as it will notify the recent trends and help them to understand the customer's interest and psychology better.

In order to build the proposed system, firstly we have collected a Twitter dataset and preprocess it. The dataset consists of short texts which contain the information related to disease and treatment. The Twitter API and python is used to prepare the dataset. We used `consumer_key`, `consumer_secret`, `oauth_token`, and `oauth_token_secret` to authenticate and import `tweepy` module. In the process, we have used manually prepared search queries related to health care to crawl the tweets, which include page name, user name, tweets, Ids, created date, and other entities. Thereafter, we have preprocessed the raw dataset and store the cleaned tweets in the database. The manually prepared search queries are represented as categories in this research.

In the following subsection, we have described the feature extraction steps for the experimental dataset to build the proposed system.

3.2 Feature Extraction

The feature extraction plays a crucial role in the process of developing the annotation system for healthcare services [9, 10]. We have extracted various features such as term frequency (TF), inverse document frequency (IDF), and TFIDF of the crawled 34,000 tweets. Besides, we have also recognized the semantic features such as the count of noun, verb, adverb, and adjective from the tweets. These features have been presented as a vector to build and validate the proposed system using machine learning classifiers.

Thereafter, we have split the dataset into two parts, namely training and test dataset. The training dataset contains 20,000 annotated tweets whereas the rest of 14,000 tweets are presenting as a test dataset. The annotated tweets are tagged with healthcare-related concepts and their categories. The extracted features of the training dataset have been applied through two well-known supervised classifiers such as multinomial Naive Bayes and support vector machine to develop the annotation system. Figure 1 describes a flow diagram of the proposed system.

3.3 Module Building

In order to design the healthcare annotation system, we have employed two well-known supervised classifiers like multinomial Naive Bayes and support vector machine (SVM). Multinomial Naïve Bayes classifier depends on Bayes' theorem with an assumption of independence among all predictor classifiers. Naive Bayes models are very efficient because looking at each feature individually they learn

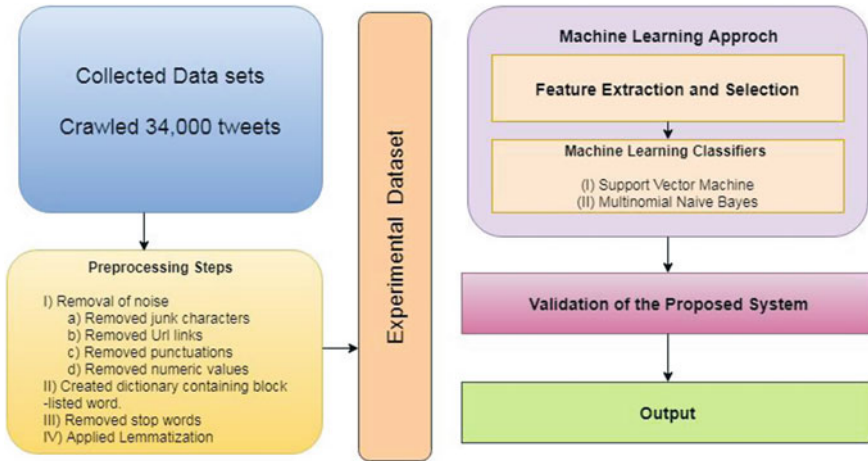


Fig. 1 A flow diagram of the proposed system

parameters and collect simple per-class statistics from each feature. In many cases involving large data sets, Naive Bayes outperforms even highly sophisticated classification methods [11]. On the other hand, SVM classifier offers a better accuracy over multinomial Naïve Bayes algorithm using ‘rbf’ kernel.

These classifiers assist in learning the module using the extracted features of training dataset. The learned modules help to predict the healthcare concepts and their categories from a provided tweet. The following algorithm illustrates the module building steps in detail.

Algorithm

- Step 1 Initially, we have prepared an experimental dataset and extracted various features $(F) = [F1, F2, F3, \dots, Fn]$. We have also labelled (L) the experimental dataset for both concepts $(L_{Concepts})$ and their categories $(L_{Categories})$.
- Step 2 We have employed multinomial Naive Bayes (M_{MNB}) and support vector machine (M_{SVM}) classifiers on the training dataset to learn the proposed system.
- Step 3 Thereafter, we have combined both of the classifiers by union operation to build another module $(M_{Combined})$ under the proposed system.
- Step 4 Finally, the test dataset has been applied on the mentioned three modules under the proposed system to validate that.

In the following section, we have discussed the evaluation process of the proposed system with three modules individually.

Table 1 An evaluation of the proposed concept identification system using precision, recall, and F-measure for all classifiers

Classifiers	Precision	Recall	F-measure
M_{MNB}	0.642	0.683	0.662
M_{SVM}	0.636	0.646	0.641
$M_{Combined}$	0.646	0.689	0.667

Table 2 An evaluation of the proposed categorization system using precision, recall, and F-measure for all classifiers

Classifiers	Precision	Recall	F-measure
M_{MNB}	0.533	0.541	0.537
M_{SVM}	0.558	0.562	0.560
$M_{Combined}$	0.565	0.568	0.566

Table 3 A confusion matrix representation for identifying medical concepts using SVM classifier

Sample: 14000		Predicted		
		Medical concepts	Non-medical concepts	Total
Original	Medical concepts	5346	2926	8272
	Non-medical concepts	3053	2675	5728
	Total	8399	5601	

4 Result Analysis

In order to validate the proposed healthcare concept identification and categorization system, we have used the test dataset as a part of experimental dataset. This dataset has been applied to the machine learning classifiers and calculated the accuracy in the form of precision, recall, and F-measure. Tables 1 and 2 present a comparative analysis between all the mentioned classifiers to the process of designing medical concept as well as categorization system.

Additionally, we have generated the confusion matrix for medical concept identification and categorization systems. Table 3 shows a confusion matrix of concept identification using SVM classifier.

The result shows that the combined classifier provides an adequate output for both of the proposed systems.

5 Conclusion

The research was primarily focused on developing an annotation system in health care. The proposed system is able to annotate concepts and categories related to healthcare services for tweets. Hence, we have employed the machine learning tech-

niques, i.e., a supervised learning method with (a) multinomial Naïve Bayes classifiers or (b) support vector machine classifiers to build the system. We have compared the results of different classifiers to better understand the utility of machine learning classifiers in this field. The obtained results show that the approach we have taken significantly outperforms the other techniques. The combined classifier comparatively provides a better alternative to other classifiers used to enhance machine learning performance. In this proposed work, using keyword we extracted the relevant tweets from Twitter and pre-processed to prepare the experimental dataset. Thereafter, we have extracted features from the experimental dataset and process through classifiers. The classifiers are predicted the output in the form of concepts and categories.

In the future, we can add a number of knowledge-based features to improve the obtained result of the proposed system. It also assists in designing various applications like relationship extraction system in the healthcare domain.

References

1. M. Craven, Learning to extract relations from medline, in *Proceedings of Association for the Advancement of Artificial Intelligence* (1999)
2. A. Mondal, E. Cambria, D. Das, A. Hussain, S. Bandyopadhyay, Relation extraction of medical concepts using categorization and sentiment analysis. *Cognitive Comput.* 1–16 (2018)
3. A. Khan, B. Baharudin, L.H. Lee, K. Khan, A review of machine learning algorithms for text-documents classification. *J. Adv. Inf. Technol.* **1**(1), 4–20 (2010)
4. O. Frunza, D. Inkpen, T. Tran, A machine learning approach for identifying disease-treatment relations in short texts. *IEEE Trans. Knowl. Data Eng.* **23**(6), 801–814 (2011)
5. V. Korde, C.N. Mahender, Text classification and classifiers: a survey. *Int. J. Artif. Intell. Appl. (IJAA)* **3**(2), 85 (2012)
6. A. Rehman, H.A. Babri, M. Saeed, Feature extraction algorithms for classification of text documents, in *International Conference on Computer and Information Technology* (2012)
7. S. Robertson, Understanding inverse document frequency: on theoretical arguments for IDF. *J. Documentation* **60**(5), 503–520 (2004)
8. M. Yetisgen-Yildiz, W. Pratt, The effect of feature representation on MEDLINE document classification, in *AMIA Annual Symposium Proceedings* (2005), pp. 849–853
9. A. Dasgupta, Feature selection methods for text classification, in *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2007), pp. 230–239
10. A. Mondal, D. Das, S. Bandyopadhyay, Relationship extraction based on category of medical concepts from lexical contexts, in *14th International Conference on Natural Language Processing (ICON 2017)* (2017), Kolkata, India, pp. 212–219
11. A. Mondal, E. Cambria, A. Feraco, D. Das, S. Bandyopadhyay, Auto-categorization of medical concepts and contexts, in *2017 IEEE Symposium Series on Computational Intelligence (SSCI)* (IEEE, 2017, November), pp. 1–7

Broad Neural Network for Change Detection in Aerial Images



Shailesh Shrivastava, Alakh Aggarwal and Pratik Chattopadhyay

Abstract A change detection system takes as input two images of a region captured at two different times and predicts which pixels in the region have undergone change over the time period. Since pixel-based analysis can be erroneous due to noise, illumination difference and other factors, contextual information is usually used to determine the class of a pixel (changed or not). This contextual information is taken into account by considering a pixel of the difference image along with its neighborhood. With the help of ground truth information, the labeled patterns are generated. Finally, Broad Learning classifier is used to get prediction about the class of each pixel. Results show that Broad Learning can classify the data set with a significantly higher F-score than that of Multilayer Perceptron. Performance comparison has also been made with other popular classifiers, namely Multilayer Perceptron and Random Forest.

Keywords Broad neural network · Change detection · Imbalance ratio · Contextual information

1 Introduction

Change detection in images has become one of the important aspects in recent years, with a wide range of applications. Various aspects include detection of changes in motion, change in vegetation cover over an area, extent of damage caused due to floods and other natural calamities, etc. The basic objective of change detection is to identify dissimilar regions by comparing a pair of the images of the same area at different points of time.

S. Shrivastava
Indian Institute of Technology, Patna, India

A. Aggarwal · P. Chattopadhyay (✉)
Indian Institute of Technology (Banaras Hindu University), Varanasi, India
e-mail: pratik.cse@iitbhu.ac.in

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_31

Over the past few years, a lot of research work has been carried out in this area, and several supervised, unsupervised and semi-supervised learning techniques have been developed, each with some pros and cons. While unsupervised learning does not need any training information, supervised learning techniques require substantial amount of ground truth knowledge, which may not be conveniently available always. On the other hand, the semi-supervised technique is efficient in a way that it requires only a few labeled data for training. The few labeled patterns of the changed class can help in training a classifier by means of self-training.

Recently, Broad Learning has been proposed in [1] as an alternative to Deep Learning. It has been seen that, Broad Learning is immensely successful in classifying image feature sets, and in some cases it performs even better than that of Deep Learning. Moreover, training using Broad Learning is significantly faster than that of Deep Learning. In this work, we use a variation of the Broad Learning algorithm proposed in [1] and compare its results with the Multilayer Perceptron and Random Forest classifiers.

2 Literature Survey

A change detection method essentially does comparison between two images to differentiate between changed and unchanged pixels. Early methods are mostly threshold-based in which a difference image D is first computed from the two input images say, X and Y , and a threshold is applied on the difference image to detect the changes. The threshold is chosen empirically as explained by Rosin [2, 3]. This is referred to as ‘simple differencing’ [4]. Two different techniques are usually employed to generate the difference image: Change Vector Analysis (CVA) [5–7] and image rationing [8]. While the former uses the modulus of the difference between the feature vectors of each pixel, the latter uses the ratio of the pixel intensities to generate the difference image. However, manual selection of threshold is erroneous and is not robust in presence of changes in illumination and noise.

In view of the two hypothesis used in change detection algorithms, i.e., the null hypothesis H_0 corresponding to a no-change decision and the alternative hypothesis H_1 corresponding to a change decision at a given pixel, the hypothesis to be chosen to best describe the extent of change at a given pixel (say, n) is based on the conditional joint probability density functions $p(X(n), Y(n)|H_0)$ and $p(X(n), Y(n)|H_1)$, based on the classical framework of hypothesis testing [9, 10].

Later work on change detection includes [11] in which pixels are softly classified into mixture components for different change models by employing the expectation–maximization algorithm. The algorithm computes the optical flow field between a pair of images. This approach seems to be an improvement over the previous techniques as it estimates mis-alignment, as well as the illumination variation between the input images. Another change detection algorithm based on the concept of ‘self-consistency’ among the regions of an image is proposed in [12]. This approach uses

the Minimum Description Length (MDL) model [13] for selecting the hypothesis that best describes the image pair.

More sophisticated change detection algorithms make use of the spatial and temporal relation between a pixel and its neighboring pixels. The spatial models consider a polynomial function of a pixel coordinate n upon which the intensity values of a block of pixels is fitted. Constant, linear and quadratic models are used on these blocks of pixels via generalized likelihood ratio tests [14, 15]. The threshold value can be obtained by t-test for constant models and F-test for linear or quadratic models. Out of the three models, the quadratic model is found to be a better option compared to the constant and linear models due to its higher confidence in prediction.

Numerous temporal models have been proposed for detecting changes in sequence of images. Variation of pixel intensities over time is modeled as an auto-regressive process [16]. Mean, variance and correlation coefficient were estimated and used in likelihood ratio tests where the null hypothesis depicts that the image intensities are dependent and the alternative hypothesis depicts them as independent [17].

Another algorithm known as the Wallflower algorithm uses a Wiener filter for predicting a pixel's value from its previous values [18]. Pixels are classified as changed if the prediction error is much more than the expected error. This algorithm can also be considered as a background estimation algorithm. Due to low accuracy of the linear models, nonlinear models are proposed to study the relationship between images [19]. The optimal nonlinear function is the conditional expected intensity value $X(n)$ given $Y(n)$. An unsupervised approach to estimate the parameters of the nonlinear predictor is proposed using adaptive neural networks [20]. Those pixels on which the predictor performed poorly are classified as changed.

The shading models are used in various change detection techniques to produce illumination-invariant algorithms. These algorithms computed ratio of the intensity of corresponding pixels in the two images as given by:

$$R(n) = \frac{X(n)}{Y(n)} \quad (1)$$

and compare it to an empirically determined threshold. Instead of directly comparing the ratio $R(n)$, a new form of the difference image was introduced in [21]:

$$D(n) = \frac{1}{N} \sum_{l \in \Omega_n} (R(n) - \mu_n)^2, \quad (2)$$

where N is the number of pixels in a block of pixels Ω_n , l is a pixel in the pixel block Ω_n , and μ_n is given by Eq.3:

$$\mu_n = \frac{1}{N} \sum_{l \in \Omega_n} R(n). \quad (3)$$

If $D(n)$ is greater than the threshold, it is considered as a changed pixel.

A change detection algorithm based on the shading model is introduced in [22] that uses the weighted combinations of difference in intensities and textures and the intensity ratio. This algorithm assumes that the texture information is less sensitive to variations in illuminations as compared to intensity difference. Instead of using the intensity ratios, the technique in [23] compares the circular shift moments to detect the changes in image. The reflectance component of the intensity is represented by these moments and is claimed to capture the details of an image more precisely.

An algorithm using semi-supervised Multilayer Perceptron was employed in [24] for detecting changes in remotely sensed images. Here, the difference image is generated by the CVA technique using the difference operator. The input pattern for each pixel is generated using the pixel intensity of the difference image as well as those within a spatial neighborhood. Thus, the input vectors contain nine components with gray values of a pixel and gray values of its eight neighboring pixels. A few initial labeled patterns were identified automatically using the K-means clustering algorithm. A neural network is trained using these labeled patterns initially. Since the input patterns have nine-pixel values, so the input layer has nine units and one bias unit. The output layer has two units for the unchanged and changed classes. The remaining unlabeled patterns are processed by the perceptron in a semi-supervised manner by obtain a soft class label for each unlabeled pattern in every iteration, and next using these soft labels along with the ground truth to train the network repeatedly till optimal condition is reached. This approach is highly time-intensive and suffers from the disadvantage that incorrect prediction at an initial iteration, propagates to the successive iterations as well.

Another change detection technique for remotely sensed images is proposed in [25] that carries out fusion of spectral and statistical indices in an unsupervised manner. The spectral changes are taken into account via the CVA technique, while the statistical changes are identified using a similarity index based on Kullback Leibler distance. The spectral change information and similarity index are fused using wavelets. A neuro-fuzzy classifier next classifies the fused image into changed and unchanged classes, overcoming mixed pixel problems of the satellite images.

In this work, we aim to study the performance of Broad Learning in classifying the difference image. It has been shown in previous studies [1] that Broad Learning is faster and also has a similar (sometimes better) level of accuracy as that of Deep Learning in classifying image feature sets. Moreover, Broad Learning algorithm is incremental, i.e., with the availability of more ground truth data, training of the classifier can be accomplished very fast.

3 Proposed Approach

A block diagram of the proposed approach is explained using the block diagram shown in Fig. 1.

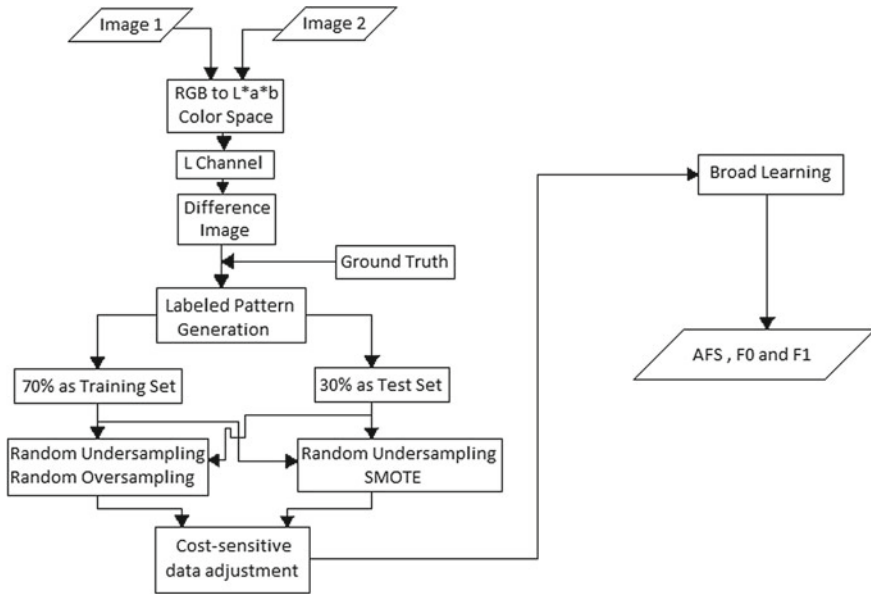


Fig. 1 System overview of change detection from a pair of aerial images captured at different time instants

3.1 Input Pattern Generation

Consider two co-registered and geometrically corrected RGB images of the same place taken at two different times. Since $L^*a^*b^*$ color space is known to be illumination invariant, the two images are initially transformed from RGB color space to $L^*a^*b^*$ color space. For generating the input pattern, a pixel is chosen along with its eight neighboring pixels. Thus, the input feature for each pixel is a vector in nine-dimensional feature space. The difference image is generated via absolute simple difference of the two images for each channel.

A general characteristic of change detection data set is that the ground truth consists of large samples from unchanged class and very few samples of changed class. Training any learning algorithm with this data set will cause a high bias for the unchanged class. To account for this problem, we study two different cost-sensitive data adjustment techniques. First the unchanged class of the training is down-sampled using random down-sampling. Next the changed class is up-sampled using a random up-sampling technique, namely *Synthetic Minority Oversampling Technique (SMOTE)*.

We construct various data sets with different *Imbalance Ratios* (i.e., ratio of the number of majority to the number of minority samples) and study the performance of the classifier for each such case.

3.2 Classification Using Broad Learning

Broad Learning network [1] is a recently proposed two-layer deep neural network which consists of input feature map nodes in the input layer and enhancement nodes in the hidden layer. The original version of Broad Learning has been seen to perform with significant accuracy in classifying some image data sets. Moreover, training using Broad Learning is very fast. Detailed architecture of the network, along with its working principle, is given in [1]. One drawback of the approach in [1] is that it does not specify the number of enhancement nodes to be added for optimum performance. This is important since we have limited memory space, and adding enhancement nodes without bounds would cause the memory to overflow.

So, we propose a modification to the existing architecture of the Broad Learning System. We continue to recursively add new layer of enhancement nodes till a desired optimum condition is reached. Sparse Autoencoder is employed to find new layer of enhancement nodes at each recursion level with a pre-specified compression rate. Finally, all the generated layers are concatenated with the input layer. The weights from the set of input and enhancement layers to the output layer are next determined using Moore-Penrose Pseudo-inverse, similar to that in [1]. The scheme is explained with the help of Fig. 2.

At each step after adding a new layer of enhancement nodes, we compute the cross-validated average F-score on classifying the training set. Addition of enhancement

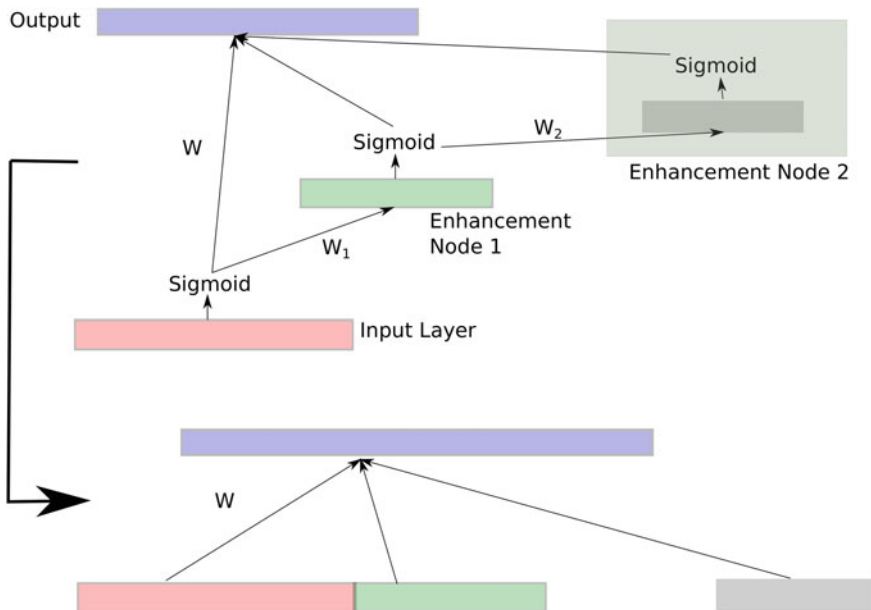


Fig. 2 Broad Learning architecture

nodes is terminated once the average F-score values in two successive iterations are less than a small threshold value. At this point, further addition of new enhancement nodes does not improve classification accuracy on training set any further, but the network tends to get over-fitted.

4 Results

Evaluation of our algorithm is done on a publicly available data set for change detection¹ which consists of 1000 pairs of 800×600 images, each pair consisting of one reference image and one test image, and the 1000 corresponding 800×600 ground truth masks. The images were rendered using the realistic rendering engine of the serious game Virtual Battle Station 2, developed by Bohemia Interactive Simulations. The data set consists of 100 different scenes containing several objects (trees, buildings, etc.) and moderate ground relief. Using the reference and test images, the difference image is created and the ground truth masks are used for labeling the data set and final evaluation of the classifier. To construct the training and test sets, from the ground truth, we randomly select 70% samples from each class as the training set and remaining 30% for test set.

We study the effect of various sampling techniques as described above, as well as the effect of decreasing imbalance ratio on the classification performance of Broad Learning. Table 1 shows the F-scores for the two classes along with the average F-score using the proposed Broad Learning neural network.

In the table, AFS represents the average F-score, F0 is the F-score of class 0, and F1 is the F-score of class 1. We observe the effect of varying the number of layers of Enhancement nodes as well as the compression rate while generating a new layer of enhancement node. The set of enhancement nodes generated at each recursion level is termed as a layer, while compression is the ratio of number of nodes in the newly generated layer to that of the previous layer.

From the table, we conclude that SMOTE gives better results and is more effective than random oversampling. Also, as expected, F-score for minority class (class 1) tends to zero as the imbalance ratio in data set increases. It is seen that the Broad Learning classifier is able to perform reasonably well even with the imbalance ratio of 10:1. Also, a higher compression results in more number of nodes in the enhancement layer and hence fits the data well. However, the effect of varying the number of layers is not well observed from the data set used in the study. Figure 3 shows the changed map obtained with two pairs of input images using the Broad Learning classifier with 10:1 imbalance ratio. The first two images shown in Fig. 3a, b are aerial images of the same place captured at different times, as obtained from the database. Figure 3c shows the ground truth, while Fig. 3d shows the change detection map obtained by using the proposed method. Visual observation shows that most of the changed regions are

¹<https://computervisiononline.com/dataset/1105138664>.

Table 1 Table showing the F-scores for the two classes along with the average F-score using the Broad Learning network

Broad learning approach		Random undersampling random oversampling									
Imbalance ratio	IR	Random undersampling random oversampling					Random undersampling SMOTE				
		Layers	Compression	AFS	F0	F1	Layers	Compression	AFS	F0	F1
1:1		3	0.9	75.8	80	72	3	0.9	81.6	81.6	81.6
		3	0.7	70	73	67	3	0.7	81.5	81.4	81.6
		5	0.9	76	80	72	5	0.9	81.6	81.6	81.6
		5	0.7	70	73	67	5	0.7	81.5	81.4	81.6
		3	0.9	79	88	63	3	0.9	79	85	66
2:1		3	0.7	74	88	55	3	0.7	77	84	63
		5	0.9	79	88	63	5	0.9	79	85	66
		5	0.7	74	88	55	5	0.7	77	84	63
		3	0.9	90	96	27	3	0.9	88	95	10
		3	0.7	88	95	10	3	0.7	87	95	4
10:1		5	0.9	90	96	27	5	0.9	88	95	10
		5	0.7	88	95	10	5	0.7	87	95	4
		3	0.9	97	99	8	3	0.9	0	99	0
		3	0.7	0	99	0	3	0.9	0	99	0
		5	0.9	97	99	8	5	0.9	0	99	0
50:1		5	0.7	0	99	0	5	0.7	0	99	0
		3	0.9	99	99	1	3	0.9	0	99	0
		3	0.7	0	100	0	3	0.7	0	100	0
		5	0.9	99	99	1	5	0.9	0	99	0
		5	0.7	0	100	0	5	0.7	0	100	0
100:1		3	0.9	0	100	0	3	0.9	0	100	0
		3	0.7	0	100	0	3	0.7	0	100	0
		5	0.9	99	99	1	5	0.9	0	99	0
		5	0.7	0	100	0	5	0.7	0	100	0
		3	0.9	0	100	0	3	0.9	0	100	0
250:1		3	0.7	0	100	0	3	0.7	0	100	0
		5	0.9	0	100	0	5	0.9	0	100	0
		5	0.7	0	100	0	5	0.7	0	100	0
		3	0.9	0	100	0	3	0.9	0	100	0
		5	0.7	0	100	0	5	0.7	0	100	0

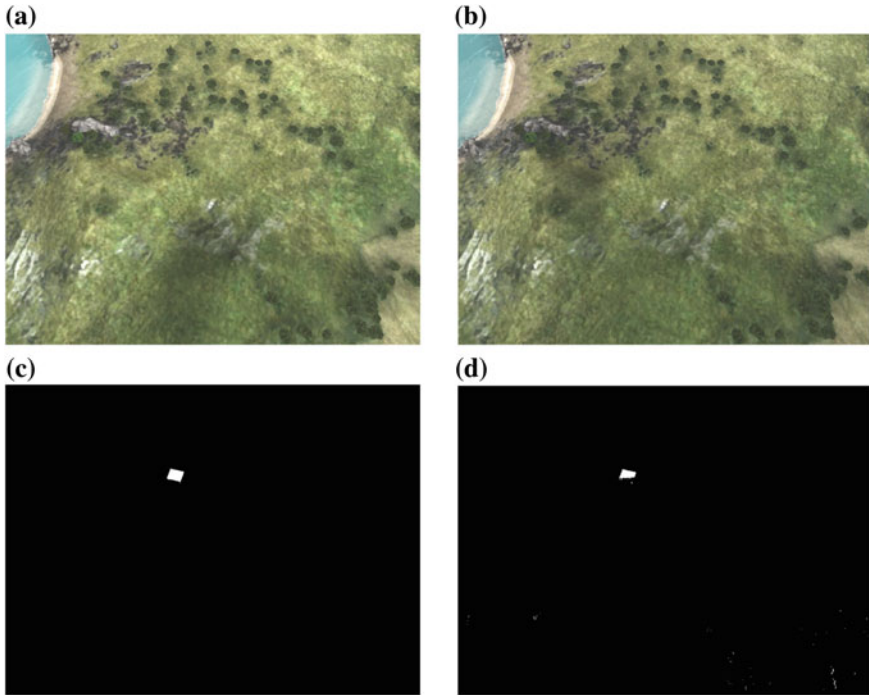


Fig. 3 a and b The two input images, c ground truth, d predicted changed map from 10:1 imbalanced ratio

correctly identified by our algorithm, although there exist few false positive cases possibly occurring due to noise/variation of illumination conditions.

We also compare the performance of our approach with two other popularly used classifiers, namely, Random Forest and Multilayer Perceptron. Table 2 shows the F-scores for the two classes along with the average F-score for both the class using the Random Forest classifier. Table 3 shows the corresponding results for Multilayer Perceptron classifier.

In each cell, value of '0' implies that there is no correct prediction for that class. The above results show that both Broad Learning and Random Forest handle imbalanced data better than a Multilayer Perceptron. Table 2 shows that performance of Random Forest classifier is almost consistent across the various imbalance ratios, but its effectiveness is too much dependent on the choice of a suitable data adjustment technique. In other words, Random Forest may not be able to perform well on any given data. On the other hand, although the effectiveness of Broad Learning depends on the choice of a suitable imbalance ratio, its performance is consistent for any given data adjustment technique, i.e., Broad Learning can robustly handle noisy data as long as the imbalance ratio is small. Similar to Random Forest, an ensemble of

Table 2 Table showing the F-scores for the two classes along with the average F-score using the Random Forest classifier

Random Forest classifier		Random undersampling random oversampling				Random undersampling SMOTE						
Imbalance ratio	Trees	AFS	F0	F1	Trees	AFS	F0	F1	Trees	AFS	F0	F1
1:1	5	43.39	60.88	17.89	5	89.29	89.99	88.59	5	89.29	89.99	88.59
	10	37.69	67.58	7.81	10	89.40	90.17	88.63	10	89.40	90.17	88.63
2:1	5	52.42	82.07	22.77	5	89.21	93.38	85.03	5	89.21	93.38	85.03
	10	46.80	81.07	12.53	10	88.07	92.88	83.26	10	88.07	92.88	83.26
10:1	5	61.76	95.92	27.60	5	82.55	97.35	67.75	5	82.55	97.35	67.75
	10	56.99	95.68	18.31	10	80.69	97.21	64.17	10	80.69	97.21	64.17
50:1	5	64.66	99.14	30.18	5	68.98	99.12	38.84	5	68.98	99.12	38.84
	10	61.99	99.12	24.85	10	66.24	99.12	33.36	10	66.24	99.12	33.36
100:1	5	62.56	99.53	25.59	5	63.77	99.48	28.06	5	63.77	99.48	28.06
	10	60.08	99.54	20.62	10	59.48	99.50	19.46	10	59.48	99.50	19.46
250:1	5	58.98	99.78	18.18	5	56.19	99.78	12.60	5	56.19	99.78	12.60
	10	56.51	99.80	13.22	10	51.71	99.79	3.64	10	51.71	99.79	3.64

Table 3 Table showing the F-scores for the two classes along with the average F-score using MLP

Multi layer perceptron		Random undersampling random oversampling						Random undersampling SMOTE					
Imbalance ratio	Alpha	AFS	F0	F1	Alpha	AFS	F0	F1	Alpha	AFS	F0	F1	
1:1	0.5	0	66.67	0	0.5	33.34	66.65	0.02	0.5	33.34	66.65	0.02	
	1	36.26	5.24	67.27	1	41.48	14.51	68.45	1	41.48	14.51	68.45	
2:1	0.5	25.03	0.06	50.01	0.5	0	80.00	0	0.5	0	80.00	0	
	1	39.01	24.85	53.18	1	0	80.00	0	1	0	80.00	0	
10:1	0.5	0	95.24	0	0.5	0	95.24	0	0.5	0	95.24	0	
	1	0	95.24	0	1	0	95.24	0	1	0	95.24	0	
50:1	0.5	0	99.01	0	0.5	0	99.01	0	0.5	0	99.01	0	
	1	0	99.01	0	1	0	99.01	0	1	0	99.01	0	
100:1	0.5	0	99.50	0	0.5	0	99.50	0	0.5	0	99.50	0	
	1	0	99.50	0	1	0	99.50	0	1	0	99.50	0	
250:1	0.5	0	99.80	0	0.5	0	99.80	0	0.5	0	99.80	0	
	1	0	99.80	0	1	0	99.80	0	1	0	99.80	0	

Broad Learners can be employed to study if the classification performance can be made robust to variation in imbalance ratio as well.

Working with deep neural network requires a large amount of training data. The feature set for change detection, as described here, is of small dimension and is not suitable for training a deep learning model. Moreover, as shown in [1], Broad Learning performs better and in a more efficient way on low-dimensional data as compared to that of Deep Learning Approaches. So, we have deliberately avoided comparison with Deep Learning in the present paper.

5 Conclusions and Future Work

In this paper, we describe a technique for detecting changes in aerial images over a period of time. Usual characteristics of a change detection ground truth data set are that the number of pixels in changed category is significantly less than that in the unchanged category. Due to this fact, a classifier trained on the available ground truth data gets highly biased toward accurately classifying the majority class. We handle this problem by applying a suitable data adjustment technique on the imbalanced ground truth data to make it balanced. Experimental results show that the Broad Learning system works quite effectively in terms of the F-score metric. Moreover, the results obtained from Random undersampling and SMOTE data adjustment technique is usually better than those obtained from a combination of random undersampling and oversampling technique for any classifier. In future, this technique can be also extended to change detection in videos that can be an efficient tool in surveillance. Good performance on the chosen data set also motivates us to test our approach on real satellite images. Broad Learning-based change detection technique can also be employed in summarizing aerial videos.

Acknowledgements The authors would like to thank Dr. Lipo Wang of Nanyang Technological University and Dr. C. L. Philip Chen of University of Macau for introducing us to the concept of Broad Learning and helping us with the initial implementation stage. Authors also sincerely acknowledge Dr. Prasenjit Banerjee, Nalanda Foundation Mentor for his support. Authors also share their sincere gratitude to HP Innovation Incubator program for giving them the opportunity toward initiation of this project.

References

1. C.L. Philip Chen, Z. Liu, Broad learning system: an effective and efficient incremental learning system without the need for deep architecture. *IEEE Trans. Neural Netw. Learn. Syst.* (2017)
2. P. Rosin, Thresholding for change detection, in *Sixth International Conference on Computer Vision, 1998* (IEEE, New York, 1998), pp. 274–279
3. P.L. Rosin, E. Ioannidis, Evaluation of global image thresholding for change detection. *Pattern Recognit. Lett.* **24**(14), 2345–2356 (2003)

4. L.D. Stefano, S. Mattoccia, M. Mola, A change-detection algorithm based on structure and colour, in *IEEE Conference on Advanced Video and Signal Based Surveillance, 2003. Proceedings* (IEEE, New York, 2003), pp. 252–259
5. L. Bruzzone, D. Fernandez Prieto, An adaptive semiparametric and context-based approach to unsupervised change detection in multitemporal remote-sensing images. *IEEE Trans. Image Process.* **11**(4), 452–466 (2002)
6. L. Bruzzone, D.F. Prieto, Automatic analysis of the difference image for unsupervised change detection. *IEEE Trans. Geosci. Remote Sens.* **38**(3), 1171–1182 (2000)
7. W.A. Malila, Change vector analysis: an approach for detecting forest changes with landsat, in *LARS symposia* (1980), p. 385
8. A. Singh, Review article digital change detection techniques using remotely-sensed data. *Int. J. Remote Sens.* **10**(6), 989–1003 (1989)
9. H. Vincent Poor, *An Introduction to Signal Detection and Estimation* (Springer Science & Business Media, Berlin, 2013)
10. S.M. Kay, *Fundamentals of Statistical Signal Processing, ser. Signal Processing Series* (Englewood Cliffs, NJ: Prentice-Hall, 1993)
11. M.J. Black, D.J. Fleet, Y. Yacoob, Robustly estimating changes in image appearance. *Comput. Vis. Image Underst.* **78**(1), 8–31 (2000)
12. Y.G. Leclerc, Q.-T. Luong, P. Fua, Self-consistency and MDL: a paradigm for evaluating point-correspondence algorithms, and its application to detecting changes in surface elevation. *Int. J. Comput. Vis.* **51**(1), 63–83 (2003)
13. J. Rissanen, *Minimum Description Length Principle* (Encyclopedia of statistical sciences, 1985)
14. Y.Z. Hsu, H.-H. Nagel, G. Rekers, New likelihood test methods for change detection in image sequences. *Comput. Vis. Graph. Image Process.* **26**(1), 73–106 (1984)
15. Y. Yakimovsky, Boundary and object detection in real world images. *J. ACM (JACM)* **23**(4), 599–618 (1976)
16. A.S. Elfishawy, S.B. Kesler, A.S. Abutaleb, Adaptive algorithms for change detection in image sequence. *Signal Process.* **23**(2), 179–191 (1991)
17. Z.-S. Jain, Y.A. Chau, Optimum multisensor data fusion for image change detection. *IEEE Trans. Syst. Man Cybern.* **25**(9), 1340–1347 (1995)
18. K. Toyama, J. Krumm, B. Brumitt, B. Meyers, Wallflower: principles and practice of background maintenance, in *The Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999*. vol. 1 (IEEE, New York, 1999), pp. 255–261
19. M.J. Carlotto, Detection and analysis of change in remotely sensed imagery with application to wide area surveillance. *IEEE Trans. Image Process.* **6**(1), 189–202 (1997)
20. C. Clifton, Change detection in overhead imagery using neural networks. *Appl. Intell.* **18**(2), 215–234 (2003)
21. K. Skifstad, R. Jain, Illumination independent change detection for real world image sequences. *Comput. Vis. Graph. Image Process.* **46**(3), 387–399 (1989)
22. L. Li, M.K.H. Leung, Integrating intensity and texture differences for robust change detection. *IEEE Trans. Image Process.* **11**(2), 105–112 (2002)
23. S.-C. Liu, C.-W. Fu, S. Chang, Statistical change detection with moments under time-varying illumination. *IEEE Trans. Image Process.* **7**(9), 1258–1268 (1998)
24. S. Patra, S. Ghosh, A. Ghosh, Change detection of remote sensing images with semi-supervised multilayer perceptron. *Fundam. Inform.* **84**(3, 4), 429–442 (2008)
25. A. Singh, K.K. Singh, Unsupervised change detection in remote sensing images using fusion of spectral and statistical indices. *Egypt. J. Remote. Sens. Space Sci.* (2018)

User-Item-Based Hybrid Recommendation System by Employing Mahout Framework



Sutanu Paul and Dipankar Das

Abstract Recommendation systems are gaining popularity nowadays. Generally, recommendations are used in multiple areas like music, movies online products, news articles, texts, study articles, search engine queries, and social networking tags. Recommender systems use machine learning techniques and data mining algorithms to predict what items should suggest to the users based on some previous information related to the users and their relations to the items. It basically offers the users a limited number of products and services which he/she would like to get among the vast amount of all available items. The growth of information on the Internet, as well as the number of visitors to Web sites, is producing many choices for a customer, but many of those are irrelevant to them. A recommendation system filters this data and refers to the filter's data to the users. The recommendation system is totally based on its training data, the performing algorithms and the recommending approach. There are many popular approaches like user-based collaborative filtering, item-based collaborative filtering, content-based filtering, and hybrid models. We have implemented three different architectures, item-based, user-based, and factor-based hybrid models in order to build recommender system for an artist. In order to cope up with the huge amount of data, we have used Apache Mahout on top of Hadoop. Mahout is basically a framework and an open source project of Apache. Generally, it is being used for creating scalable machine learning algorithms. Using Mahout, we can fasten the process. Mahout is ready-to-use framework, helps programmers for doing data mining tasks on a large scale of data.

Keywords Recommendation system · Hybrid approach · Machine learning · Apache Mahout · Hadoop

S. Paul (✉)
Computer Science & Engineering, Jadavpur University,
Kolkata, India
e-mail: sutanupaul27@gmail.com

D. Das
Faculty of Computer Science & Engineering, Jadavpur University, Kolkata, India
e-mail: dipankar.dipnil2005@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_32

1 Introduction

The present attempt to focus on developing recommendation systems mostly used in the digital domain. Majority of today's e-commerce sites like eBay, Amazon, Alibaba, etc. make use of their proprietary recommendation algorithms in order to better serve the customers with the products they are supposed to like. From a business perspective, recommendation systems are typically used to enhance the business and increase the sale by helping users to discover items which are not have been reviewed by them yet and promote sales to potential customers based on their previous selections over the other similar data. If a recommendation system is able to promote the most relevant products to the customer, then we can call it is a good recommendation system. That's why recommendation systems become the smart approach for advertisements. So we are trying to study here how to build effective recommendation systems which able to predict items or products that customers like the most and interested to acquire. We will try to solve this problem and analyzing some preexisting models and algorithms like user-based collaborative filtering, item-based collaborative filtering, and hybrid recommendation algorithm; we try to build a completely new model which is just combination of other algorithms and some other approaches. The new system can also be able to predict the rating for a product which is completely unknown to the customer, by analyzing the historical data of all other users and all of their preference values or ratings. We are going to implement these algorithms and then experiment them on Last.fm-360k-datasets to do comparisons and produce results. This dataset has 17,000,000+ ratings and 360,000+ users. Our goal is to build a recommendation system which recommends by analyzing the historical data of the user's previous activity, item similarity, and user profile. We use a machine learning framework of Apache called Mahout coupled with Hadoop to analyze the huge amount of data. There is another tsv dataset which contains the user profile of all the users like their sex, age, country, and their registration date. Such a dataset has been further used in the hybrid model to improve the performance of the system.

2 Related Work

The content-based filtering approach and the collaborative filtering approach are two traditional and main approaches to build a recommendation system have been used for years. These are the popular algorithms. For the content-based CF approach, the item profile and user profile data based on their attributes must be stored at the beginning. Then the system tries to identify the similar items and recommends only those similar items which are highly relevant to the similar user profiles. Examples of such systems are NewsWeeder [1], Infofinder [2], and News Dude [3].

On the other hand, the collaborative filtering has two subtypes one is user-based CF other is item-based CF; these algorithms cluster the data according to the user-

based or item-based approach and group them as based on similarity measurement and recommend them accordingly. This type of recommendation system is mentioned by surendra [4] and Hintone [5]. Now, the collaborative approach works in completely different. It does not measure the similarities between the item or user profiles; it measures the similarity by using the data of the past experience of the user-item relationship. As described above, the purpose of the collaborative approach is to recommend among the users with similar choices or similar taste. The recommendation system with collaborative method finds out the taste of the user by finding the user interests which are determined from the users' past access histories.

Some Web sites even track the item Web pages searched or viewed by the user and those data taken account to derive preference value of the content of the page. There comes a backdrop of the content-based method. Suppose users might like a music from a group but he never ever reviewed any music from that group yet. So, he will never get a recommendation for that music. So, the collaborative method always to provide anticipated and entertaining products to the customers but the CF method is to provide surprising findings because it shares the information among similar users. To refer to the knowledge data from the separate users, we find the users with the same interest first. These are the technique proposed in [6] for user grouping. Examples of such systems are Ringo [7] and SiteSeer [8]. So because of the information sharing between the users, this collaborative filtering approach is highly probable to recommend the product which is unexpected.

Content-based filtering methods. For example, Tapestry [9] and GroupLens [10] use the users comment on Netnews then it estimating the similarities based on their ratings and reviews in newsgroups and based on those similarities it groups them. In the beginning, the user must specify their interests by defining their profile and describing the products and its features for which they are interested in. From the outlines of past accessed items, it determines the user interests. The users must define at the beginning satisfactory degrees of the accessed items which are being used to determine the user interest. Here grouping should be done based on those users who define similar satisfactory degree with respect to the products.

Now, the personalized television system [11] presents a personalized catalog of recommended programs. The FAB system [12] analyzes the accessed Web pages and based on that it determines the user profiles. In a collaborative approach, it groups the users by comparing their profiles.

Now, Mahout is an open source machine learning framework from Apache. Using Mahout, we can combine the traditional algorithms. Mahout mainly performs machine learning methods like recommendation engines, clustering, and classification. Mahout's algorithms are written on java and it coupled with Hadoop which helps it to deal with the large scale of datasets [13]. Basically, we can say it combines big data with machine learning.

3 Dataset Preparation

The recommendation Web sites store the data generated per day. It stores the details of the user like their age, sex, country, etc. and along with the data related to user's relationship with content. The Web sites usually store these data which is supposed to be used for recommendation.

3.1 Data Source

Last.fm, founded in the UK in 2002 is a Web site for songs. It is already using a music recommender system ("Audioscrobbler") to recommend music to the users. Audioscrobbler is invented by Richard Jones. The Web site develops a comprehensive profile of each user's musical interest by recording features of the tracks the user already reviewed, either from online radio stations or the user's computer or other transportable music engines. This data is transported ("scrobbled") through the music player itself (including, among others "Spotify", "Deezer", "Tidal", and "MusicBee") or via a plug-in connected into the music player to the database of the Web site. This data is used to build reference pages for different artists as per wiki description. I have chosen the last.fm-360k-dataset. I have downloaded the dataset contained in a tar file from Last.fm Dataset—UPF.

The zip file contains three files "usersha1-artmbid-artname-plays.tsv", "usersha1-profile.tsv", and "mbox_sha1sum.py". File "usersha1-artmbid-artname-plays.tsv" is having total 17,559,530 lines with 359,347 unique users among them 186,642 artists are having MBID and 107,373 does not have MBID. The data in "artmbid-artname-plays.tsv" is formatted in tab separated format where per line contains "user-mboxsha1" (user id), "musicbrainz-artist-id" (artist id), artist-name, number of plays and "usersha1-profile.tsv" contains "user-mboxsha1", "gender" (mlfempty), "age" (intempty), "country" (strempty), "signup" (dateempty) per line with the same format. We don't need "mbox_sha1sum.py" file.

As we know from the above-mentioned information that "usersha1-artmbid-artname-plays.tsv" is the main file which contains the main file which has the data based on previous user-item relation, i.e., users past experiences. Another file contains is the dataset based on the user's profile. Our ultimate aim is to build a hybrid recommendation system which performs better than the traditional systems. CF recommender does not require the user profiles. It only needs the user-item relationships to compute the recommendation algorithm.

3.2 Preprocessing of Data

Now usersha1-artmbid-artname-plays.tsv dataset contains four tab separated columns as mentioned above:

```
user-mboxsha1 \t musicbrainz-artist-id \t artist-name \t plays
```

We preprocess the data in the following steps.

1. The “usersha1-artmbid-artname-plays.tsv” dataset contains user id, artist id, artist name, or plays. There may be some incomplete information like any of user id, artist id, artist name, or plays is missing. We simply discard that rating information and consider the remaining. So, if all the four attributes are not available then we will simply discard that rating by applying a filter condition over the dataset we do that.
2. A preference point is a number which gives a value to the item for a user that how much he/she likes it. Since we have any preference value but the number of plays of an artist is available. For a user, if a number of hits or plays for a song are higher than that of the other song then its preference point will be high. So, we can define a preference value which is based on a user’s number of plays but every user’s liking should be rate in a same range of interval. Because some user listens to music frequently and some user listens to music often. Let us take an example:- a user listens to music frequently but listens to a particular song a certain number of times and some other user listens to that song the same number of times but this user is less interested in music so he is usually less frequent to the Web site. But does not mean that the first user likes the song same the second user. Because second users’ listening percentage to that song is more than that of the first user so, we can assume that the second user likes the song more than the first singer. So, we can solve this problem by normalizing the number of plays and derive a preference value out of it which ranges in an interval say 1 to 5 per user which will be based on the listening percentage. That is, the artist he/she listens most is the item with the highest rating say 5 and the artist he/she listens least is the item with the lowest rating. That is, if the number of plays is high then that item is a highly preferable item for the user and if it is less listened artist then it is less preferable to the user. Let a and b are the number of highest and lowest number plays, respectively, of a user among his/her all past listened to artists. Now suppose x is the number of plays for any random artist then the rating generating function for that artist will be:

$$\text{Rate}(x) = 1 + \frac{x - a}{b - a} * 4$$

This value ranges from 1 to 5.

3. The user-mboxsha1 is a unique identifier of users, is a 40 character string, “musicbrainz-artist-id” is a unique identifier for artists, is a 36 character string, artist-name is a character string and the number of plays is an integer which

counts the number of times the artist is hit by the user. But this dataset should be a model in a meaningful format which can be identified by the Mahout framework. As per Mahout framework, the data model for input data should be a CSV file whose first attribute should be user id and data type must be long and the second attribute is item-id and this is also a long data type and third data type is a double data type which is called the preference id and all should be formatted in comma separated format like example-1 dataset. So, we need to process the data in such a way that 17,000,000 ratings converted to the CSV format. To do that, I have used a hash map and which maps the string to a long data type and compute the preference value from the number of plays and keep the mapping until the recommendation is done. I.e., unique integers for per “user-mboxsha1” (user id) and “musicbrainz-artist-id” (artist id) will be generated and a new dataset will be formed with those new integer ids. With the new dataset recommendation will be done and end of the recommendation, in the result set it maps back the String id’s or artist name by using the hash map. We built a StringToInt class to handle this mapping.

The dataset is huge and tough to handle it we use Hadoop’s Map Reduce technique. I have used a Hadoop’s single node cluster if we can increase the number of node in future then we can test the recommending time is being reduced or not and if reduce then by how much. This mapper takes the first attribute, i.e., the user id as the key and the remaining as value and at reducer phase, it computes the necessary values for each key.

After all the preprocessing part, the dataset transformed into a new dataset in Hadoop’s output file named as “part-00000”. We store that file in a separate space and the recommendation will be done on this file. It’s a CSV file with three attributes and comma separated. This is the specific file format for Mahout’s recommendation system.

4 System Design

Let us understand a user-item rating matrix in order to understand CF Recommender systems. It is a matrix with n rows and m columns where n is number of users rated m is number of items. Here, the (i, x) cell contains the ratings given to item “ i ” by the user “ x ”, i.e., each row represents the rating vector corresponding to a user. This user rating matrix is typically sparse, as most users do not rate most of the items. Thus, the goal of the present system is to recommend those unrated cells, i.e., those items which are not reviewed by the user. The user under current consideration is referred to as the active user. The three models are given below:

1. User-based collaborative filtering;
2. Item-based collaborative filtering;
3. Hybrid approaches: These methods combine both collaborative-based approaches along with the factor models based on user profiles.

4.1 Mahout Framework

Apache Mahout is a distributed linear algebra framework implemented on Java and coupled with Hadoop to boost up machine learning algorithms. Generally works on the areas of machine learning like collaborative filtering, clustering, and classification. Mahout also gives us various Java libraries for linear algebra and statistics together with primitive Java collections to perform machine learning operations. Since Mahout is an in-progress project, the number of algorithms has been developed quickly, but several algorithms are still not available. Still, it is a powerful tool in the data science research field to deal with machine learning. Mahout is providing in-built ready to use basic machine learning algorithms.

The weakness of Mahout includes poor visualization and less support for scientific libraries compared to Spark for Python and what Python has inherited from R. Mahout has a lot of dependencies which can be a drag if one is simply using it for Spark jobs.

4.2 User-Based Recommendation

The user-based recommendation is a type of collaborative filtering where user-based algorithm generates a recommendation list for the user corresponding to the review of other users with similar interest. The underlying concept is, if the ratings given by some users to some items are similar, then rating by these users to different items will be similar too. Refer to the previous example-1, we make this table (Fig. 1).

Now let us look at the ratings among user-1, user-2, and user-3. Since for most of the cases their ratings are similar, we can put them in a same group. If we want to refer some item to user-2, then it must be item-03 and item-04 because of high rating from user-1 and user-3.

There are some metric functions in Mahout which measures the similarity measures between the users and based on that value it group the users. Based on a threshold value, it will allow other user to be in same neighborhood. Collaborative filtering algorithm that uses a threshold value to determine the neighborhood is discussed here. The algorithm can be summarized in the following steps:

Fig. 1 User-based approach

	ItemID (ArtistID)						
UserID	00	01	02	03	04	05	06
1	1.0	2.0	5.0	5.0	5.0	NA	NA
2	1.0	2.0	5.0	NA	NA	5.0	4.5
3	NA	2.5	5.0	4.0	3.0	NA	NA
4	5.0	5.0	5.0	0.0	NA	NA	NA

- Find the similarity value between the users. This similarity metric is Pearson correlation similarity according to our problem.

$$S(x, y) = \frac{\sum_{i=1}^n (r(x_i) - \overline{r(x)})(r(y_i) - \overline{r(y)})}{\sqrt{\sum_{i=1}^n (r(x_i) - \overline{r(x)})^2 \sum_{i=1}^n (r(y_i) - \overline{r(y)})^2}}$$

- Select n active users who have the top similarity measure between them up to a threshold value.
- Find the items which are not reviewed by the user x yet and reviewed by most of other users among n users.
- Compute a predicted rating, from the top similarity. Let x is the current user, y is any of the similar users, and i is the item then formula for prediction for user x over item i will be:

$$P(x, i) = \overline{r(x)} + \frac{\sum_{y=1}^n (r(y_i) - \overline{r(x)}) * S(x, y)}{\sum_{y=1}^n S(x, y)}$$

where $\overline{r(x)}$ and $\overline{r(y)}$ is the mean rating of user x and user y respectively.

$r(x_i)$ and $r(y_i)$ are rating of user y and user x over item i respectively.

$S(x, y)$ is the similarity between user x and user y .

Since Mahout has in-built ready to use functions and classes for recommendation. So, we don't need to work hard for coding part. The recommendation system is implemented in Java and computes 20 recommendations for particular user.

4.3 Item-Based Recommendation

The item-based recommendation is another type of collaborative filtering. The idea is the ratings by some users to some items are similar. The rating by the other users to those items will also be similar, i.e., if some user rates some items with similar rating then these items will be a similar kind of item. We again refer to the previous example-1 and make the user-item rating matrix table to understand the concept (Fig. 2).

Now, let us look at the ratings among item-02, item-03, and item-04. Since for most of the cases their ratings are similar, we can put them in a same group. If we want to refer to some item user-2 then it must be item-03 and item-04 because they are in a same group with item-02.

The metric functions in Mahout are used to measure the similarity measures between the items and based on the rating by the user. Collaborative filtering does not need a threshold value for neighborhood. The algorithm can be summarized in the following steps:

Fig. 2 Item-based approach

UserID	ItemID (ArtistID)						
	00	01	02	03	04	05	06
1	1.0	2.0	5.0	5.0	5.0	NA	NA
2	1.0	2.0	5.0	NA	NA	5.0	4.5
3	NA	2.5	5.0	4.0	3.0	NA	NA
4	5.0	5.0	5.0	0.0	NA	NA	NA

- Find the similarity value between the items. This similarity metric is Pearson correlation similarity according to our problem.

$$S(i, j) = \frac{\sum_{x=1}^n (r(x_i) - \overline{r(i)})(r(x_j) - \overline{r(j)})}{\sqrt{\sum_{x=1}^n (r(x_i) - \overline{r(i)})^2 \sum_{x=1}^n (r(x_j) - \overline{r(j)})^2}}$$

- Select active users that have the highest similarity.
- Find the items which are not reviewed by the user x yet and reviewed by most of other users among n users.
- Compute a prediction, from the top similarity. Let x is the user and i is the item then formula for prediction for user x over item i will be:

$$P(x, i) = \overline{r(i)} + \frac{\sum_{j=1}^n (r(x_j) - \overline{r(j)}) * S(i, j)}{\sum_{y=1}^n S(i, j)}$$

Where $\overline{r(i)}$ and $\overline{r(j)}$ is the mean rating of item i and item j respectively.

$r(x_i)$ and $r(y_i)$ is rating of user y and user x over item i respectively.
 $S(i, j)$ is the similarity between item i and item j .

Now, we can proceed to our main goal, the hybrid model. This model combines both the recommendations above and user profile to recommend better.

4.4 Hybrid Recommendation

This hybrid recommendation system is a combination of the recommendation systems mentioned above and also it associates the measurement of the user profile data. There are mainly seven kinds of hybrid recommendation systems: weighted, switching, mixed, feature combination, feature augmentation, cascade, and meta-level.

This recommendation system is a type of mix-weighted hybrid recommendation together with the factor models based on user profile data. We combine some the

recommendation from user-based CF and some from item-based CF and combine them and then we put them in a single list. Then divide their preference value into some parts.

- P_1 : 70% preference value based on their collaborative algorithms.
- P_2 : 10% preference value based on their sex-based model.
- P_3 : 10% preference value will be based on their age.
- P_4 : 10% preference value will be based on their location.

Now after data cleaning, all the parts are same but after that we combine all the approaches on the preprocessed dataset like previous examples we determine the similarity between two users or items based on that similarity.

4.4.1 Sex-Based Factor Model

To optimize our recommendation system we use user profile data; the underlying concept being, if an artist is famous among a particular gender then it is obvious that the artist is highly preferable to that gender. For example, Justin Bieber is famous among girls or ladies then he is highly preferable to a female user.

Let, \mathbf{i} is an artist and \mathbf{u}_i are the users who have already reviewed \mathbf{i} . Let \mathbf{x} is an active user then according to \mathbf{x} , the preference value of the artist \mathbf{i} will say $P_2(\mathbf{x}, \mathbf{i})$. Let total male count is \mathbf{M} and total female count is \mathbf{F} . and the song is preferred by \mathbf{f} female user and \mathbf{m} male users.

Then female percentage is $\frac{f*100}{F}$ and male percentage is $\frac{m*100}{M}$ who performed the artist.

Case 1: If the user is female, then her preference will be high if $\frac{\frac{f*100}{F}}{\frac{m*100}{M}} = \frac{f*M}{m*F} > 1$ and

Preference will be low if $\frac{f*M}{m*F} < 1$

We call it advantage ratio over sex, $r = \frac{f*M}{m*F}$.

Case 2: Similarly, for a male user it will be $r = \frac{m*F}{f*M}$.

Then for our recommendation system, the preference value based on sex will be determined by the following function.

$$P_2(\mathbf{x}, \mathbf{i}) = \frac{r^2}{1 + r^2}$$

This value lies between 0 and 1 and for a neutral case, it will give 0.5.

4.4.2 Country-Based Factor Model

In our recommendation system, we use user's location data. Here the concept is if a artist is famous in some country then it is much likely that an user belonging to that

country would prefer that artist. For example, being famous in India, Arijit Singh is most likely to be preferred by an Indian user.

Let, \mathbf{i} is an artist and \mathbf{u}_i are the users who have already reviewed \mathbf{u}_i . Let \mathbf{x} is an active user then according to \mathbf{x} the preference value of the artist \mathbf{i} will say $P_3(\mathbf{x}, \mathbf{i})$. Let count of users from all countries is \mathbf{T} and count of users from user \mathbf{x} 's country among all users is \mathbf{C} . Count of users among \mathbf{u}_i is t and count of users among \mathbf{u}_i is c .

Then percentage users from the country of the user x among all users is $\frac{C*100}{T}$ and Percentage users from the country of the user x among \mathbf{u}_a is $\frac{c*100}{t}$. Then,

Her preference will be high if $\frac{\frac{c*100}{t}}{\frac{C*100}{T}} = \frac{c*T}{t*C} > 1$ and

Preference will be low if $\frac{c*T}{t*C} < 1$

We call it advantage ratio over country, $r = \frac{c*T}{t*C}$

Then for our recommendation system the preference value based on country will be determined by the following function.

$$P_2(\mathbf{x}, \mathbf{i}) = \frac{r^2}{1 + r^2}$$

This value lies between 0 and 1 and for a neutral case it will give 0.5.

4.4.3 Age-Based Factor Model

The concept behind incorporating a user's age in our recommendation system is, if an artist is famous among a certain age group then that artist is highly preferable to a user from that age group. So, if the variance of the difference between the ages of the users who prefer that item and the current user is smaller, then the artist is highly preferred by the person.

Let, \mathbf{i} is an artist and u_i are the users who have already reviewed and let their age is l_u . Let \mathbf{x} is an active user then according to \mathbf{x} the preference value of the artist \mathbf{i} will say $P_4(x, a)$. let \mathbf{x} 's age is l ,

$$D = \sqrt{\frac{1}{n-1} \sum_{u=0}^n (l_u - l)^2}$$

$\bar{l} = \frac{\sum_{u=1}^n l_u}{n}$, \bar{l} is the mean

Then, the coefficient of variant say, $v = \frac{D}{\bar{l}}$.

v lies between 0 and 1 and it is minimum when variation of ages is less, i.e., for that case it's a highly preferable item. On the other hand, if v is near to 1 then this is less preferable item. Then for our recommendation system, the preference value based on country will be determined by the following function:

$$P_3(\mathbf{x}, \mathbf{i}) = 1 - v.$$

This value also lies between 0 and 1.

Lets summarized the main algorithm in the following steps which combines all the functions together:

- Find the similarity value between the users and similarity between items using the similarity metric mentioned above

$$S(i, j) = \frac{\sum_{x=1}^n (r(x_i) - \overline{r(i)})(r(x_j) - \overline{r(j)})}{\sqrt{\sum_{x=1}^n (r(x_i) - \overline{r(i)})^2 \sum_{x=1}^n (r(x_j) - \overline{r(j)})^2}}$$

and

$$S(i, j) = \frac{\sum_{x=1}^n (r(x_i) - \overline{r(i)})(r(x_j) - \overline{r(j)})}{\sqrt{\sum_{x=1}^n (r(x_i) - \overline{r(i)})^2 \sum_{x=1}^n (r(x_j) - \overline{r(j)})^2}}$$

- Using the traditional recommenders (user-based CB and user-based CF) compute the recommend item and merge them.
- Compute a predicted rating using Eqs.

$$P(x, i) = \overline{r(x)} + \frac{\sum_{y=1}^n (r(y_i) - \overline{r(x)}) * S(x, y)}{\sum_{y=1}^n S(x, y)}$$

and

$$P(x, i) = \overline{r(i)} + \frac{\sum_{j=1}^n (r(x_j) - \overline{r(j)}) * S(i, j)}{\sum_{j=1}^n S(i, j)},$$

respectively, say it is R. Preference will depend on this rating then P1 will be 70% of R.

i.e., $P_1 = R * 0.7$

- Using the sex-based factor model, find the preference value based on user's sex say it is P_2 .
- Using the country-based factor model, find the preference value based on user's country say it's P_3 .
- Using the age-based factor model, find the preference value based on user's age say it's P_4 .
- Now calculate the new predicted rating of user **X** over item **i**.

$$R(x, a) = W_1 * P_1 + W_2 * P_2 + W_3 * P_3 + W_4 * P_4$$

Where $W_1, W_2, W_3,$ and W_4 are the weightage.

$W_1 = 3.5, W_2 = 0.5, W_3 = 0.5$ and $W_4 = 0.5$

$$0 \leq P_1, P_2, P_3, P_4 \leq 1$$

$$\text{So } 0 \leq R(\mathbf{x}, \mathbf{a}) \leq 5.$$

- Now calculate the new predicted rating of user \mathbf{x} over item \mathbf{i} . Sort it based on this predicted rating and recommend them best among them.

We have also evaluated the recommendation system above. For that, we have created another evaluator class. Mahout provides us these various features to create, implement, and test the results so, it reduces the effort and complex algorithms to implement. Because all in there in Mahout.

4.5 Evaluator Model

Mahout provides us various evaluators which we can employ upon our recommendation systems and verify the results. We use IRStatisticsEvaluator to evaluate our recommendation system it takes a small percentage for training and recommend based on that and calculates that value is how much correct. To evaluate that, we are going to measure the precision, recall, and f1 score.

5 Evaluation

The experiments regarding the recommendation systems, their results, and analysis on their result set are discussed in this section.

5.1 Experiment and Results

We have run our three recommendation systems over the 17,000,000 ratings and we observe the different recommendations recommended by these three recommendation engines for a particular user. The output consists of 20 artists and their predicted ratings provided by the different recommendation systems.

The user id is: 00000c289a1829a808ac09c00daf10bc3c4e223b.

Here, we can see that the rating of item-based system and user-based system is comparatively higher than the hybrid system. But this does not mean that hybrid system works poor. Because the rating is a predicted value and that depends on the rating function.

In order to evaluate the recommendation performance of our traditional systems and to assess the performance contribution of its hybrid model, we used the Mahout's evaluator function. It is not possible to calculate the same by Mahout for the hybrid system because there are too many post calculations. Since the hybrid model is the

combination of these two models so we can look the precision and recall value of the traditional systems.

Evaluation result for User-based system:

Precision: 0.08805031446540884

Recall: 0.08720930232558138

F1 Measure: 0.08762779052785313

Evaluation result for Item-based system:

Precision: 0.0

Recall: 0.0

Here, Precision and Recall value is very small because we have used a very small percentage of the dataset for training, i.e., 0.05%. If we could use more nodes of the Hadoop cluster, we could have used more training data and expect a good precision and recall value (Table 1).

5.2 Comparative Analysis

Look at the recommendations for user-based CF recommendation systems the rating ranges from 5.0 to 3.9466 and Item-based system's predicted rating is constant, 5.0. The ratings from hybrid system are ranges from 4.7121 to 4.441. Here the user-based recommendation rating ranges more than the hybrid model. So, we can say that the recommended items are more closely clustered for hybrid system. On the other hand, if we compare the item-based CF result and hybrid model then we can see that the item-based CF recommendation results do not have variation for rating it always rate the items the same it's a backdrop for this item-based system.

Also, the recommended items are different from each other. That's because the hybrid models not only combines the results of two systems it also combines the factor models which results to change of preference value.

5.3 Error Analysis

We have found some errors regarding our recommender systems. Firstly, the time taken by the system to recommend is very high. The user-based approach takes around 10–15 min to recommend an user; the item-based recommender takes around 5–10 min to recommend an user and the hybrid system takes 15–20 min to recommend an user. Reason of this problem is maybe we are using a huge dataset to train the engine. The dataset size is 1.7 GB and contains 17,000,000+ rating. In the future, it can be resolved by utilizing Hadoop with multinode cluster.

Also, the predicted rating computed by item-based recommendation is 5.0. The reason is not clear but it might be Mahout's framework related problem. We need to explore Mahout more to resolve this matter.

Table 1 Result sets

User-based recommender		Item-based recommender		Hybrid recommender	
Artist	Ratings	Artist	Ratings	Artist	Ratings
Messiah J & the Expert	5.0	The Stone Roses	5.0	Richard Elliot	4.7121
Ion Storm	4.8691	Deep Insight	5.0	Jazzmasters	4.6394
Steinar Albrigtsen	4.6045	Der Plan	5.0	Cartola	4.5940
Joaquin Phoenix & Reese Witherspoon	4.5824	Bill Evans	5.0	Ana Cañas	4.5798
D. Batistatos	4.4402	Fefe Dobson	5.0	Gojira	4.5726
Dumonde aka De Leon and Jamx	4.4293	Silverstein	5.0	Acidman	4.5530
Femme Fatale	4.3571	R.E.M.	5.0	Louis Prima	4.5449
Kerbenok	4.2243	Clawfinger	5.0	Avril	4.5420
The Apostles	4.1774	The Go! Team	5.0	Am	4.5318
DAB (Digital Analog Band)	4.1538	Laura Branigan	5.0	Entombed	4.5313
Ross Parsley	4.1098	Marissa Nadler	5.0	The Adolescents	4.5003
David Rovics	4.1079	Why?	5.0	Nico	4.4919
Druhá Tráva	4.0957	Brujeria	5.0	Paulinho da Viola	4.4850
Shael Riley	4.0909	Lars Winnerbäck	5.0	Everette Harp	4.4809
Static Revenger	4.0565	Amanda Rogers	5.0	Arvo Part	4.4778
Schola Hungarica	4.0360	Dredg	5.0	Lee Ryan	4.4776
Tanzmuzik	4.0294	Pink Floyd	5.0	Dire Straits	4.4668
Yellow blackbird	3.9802	White Denim	5.0	Inmigrantes	4.4665
Tom Novy feat Abigail Bailey	3.9686	Atmosphere	5.0	Junior Senior	4.4508
Woods of Ypres	3.9466	Festival	5.0	The Paul Butterfield blues band	4.4410

The evaluator model built to evaluate the recommender systems is working properly for user-based system but for item-based problem it gave no value. Maybe it's also an error of the system. Apart from that, the evaluator models consume along time to evaluate the recommenders. Because of having too much post calculations, we are not able to evaluate the hybrid system.

6 Conclusion

We have built three recommender systems one follows user-based collaborative filtering approach, one follows item-based recommendation approach, and the last one follows hybrid model. The hybrid model combines the results from user-based and item-based approach and also associates some factor model based on user profile to improve the recommendation mechanism. These three factor models are based on user's age, sex, and country. We have recommended a user by all three recommendation system and also checked the results and compared the recommendations by all three systems.

In the future, we can improve the system by assigning more nodes to the system. Since the system takes a bit time to compute the results, more nodes to Hadoop can help out this situation. Also, we can add more recommendation models to the hybrid system which and improve the performance. Also we can search for some item profile-based data to implement a content-based model.

References

1. K. Lang, Newsweeder: learning to filter netnews, in *Machine Learning Proceedings 1995* (1995), pp. 331–339
2. B. Krulwich, C. Burkey, Learning user information interests through extraction of semantically significant phrases, in *Proceedings of the AAAI Spring Symposium on Machine Learning in Information Access*, vol. 25, No. 27 (1996 March), p. 110
3. D. Billsus, M.J. Pazzani, A hybrid user model for news story classification, in *UM99 User Modeling* (Springer, Vienna, 1999), pp. 99–108
4. M.S.P. Babu, B.R.S. Kumar, An implementation of the user-based collaborative filtering algorithm. *Int. J. Comput. Sci. Inf. Technol.* **2**(3), 1283–1286 (2011)
5. R. Salakhutdinov, A. Mnih, G. Hinton, Restricted Boltzmann machines for collaborative filtering, in *Proceedings of the 24th International Conference on Machine Learning* (ACM, New York, 2007 June), pp. 791–798
6. Y.H. Wu, Y.C. Chen, A.L. Chen, Enabling personalized recommendation on the web based on user interests and behaviors, in *Eleventh International Workshop on Research Issues in Data Engineering, 2001. Proceedings* (IEEE, New York, 2001), pp. 17–24
7. U. Shardanand, P. Maes, Social information filtering: algorithms for automating “word of mouth”, in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (ACM Press/Addison-Wesley Publishing Co, New York, 1995 May), pp. 210–217
8. J. Rucker, M.J. Polanco, Siteminer: personalized navigation for the Web. *Commun. ACM* **40**(3), 73–76 (1997)
9. D. Goldberg, D. Nichols, B.M. Oki, D. Terry, Using collaborative filtering to weave an information tapestry. *Commun. ACM* **35**(12), 61–70 (1992)
10. J.A. Konstan, B.N. Miller, D. Maltz, J.L. Herlocker, L.R. Gordon, J. Riedl, GroupLens: applying collaborative filtering to Usenet news. *Commun. ACM* **40**(3), 77–87 (1997)
11. P. Cotter, B. Smyth, A personalized television listing service. *Commun. ACM* **43**(8), 107–111 (2000)

12. M. Balabanović, Y. Shoham, Fab: content-based, collaborative recommendation. *Commun. ACM* **40**(3), 66–72 (1997)
13. S. Walunj, K. Sadafale, An online recommendation system for e-commerce based on apache Mahout framework, in *Proceedings of the 2013 Annual Conference on Computers and People Research* (ACM, New York, 2013), pp. 153–158

Gray Matter Segmentation and Delineation from Positron Emission Tomography (PET) Image



Abhishek Bal, Minakshi Banerjee, Punit Sharma and Mausumi Maitra

Abstract Gray matter segmentation and delineation from positron emission tomography (PET) are a very essential requirement in medical applications due to low spatial resolution, presence of noise, and larger variability in the shape as well as in texture. PET images provide the functional details of the brain which are very much essential to detect brain disorders. The diagnosis of dementia, particularly in the early stages, are very much helpful with PET image processing. In this present work, we implement three essential steps, which significantly help during dementia analysis. The first step is skull stripping from PET image based on K-means and connected component analysis, the second step is to segment gray matter from skull-stripped PET image using fuzzy possibilistic C-means (FPCM), and the last step is to delineate the gray matter based on cellular automata. The Indian PET dataset collected from Apollo Gleneagles Hospitals, Kolkata, India, has been used in our experiment as a standard dataset. Experimental results show that the proposed method has achieved promising results for automatic segmentation and delineation of gray matter on PET brain images.

Keywords Segmentation · Delineation · PET · FPCM · Cellular automata

1 Introduction

Segmentation process can provide the meaningful information about the image from different segmented regions that can be useful for different types of image analysis [1–5]. Image segmentation has two tasks that are recognition and delineation.

A. Bal (✉) · M. Banerjee
RCC Institute of Information Technology, Kolkata, India
e-mail: abhisheknew1991@gmail.com

P. Sharma
Apollo Gleneagles Hospital, Kolkata, India

M. Maitra
Government College of Engineering and Ceramic Technology, Kolkata, India

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_33

The recognition process is generally used to determine where the required object is and how to distinguish the required object from the others. Whereas the delineation process separates the detected object from the irrelevant portion of the image.

The brain is the origin of the human body that has a very complex structure. The inside brain regions can be affected by the several diseases that cause the abnormal changes in brain cells. One kind of disease is dementia. The leading cause of dementia is primary neurodegenerative disorders that damage neuronal structure and interconnectivity of brain tissues which results in memory loss and progressive impairment of higher cognitive functions.

Positron emission tomography (PET) technology is used in medical studies, and several research purposes involved in living the organs like the brain. PET [6, 7] focuses on the functional details (blood flow) inside the brain. Identification of brain volume from the PET images is the basic task of analyzing the functional images. It also allows identifying the sub-regions of the brain like thalamus, hypothalamus, posterior cingulate gyrus, and temporal lobe.

Computed tomography (CT) and magnetic resonance imaging (MRI) are mostly used in clinical and research purpose for analyzing different kinds of diseases. But, to analyze dementia, positron emission tomography (PET) plays vital roles than the CT or MRI. CT and MRI scans look at structural details in the body, whereas positron emission tomography (PET) [8] provides the functional details that are very much essential in the diagnosis of dementia at an early stage.

Generally, in the clinic, the recognition process is done manually by analyzing and observing the regions where the required object is residing, which is considered as a region of interests (ROIs). Then, the target of the clinicians is to separate the identified uptake regions from the other irrelevant parts of the image. The contrast between objects in PET image is decreased due to the presence of low spatial resolution and high smoothing, for that reasons boundaries of two successive objects are becoming unclear. There are also some other factors that can significantly affect the resolution quality of the image, such as if a patient moves their body during the scanning process, then it will affect the image quality. Due to several intrinsic and extrinsic factors in PET images, manual segmentation on PET is a very difficult task. The general procedure to extract and delineate the region of interests (ROIs) on the PET brain image is to extract the corresponding ROIs from the structural images like MRI [9] or CT images and then superimpose the extracted anatomical part on the PET images. Although this procedure will produce the most promising results, it will require an accurate registration between PET and MRI/CT images. Due to the fundamental differences between PET and MRI/CT images, image registration between two different types of images is done manually most of the time for better accuracy. More information about medical image registration is described in Ref. [10].

Computer-aided diagnosis (CAD) methods have been investigated for several years to deal with several types of problem. Automated computer-aided diagnosis (CAD) system should be designed carefully such that it can capable to handle noise, heterogeneous shape, and partial volume effects that can vary from patient to

patient. The target of the dementia analysis is to detect the metabolic deficits (hypo-metabolism) related to neurodegenerative changes that occur early stage of dementia by the statistical evolution of FDG-PET brain images. Basically, the metabolism of the human brain resides in the gray matter portion. The term metabolism refers to the chemical reactions that involved in maintaining the living cell of the human body, whereas hypo-metabolism refers the abnormal metabolism; in this case, the human body cannot able to generate required energy. Detection of dementia through automated computer-aided diagnosis (CAD) system is required to initially segment gray matter from brain PET image; then, the analysis is carried out on the extracted gray matter to identify hypo-metabolism areas.

Several methods have been proposed for dementia analysis from the past few years, but these all analysis are based on foreign PET dataset. It is clinically tested that the researches which are based on foreign PET datasets are not well performed or applicable on Indian PET datasets. This type of misclassification is occurring because the metabolic patterns of Indian PET datasets (patients) are different from the foreign PET datasets due to several factors. As per our knowledge, there is very less amount of researches is going on Indian patient PET datasets.

Our target is to analyze the Indian PET datasets for identifying the early stage of dementia. Toward our target, the present work implements three steps, which significantly help during dementia analysis. The first step is to remove the skull from PET image based on K-means and connected component analysis, the next step is to segment gray matter from skull-stripped PET image based on fuzzy possibilistic C-means with several important features, and the last step is to delineate the gray matter using cellular automata. In this present work, recognition and delineation [11] of gray matter have been taken place only on the PET image without any interaction with structural images like MRI or CT.

The paper is organized into six sections. Sections 2 and 3 describe the brief details of fuzzy possibilistic C-means and cellular automata, respectively. The proposed work and the algorithm are introduced in Sects. 4 and 5, respectively. Section 6 describes the performance analysis of the present work on PET dataset. Section 7 describes the conclusion and future scope.

2 Fuzzy Possibilistic C-Means

The problem of fuzzy C-means (FCM) and possibilistic C-means (PCM) such as noise sensitivity and coincident cluster is resolved by the combination of fuzzy membership [12] and possibilistic membership [13] called fuzzy possibilistic C-means (FPCM)

[14, 15]. The objective function and centroid value calculation function in FPCM are formulated using Eqs. 1 and 2, respectively.

$$J = \sum_{i=1}^c \sum_{j=1}^n \{a(\mu_{ij})^{m_1} + b(v_{ij})^{m_2}\} \|x_j - v_i\|^2 + \sum_{i=1}^c \eta_i \sum_{j=1}^n (1 - v_{ij})^{m_2} \quad (1)$$

$$v_i = \frac{1}{n_i} \sum_{j=1}^n \{a(\mu_{ij})^{m_1} + b(v_{ij})^{m_2}\} x_j \quad (2)$$

where n_i is represented as

$$n_i = \sum_{j=1}^n \{a(\mu_{ij})^{m_1} + b(v_{ij})^{m_2}\} \quad (3)$$

Two fuzzifier are represented by $m_1 (1 \leq m_1 < \infty)$ and $m_2 (1 \leq m_2 < \infty)$. The cluster centroid of the FPCM is significantly controlled by two constants a and b . The centroid value is more controlled by the fuzzy membership using the greater value of b over a , whereas possibilistic membership controls the centroid value using the greater value of a and b . The membership of x_j to cluster β_i for fuzzy and possibilistic clustering is represented as μ_{ij} and v_{ij} , respectively. The i th cluster (β_i) size is represented by the scale parameter η_i and formulated by Eq. 4.

$$\eta_i = k \frac{P}{Q} \quad (4)$$

$$P = \sum_{j=1}^n (v_{ij})^{m_2} \|x_j - v_i\|^2 \quad (5)$$

$$Q = \sum_{j=1}^n (v_{ij})^{m_2} \quad (6)$$

In FCM [16], membership (Eq. 7) value must maintain two conditions, which are formulated in Eqs. 8 and 9.

$$\mu_{ij} = \left(\sum_{k=1}^c \left(\frac{d_{ij}}{d_{kj}} \right)^{\frac{2}{m_1-1}} \right)^{-1} \quad (7)$$

$$\sum_{i=1}^c \mu_{ij} = 1, \forall j \quad (8)$$

$$0 < \sum_{j=1}^n \mu_{ij} \leq n, \forall i \quad (9)$$

whereas the membership (Eq. 10) in PCM must maintain the conditions which are formulated in Eqs. 11 and 12.

$$v_{ij} = \frac{1}{1 + E}, \quad \text{where } E = \left\{ \frac{\|x_j - v_i\|^2}{\eta_i} \right\}^{\frac{1}{m_2-1}} \tag{10}$$

$$0 < \sum_{j=1}^n v_{ij} \leq n, \forall i \tag{11}$$

$$\max_i(v_{ij}) > 0, \forall j \tag{12}$$

3 Cellular Automata

A bidimensional cellular automata (CA) [17, 18] are represented by $\mathcal{A} = (S, N, \delta)$. S represents a non-empty state set, N represents number of neighborhoods, and δ represents the local transition function, denoted by $\delta : S^N \rightarrow S$. The basic element of the cellular automata is cell or memory element. Each cell in cellular automata is represented either by 1 or 0. The state of a cell in cellular automata is determined by the rules where the rules are generated from the neighborhood of that cell. The three well-known neighborhoods are

3.1 Von Neumann Neighborhood

In this neighborhood pattern, the neighborhood of a cell is defined as the cells of the left, right, the top and bottom position of it. So, the total number of neighborhoods (N) including the cell itself is 5. The local transition function of the von Neumann neighborhood is represented as $\delta : S^5 \rightarrow S$. The pictorial representation of the von Neumann neighborhood is shown in Fig. 1a.

3.2 Moore Neighborhood

It is a extension of von Neumann neighborhood by including the diagonal cells. So, the total number of neighborhoods (N) including the cell itself is 9. The local transition function of the Moore neighborhood is represented as $\delta : S^9 \rightarrow S$. The pictorial representation of the Moore neighborhood is shown in Fig. 1b.

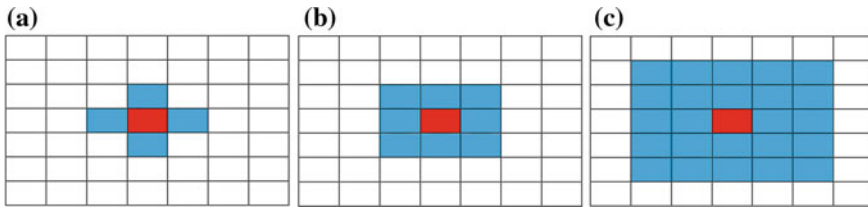


Fig. 1 Three different neighborhoods pattern. All blue pixels (cells) are the neighborhood of red pixel (cell). **a** von Neumann neighborhood, **b** Moore neighborhood, and **c** extended Moore neighborhood

3.3 Extended Moore Neighborhood

This neighborhood pattern is similar to Moore neighborhood except it includes the neighborhood cells over the distance of next adjacent cells. The total number of neighborhoods (N) including the cell itself is 25. The local transition function of the extended Moore neighborhood is represented as $\delta : S^{25} \rightarrow S$. The pictorial representation of the extended Moore neighborhood is shown in Fig. 1c.

4 Proposed Method

The pictorial representation of the proposed method is shown in Fig. 2, and the steps are:

- At first K-means along with connected component analysis (Sect. 4.1) is performed for removing the skull from PET image.
- Perform denoising on skull-stripped PET image and gray matter region (Sect. 4.2) of the brain tissues is extracted using FPCM.
- Delineate segmented gray matter using cellular automata (Sect. 4.3).

The details of the above steps are described in next sub-sections.

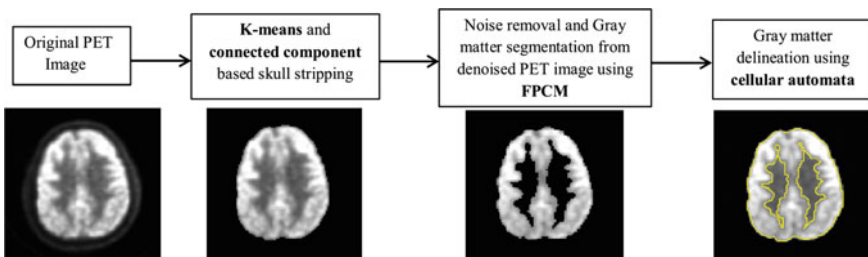


Fig. 2 Block diagram of the proposed model

4.1 *K-Means and Connected Component Analysis for Skull Stripping*

The human brain can be visualized or analyzed with three different planes, namely axial, sagittal, and coronal, which are shown in Fig. 3. In this present work, K-mean and connected component-based skull-stripping methods are implemented that can remove the skull from three different planes of PET brain image. The steps of skull stripping are:

- Initially, K-means clustering is applied to the PET image (Fig. 4a) with two clusters, and after clustering, brain tissues are initialized by 1 and the background along with non-brain tissues are initialized by 0 (Fig. 4b).
- On the outcome of the binarization, hole filling operation is performed within the four connected components (Fig. 4c). Chose the connected component with solidity >0.75 and roundness >0.6 and remove the other components. The equations of solidity (*S*) and roundness (*R*) are shown in Eqs. 13 and 14, respectively. Basically, small connected components denote some skull areas in sagittal and coronal planes. The extracted large connected component (Fig. 4d) denotes the brain mask, and brain tissues are extracted from the original PET image depending on the brain mask, which is shown in Fig. 4e. The delineation of brain tissues in the original PET brain is shown in Fig. 4f.

$$S = \frac{\text{Area}}{\text{Convex Area}} \tag{13}$$

$$R = \frac{4 \times \text{Area}}{\text{Perimeter}^2} \tag{14}$$

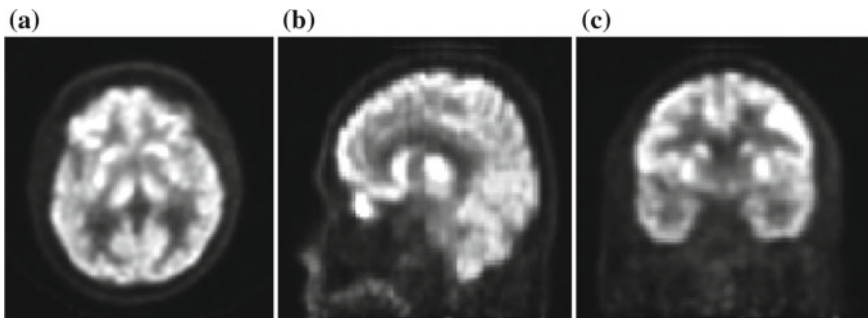


Fig. 3 Three different planes of a patient PET image. **a** Axial plane, **b** sagittal lane, and **c** coronal plane

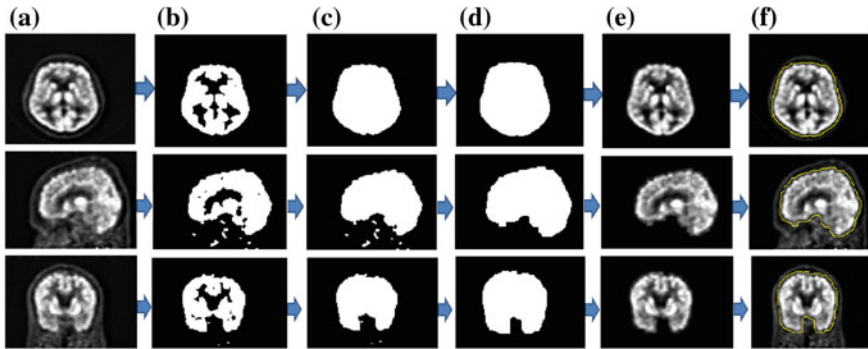


Fig. 4 The steps of the skull removal of PET image from three different planes of brain PET images. The first row shows the axial plane, the second row shows the sagittal plane, and the third row shows the coronal plane of the brain

4.2 Noise Removal and Gray Matter Segmentation

After the skull stripping, the present work applied BayesShrink [19] denoising method on skull-stripped PET image for noise removal. Then, the gray matter of the brain has been segmented using the FPCM (see Sect. 2). Due to the poor structure of the PET image, it is not possible to extract the CSF region from PET brain image. Due to this factor, the present work uses three clusters ($c = 3$) to partition the skull-stripped PET image into the background, gray matter along with CSF and white matter.

A set of sixteen (16) features is used in this FPCM for segmenting the gray matter from skull-stripped PET image. These features include the pixel gray value, ROAD factor, mean, standard deviation, the median of absolute deviation (MAD), median, min, max, probability, variance using probability, absolute energy, skewness, kurtosis, homogeneity, and edge value. The details about these features are described in related papers [20–23]. It is experimentally noticed that these features are very much useful for characterizing a pixel of the PET image. To find out the features of each pixel, we choose 3×3 neighborhood of that pixel. The highlighted feature-based FPCM achieved very promising results for segmenting the gray matter from the skull-stripped PET image, which is shown in Fig. 9.

4.3 Cellular Automata-Based Gray Matter Boundary Delineation

In this present method, we use the Moore neighborhood concept to find out the boundary of the gray matter of skull-stripped PET images. The steps of cellular automata-based gray matter boundary delineation are:

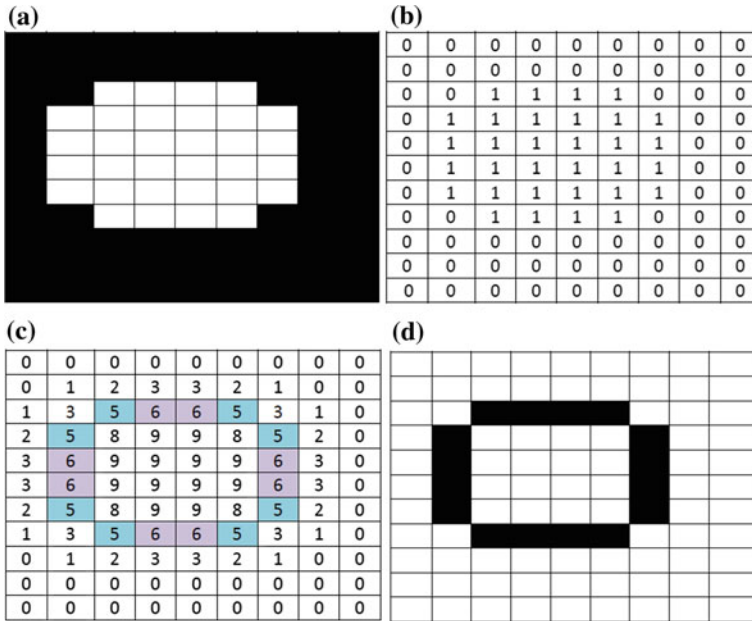


Fig. 5 Pictorial steps of boundary delineation using Moore neighborhood. **a** Binary image with object (white color), **b** object pixels represented by 1, **c** number of Moore neighborhood of each pixel, and **d** boundary delineation (black color) the object

- After segmentation of the gray matter from skull-stripped PET image, present work constructs a binary image, where gray matter regions are represented by 1 and remaining regions are represented by 0 (Fig. 5b).
- Then, Moore neighborhood (Sect. 3.2) concept is applied to find out the number of neighborhoods of each cell (pixel).
- According to the concept of Moore neighborhood, a cell (pixel) has minimum zero (0) neighborhood and maximum nine (9) neighborhood. So, the total number of neighborhoods is ten (10). Therefore, the total number of possible rules is 2^{10} (1024).
- After that, according to our problem domain, we select a suitable rule to find out the boundaries of the gray matter regions.

Let, we want to identify the object (white color) boundary in Fig. 5a. The object portions are initialized by 1, and the remaining portions are initialized by 0 (Fig. 5b). Then, we calculate the neighborhood of each pixel using the Moore neighborhood concept, and each pixel value is replaced by its number of neighborhoods. In this particular example, we select those pixels (active) which have five or six neighborhoods and consider the remaining pixel as a dead (inactive) pixel. The first row of Fig. 6 shows the neighborhood count, and the second row determines the corresponding selected rule. In this example, the suitable rule (DR) is 0000011000 that means 24.

NC	0	1	2	3	4	5	6	7	8	9
DR	0	0	0	0	0	1	1	0	0	0

Fig. 6 Neighborhood count (NC) and corresponding selected rule in binary (DR)

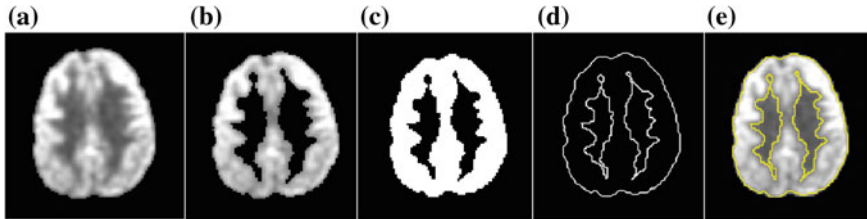


Fig. 7 Pictorial steps of gray matter boundary delineation of PET image. **a** Skull-stripped PET, **b** segmented gray matter, **c** gray matter mask, **d** boundary of segmented gray matter, and **e** gray matter boundary delineation (yellow color) on skull-stripped image

In our application, we have applied several rules to find out the most suitable one to delineate the boundary region of the gray matter. It is experimentally observed that the rule 56 (0000111000) is the most suitable one that can efficiently find out the boundary of the gray matter region of the PET image. The pictorial representation of the overall delineation steps is shown in Fig. 7.

5 Algorithm

The proposed algorithm is stated as follows:

Algorithm 1: Gray matter segmentation and delineation on PET image

Procedure: $GMD(I_{PET})$

INPUT:

- PET image, I_{PET}

OUTPUT:

- Segmented gray matter, S_{GM}

Steps:

- 1: Perform K-means and connected component analysis on original PET (I_{PET}) and extract brain tissues (I_{BT}).
 - 2: Apply BayesShrink-based noise removal method on skull-stripped PET image (I_{BT}) to estimate denoised image (I_{DBT}).
 - 3: Sixteen (16)-feature-based FPCM with $c = 3$ has been applied on the denoised brain tissues (I_{DBT}).
 - 4: Once FPCM is done, gray matter region (I_{GM}) is extracted from denoised brain tissues (I_{BT}) using the connected component analysis.
 - 5: Apply cellular automata-based delineation method to delineate the extracted gray matter and distinguish the gray matter from other brain regions.
-

6 Experimental Results

Standard clinical Indian PET dataset provided by Apollo Gleneagles Hospital, Kolkata, India, has been used in our proposed method. These clinical PET images are constructed by Ingenuity TF PET/CT system with Astonish third generation time-of-flight (TOF) technology manufactured by Philips Company. The dimension of each patient PET brain is $128 \times 128 \times 90$. All the images are in DICOM format with 16-bit gray level. More information about this PET scanning is available at <http://www.philips.co.in/healthcare/product/HC882456/>.

In this present work, MATLAB 8.6 is used as a software tool. Original PET images of a patient with three different planes are shown in Fig. 3. Few consecutive PET brain slices (with three different planes) and the corresponding skull-stripped PET slices are shown in Fig. 8.

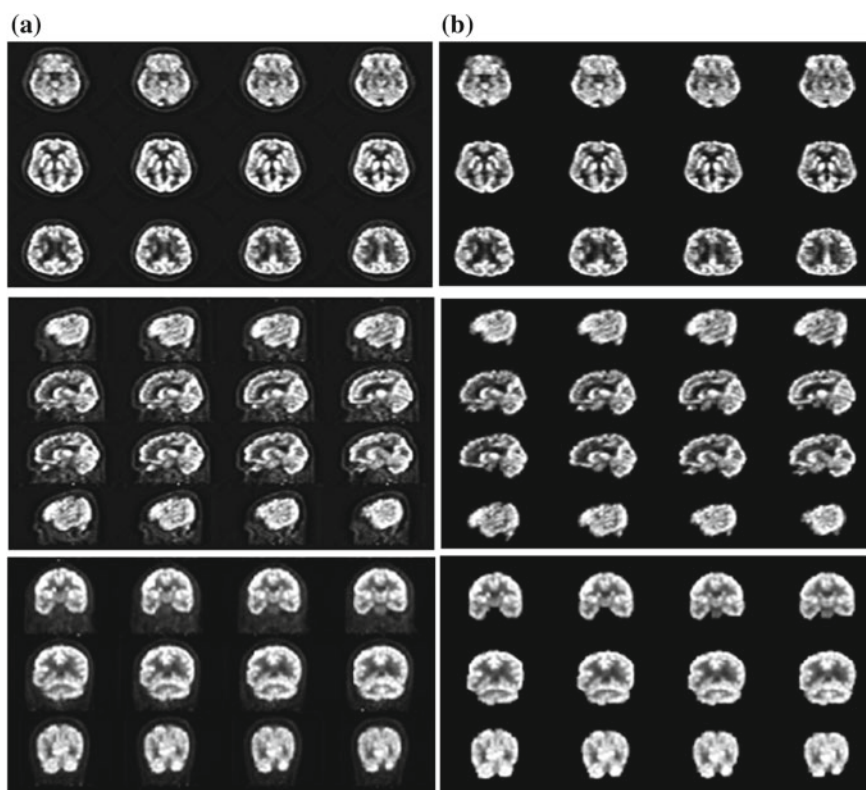


Fig. 8 Three different planes of a patient PET image. **a** Axial, sagittal, and coronal plane with skull, **b** axial, sagittal, and coronal plane without skull

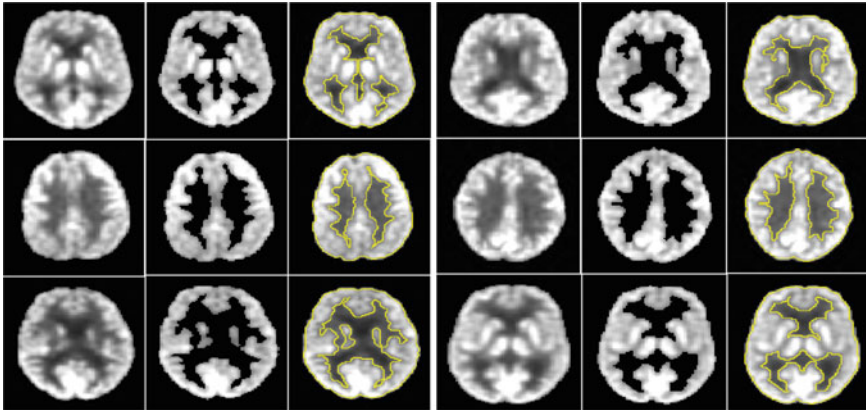


Fig. 9 Gray matter segmentation and delineation on PET axial slices. First and fourth columns show the skull-stripped axial PET slices, second and fifth columns show segmented gray matter, and third and sixth columns show the gray matter delineation

In this present work, we use three clusters ($c = 3$) for FPCM along with the fuzzifier $m_1 = m_2 = 2.0$. The value of two constants (a and b) is set to 0.5. In FPCM, the initial prototype is set by the final prototype of the FCM that can remove the problem of coincident cluster.

Figure 9 shows the segmented results of a patient PET image for different axial slices. The pictorial results in Fig. 9 signify that FPCM with sixteen feature properly segment the gray matter region from the skull-stripped PET image. The delineation results in third and sixth columns in Fig. 9 shows the cellular automata-based boundary delineation achieves promising results. In the presented delineation method, the Moore neighborhood concept is used for finding out the neighborhood of each pixel. For the Moore neighborhood, the total 1024 rules are possible. It is experimentally observed that in this particular application, rule 56 (0000111000) is the most suitable one. Selection of the rule may be varied depending on the application or the structure of the required object.

7 Conclusion and Future Work

In this present work, we propose a method that can segment gray matter of the brain and also able to delineate the segmented gray matter on standard Indian PET dataset. The visual interpretation of the evaluation results show that the proposed method achieves promising performance. During skull stripping, the skull of the PET images is removed using K-means and connected component analysis and then gray matter regions are segmented from PET brain tissues using FPCM with sixteen features. At last, segmented gray matter is delineated by the concept of cellular automata.

In the future, we want to propose an efficient PET-based machine learning method that can able to analyze the different kinds of dementia, such as Alzheimer's disease, Dementia with Lewy bodies (DLB), and Parkinson's disease.

Acknowledgements This research work is supported by the Board of Research in Nuclear Sciences (BRNS), DAE, Government of India, under the Reference No. 34/14/13/2016-BRNS/34044.

References

1. A. Bal, M. Banerjee, P. Sharma, M. Maitra, Brain tumor segmentation on MR image using k-means and fuzzy-possibilistic clustering, in *2018 2nd International Conference on Electronics, Materials Engineering Nano-Technology (IEMENTech)* (May 2018), pp. 1–8
2. P. Rusjan, D. Mamo, N. Ginovart, D. Hussey, I. Vitcu, F. Yasuno, S. Tetsuya, S. Houle, S. Kapur, An automated method for the extraction of regional data from PET images. *Psychiatry Res.: Neuroimaging* **147**(1), 79–89 (2006)
3. A. Bal, R. Saha, An improved method for text segmentation and skew normalization of hand-writing image (2018), pp. 181–196
4. B. Fischl, D.H. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. Van Der Kouwe, R. Killiany, D. Kennedy, S. Klaveness et al., Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**(3), 341–355 (2002)
5. A. Bal, R. Saha, An improved method for handwritten document analysis using segmentation, baseline recognition and writing pressure detection. *Procedia Comput. Sci.* **93**, 403–415 (2016)
6. J.M. Mykkänen, M. Juhola, U. Ruotsalainen, Extracting VOIs from brain PET images. *Int. J. Med. Inform.* **58**, 51–57 (2000)
7. R.H. Huesman, G.J. Klein, B.W. Reutter, X. Teng, Multislice PET quantitation using three-dimensional volumes of interest, in *Quantitative Functional Brain Imaging with Positron Emission Tomography* (Elsevier, Amsterdam, 1998), pp. 51–58
8. J. Mykkänen, J. Tohka, J. Luoma, U. Ruotsalainen, Automatic extraction of brain surface and mid-sagittal plane from PET images applying deformable models. *Comput. Methods Programs Biomed.* **79**(1), 1–17 (2005)
9. C. Svarer, K. Madsen, S.G. Hasselbalch, L.H. Pinborg, S. Haugbøl, V.G. Frøkjær, S. Holm, O.B. Paulson, G.M. Knudsen, MR-based automatic delineation of volumes of interest in human brain PET images using probability maps. *Neuroimage* **24**(4), 969–979 (2005)
10. B.F. Hutton, M. Braun, L. Thurfjell, D.Y. Lau, Image registration: an essential tool for nuclear medicine. *Eur. J. Nucl. Med. Mol. Imaging* **29**(4), 559–577 (2002)
11. J. Mykkanen, J. Tohka, U. Ruotsalainen, Delineation of brain structures from positron emission tomography images with deformable models. *Stud. Health Technol. Inform.* **95**, 33–38 (2003)
12. J.C. Bezdek, R. Ehrlich, W. Full, FCM: the fuzzy c-means clustering algorithm. *Comput. Geosci.* **10**(2–3), 191–203 (1984)
13. R. Krishnapuram, J.M. Keller, A possibilistic approach to clustering. *IEEE Trans. Fuzzy Syst.* **1**(2), 98–110 (1993)
14. N.R. Pal, K. Pal, J.C. Bezdek, A mixed c-means clustering model, in *Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, 1997*, vol. 1 (IEEE, New York, 1997), pp. 11–21
15. N.R. Pal, K. Pal, J.M. Keller, J.C. Bezdek, A possibilistic fuzzy c-means clustering algorithm. *IEEE Trans. Fuzzy Syst.* **13**(4), 517–530 (2005)
16. J.C. Bezdek, *Pattern Recognition with Fuzzy Objective Function Algorithms* (Springer Science & Business Media, Berlin, 2013)
17. A. Popovici, D. Popovici, Cellular automata in image processing, in *Fifteenth International Symposium on Mathematical Theory of Networks and Systems*, vol. 1 (Citeseer, 2002), pp. 1–6

18. R. Saha, A. Bal, M. Bose, Gray scale image recognition using finite state automata, in *National Conference on Recent Innovations in Computer Science & Communication Engineering* (2016)
19. S. Grace Chang, B. Yu, M. Vetterli, Adaptive wavelet thresholding for image denoising and compression. *IEEE Trans. Image Process.* **9**(9), 1532–1546 (2000)
20. P. Maji, S.K. Pal, *Rough-Fuzzy Pattern Recognition: Applications in Bioinformatics and Medical Imaging*, vol. 3 (Wiley, New York, 2011)
21. H.B. Kekre, S. Gharge, Texture based segmentation using statistical properties for mammographic images. *Entropy* **1**, 2 (2010)
22. S. Masood, I. Soto, A. Hussain, M. Arfan Jaffar, Statistical features based noise type identification, in *Mexican International Conference on Artificial Intelligence* (Springer, Berlin, 2014), pp. 231–241
23. H.-Y. Yang, X.-J. Zhang, X.-Y. Wang, LS-SVM-based image segmentation using pixel color-texture descriptors. *Pattern Anal. Appl.* **17**(2), 341–359 (2014)

A Joint Image Compression–Encryption Algorithm Based on SPIHT Coding and 3D Chaotic Map



Ramkrishna Paira

Abstract In this paper, the author has proposed secure and fast joint image compression–encryption algorithm using discrete wavelet transform (DWT), set partitioning in hierarchical trees (SPIHT) algorithm, and 3D chaotic map. The proposed approach is divided into three phases for partial encryption, full encryption, and permutation in order to elaborate the cryptographic properties. Compression part is carried out via secure SPIHT. This inclusion of security in SPIHT has no reflex on compression properties. Intensive experiments are performed to justify and validate the proposed approach.

Keywords Image compression · Image encryption · SPIHT · DWT · Chaotic map

1 Introduction

In this era of Internet and smartphones, an enormous amount of data, especially images are produced. Thus, it has become a challenge for us to store and circulate image securely over the Internet. Hence, the two important factors are considered.

- (1) Image compression—It means to reduce the size of an image without degrading the image quality. This allows us to store images in limited disk or memory space. It is also useful in transmitting the image in low bandwidth. For instance, normally the pictures of size 15 MB from the smartphones are taken with an Internet speed of 64 k bit/s. It would take one hour to transfer but on applying image compression, it can be done in few minutes. The basic steps for image compression include transformation, followed by quantization and entropy coding (Fig. 1).

The most frequently used transformations in digital image processing are discrete Fourier transform (DFT), discrete cosine transform (DCT), and discrete wavelet

R. Paira (✉)

Department of Computer Engineering, Smt. S. R. Patel Engineering College,
Unjha 384170, Gujarat, India
e-mail: pairaram@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_34

373

Fig. 1 Image compression process

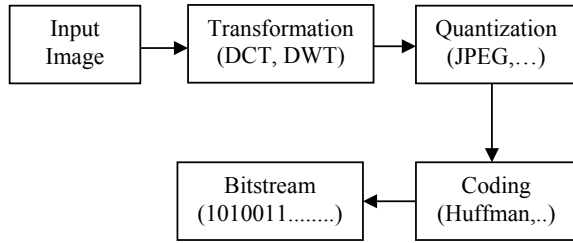
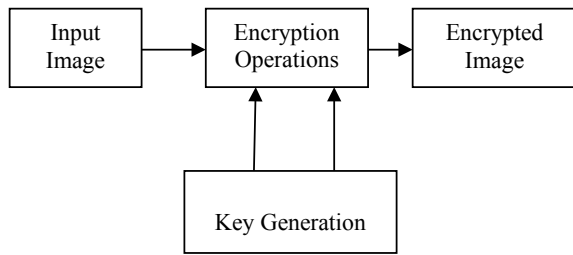


Fig. 2 Image encryption process



transform (DWT). DCT is basic in many lossy image compression methods. The image is divided into 8×8 blocks and then, DCT is performed. The motive of DCT is to assemble energy in terms of a sum of sinusoids with different frequencies and amplitudes. However, image compression algorithms based on DWT are highly efficient and diminish blocking artifacts [1]. In DWT, images are divided into sub-bands (LL, HL, LH, HH) by using wavelet filters [2]. The classic compression standard, i.e., JPEG 2000 is based on DWT while old JPEG is based on DCT. After transformation, quantization is applied. There are several quantization methods like embedded zerotree of wavelet (EZW), SPIHT, etc. The SPIHT coder is an efficient and fast algorithm for image compression [3]. In the final step of entropy coding, the output of the SPIHT algorithm is combined with Arithmetic, Huffman Coding, etc.

(2) Image encryption—The process using encryption algorithms to convert an image into a secret image so that unauthorized user can't access it (Fig. 2).

From security point of view, few suggested the direct transmission of an encrypted image without compression, but it requires a lot of bandwidth. Other proposed algorithms which compress the image in the first step and then encrypt the image in the second step, but it needs more time to execute. However, very few proposed the encryption of image before the compression. While some suggested novel methods have a joint image compression and encryption process.

2 Study on DWT and SPIHT

- (1) *Discrete wavelet transform (DWT)*: The wavelet transform is practically used in image and signal processing. The Joint Photographic Experts Group (JPEG) has unleashed its new image compression standard, i.e., JPEG 2000, which is based upon DWT. This transform break signals into set of basis functions, called wavelets and which are obtained from mother wavelets [4]. DWT has various transforms such as Haar wavelet, Daubechies wavelets, Biorthogonal wavelets, and Meyer wavelet. In this process, image is analyzed by passing through filter banks at each decomposition stage. A 2D transform takes place by performing two isolated 1D transforms. In first stage, image is filtered along x-dimension, then sub-image along y-dimension. Finally, the input image is splitted into four sub-bands [1, 5]. The four sub-bands, i.e., LL, LH, HL, and HH images which contain all the information in the original image. The LL sub-band has the significant information called approximation sub-band of the original image. The LH, HL, HH has vertical, horizontal, and diagonal information of the original image, respectively (Fig. 3).
- (2) *Set partitioning in hierarchical trees (SPIHT)*: It is the improved version of embedded zerotree wavelet (EZW) algorithm [6]. It is wavelet based on image encoding technique [7]. It is suitable for both lossy as well as lossless image compression [3].

For each coordinate value, there is a significance test function to valuate the significance. Let, t be a coordinate on which significance has to be checked and n is the integer value, then the significance test function is $S_n(t)$.

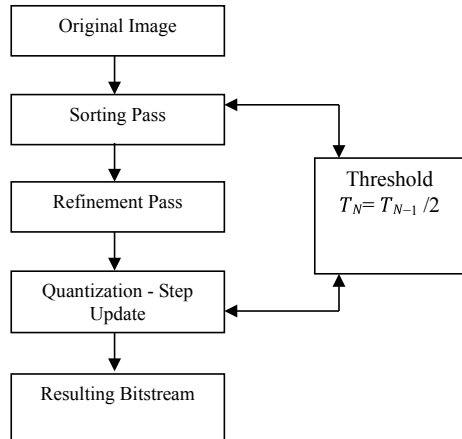
$$\begin{aligned}
 &\text{If}(\max(i, j)\{|c_{i,j}|\} \geq 2^n) \\
 &\quad S_n(t) = 1 \\
 &\text{else} \\
 &\quad S_n(t) = 0
 \end{aligned}$$

where $c_{i,j}$ defines the coefficient value at location (i, j) . If $S_n(t) = 1$, then coordinate t is said to be significant at the threshold value $T = 2^n$ otherwise, insignificant [3].

Fig. 3 Two-level decomposition

LL ₂	HL ₂	HL ₁
LH ₂	HH ₂	
LH ₁		HH ₁

Fig. 4 Flowchart of SPIHT



This algorithm has three lists to store information namely, list of insignificant pixels (LIP), list of significant pixels (LSP), list of insignificant sets (LIS). All the entries in the list are identified by the pixel value (i, j) (Fig. 4).

Algorithm:

1. Input image.
2. Perform transformation using DWT.
3. Initialization of LSP, LIP, LIS using significant test where threshold value $T = 2^n$ and n is given by:

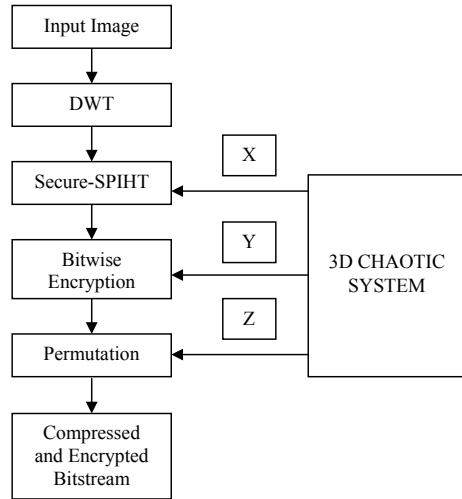
$$n = \lfloor \log_2(\max(i, j)\{|c_{i,j}|\}) \rfloor.$$

4. Output n sent this output value to the decoder.
5. Output μn and sign of each μn pixel coordinates the such that $2^n \leq |c_{\eta}(k)| \leq 2^{n+1}$ (Sorting Pass).
6. Output the n th most significant bit of all the coefficients with $|c_{i,j}| \geq 2^{n+1}$ (i.e., those coordinates were transmitted in the past go), in a similar request used to send the pixel coordinates (Refinement Pass).
7. $n = n-1$ (Quantization-step Update), and go to (Step 2).

3 Proposed Algorithm

The approach, “Joint image compression & encryption based on SPIHT algorithm & 3D chaotic map” is grounded on wavelet transformation, SPIHT algorithm, bitwise encryption and permutation of the output bitstream using chaotic secure system. It

Fig. 5 Flowchart of proposed algorithm



is a fast algorithm for an image compression and encryption process. The proposed method has good PSNR ratio which is required for proper reconstruction of the image (Fig. 5).

Proposed Algorithm:

1. Input image.
2. Apply DWT on the input image.
3. Perform secure SPIHT for joint image compression and partial encryption process.
4. Bitwise encryption on output bitstream of secure SPIHT encoding.
5. Permutation.
6. Compressed and encrypted output.

(1) *Chaotic System*: The simplest way of chaos generation is defined by the logistic map. The equation for the logistic map is given by:

$$X_{n+1} = \mu X_n(1 - X_n)$$

$0 < X_n < 1$ and $\mu = 4$ are the basic conditions to make the chaotic equation [8]. Hongjuan Liu gives us the 2D logistic map employing quadratic coupling for security enhancing and it is upgraded to 3D version, which is given by the following equations:

$$\begin{aligned} X_{n+1} &= \gamma X_n(1 - X_n) + \beta Y_n^2 X_n + \alpha Z_n^3 \\ Y_{n+1} &= \gamma Y_n(1 - Y_n) + \beta Z_n^2 Y_n + \alpha X_n^3 \\ Z_{n+1} &= \gamma Z_n(1 - Z_n) + \beta X_n^2 Z_n + \alpha Y_n^3 \end{aligned}$$

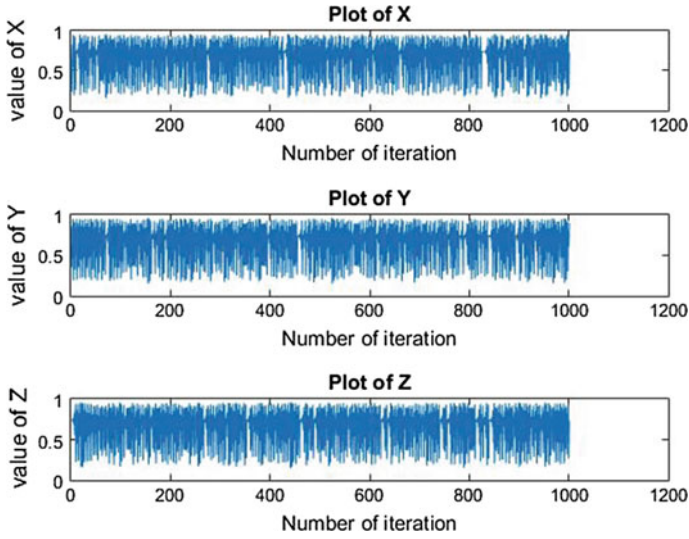


Fig. 6 3D chaotic sequence

where $3.53 < \gamma < 3.81$, $0 < \beta < 0.022$, and $0 < \alpha < 0.015$. The initial input values of the X , Y , and Z are lying between zero and one. The secure chaotic system is hypersensitive to the initial conditions. The two very close initial values will never follow the same path [9].

Figure 6 shows the generated chaotic sequence for the initial values $X(1) = 0.235$, $Y(1) = 0.350$, $Z(1) = 0.735$, $\alpha(1) = 0.0125$, $\beta(1) = 0.0157$, $\gamma(1) = 3.7700$ [10].

- (2) *Secure SPIHT*: This process encapsulated the encryption in the SPIHT algorithm to enhance the security. The data is not only compressed but also encrypted, which will enhance the cryptographic property of the bitstream. For each coefficient, there is a need to check if its value is greater than threshold (T) value, where $T = 2^n$ and n is obtained by

$$\lfloor n = \log_2(\max(i, j)\{c_{i,j}\}) \rfloor.$$

In this case, the output is 1 else the output is 0. These output bits are encrypted using the key value [11].

Modified LIP sorting pass process is shown below.

```

For each entry in the LIP
  If (current coefficient ≥ threshold( T ))
    Output = bitxor( 1, key )
    If ( sign( current coefficient ) > 0)
      Output = bitxor( 1, key )
    Else
      Output = bitxor( 0, key)
  Else
    Output = bitxor( 0, key )
End
    
```

Encryption of LIP sorting pass at encoder end.

```

For bitstream at decoder
  If ( bitxor ( bitstream, key ) == 1)
    Update the coefficient with regards sign and
    magnitude
    If ( bitxor ( bitstream, key ) == 0)
      Coefficient ≥ threshold( T )
    Else
      Coefficient ≤ threshold( T )
  Else
    Insignificant coefficient
End
    
```

Decryption of LIP sorting pass at decoder end.

- (3) *Bitwise Encryption*: This process encrypts the output bitstream of secure SPIHT algorithm. Encryption process is based on the key value. The number of key values depends on the number of compressed code stream generated from the secure SPIHT. In this process, the initial four bitstreams are ignored as they carry the important information to the decoder, which includes T value, number of rows, columns, and levels [6] (Fig. 7).
- (4) *Permutation*: For the last stage of security enhancement, this step permutes the encrypted output to give new location based on the chaotic sequence [12] (Fig. 8).

Fig. 7 Bitwise encryption of code stream

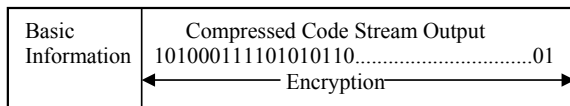


Fig. 8 Permutation of code stream

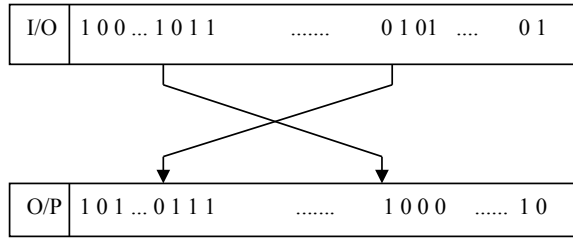
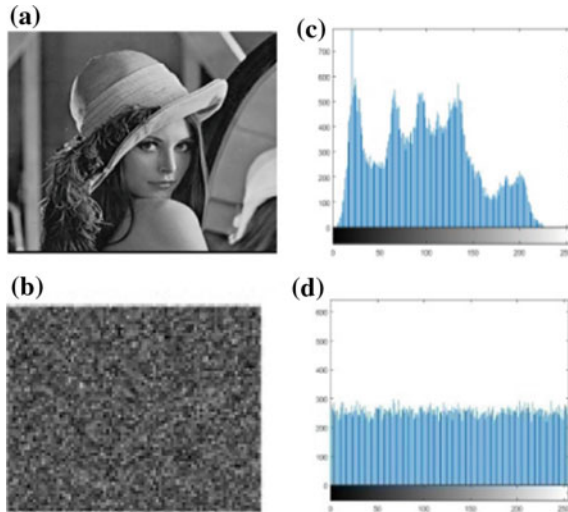


Fig. 9 **a** Lena image, **b** Encrypted image, **c** Histogram of (a), **d** Histogram of (b)



4 Experimental Results

(1) *Histogram*:

All Fig. 9 experiments have been performed on MATLAB. For ensuring encryption quality, there is a need to perform histogram test of the input and encrypted image.

(2) *Peak Signal-to-Noise Ratio (PSNR)*: It is given by the formula mentioned below

$$PSNR = 20 \log_{10} \frac{255}{\sqrt{MSE}}$$

where MSE is given by $MSE = \frac{1}{LL} \sum_{i=1}^L \sum_{j=1}^L (m_{ij} - n_{ij})^2$, m and n are output and input image coordinates, respectively at location (i, j) .

Table 1 shows the PSNR value for the proper reconstructed image at the decoder end. For the quality of image to be better, PSNR value should be high.

Table 1 PSNR table

Image	PSNR
Lena	32.61
Cameraman	28.15
Airplane	32.16
Rice	31.87
Moon	31.21

Fig. 10 Compression ratio chart

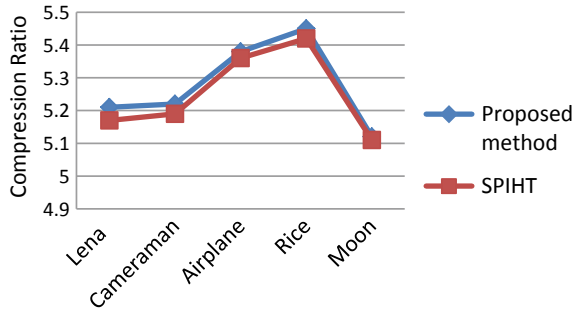


Table 2 Key sensitivity test table

Image	PNSR (1 bit modification)
Lena	1.15
Cameraman	2.17
Airplane	1.28
Rice	1.38
Moon	2.16

(3) *Compression Ratio*: By definition, it is the ratio of uncompressed image to the compressed image size (Fig. 10).

The above chart shows that the compression ratio is not affected by the encryption process.

(4) *Key Sensitivity Test*: An algorithm is said to be good if it is key sensitive. A minute modification in the key value should make a huge difference. In this, one bit in the key value is modified, which leads to the failure of decryption process at the decoder end (Table 2).

5 Conclusion

This paper presented secure and fast joint image compression and encryption algorithm which is based on DWT, secure SPIHT, and 3D chaotic system. This chaos values are very secure and sensitive in nature, which provides excellent encryption

process, while compression part is carried out by basic SPIHT. By employing various evaluation parameters, security tests and performance tests are performed to validate the goals.

References

1. M. Vetterli, J. Kovacevic, *Wavelets and Subband Coding* (Prentice Hall, Englewood Cliffs, NJ, 1995)
2. M. Zhang, X. Tong, Joint image encryption and compression scheme based on IWT and SPIHT. *Opt. Lasers Eng.* (2017)
3. A. Said, W.A. Pearlman, A new fast and efficient image codec based on set partitioning in hierarchical trees (1996)
4. A. Alice Blessie, J. Nalini, S.C. Ramesh, Image compression using wavelet transform based on the lifting scheme and its implementation. *IJCSI Int. J. Comput. Sci. Issues* **8**(1) (2011)
5. *Wavelet Methods, Data Compression* (2007)
6. M. Hamdi, R. Rhouma, S. Belghith, A selective compression-encryption of images based on SPIHT coding and Chirikov Standard Map. *Signal Process.* (2017)
7. R. Sudhakar, R. Karthiga, S. Jayaraman, Image compression using coding of wavelet coefficients—A survey. *ICGSTGVIP J.*
8. S. Al-Maadeed, A new chaos-based image-encryption and compression algorithm. *J. Electr. Comput. Eng.* (2012)
9. Md. Billal Hossain, Md. Toufikur Rahman, A.B.M. Saadmaan Rahman, S. Islam, A new approach of image encryption using 3D chaotic map to enhance security of multimedia component, in *2014 International Conference on Informatics, Electronics & Vision (ICIEV)*
10. P.H. Kharat, S.S. Shriramwar, A secured transmission of data using 3D chaotic map encryption and data hiding technique, in *2015 International Conference on Industrial Instrumentation and Control (ICIC)* (2015)
11. T. Xiang, J. Qu, C. Yu, X. Fu, Degradative encryption: an efficient way to protect SPIHT compressed images. *Opt. Commun.* (2012)
12. L. Bao, Y. Zhou, C.L. Philip Chen, H. Liu, A new chaotic system for image encryption, in *2012 International Conference on System Science and Engineering (ICSSSE)* (2012)

Improved Multi-feature Computer Vision for Video Surveillance



Ashutosh Upadhyay and Jeevanandam Jotheeswaran

Abstract Computer vision deals with automatic extraction, analysis, and understanding of information from image or video. Recent research in the area of computer vision mainly focused on developing intelligent systems for detecting and observing human behaviors. Human detection is the process of locating human automatically in an image or video sequence. This paper presents an efficient intelligent surveillance system using features, namely HAAR, local binary patterns, and histogram of oriented gradients to detect the presence of human being. Support vector machine is used to train the system. In most of the work in computer vision, researchers used different datasets and different features. This paper compares the result of the mentioned features using a common high-definition dataset. To reduce the time complexity of high-definition video, the proposed method normalized the extracted video frames into equal size. The experiments are performed on images and videos and the performance of the system is presented term of true positive rate, false positive rate, accuracy, and computation time.

Keywords Surveillance · Intelligent · Anomaly · HAAR · Wavelet · LBP · HOG

1 Introduction

The understanding of the human behavior is an important research in the area of computer vision which becomes more popular in the last few years. Human beings are capable of perceiving humans by only slight indication, but intelligent human detection algorithms are still away from human being capability of observing and detecting the human beings.

A. Upadhyay (✉) · J. Jotheeswaran
Galgotias University, Greater Noida, India
e-mail: ashutoshup90@gmail.com

J. Jotheeswaran
e-mail: jeevanandamj@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_35

Since the last decade, the human detection in video and images has attracted the attention of researchers in computer vision and pattern recognition mainly due to the various applications. The applications of human detection include visual content management, which requires the tagging of the objects, especially humans in image and video to facilitate efficient search. In video surveillance, the key tasks are to detect the objects, identify its behaviors or patterns and monitor the detected objects in crowded as well as public places that include airports, railway stations, bus stops, and supermarkets. The autonomous vehicle also uses the human detection for detecting the presence of pedestrians to alert the vehicle driver to avoid accidents.

The use of automated surveillance also gaining the attention to ensure the security for public places, banks and schools etc. To fulfill the requirement of different applications, many networked intelligent video surveillance systems have been placed, which not only ensure the security of citizens, but also effectively helping to transform the cities to smart cities. With the advancement in the technology, it becomes easy to install closed-circuit television (CCTV) for systematic observation on a group of people, sensitive places, suspicious persons, and many other anomalous and dynamically changing objects in the environment. We can gather information from the devices for analysis of the behavior and activities of the human [1]. Due to the increase in the crime rates, handling of critical issue using visual surveillance [2] becomes a tough job to locate the antisocial activities. The need of an intelligent surveillance system is crucial to automate surveillance system and make the system capable of detecting and monitoring anomalous activities in video without human interaction, so that we can reduce the burden of operator and make it more accurate to control the increasing crime rates.

Automated human detection in video and images remains an active research area in computer vision because of the following [2] challenges:

- **Dynamic shape of the object:** The shape of the human is not fixed due to the flexible nature of the human body. So developing a mathematical model to train a system for capturing human is a challenging task.
- **Changes in illumination:** The sudden changes in the illumination of the images or video may happen in indoor as well as environment. The appearance of the background strongly affected by the by the illumination and may lead to false-positive detections.
- **Orientation of the objects:** Gradual changes in the orientation of the object leads to the failure of temporal differencing method to detect uniform region the object is preserving.
- **Partial occlusion:** Partial or full occlusion affects computing of background frames. Occlusion is temporary and may occur at any time if subject passes behind an object in front of a camera.
- **Shadowing effect:** The shadowing region of targets signals is too weak to be detected. Sometimes shadows may be classified as foreground components. This is because shadows have similar patterns and similar intensity changes as foreground objects have.

There are limits on the human efforts to analyze the data collected during the image and video acquisition. The analysis includes identification of the object in the video and images and later analysis of the detected object is performed. Object behavior understanding is a tough task because of the wide range of features, real-time changing pose, and orientation. The complexity of backgrounds makes it more difficult to identify the objects in real time in images and video. Human detection has many application in the in military as well as industry, so human may be considered as an essential component of automated visual surveillance. The human detection and identification include analysis of human behavior, action, and movement and tracking of human in real time in the images or videos.

The paper is organized in six sections. Section 2 describes the related work to show the recent algorithm and techniques for the human detection. In Sect. 3, paper describes the terminology used in human detection. And in Sect. 4, paper presents the steps involved in the proposed methodology and experiments for analyzing the performance. Results are shown in Sect. 5. Finally, conclusion is given in Sect. 6.

2 Literature Survey

The flexible nature of the human body capable of producing many possible poses and gestures.

There are many ways to extract the object regions. The general approach is to enclose human object in a detection window. It is also difficult to model orientation and a shape as well as size variation arises due to the change of the position and direction of the camera. Unlike other objects, humans can be clothed himself with varying colors which increases the dimension of complexity.

In general, human detection process from images and videos can be explained in the mentioned steps: Extracting the region which is covered by human object, classifying the extracted regions into two distinct groups that are human or non-human, perform post-processing for example merging the similar regions [3] or performing the size adjustment of those regions. This process is illustrated in Fig. 1.

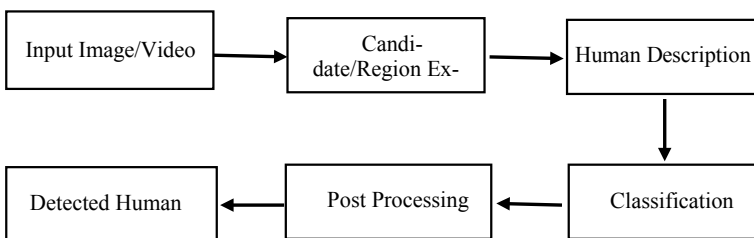


Fig. 1 A human detection framework. Image or video preprocessing applied to enhance the quality of input before candidate/region extraction

There are many ways to extract the object regions. The general approach is to enclose human object after detection in a rectangular window.

To extract the candidate in the input video sequence, the popular technique known as background subtraction proposed in paper [4] can be used to obtain human candidates. In this method, the objects are separated from the background by subtracting the current input image having object from the reference background image without objects. Moreover, this method often needs a static camera and background without human for reference. The background subtraction technique is used in paper [5] for detecting object in motion.

Authors [6] present background subtraction and human detection for outdoor captured video by using fuzzy logic approach. The proposed system is capable enough to detect human in the presence of other animals in the frame. The proposed approach is scale and view invariant. Researchers [7] proposed the approach by focusing on real-time anomalous activities. The authors collected the data of shopping malls for experiments. Authors used supervised classifier, i.e., support vector machine (SVM). Authors achieved success near to 85% in case of occlusion situations.

Paper [8] presents an algorithm for the detection of the vehicles in real time. The authors used Haar-like features to train the system and used neural network for the features classification. The proposed system is efficient and robust for detecting the vehicle in real time. Paper [9] presents the pedestrians detection algorithms using histogram of oriented gradient (HOG) features for the template matching. The proposed technique locates and tracks all the human being present in the sequence of frames. The presented algorithm achieves 80% human detection rate.

Authors [10] proposed new features sector-ring HOG (SRHOG) to detect rotating human. The proposed method is rotation invariant. Authors used INRIA dataset with HOG features and compared the results with SRHOG feature. Paper [11] presented a human detection approach which uses multi-instance learning (MIL) framework to detect human in any arbitrary pose. Authors used LSP and MPII datasets to train the system to detect the human under any arbitrary pose.

In paper [12], authors proposed a hybrid human detection model which utilizes the advantages of both the features that are Haar-like features and the HoG features to make object detector more effective. For human recognition, authors proposed a new approach by Overlapping Local Phase Features (OLPF) to locate the human face region which improves the robustness against the change in position of human and blurring effect. Authors used INRIA, Caltech, and ETHZ datasets. In paper [13], authors proposed a cascade-structured approach to detect human in dynamic and cluttered environment. The authors used three datasets, i.e., clothing store datasets, office dataset, and mobile dataset for mobile platform.

From the discussion, we can conclude that the authors used different features and different datasets to train system with different trainer. In the proposed approach, we have used the high-definition datasets to measure the performance of the different features, namely HOG, LBP, and HAAR. The proposed human detection approach uses high-definition datasets because the high-definition datasets need more processing time. We reduce the number of frames in HDV to reduce the processing time.

3 Terminology Used in Human Detection

In this section, the explanation of different algorithms and terminologies used in automated surveillance systems are presented. The algorithms used for in visual surveillance are HAAR wavelet, LBP, HOG.

3.1 HAAR-Like Features

HAAR-like features are similar to DWT. DWT captures both the location and frequency information efficiently. HAAR-like feature mainly considers a rectangular window near the location.

In Paper [14], authors presented a survey on HAAR wavelet features. HAAR features are calculated by Eqs. 1 and 2.

$$F_{(P,D)} = \sum_{i=0}^{m-1} \left(\sum r_0 - \sum r_n \right) \tag{1}$$

where $\sum r_0$ represents the total intensity of the pixels. The pixels are selected in neighborhood of the current pixel that is N8 pixels.

$$S(x) = \begin{cases} F_{(P,D)}, & X < 0 \\ 1, & X \geq 0 \end{cases} \tag{2}$$

$\sum r_n$ represents the total values of neighboring N8 pixels. The pixels are taken in circular fashion that is p pixels in the diameter range D. The HAAR-like feature classification is done by calculating the sum and difference in the intensity values only in the selected window region. During the detection of the object, the selected window is moved over the input image over the each selected subsection. The feature vector in HAAR-like feature is calculated by moving the window over each subsection. Finally, the calculated feature vector is then matched with the threshold to finalize the location of objects in the image and videos. It has been observed by the researchers that the HAAR-like features are more robust to shape, size, and position of an object.

3.2 Local Binary Pattern (LBP)

LBP is another feature descriptor like HAAR features to locate the objects in images and videos. LBP initially used to classify texture pattern. LBP features are calculated using Eqs. 3 and 4. LBP considers the result as a binary number by thresholding the neighboring pixels in a circular manner. LBP is popular feature descriptor because of the efficiency in term of computation overhead.

The (x_a, y_b) pixel value of LBP is given by:

$$F_{(P,D)} = \sum_{i=0}^{m-1} s(r_0 - r_n)2^i \quad (3)$$

where r_0 denotes the intensity value of a pixel at location (x, y) and r_n represents the pixel in neighborhood of P pixels in the range of diameter D .

$$S(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (4)$$

LBP feature descriptor calculates the feature vector with respect to its neighbors by comparing pixel value to its N8 pixel. If the intensity value of the selected center pixel is less than neighbor's pixel value, then center pixel value is set to one. If the center value is greater, then set to zero. The calculation is done for each pixel by moving the window to produce entire feature vector.

3.3 Histogram of Gradient (HOG)

The HOG feature is another feature to be used widely in human detection algorithms. The edge information is another important feature in any object detection. The HOG feature mainly computes the edge direction and local intensity distribution. To calculate the hog features, we can use Eqs. 5 and 6.

$$g(x, y) = \sqrt{g_x^2 + g_y^2} \quad (5)$$

$$\Theta(x, y) = \tan^{-1}\left(\frac{g_y}{g_x}\right) \quad (6)$$

where in Eq. 5 g_x and g_y is calculated by using the mask $[1 \ 0 \ -1]$ and $[1 \ 0 \ -1]^T$ mask, respectively. For calculating each parameter, the whole input image is divided into small cell and histogram of gradient is calculated for each pixel in the divided cells.

3.4 Support Vector Machine (SVM)

SVM is supervised learning models which analyze the data with the help of associated learning algorithms for the classification. The SVM assigns the input to the one or the other groups based on the learned data. This is a binary classifier and given by the function.

$$g: R_n \rightarrow \pm 1$$

where g is function which maps patterns x to their correct classification y as given by $y = f(x)$.

The function f in case of SVM is given in Eq. (7).

$$f(x) = \sum_{i=1}^n y_i \alpha_i k(x, x_i) + b \quad (7)$$

where (x_i, y_i) is i th training pattern. α_i and b are learned weights and $k(x, x_i)$ here is a kernel function.

4 Proposed Methodology

In this section, paper presents detailed steps involved in the procedure of identifying human in real-time images and videos. The technique is capable of identifying the human being when they appear in the video frame. After detecting the human in the frame, validation of object is perform on the basis of the number of the consecutive frames the human is present. In the proposed algorithm if the human is present in the five consecutive frames, declare it as human to reduce the computational overhead and to maintain the accuracy in high-quality videos.

4.1 System Initialization

System initialization is the very first step in the process of human detection. System initialization phase loads the video or an image captured from the environment through cameras or sensors and also loads the output file classifier which is generated during the training phase. The classifier classifies the human and non-human part in the input video frames and identifies the presence of the human in the video frames.

4.2 Video Preprocessing

Video preprocessing is mainly used to extract the frames from the input video and removes the present noises in the selected frames. The cause of presence of the noise in the video and image during video and image acquisition from CCTV is mainly due to the change in environmental conditions and due to unreliable network condition during the transmission of the acquired video. To address the noise issue, apply the image filter to enhance the received signals.

 Algorithm: Human Detection

Input : $V_i \rightarrow$ input video
 $Fr_i \rightarrow$ i^{th} frame extracted from input video
 $COUNT \rightarrow$ counter variable (for counting the frame)

Initialization: Load the file which is output of training phase Initialize $NUM=0$,
 detected object=nil.

Begin

For each frame Fr_i
 From frame Extract the features (one by one for
 HAAR, LBP and HOG using equation 1-6).
 Load the feature vector and match the features
 if (features matched more than threshold) then
 Draw the rectangular boundary box over detected object.
 Increase COUNT.
 else
 Reset COUNT.
 end if
 if $COUNT > 0$ then
 Raise the Alarm.
 end if

End

4.3 Features Matching and Human Detection

The extracted frame sizes for high-quality video are larger in size, i.e., high resolution. Because of greater resolution, the computation time is also large. Therefore, it is required to normalize high-quality video and images into reduce computation overhead. This paper removes unwanted noise and avoids processing in entire frame for detection if object detected in consecutive five frame processing ends. In this process, we aim to make all frames equivalent in size. In order to resize, we have rescaled the resolution.

In this paper, different features like HAAR feature, LBP, and HOG calculated separately from input video after extracting frame and rescaling. This paper used SVM supervised classifier and the comparative study among these features is presented. The matching of the extracted features of the frames with already trained classifier's output. A human is recognized if more than half of the selected features are classified into two groups that are human or non-human. The SVM supervised classifier produces a binary kind of categorization which assigns simply a binary value to categorize the features into two distinct classes either 0 or 1. The selected feature vector is said to be matched with the stored vector only if they satisfy some given adaptive threshold condition, otherwise it considers as the ambiguous.

5 Result Analysis and Comparisons

A dataset is a collection of images and videos on which the system is trained. There are many available standard datasets for the low-quality videos. We created high-definition video (HDV) datasets. The training dataset contains total 500 positive and negative images. In this paper, we performed eight tests as real-time surveillance videos in restricted background. The size of the video taken for experiment is HD with 1920×1200 pixels with 30 frames per second. This section presents the comparative study of human detection results for complex background by using three features, namely HAAR, HOG, and LBP. A green color rectangular window is drawn over the detected human after detection is confirmed in all the frames demonstrated in Fig. 2a–c, respectively.

In order to detect the human, the evaluation is done on many parameters to measure the efficiency, accuracy, and effectiveness model. The comparative study is presented using features HAAR, HOG, and LBP by taking eight test cases. On an average 83.5, 81, and 89.5%, TPR is achieved by using HAAR, LBP, and HOG, respectively, shown in Fig. 3 and FPR is approximately 3.2, 13, and 2.1% FPR is achieved for HAAR, LBP, and HOG, respectively, as shown in Fig. 4.

The average accuracy achieved is 89.375, 86.375, and 92.5% as shown in Fig. 5.

The data shows that accuracy is more for HOG features. As shown in Fig. 6, the computation time is almost the same for HAAR, LBP, and HOG, respectively, that is 68.75, 66.75, and 67.25, respectively.



Fig. 2 Sequence of frames with blob using a HAAR, b LBP, and c HOG

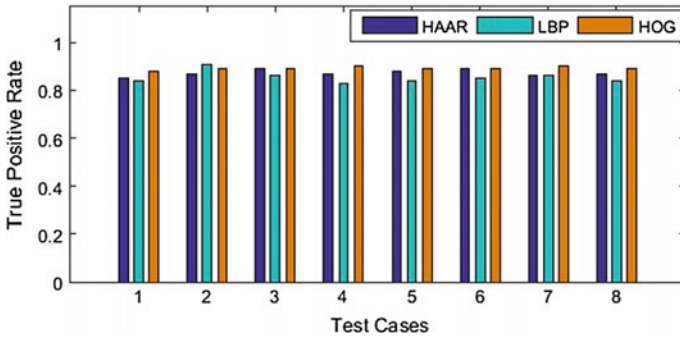


Fig. 3 TPR for HAAR, LBP, HOG

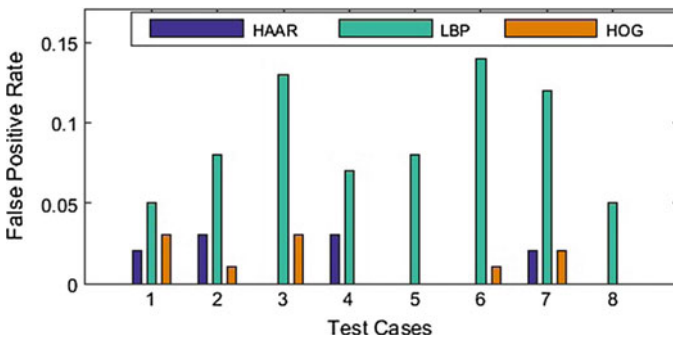


Fig. 4 FPR for HAAR, LBP, and HOG

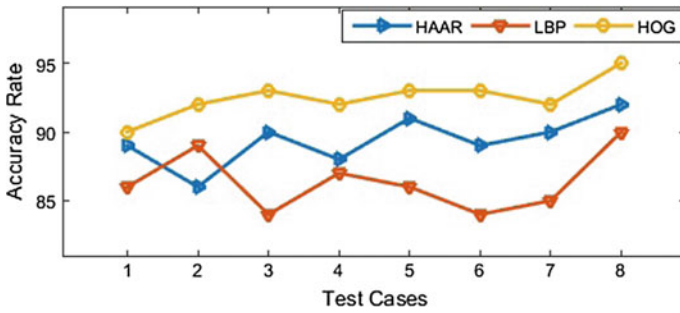


Fig. 5 Accuracy for HAAR, LBP, and HOG

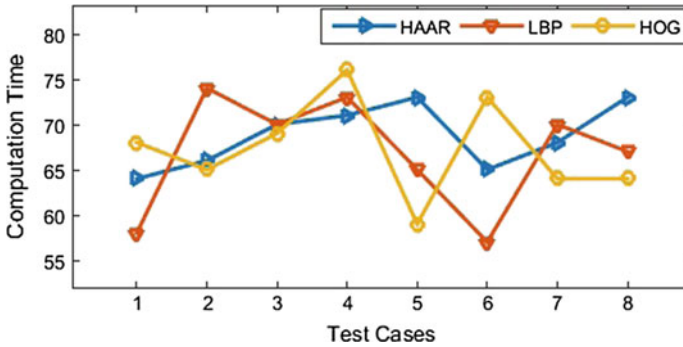


Fig. 6 Computation time for HAAR, LBP, and HOG

6 Conclusions and Future Scope

This paper discussed and demonstrated an intelligent human surveillance system that is able to detect human being in constraint environment. The paper uses the HAAR, LBP, and HOG algorithms for features. The proposed approach is robust enough to tackle issues like shape of the object, partial occlusion, shape transformation, illumination or lighting condition, shadowing effect, and objects multiplicity. Based on the experimental discussion, advantage of the proposed model is that the overall processing time is less for HDV. From result analysis on the same datasets, we can conclude that HOG feature gives better result for human detection in the constraints environment. We have used supervised SVM classifier to train our system. In the future, the system can be enhanced to achieve better results for night and one can utilize the unsupervised classifiers to improve the accuracy and efficiency.

References

1. D. Lane Nicholas, M. Lin, M. Mohammad, X. Yang, H. Lu, G. Cardone, S. Ali, A. Doryab, E. Berke, A.T. Campbel, T. Choudhury, BeWell: sensing sleep, physical activities and social interactions to promote wellbeing. *Mob. Netw. Appl.* **19**(3), 345–359 (2014)
2. E.L. Piza, J.M. Caplan, L.W. Kenned, Analyzing the influence of micro-level factors on CCTV camera effect. *J. Quant. Criminol.* **30**(2), 237–264 (2014)
3. M. Paul, S.M.E. Haque, S. Chakraborty, Human detection in surveillance videos and its applications—A review. *EURASIP J. Adv. Signal Process.* **1** (2013)
4. A. Sobral, A. Vacavant, A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos. *Comput. Vis. Image Underst.* **122**, 4–21 (2014)
5. P. Dewan, R. Kumar, Detection of object in motion using improvised background subtraction algorithm, in *International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli* (2017)
6. A. Mahapatra, T.K. Mishra, P.K. Sa, B. Majhi, Background subtraction and human detection in outdoor videos using fuzzy logic, in *IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), Hyderabad* (2013)

7. R. Arroyo, J.J. Yebes, L.M. Bergasa, I.G. Daza, J. Almazan, Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls. *Expert Syst. Appl.* **42**(21), 7991–8005 (2015)
8. A. Mohamed, A. Issam, B. Mohamed, B. Abdellatif, Real-time detection of vehicles using the haar-like features and artificial neuron networks. *Procedia Comput. Sci.* **73**, 24–31 (2015)
9. V. Barbu, Pedestrian detection and tracking using temporal differencing and HOG features. *Comput. Electr. Eng.* **40**(4), 1072–1079 (2014)
10. B. Liu, H. Wu, W. Su, J. Sun, Sector-ring HOG for rotation-invariant human detection. *Sig. Process. Image Commun.* **54**, 1–10 (2017)
11. Y. Cai, X. Tan, Selective weakly supervised human detection under arbitrary poses. *Pattern Recogn.* **65**, 223–237 (2017)
12. Q. Liu, V. Zhang, H. Li, K.N. Ngan, Hybrid human detection and recognition in surveillance. *Neurocomputing* **194**, 10–23 (2016)
13. J. Liua, G. Zhang, Y. Liu, L. Tian, Y.Q. Chen, An ultra-fast human detection method for color-depth. *J. Vis. Commun. Image Represent.* **31**, 177–185 (2015)
14. M. Oualla, A. Sadiqu, S. Mbarki, A survey of Haar-like feature representation, in *International Conference on Multimedia Computing and Systems (ICMCS), Marrakech* (2014)

Topic Modeling for Text Classification



**Pinaki Prasad Guha Neogi, Amit Kumar Das, Saptarsi Goswami
and Joy Mustafi**

Abstract Topic models allude to statistical algorithms for finding out an extensive text body's latent semantic structures. Standing here in today's world, the measure of the textual data and information we come across in our day-to-day lives is basically beyond our handling limit. Topic models can provide a way out for us to understand and manage the vast accumulations of unstructured textual data and information. Initially emerged as a text-mining instrument, topic models have found applications in various other fields. This paper makes a thorough comparative study of LSA with that of commonly used TF-IDF approach for text classification and proves that LSA yields better accuracy in classifying texts. The novelty of the paper lies in the fact that we are using a much sparser representation than usual TF-IDF and also, LSA can get from the topic if there are any synonym words. This paper proposes a method, using the concept of entropy, which further increases the accuracy of text classification.

Keywords Text classification · Topic modeling · Latent Semantic Analysis (LSA) · TF-IDF · Entropy

P. P. G. Neogi (✉)

Department of CSE, Meghnad Saha Institute of Technology, Kolkata 700150, India

e-mail: ratul.ng@gmail.com

A. K. Das · S. Goswami

A.K. Choudhury School of Information Technology, Calcutta University,

Kolkata 700106, India

e-mail: amitkrdas.kol@gmail.com

S. Goswami

e-mail: saptarsi007@gmail.com

J. Mustafi

MUST Research Club, Hyderabad 500107, India

e-mail: mustafi.joy@live.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling*

and Graphics, Advances in Intelligent Systems and Computing 937,

https://doi.org/10.1007/978-981-13-7403-6_36

1 Introduction

We live and experience a daily reality where floods of information are generated relentlessly. Therefore, hunting down intuitions from the gathered data can turn out to be exceptionally monotonous and tedious. News can be considered as a contemporary observer of the society and can reveal to us a great deal about things that went right or off-base. However, the quantity of news articles being delivered on the planet even in a single day is remarkably prodigious and this unthinkable vast quantity of data is beyond manual or human processing capability. As indicated by Chartbeat, every 24 h, more than 92,000 articles are published on the web. But, of course, utilization of computers can be made to segregate helpful data from news writings, for example, names of individuals, associations, geological locations, political gatherings and basically any word that has got a meaningful denotation in Wikipedia. However, it is extremely precarious for a computer to nail down “What is this article about?” in a couple of words, simply like any human being would answer naturally. Topic modeling is a technique of employing computers for this labyrinthine task. Topic modeling was delineated as an instrument for organizing, comprehending and searching for tremendous amounts of textual data. Thus, the idea is to make application of computational categorization techniques for answering the inquiry “What is this news article about?” on a substantial scale. Fortunately for us, the BBC is always ready with hundreds of fresh news every minute, so data are definitely not a limiting factor.

In machine learning, topic model can thus be particularly elucidated as a natural language processing tool, used for detecting concealed semantic structures of textual information in an accumulation of textual records, termed as corpus. Usually, each record alludes to a continuous arrangement of words, similar to an article or a passage, where each passage or article comprises an arrangement of words. Topic modeling is basically an unsupervised learning algorithm to deal with clumping of textual reports, by finding out topics based on the analysis of their substance. Its fundamental working concept is much analogous to that of expectation-maximization work and K-means algorithm. Since we are clumping documents, in order to find out topics, we will have to take into consideration and process every single word in the document and allocate values to each depending on its distribution. This swells the volume of information we are working with and thus to handle the complex processing necessitated for grouping documents, we should have to make utilization of well-organized sparse data structures.

This paper is organized as follows: Sect. 2 gives a brief overview of the related works in the fields we are focusing upon; Sect. 3 narrates the TF-IDF approach for classifying text documents; Sect. 4 gives a concise account of Latent Semantic Analysis (LSA); Sect. 5 explains the concept of entropy in a text document; the proposed method is described under Sect. 6, including the algorithm (Sect. 6.3); Sect. 7 gives a detailed account of the dataset and coding environment; experimental result, along with comparison tables and confusion matrices, is presented under Sect. 8; and lastly, the conclusion is given in Sect. 9.

2 Related Works

Zelikovitz [1] applied the semantic analysis algorithms [2–4] for the purpose of classifying the short texts. A progression of latent semantic allocation (LSA) for the classification of short texts is the Transductive LSA. Transduction makes the utilization of the test examples for the purpose of selecting the hypothesis of the trainee to settle on decisions contrapose the test cases. Pu et al. [2] amalgamated independent component analysis (ICA) and latent semantic allocation (LSA) together. A large-scale classification framework of short text documents was established by Phan et al. [5], which is essentially in light of machine learning methods such as SVMs and maximum entropy and that of latent dirichlet allocation (LDA). In any case, their work primarily centered on how to apply it to Wikipedia and no intuition was given on if there exists a different approach to train the same model. Generally, what happens in case of web search is that, the search engine gets employed directly in this line of research. For instance, a kernel function was proposed by Sahami et al. [6] in light of search engine outcomes. The method was extended even more by the application of some machine learning algorithm by Yih et al. [7].

LDA was stretched out by Ramage et al. [8] to a supervised form and its applications were analyzed in micro-blogging environment. A strategy in light of labeled LDA was built up by Denial Ramage et al. for multi-labeled corpora [9].

In [10], using Naïve Bayes algorithm for text classification, a novel approach has been proposed for feature selection. In light of latent dirichlet allocation for topic extraction from source code, an approach was proposed by Maskeri et al. [11]. A technique in text mining, based on partly or incompletely labeled topics was proposed by Manning et al. [12]. In these models, implementation of unsupervised learning algorithms is made for topic models so as to find out the unrevealed topics within every single label, as well as unlabeled latent topics. A semi-supervised hierarchical topic model (SSHLDA) has been proposed in [13], where the newer topics are spontaneously expected to get explored.

3 TF-IDF for Text Classification

In the process of data retrieval, term frequency–inverse document frequency (TF-IDF) is a numerical approach to indicate the importance of any word in a document present in a corpus [14]. A rise in the TF-IDF value corresponding to a specific word is observed with the increase in the frequency of that word in the text document and is offset by the word frequency of the corpus. This adjusts the fact that the frequency of few words in a document is much higher compared to others.

On account of the term recurrence $tf(t, d)$, the most commonly utilized approach is to utilize the raw count of a term t in a document d , i.e., the number of occurrences of the term t in the given document d . If the raw count is denoted as $f_{t,d}$, then the most straightforward tf scheme is $tf(t, d) = f_{t,d}$. The inverse document frequency or IDF

provides a measure of the amount of information retrievable from a specific word in a document, regardless of whether the term is rare or common over all documents. It is the logarithmically scaled inverse fraction of the records containing the term, acquired by dividing the total quantity of reports by the quantity of reports containing the term, and afterward taking the logarithm of the quotient thus obtained.

$$idf(t, D) = \log[N \div |\{d \in D : t \in d\}|] \quad (1)$$

where N denotes the total quantity of documents in the corpus, $N = |D|$ and $|\{d \in D : t \in d\}|$ refers to the quantity of documents where the term t is present which means $tf(t, d) \neq 0$. If the term is missing from the corpus, the consequence will be a division by zero. Thus, the denominator is modified as $1 + |\{d \in D : t \in d\}|$.

Then, TF-IDF is computed as

$$\text{TF-IDF}(t, d, D) = \text{TF}(t, d) * \text{IDF}(t, D) \quad (2)$$

4 Latent Semantic Analysis (LSA)

To discover a topic in a gathering of textual files and documents, different techniques are utilized. Topic modeling algorithms, for the most part, are used to build up a model for browsing, searching and outlining extensive corpus of writings. Issue to consider being given a huge arrangement of messages, news articles, diary papers, and reports are to comprehension of the key data contained in set of records. The ultimate objective is to understand the key information contained in a huge corpus of news articles, emails, journal papers, etc. and keeping aside the unnecessary detailing. To extricate topics from huge corpus, various generative models are utilized for topic modeling, of which LSA is a frequently used one.

In natural language processing, specifically distributional semantics, Latent Semantic Analysis (LSA) is a way of examining the connections between an arrangement of documents and the terms contained by them by creating an arrangement of ideas associated with the records and terms. In case of LSA, it is assumed that words those are near in their meaning appear in related pieces of texts (the distributional hypothesis). A matrix that keeps record of the word count in each paragraph (each of the paragraphs is represented by the columns and the rows represent the unique words) is built from a vast section of text and in order to minimize the row count, a mathematical approach called singular value decomposition or SVD (in which a term-by-document matrix, say X , is decomposed into three other matrices W , S and P , such that when multiplied together, they give back the matrix X with $\{X\} = \{W\} \{S\} \{P\}$). Refer to Fig. 1) is implemented, while conserving the comparability structure among columns.

Then, comparison between the words is made by taking into consideration the cosine of the angles between any two vectors formed by rows (or by taking the dot

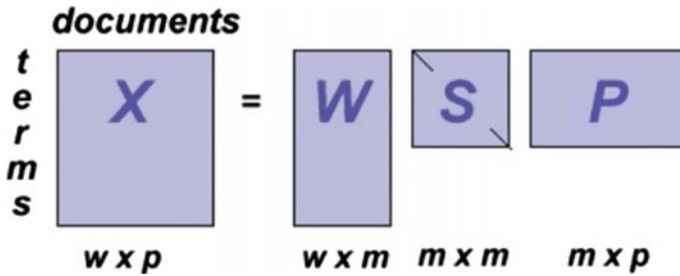


Fig. 1 Solitary Value Decomposition (SVD) of Latent Semantic Analysis (LSA)

product of the normalizations of the two vectors). If the value is near about 0, it means that the words are dissimilar, whereas if the value is near about 1, it means they are very much similar.

5 Entropy of Text Document

Specific conditions are required to be met for legitimate estimation of entropy of which ergodicity and stationarity are typical. Stationarity implies the fact that the statistical properties of a collection of words are totally independent of the position of the words in the text sequence, whereas ergodicity implies that the average properties of the troupe of all conceivable text sequences are matched with the statistical properties of an adequately large text series (e.g., [15–17]). Fitness is another such condition (i.e., a corpus of text documents has a limited arrangement of word types). While, in research articles, ergodicity and stationarity have a tendency to be exhibited and examined together, fitness is generally expressed independently, ordinarily as a component of the underlying setting of the issue (e.g., [15, 17]). A fourth general condition is the supposition that types are independent and identically distributed (i.i.d.) [15, 17, 18]. Consequently, for legitimate entropy estimation the normal pre-requisites are either (a) ergodicity, stationarity and fitness, or (b) just fitness and i.i.d., as ergodicity and stationarity take after inconsequentially from the i.i.d. supposition.

For entropy estimation, a vital pre-requisite is the guess of the probabilities of word types. In a text document, each word of the types w_i has a token frequency given by $f_i = \text{freq}(w_i)$. The probability of a word w_i , say $p(w_i)$ can be estimated by the so-called maximum likelihood method as

$$\hat{p}(w_i) = \frac{f_i}{\sum_{j=1}^V f_j} \tag{3}$$

where the numerator denotes the frequency of the word type w_i taken into consideration (denoted by f_i) while the denominator denotes the summation of the frequencies

of each word type over an empirical vocabulary having size of $V = |\mathcal{V}|$. This gives the probability of a word type w_i .

It is assumed that a text is an arbitrary variable T , which is formed by method of drawing and concatenating tokens from a vocabulary of word types $\mathcal{V} = \{w_1, w_2, \dots, w_W\}$, where the letter W represents the theoretical size of the vocabulary set and for every $w \in \mathcal{V}$, a probability mass function is obtained by $p(w) = Pr\{T = w\}$. From all these, the theoretical entropy T can be computed using the following relation [19]:

$$H(T) = - \sum_{i=1}^W p(w_i) \cdot \log_2 p(w_i) \quad (4)$$

where the term $p(w_i)$ presents the probability of the word type w_i , such that the summation of the probabilities of every word type is 1, i.e., $\sum_{i=1}^W p(w_i) = 1$. As seen from Eq. (4), each of the probability term is being multiplied with the logarithmic term $\log_2 p(w_i)$ according to Shannon's Entropy Equation. The resultant outcome that is obtained is negative and that is why the preceding ($-$) sign is used so as to make the final outcome positive. For this situation, $H(T)$ can be viewed as the average data substance of word types. A pivotal walk toward assessing $H(T)$ is to dependably estimated the probabilities of word types $p(w_i)$.

6 Proposed Method

6.1 Text Classification Using TF-IDF Versus Text Classification Using Topic Modeling

TF-IDF can be utilized as attributes in a supervised learning setting (i.e., depicting the data/information of a word in a record relating to some suitable outcome) whereas topic modeling is generally an unsupervised learning problem (basically endeavoring to comprehend the topics of a corpus). One noteworthy contrast is that TF-IDF is at the word level, so a textual report that is about *car* may be afar from a report about *tire* when TF-IDF is utilized to represent them. On the contrary, since the words *car* and *tire* very often appear simultaneously in articles, they are likely to arrive from the same topic and as a result the topic modeling portrayal of these reports would be close. In case of TF-IDF, though the matrix has number of rows equal to the number of documents in the corpus set and number of columns equal to the size of the vocabulary set, most of the cells in the matrix are empty. In our case, the sparsity i.e., percentage of non-zero cells is around 2.37%. That means remaining 97.63% of the cells is zero, i.e., contains no information. Whereas in case of topic modeling (say LSA), the number of columns is drastically reduced (the number of columns is equal to the number of categories or topics in the corpus set). So, from all

these, we can infer that topic modeling is supposed to give better text classification accuracy, compared to TF-IDF. We made use of Latent Semantic Analysis (LSA) in our experiment.

6.2 *Human Intervention in Deciding Categories of Document Having Very High Entropy Value*

The root of the word entropy is in the Greek *entropia*, which signifies “a moving in the direction of” or “change.” Up in the year 1868, the word was utilized to depict the estimation of disorder by the German physicist Rudolph Clausius. It basically refers to a numerical measure of the uncertainty of an outcome. Sometimes, it happens that a single textual document has almost equal inclination to more than one topic categories because the document contains words from different topics in almost equal proportions.

For example, suppose an illegal or criminal activity takes place in the educational sector. Now, a textual document related to this contains keywords related to both “Crime” and “Educational Sector” and sometimes it becomes difficult for machines to identify the most appropriate category and ends up giving unexpected outcomes. Such documents generally have a very high textual entropy value and hence, if the textual entropy of a specific document is beyond a certain threshold value (in our experiment, the threshold value is taken as 3.80), then human assistance is asked by the machine to help it decide the most appropriate category for the document. And moreover, text classification is very subjective, very difficult to match the classification done by human beings with that of the models. However, getting these tremendous volumes of textual data classified by humans is not possible. Here, we only route those observations which have a high uncertainty associated with the decision.

For instance, let us examine one such text document from the dataset of this experiment “*M. Aswini of Madhuravoyal, a first year B.Com student at the Meenakshi Academy of Higher Education and Research was brutally murdered near her college in K.K. Nagar on Friday. According to the police report, the murderer had been stalking Aswini, asking her to marry him.*” Here, we can see that the document has words belonging to Topic 9 (Education Sector) like “*B.Com,*” “*student,*” “*Education,*” “*Research,*” “*college,*” etc. and at the same time has words like “*brutally,*” “*murdered,*” “*police,*” “*murderer,*” etc. that fit in Topic 0 (Crime). In spite of this being a clear report of crime (Topic 0), in order to describe the identity and whereabouts of the victim, references to education sector (Topic 9) is made. But for machine, it is rather hard to identify the most appropriate category and ended up giving unexpected outcome. When this very document was tested with only LSA model, 4 out of 10 times the outcome says that it belongs to Topic 9 and 6 out of 10 times the outcome says that it belongs to Topic 0. And when the entropy of the document is computed it is found to be 3.85 which is quite high indicating that the document has

high uncertainty. Hence, under such circumstances, the best way to classify the document under the most appropriate category is to ask for human decision. And using human intervention for documents having very high entropy improves the accuracy as every time such ambiguity is faced, human decision is taken into account and the category decided by human beings are taken as the ultimate category.

6.3 Proposed Algorithm

Input: Train Data and test Data

Output: Labels of Test Data

Step 1: Represent the documents in terms of topic coverage distribution after applying LSA.
 Step 2: The dataset is being split into test and train set and the model is trained using the train set data consisting of 80% of the data elements.
 Step 3: For a test set textual document, predict label, say 'P' (the dominant topic).
 Step 4: For the test set textual document, compute the entropy.
 Step 5: a) If the entropy value is beyond a threshold value then human intervention is asked by the machine to help it deciding the most appropriate category.
 b) Else, no human intervention is required and the predicted label 'P' (as obtained in Step 3) is taken into consideration.

The proposed algorithm is diagrammatically depicted in Fig. 2.

7 Dataset and Coding Environment

The dataset that is being used in this experiment is a self-made dataset of 870 news articles belonging to 10 different news categories (accidents, business, crime, education sector, entertainment, health and medicine, politics, science and technology, sports, travel and tourism). The corpus set consists of news reports of various different fields from across India, (news sources—The Times of India, The Indian Express and www.mid-day.com). The number of documents under each category is taken in varying proportions so as to study the effect of the number of documents in determining the accuracy in text classification. For example, under the category “Crime” we have taken 128 sample documents in our dataset, whereas the category “Business” has only 65 sample documents. All other categories contain number of documents in between 65 and 128. Statistical representation of the dataset, with different topics and their distribution, is depicted in Fig. 3.

The code has been generated using Python programming language. Various libraries of Python have found applications in this code, including *scikit-learn* and its

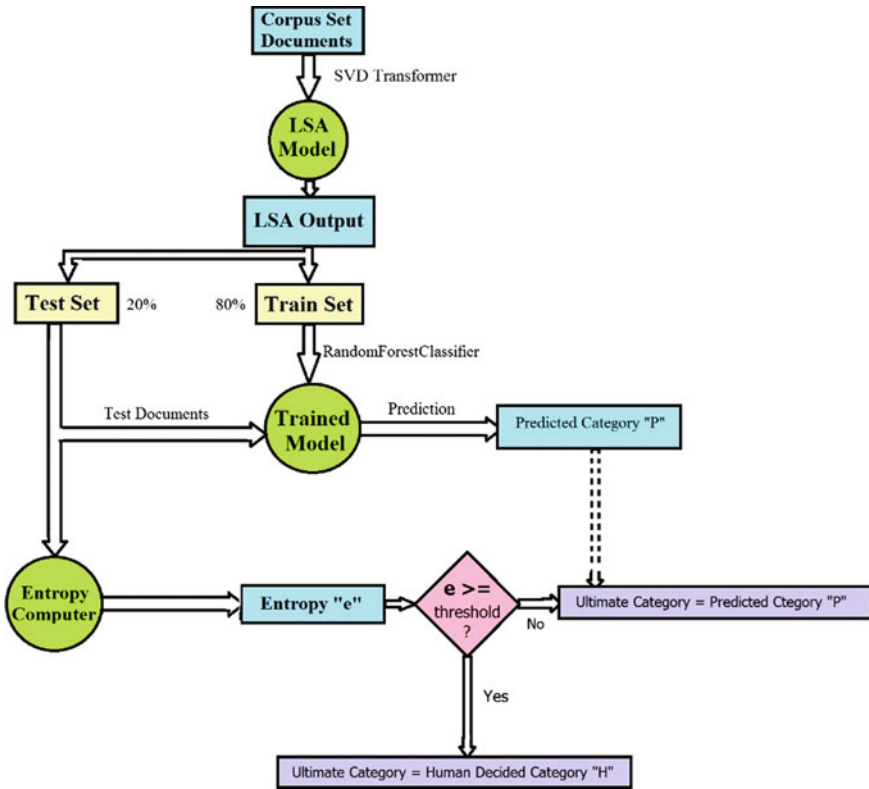
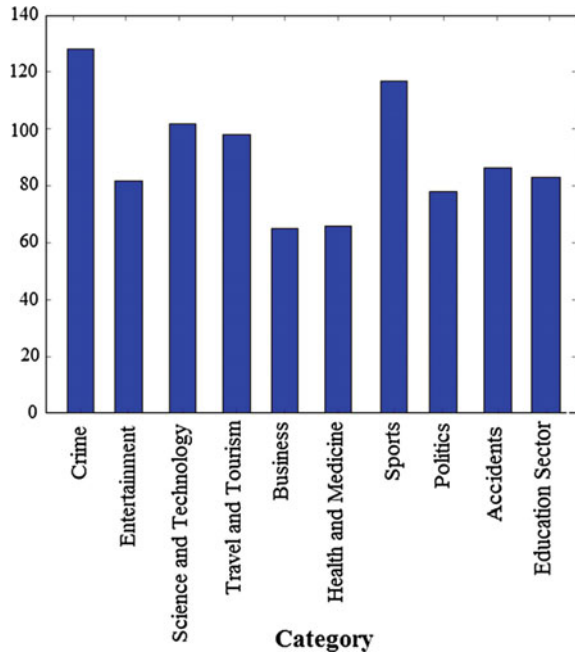


Fig. 2 Diagrammatic representation of the algorithm being proposed in this paper

various sub-libraries, *numpy*, *pandas*, *matplotlib.pyplot*, *re*, *spacy*, *keras*, *seaborn*, etc. Random Forest Classifier is the classification algorithm that has been used for this experiment for training the model based on the training set (80%) and the accuracy is then checked on the test set (20%). Random Forest Classifier helps to conquer numerous problems associated with many other classification algorithms (like that of decision trees), including Reduction in over-fitting (there is an impressively lower risk of over-fitting due to the reason that average of several trees are being taken into consideration) and less variance (By utilizing numerous trees, the likelihood of stumbling across a classifier that performs poorly is minimized due to the connection between the test and the train data.). Besides, it also has high accuracy and runs efficiently on large data sets.

Fig. 3 Statistical representation of the dataset, with different topics and their distribution



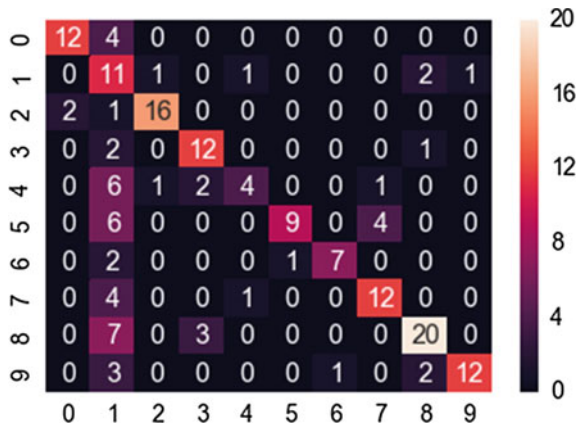
8 Experimental Result

The proposed algorithm is given Sect. 6.3. At the beginning, the topic distribution (in percentage) of each document is presented in the form of a matrix using the Latent Semantic Analysis. After training the classification model (Random Forest Classifier) with 80% of the sample data, a prediction is made for the test set (20%) and at the same time, entropy of the text document is also computed. If the entropy value is beyond a threshold value, then human intervention is asked by the machine to help it deciding the most appropriate category. Else, the predicted category is taken as the outcome. Accuracy is calculated by dividing the number of documents correctly predicted by the model on the test set by the actual number of documents in that very category in the test set. The average accuracy obtained in case of TF-IDF text classification is around 65.4%. Whereas using Latent Semantic Analysis (LSA), the accuracy shows a drastic improvement (around 78.8%), which is further enhanced implementing the concept of entropy. When the entropy of the text is greater than or equal to 3.8, then based on human decision, the categories are chosen for better accuracy and surety. This gives an accuracy of around 87.6%. The comparison of the topic wise accuracies in all the three cases is depicted in Table 1. From Table 1, we can see that the topic wise, as well as the average accuracy in case of LSA is more as compared to TF-IDF, which is even increased implementing the concept of entropy. From Table 1, we can also see that Topic 0 having 128 sample documents has higher accuracy compared to that of Topic 4 which has only 65 sample documents. This

Table 1 Accuracy of text classification using TF-IDF, LSA and LSA with human decision based on entropy value

Topics	Accuracy		
	TF-IDF	LSA	LSA (with human decision based on entropy value)
Topic 0	0.75	0.93	1.00
Topic 1	0.68	0.75	0.81
Topic 2	0.84	0.89	0.95
Topic 3	0.80	0.86	0.93
Topic 4	0.28	0.5	0.64
Topic 5	0.47	0.68	0.79
Topic 6	0.70	0.9	1.00
Topic 7	0.70	0.82	0.88
Topic 8	0.66	0.83	0.87
Topic 9	0.66	0.72	0.89
Average	0.654	0.788	0.876

Fig. 4 Confusion matrix of text classification using TF-IDF



also gives a hint toward the fact that increasing the number of documents leads to accuracy gains for LSA models. The confusion matrices of text classification using term frequency–inverse document frequency (TF-IDF), Latent Semantic Analysis (LSA) and that of LSA with human decision based on entropy value are shown in Figs. 4 and 5a, b, respectively.

9 Conclusion

The proposed method suggested in this paper presented the utilization of entropy in enhancing the accuracy of text classification. In this paper, Latent Semantic Anal-

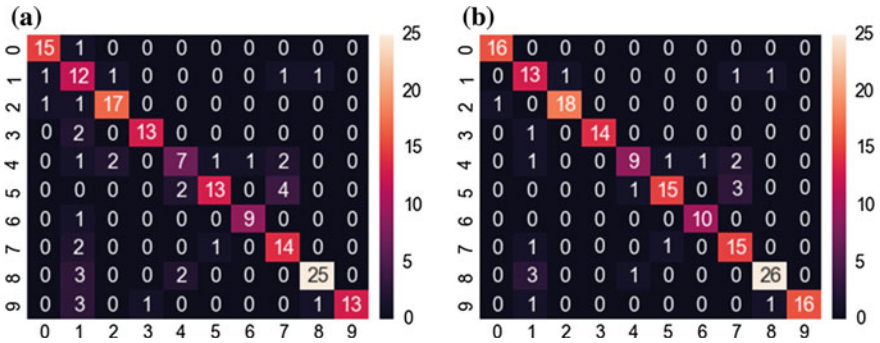


Fig. 5 **a** Confusion matrix of text classification using LSA, **b** Confusion matrix of text classification using LSA with human decision based on entropy value (the proposed method)

ysis has been utilized for text classification, which uses a much lower-dimensional representation of documents and words than usual TF-IDF. Though the number of rows in case of both TF-IDF and LSA remains the same, being the number of textual documents in the corpus, the number of columns is huge in case of TF-IDF (equal to the size of the vocabulary set), whereas in case of LSA it is equal to only the number of topics/categories. Additionally, in spite of having so many columns, most of the cells in case of TF-IDF are empty; in case of our dataset, the sparsity is only 2.37%, remaining 97.63% cells are empty (for TF-IDF). And moreover, text classification is very subjective, very difficult to attain the same accuracy for models as that of the classification done by human beings. However, getting these tremendous volumes of textual data classified by humans is also impractical. Here, only those documents are sent for human assessment which has a high uncertainty associated with the machine’s decision, thus boosting up the accuracy in text classification.

Acknowledgements We gratefully acknowledge the contribution of MUST Research Club and CU Data Science Group. MUST Research Club is a non-profit organization registered under Society Act of India. MUST Research Club is dedicated to promote excellence and competence in the field of data science, cognitive computing, artificial intelligence, machine learning and advanced analytics for the benefit of the society. Calcutta University Data Science Group is a group to create solutions to solve societal problems, as well as form generic solutions to common data science and engineering issue. This is a forum which brings together researchers, industry practitioners, scholars, interns and form groups.

References

1. S. Zelikovitz, Transductive LSI for short text classification problems, in *Proceedings of the 17th International Flairs Conference* (2004)
2. Q. Pu, G. Yang, Short-text classification based on ICA and LSA, in *Advances in Neural Networks ISNN 2006* (2006), pp. 265–270

3. S. Zelikovitz, H. Hirsh, Using LSI for text classification in the presence of background text, in *Proceedings of 10th International Conference on Information and Knowledge Management* (2001), pp. 113–118
4. B. Wang, Y. Huang, W. Yang, X. Li, Short text classification based on strong feature thesaurus. *J. Zhejiang Univ. Sci. C* **13**(9), 649–659 (2012)
5. X. Phan, L. Nguyen, S. Horiguchi, Learning to classify short and sparse text & web with hidden topics from large-scale data collections, in *IW3C2* (2008)
6. M. Sahami, T.D. Heilman, A web-based kernel function for measuring the similarity of short text snippets, in *WWW '06: Proceedings of the 15th International Conference on World Wide Web* (2006), pp. 377–386
7. W.-T. Yih, C. Meek, Improving similarity measures for short segments of text, in *AAAI'07: Proceedings of the 22nd National Conference on Artificial Intelligence* (2007), pp. 1489–1494
8. D. Ramage, D. Hall, R. Nallapati, C.D. Manning, Labeled LDA: a supervised topic model for credit attribution in multi-labeled corpora, in *EMNLP '09: Proceedings of the Conference on Empirical Methods in Natural Language Processing* (Association for Computational Linguistics, 2009), pp. 248–256
9. D. Ramage, D. Hall, R. Nallapati, C.D. Manning, Labeled LDA: a supervised topic model for credit attribution in multilabeled corpora, in *Proceeding of the 2009 Conference on Empirical Methods in Natural Language Processing* (2009)
10. S. Dey Sarkar, S. Goswami, A. Agarwal, J. Aktar, A novel feature selection technique for text classification using Naïve Bayes. *Int. Sch. Res. Not.* **2014**, Article ID 717092. <https://doi.org/10.1155/2014/717092>
11. G. Maskeri, S. Sarkar, K. Heafield, *Mining Business Topics in Source Code using Latent Dirichlet Allocation* (ACM, 2008)
12. D. Ramage, C.D. Manning, S. Dumais, Partially labeled topic models for interpretable text mining, San Diego, California, USA (2011)
13. D. Ramage, E. Rosen, Stanford Topic modelling Toolbox, Dec 2011. [Online]. Available: <http://nlp.stanford.edu/software/tmt/tmt-0.4>
14. A. Rajaraman, J.D. Ullman, *Data Mining. Mining of Massive Datasets* (PDF) pp. 1–17 (2011). <https://doi.org/10.1017/cbo9781139058452.002>. ISBN 978-1-139-05845-2
15. Ł. Dębowski, Consistency of the plug-in estimator of the entropy rate for ergodic processes, in *Proceedings of the 2016 IEEE International Symposium on Information Theory (ISIT)*, Barcelona, Spain 10–15 July 2016, pp. 1651–1655
16. J. Jiao, K. Venkat, Y. Han, T. Weissman, Minimax estimation of functionals of discrete distributions. *IEEE Trans. Inf. Theory* **61**, 2835–2885 (2015)
17. A. Lesne, J.L. Blanc, L. Pezard, Entropy estimation of very short symbolic sequences. *Phys. Rev. E* **79**, 046208 (2009)
18. G.P. Basharin, On a statistical estimate for the entropy of a sequence of independent random variables. *Theory Probab. Appl.* **4**, 333–336 (1959)
19. C.E. Shannon, W. Weaver, *The Mathematical Theory of Communication* (The University of Illinois Press, Urbana, IL, 1949)

Low Frequency Noise Analysis in Strained-Si Devices



Sanghamitra Das, Tara Prasanna Dash and Chinmay Kumar Maiti

Abstract In this paper, the low frequency ($1/f$) noise (LFN) in strained-Si p-MOSFETs has been modelled by modifying the Hooge's parameter and implemented using TCAD tools. The $1/f$ drain current noise characteristics have been investigated with the variation of strain and gate bias and higher noise level is observed in the strained-Si devices. The results obtained from simulation closely matches with reported experimental data.

Keywords Strained-Si · Low frequency noise · Hooge's parameter

1 Introduction

Strain engineering has proven to be a promising performance booster over the conventional CMOS beyond 90 nm technology nodes due to their superior physical and electrical characteristics in microelectronic devices [1]. In a biaxial tensile strained-Si layer over a thick relaxed $\text{Si}_{1-x}\text{Ge}_x$ substrate, carrier mobility improvements can be achieved due to band offsets occurred in their both valence and conduction bands [2, 3]. But, the interfaces involved in these heterostructure MOSFETs come with additional traps and defects which may generate unwanted effects such as higher low frequency noise ($1/f$ noise) in the device.

Due to higher carrier mobilities, strained-Si (SS) MOSFETs are in great demand for their use in microwave and RF communication circuits [4]. However, LFN affects

S. Das · T. P. Dash (✉) · C. K. Maiti
Department of Electronics and Communication Engineering,
Siksha 'O' Anusandhan (Deemed to be University),
Bhubaneswar 751030, Odisha, India
e-mail: taradash@soa.ac.in

S. Das
e-mail: sanghmitra.das01@gmail.com

C. K. Maiti
e-mail: ckmaiti@soa.ac.in

the performance of digital, analogue, RF and mixed-signal circuits in a large scale [5]. The phase of the signal changes due to addition of noise in these circuits and thus degrades the performance [6]. As the implementation of strain engineering is gradually increasing in advanced CMOS fabrication, it is highly essential to investigate the $1/f$ noise characteristics in strained-Si devices [7].

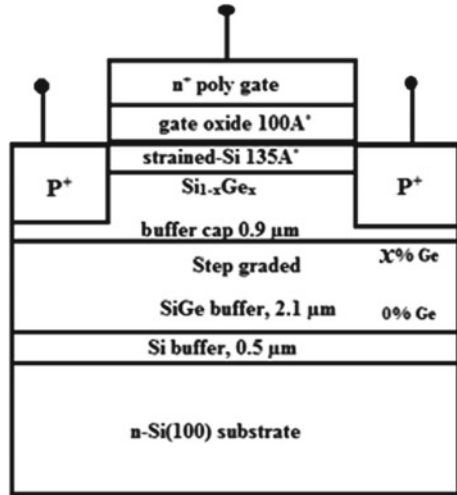
Several researchers have reported the LFN performance mostly for SS n-MOS and SiGe p-MOS devices [8–13]. Higher LFN and oxide trap density have been found for SS n-MOSFETs [8, 13]. The low frequency noise study of SS n-channel along with p-channel devices have been investigated by Haartman et al. [9] on thick and thin relaxed SiGe substrates. Similar study has also been performed for SiGe channel p-MOSFETs. Haartman et al. [10] has presented the $1/f$ noise in SiGe channel p-MOSFETs with a Si cap layer on it and reported reduced LFN levels in the SiGe devices than the bulk-Si channel device. The physical separation of the carriers from the SiO₂/Si interface is the main cause of lower noise levels in these devices. Improved low frequency or Flicker noise performance of 100 nm strained Si_{0.5}Ge_{0.5} channel MOSFETs has also been shown by Li et al. [11]. But these devices act as buried channel devices because of presence of the Si cap layer whereas improved electrical performance is also possible with surface channel devices. Although LFN performance for SS n-MOSFETs and SiGe p-MOSFETs have been rigorously studied by various researchers, a very few literatures are available on the noise behaviour of short channel SS p-MOS devices. Hence, the $1/f$ noise characteristics of a SS p-MOSFET have been focused here exclusively.

In this work, we present the simulation study of low frequency noise performance of a SS p-MOSFET having 500 nm gate length. The standard Hooge's model for $1/f$ noise has been modified as a function of Ge fraction to be implemented for the strained-Si MOSFETs. The drain current noise characteristics of the device have been studied as a function of frequency and gate bias for different levels of strain. There is good agreement between simulation result and the experimental data. The simulations have been performed by using the device simulator ATLAS from SILVACO [14]. Also the enhancement in electrical properties of the strained-Si device has been shown with increase in strain.

2 Modelling of Device Parameters

A 500 nm long biaxially tensile SS p-MOSFET with width (10 μm) has been considered for simulation. The schematic diagram of the device is shown in Fig. 1. The device structure is similar to the experimental device [15] which shows 40% hole mobility enhancement at 300 K. The source and drain regions have been uniformly doped with Boron of concentration $1 \times 10^{20} \text{ cm}^{-3}$ with n-type doping level of $6 \times 10^{16} \text{ cm}^{-3}$ in the bulk regions. Simulation was done above 50 Å, so that quantum effect can be neglected.

Fig. 1 Structure of the SS p-MOSFET considered for simulation [15]



The noise characteristics of MOSFETs can be defined by the widely accepted Hooge’s model [16] according to which, the noise spectral density is given as:

$$S_{1/f} = \frac{\alpha_H J^2}{f N} \tag{1}$$

- α_H is the Hooge’s constant,
- J represents the current density and
- N represents the number of free charge carriers in the bulk semiconductor.

As Hooge’s model is followed as a standard model for the $1/f$ noise performance study in most of the silicon devices, it can be modified and suitable fitting parameters can be chosen for strained-Si device for simulation purpose. Here, the Hooge’s parameter α_H has been modified by a quadratic expression considering the variation of strain (Ge content) which has been derived following Ref. [17] as discussed in the last paragraph. The modified Hooge’s parameter as a function of Ge content is given as [18]:

$$\alpha_{H,SS} = \alpha_{H,Si}(5.12 - 43.2x + 140x^2) \tag{3}$$

where $\alpha_H = 2 \times 10^{-3}$ and x is the Ge content in the relaxed $Si_{1-x}Ge_x$ substrate.

The above equation for $\alpha_{H,SS}$ has been implemented in Eq. (1) for the low frequency noise simulation of SS p-MOSFETs. The percentage of Ge in the relaxed $Si_{1-x}Ge_x$ buffer layer has been varied to investigate the low frequency noise performance of the device at different levels of strain.

3 Results and Discussions

3.1 Low Frequency Noise Variation with Strain

To investigate the effect of strain on the LFN performance, the Ge content in the relaxed SiGe buffer layer has been varied. The variation of low frequency ($1/f$) noise for different Ge contents (0.20, 0.25, 0.30) with frequency is shown in Fig. 2. Also the noise for the Si channel p-MOSFET device is shown for comparison.

The $1/f$ noise behaviour of the drain current noise is as per the expectation. Higher level of LFN is observed in case of the SS devices in comparison to the bulk-Si devices. Also the noise level increases with increase in Ge fraction in the relaxed SiGe substrate. This is attributed to the increase in the defect or trap density (D_{it}) at the interface of the strained-Si channel and the oxide layer with increase in Ge content than that of the bulk-Si channel MOSFETs as mentioned clearly in Sect. 2. For 30% Ge, the drain current noise is almost one order higher than that of the bulk-Si device. There are experimental results showing higher low frequency

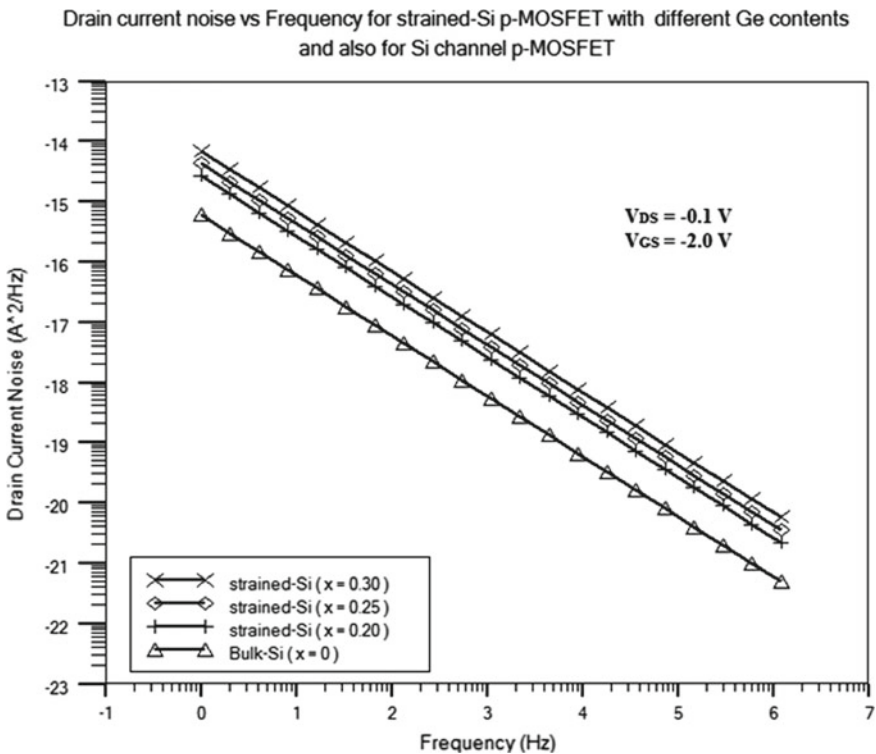
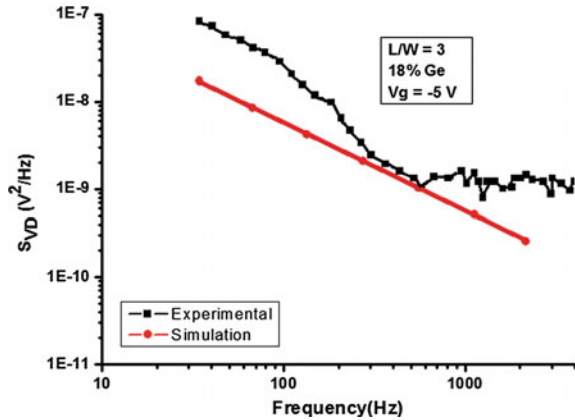


Fig. 2 Simulated drain current noise of SS p-MOSFETs for Ge percentage (x) = 0.20, 0.25, 0.30 compared with bulk Si

Fig. 3 Comparison of simulated drain current noise with the experimental reported data



noise for strained-Si n-MOSFET devices than the bulk-Si MOSFETs [19]. Hence, these simulation results can be useful for predicting the impact of strain on the low frequency noise behaviour of p-MOSFETs.

In order to compare the low frequency noise simulation results with the reported experimental results [20], a strained-Si device with $L/W = 3$ and 18% Ge content has been simulated. Figure 3 shows the drain voltage noise comparison of the simulated device with the above experimental data at $V_g = -5$ V. As it can be seen, the simulated noise in the device is in good agreement with the experimental result.

3.2 Effect of Vertical Field on Low Frequency Noise

The low frequency noise level also changes with the gate bias. The variation of the $1/f$ noise at different gate bias for a constant drain voltage of -0.1 V for 20% Ge is shown in Fig. 4. The result shows that the low frequency drain current noise increases with gate voltage. The increase in electric field across the channel at high gate bias causes velocity saturation of the carriers because of the scattering due to oxide traps, interface defects and surface roughness which causes mobility degradation at high field values. The increased fluctuation in mobility of the carriers at higher gate voltages due to these scattering leads to higher Flicker noise values. The result shown below also justifies the fact that fluctuation in mobility is a major cause of LFN or $1/f$ noise in MOSFETs [21] which is described by the Hooge's model (Eq. 1).

The drain current noise as a function of gate voltage for the strained-Si p-MOSFETs with different Ge contents of 20, 25 and 30% at a low frequency value of 30 Hz is shown in Fig. 5. Also the noise in Si channel MOSFET is shown for comparison.

From the above figure, we note that the $1/f$ noise increases with strain or Ge content for all the gate voltages. As described before (in Sect. 2), the low frequency

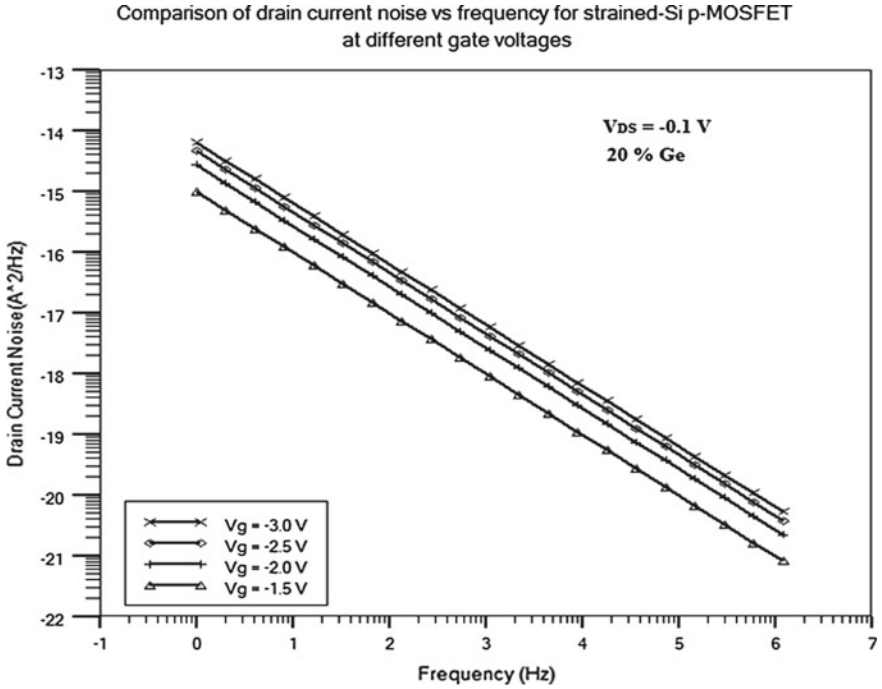


Fig. 4 Simulated drain current noise of a SS p-MOS device at different gate voltages for Ge percentage (x) = 0.20

drain current noise in the MOSFETs is primarily dependent on two parameters which are the fluctuation in the number of charge carriers due to traps and the fluctuation in the mobility of the carriers. The increase in the number of trap sites at the SiO₂/Si interface with increase in the Ge fraction of the device causes the enhancement in the noise level with strain. However, the noise level also shows an increasing trend with increase in the gate bias which was also shown in Fig. 4. At lower gate voltages, the carrier mobility is relatively higher and hence the source of noise is mainly the fluctuation in carrier number due to traps and defects. However, at higher gate voltages, the mobility fluctuation becomes the dominant source of low frequency noise. From Eq. (1), it is evident that:

$$s_{\frac{1}{f}} \propto I_d^2 \tag{4}$$

As the biasing has been kept in the linear region,

$$I_d^2 \propto [2(V_{GS} - V_T)V_{DS} - V_{DS}^2]^2 \tag{5}$$

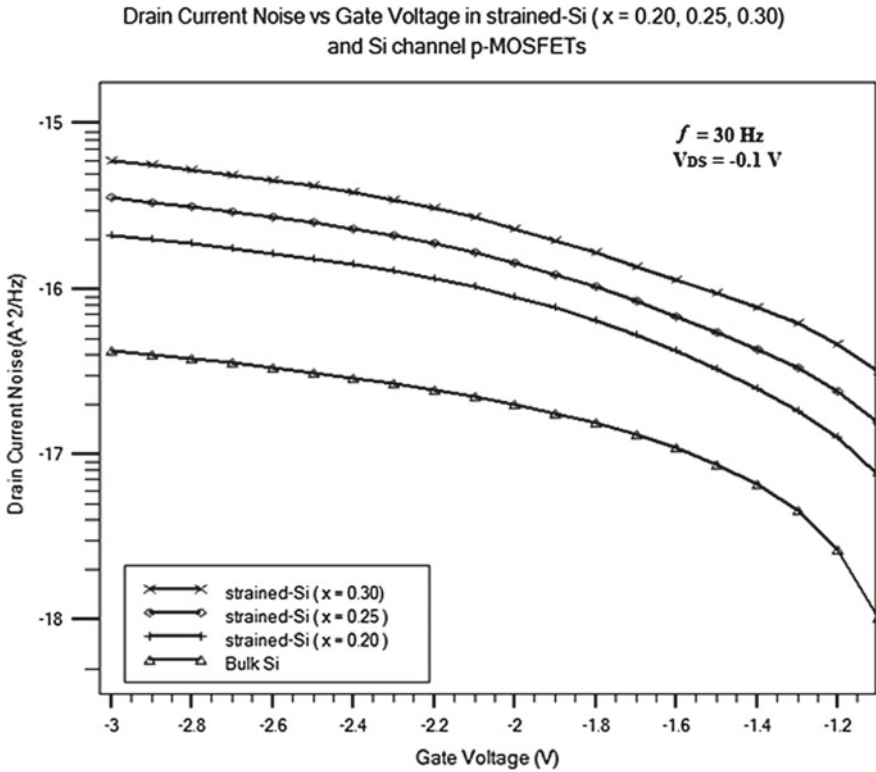


Fig. 5 Simulated drain current noise versus gate voltage for Si channel and strained-Si channel p-MOSFETs with variation of Ge percentage at frequency of 30 Hz

Hence, for a constant drain voltage, the low frequency noise power follows a quadratic relationship with the gate overdrive. The nature of the simulated result shown above justifies this fact.

4 Conclusion

The LFN for SS p-MOSFETs has been modelled by modifying the Hooge parameter. The drain current noise has been investigated as a function of frequency and gate bias. The simulation results show increase in the noise level for the strained-Si devices with increase in Ge content. For 30% Ge, the noise level is one order higher than the Si channel device. With increase in the gate bias, the low frequency noise increases due to the high field effects. These simulation results can be useful for optimizing the effect of various device parameters on the low frequency noise performance of strained-Si p-MOSFETs.

References

1. D. Pradhan, S. Das, T.P. Dash, Study of strained-Si p channel MOSFETs with HfO₂ gate dielectric. *Superlattice Microstruct.* **98**, 203–207 (2016)
2. K. Rim, J. Chu, H. Chen, K.A. Jenkins, T. Kanarsky, K. Lee, A. Mocuta, H. Zhu, R. Roy, J. Newbury, J. Ott, K. Petrarca, P. Mooney, D. Lacey, S. Koester, K. Chan, D. Boyd, M. Jeong, H.-S. Wong, Characteristics and device design of sub-100 nm strained Si N- and P-MOSFETs, in *Symposium VLSI Technology Digital Technical Paper* (2002), pp. 98–99
3. S.I. Takagi, T. Mizuno, T. Tezuka, N. Sugiyama, S. Nakaharai, T. Numata, J. Koga, K. Uchida, Sub-band structure engineering for advanced CMOS channels. *Solid State Electron.* **49**, 684–694 (2005)
4. R.K. Nanda, T.P. Dash, S. Das, C.K. Maiti, Noise characterization of Silicon-Germanium HBTs, in *International Conference on Microwave, Optical and Communication Engineering* (2015), pp. 284–287
5. Y. Nemirovsky, I. Brouk, C.G. Jakobson, 1/f noise in CMOS transistors for analog applications. *IEEE Trans. Electron Devices* **48**, 921–927 (2001)
6. P.W. Li, W.M. Liao, Low-frequency noise analysis of Si/SiGe channel pMOSFETs. *Solid State Electron.* **46**, 2281–2285 (2002)
7. P.C. Huang, C.Y. Chang, O. Cheng, S.L. Wu, S.J. Chang, Temperature dependence of low-frequency noise characteristics in uniaxial tensile strained nMOSFETs. *Jpn. J. Appl. Phys.* **54**, 100301(1–4) (2015)
8. T. Contaret, T. Boutchacha, G. Ghibaudo, F. Boeuf, T. Skotnicki, Low frequency noise in biaxially strained silicon n-MOSFETs with ultrathin gate oxides. *Solid State Electron.* **51**, 633–637 (2007)
9. M. von Haartman, B.G. Malm, P.E. Hellstrom, M. Ostling, T.J. Grasby, T.E. Whall, E.H.C. Parker, K. Lyutovich, M. Oehme, E. Kasper, Impact of strain and channel orientation on the low-frequency noise performance of Si n- and pMOSFETs. *Solid State Electron.* **51**, 771–777 (2007)
10. M. Von Haartman, A.C. Lindgren, P.E. Hellstrom, B. Gunnar Malm, S.L. Zhang, M. Ostling, 1/f Noise in Si and Si_{0.7}Ge_{0.3} pMOSFETs. *IEEE Trans. Electron Devices* **50**, 2513–2519 (2003)
11. F. Li, S. Lee, Z. Fang, P. Majhi, Q. Zhang, S.K. Banerjee, S. Datta, Flicker-noise improvement in 100 nm Lg Si_{0.5}Ge_{0.5} strained quantum-well transistors using ultrathin Si cap layer. *IEEE Electron Device Lett.* **31**, 47–49 (2010)
12. E. Simoen, C. Claeys, Impact of strain-engineering on the low-frequency noise performance of advanced MOSFETs, in *International Conference on Solid-State and Integrated Circuit Technology Proceedings* (2006), pp. 120–123
13. M.H. Lee, P.S. Chen, W.C. Hua, C.Y. Yu, Y.T. Tseng, S. Maikap, Y.M. Hsu, C.W. Liu, S.C. Lu, W.Y. Hsieh, M.J. Tsai, Comprehensive low-frequency and RF noise characteristics in strained-Si NMOSFETs, in *International Electron Devices Meeting Technology Digital* (2003), pp. 69–72
14. Silvaco Inc., ATLAS User's Manual (2013)
15. D.K. Nayak, K. Goto, A. Yutani, J. Murota, Y. Shiraki, High-mobility strained-Si PMOSFET's. *IEEE Trans. Electron Devices* **43**, 1709–1716 (1996)
16. F.N. Hooge, 1/F noise sources. *IEEE Trans. Electron Devices* **41**, 1926–1935 (1994)
17. G.K. Dalapati, S. Chattopadhyay, K.S.K. Kwa, S.H. Olsen, Y.L. Tsang, R. Agaiby, A.G. O'Neill, P. Dobrosz, S.J. Bull, Impact of strained-Si thickness and Ge out-diffusion on gate oxide quality for strained-Si surface channel n-MOSFETs. *IEEE Electron Device Lett.* **53**, 1142–1152 (2006)
18. S. Das, T.P. Dash, C.K. Maiti, Effects of hot-carrier degradation on the low frequency noise in strained-Si p-MOSFETs. *Int. J. Nanoparticle* **10**, 58–76 (2018)
19. W.C. Hua, M.H. Lee, P.S. Chen, S. Maikap, C.W. Liu, K.M. Chen, Ge outdiffusion effect on flicker noise in strained-Si nMOSFETs. *IEEE Electron Device Lett.* **25**, 693–695 (2004)

20. C.K. Maiti, T.K. Maiti, *Strain engineered MOSFETs* (CRC Press, Taylor and Francis Group, 2013)
21. C. Mukherjee, C.K. Maiti, Stress-induced degradation and defect studies in strained-Si/SiGe MOSFETs, in *Physics of Semiconductor Devices* (2014), pp. 17–20

Employing Cross-genre Unstructured Texts to Extract Entities in Adapting Sister Domains



Promita Maitra and Dipankar Das

Abstract This paper presents a method to employ unstructured texts of different domains for extracting entities from the closely related domains. To investigate the domain closeness, we trained our model in one domain and test in rest of the domains. The second challenge lies in the fact that if we retrain our model using a pair of closely related domains, we achieve better results while testing on texts from mixed genre in contrast to those individual genres. Working on the idea of domain adaptation, we carried out experiments with single-domain and mixed-domain training data and observed the precision, recall, and f-score of our entity extraction system. The performance of our system was optimal in a semi-supervised dataset with closely related or sister domains.

Keywords Cross-genre entity extraction · Text analytics

1 Introduction

Entity identification, a globally acknowledged task in the field of text analytics, involves automatic discovery of single or multi-word proper nouns and further categorization of those into appropriate predefined classes or entity types.

Recent researches have advanced to address cross-domain, cross-lingual, and code-mixed entity detection challenges. Our present task at hand is focused on building a system that will extract entities from unstructured texts from domains of different genres. The datasets we are considering for our experiment are user-generated data for Web-based platforms like Twitter, blog, review and Wikipedia articles, news reports, and texts from contemporary literature as well as mythology-based texts. These texts talk about theme or domain completely different from each other, which

P. Maitra (✉) · D. Das
Jadavpur University, Kolkata, India
e-mail: promita.maitra@gmail.com

D. Das
e-mail: dipankar.dipnil05@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_38

makes it absolutely impossible to consider domain-specific entity types based on the genre of text. Working on two hypotheses based on the concept of domain adaptation, we carried out experiments with a single-domain and mixed-domain training data and including only the genre-independent features once and all features together in the second phase. We observe that our system achieved an overall better result with the mixed domain training module.

Section 2 talks about the related works carried out in this domain in brief. Then, we move on to describe the datasets we have used in Sect. 3. We describe our approach and features used in detail in Sect. 4, while Sect. 5 contains the experimental results and a brief discussion of them. Finally, we conclude the present work in Sect. 6.

2 Related Works

Named entity identification is a method for recognizing mostly proper nouns in text and associating them with appropriate predefined types. Such common types may include location, person name, date, and address. Some of these systems are incorporated into parts-of-speech (POS) taggers, though there are also many stand-alone applications. There are three main methods of learning NE, which are supervised learning, semi-supervised learning, and unsupervised learning. The main disadvantage in adopting a supervised learning approach is the requirement of a large annotated corpus. Limited availability of such resources in some of the domains or genres and the prohibitive cost of creating them make the other alternative learning methods more effective for domain adaptation techniques.

Several sequence labeling tasks in the field of text analysis have adopted different domain adaptation techniques. DaumIII and Marcu [1] suggested distinguishing between general domain-independent features and domain-dependent features. They empirically showed the effectiveness of their method on type classification, tagging, and recapitalization systems. Jiang and Zhai [2] worked on instance weighting method for semi-supervised domain adaptation. They assigned more weights to labeled source and target data, removing misleading training instances in the source domain, and extending target training instances with predicted labels. They empirically evaluated their method for cross-domain part-of-speech tagging and named entity recognition to evaluate the system's efficiency. DaumIII [3] later came up with an easy adaptation learning method (EA) by using feature replication, which was extended into a semi-supervised version (EA++) by incorporating unlabeled data through co-regularization [4]. A bootstrapping-based approach was proposed in 2011 for cross-domain named entity recognition system which showed 72.45 as best achieved f-score [5].

Compared to these systems, our present work deals with a large variety of texts with contrasting literary characteristics. Supervised domain adaptation for NER works when annotated data is available from both the source and the target domains; whereas training set with limited data from the target domain is called semi-

supervised approach. We have considered both in current work in first and second hypotheses, respectively.

3 Dataset Description

To carry on the task at hand successfully, we need to have unstructured textual data from many different genres of writing such as social media texts and mythological texts, etc., which in turn, can talk about multiple topics in a single dataset. The complexity of our case is twofold. First, it does not revolve around a particular theme or topic, and second, it does not deal with a regular structure of the text. We noticed that the social media texts of recent times can contribute some important insights as they are vastly irregular in structure, follow no standard grammatical rules, and incorporate various foreign words and symbols and emoticons. These unusual characteristics make it quite difficult to extract relevant information from these kinds of noisy texts. For our work, we have mainly considered data from three domains: Twitter, blog, and movie review. As each of them serves different purposes, the way they are written also changes significantly. On the other hand, the mythological texts that we are considering diverge manifold from the contemporary style of writing that we are acquainted with. Hence, we have considered both of them to observe the performance of our system in these two close genres of writing. In addition to these, a Wikipedia article corpus and one News corpus have also been included in our experiment as News and Wikipedia articles ideally follow a neutral reporting manner of writing without many regional or cultural influences. The sources of the data collected and brief descriptions of them are mentioned below.

Twitter: We participated in the shared task titled named entity recognition and linking (NEEL), organized by Microposts-2016 [6] workshop, which was a part of World Wide Web (WWW)-2016 conference and received the named entity annotated tweet dataset for training our model. The dataset consists of tweets extracted from a collection of over 18 million tweets. The gold standard was generated with the help of three annotators. It contains seven different entity types in total: PERSON, LOCATION, ORGANIZATION, THING, EVENT, CHARACTER, and PRODUCT. However, due to the ambiguity present among the entity types PERSON-CHARACTER and THING-PRODUCT, we decided to merge these pairs into two, PERSON and THING.

Blog: Texts found in blogs do not conform to the style of any particular genre per se varying from person to person and thus offers a variety in writing styles, choice, and combination of words, as well as topics. We used the dataset built by Aman and Szpakowicz for their emotion analysis task on blog posts (Aman and Szpakowicz 2007, 2008) excluding the emotions labels associated with that dataset, where the blog posts are collected directly from the Web.

Review: Pang and Lee built a sentiment annotated IMDb movie review corpus for their sentiment-based on subjectivity analysis task [7]. We used the 2004 release of their dataset named ‘polarity dataset v2.0’ made public in June, keeping only

the unprocessed movie review source files for our work and ignoring the sentiment annotations.

Wikipedia Articles: The types of user-generated documents in the Web we discussed so far follow a particular style of writing which gets affected heavily by various factors such as time, trend, region, culture, and language knowledge of user. But Wikipedia articles, on the contrary, maintain an impersonal flow of words, with a formal writing approach so that it can be understood by a larger number of people across the globe. Standard and non-complex sentence structure and moderate length of articles make it an unbiased, balanced dataset to work with. We use the NE annotated English dataset built and released by Nothman et al. in 2012 for their work of multilingual named entity recognition from Wikipedia [8]. The labels that they have used for entity annotation are PERSON, LOCATION, ORGANIZATION and MISCELLANEOUS. However, to maintain the uniformity of our approach, we disregarded the MISCELLANEOUS tag and considered the rest.

News: We collected the Press Reportage articles from the Brown Corpus¹ of Standard American English, compiled by W.N. Francis and H. Kucera, Brown University. The corpus consists of one million words of American English texts printed in 1961. The texts for the corpus were sampled from 15 different text categories to make the corpus a good standard reference.

Contemporary Literature: A text corpus collected from the training set of PAN-2015 authorship verification task [9] was used. It includes written documents from 100 different authors, two documents each. The dataset being cross-genre and cross-topic contains a large variety of entities for each type. In addition to these texts, we have collected a few famous short stories considered classics in English literature from the Web and included them in our experimental dataset as well.

Mythology: Mythological texts are very different from the kind of writing manner we are familiar with. Long sentences, unusual structure, usage of obsolete words, and complex phrases make it more difficult to analyze these texts using off-the-shelf tools or lexico-syntactic patterns used for general texts. We found an English version of great Indian epic Mahabharat translated from the original Sanskrit text by Kisari Mohan Ganguli and included the same in our dataset. A translated text was chosen over mythologies or folklores originally written in English because it has a diverse set of entities compared to what we have included in our dataset till now.

We had named entity annotations of only the Twitter and Wikipedia dataset. In order to evaluate our systems performance, a gold standard tagging of entities was needed for all the genres of text. Stanford CoreNLP,² being one of the best natural language processing tool, was used to tag named entities from all other documents blog, review, News, contemporary literature, and mythology belonging to three classes—PERSON, LOCATION, and ORGANIZATION. As the forms are different, the distribution of entities throughout the text is also irregular. As it is shown in Table 1, while the Twitter or Wikipedia dataset can be called an entity-rich corpus with more than 10% words as entities, datasets like blog, review, and literature

¹<http://icame.uib.no/brown/bcm.html>.

²<http://nlp.stanford.edu/software/CRF-NER.shtml>.

Table 1 Number of entities and non-entities in each domain

Genre	#Words	#Entities	#Non-entities
Blog	43,660	622	43,038
Review	51,175	974	50,201
Twitter	99,629	11,010	88,259
Wikipedia	64,730	7048	57,682
News	59,539	5672	53,867
Literature	64,977	1839	63,138
Mythology	66,387	4849	61,538

have only 2–3% of words as entities. This makes it difficult to consider one domain as training and a different one for evaluation, as the system will be heavily biased.

4 System Description

Although the number of digitized texts is rapidly increasing, as the language used in different domains varies so widely, collecting and curating training sets for each domain are prohibitively expensive. At the same time, the differences in vocabulary and writing style across domains can cause the state-of-the-art supervised models to dramatically increase the error. Domain adaptation methods provide a way to alleviate such problems by generalizing models from a resource-rich source domain to a different, resource-poor target domain. We have considered two scenarios in this context, a supervised, and a semi-supervised approach.

4.1 Feature Extraction Module

Experimental results by contemporary researchers showed that latent domain-independent features can increase the accuracy for out-of-domain prediction performance [10]. It has also been supported by a recent theoretic study that a proper feature representation is essential to domain adaptation for its contribution in reducing effective domain divergence [11, 12].

We have incorporated various orthographic, as well as contextual features in our approach to recognize named entities from unstructured texts. Present work deals with several datasets that vary enormously from one genre to other, both in terms of content and inscription style. We have experimented with various features, some of which are valid for all the domains, while some are helpful just for some specific domains. For example, features like ‘if the word is a hashtag’ will only work for tweets; for the other datasets, it is just an added overhead. Hence, our features can be

classified into two distinct categories: domain-independent and domain-dependent features.

4.1.1 Domain-Independent Features

- **Bidirectional POS-based features:** All the texts were POS tagged using Stanford CoreNLP parts-of-speech tagger during the preprocessing phase. We use three features, i.e., POS tag of the current word, previous word, and next word as features to identify whether or not the current word is a named entity.
- **Bidirectional N-gram-based features:** In the fields of computational linguistics, an n-gram is a contiguous sequence of n items from a given sequence of text or speech. The items can be phonemes, syllables, letters, words, or base pairs according to the application. For this specific task, we consider words as grams. An n-gram of size 1 is referred to as a ‘unigram’; size 2 is a ‘bigram’; size 3 is a ‘trigram.’ Larger sizes are sometimes referred to by the value of n , e.g., ‘four-gram,’ ‘five-gram,’ and so on. We only considered unigram, bigram, and trigram of current word in both the directions, i.e., previous and next bigram and trigram.
- **Uppercase-based feature:** This Boolean feature was kept to catch the abbreviated named entities like the USA or IBM. It checks if all the characters of the present word are in capital letters.
- **Lowercase-based feature:** Contrary to the previous one, this feature checks if all the characters of the current word are in lowercase. In such cases, the probability of the current word being a non-entity is high. This is also a Boolean feature like the previous one.
- **Mixed case-based feature:** This feature returns true if the current word is a combination of uppercase and lowercase characters.
- **Initial character case:** This feature notes in only the initial letter of the current word is capitalized. We can consider this as domain-independent because, for all the domains, it is quite a regular practice for named entities to have their 0th position character in capital-case.
- **Roman characters based feature:** This feature needs to be added because a few of the PERSON-typed named entities used by authors contain Roman symbols and/or digits.
- **Function word based binary feature:** This feature holds true if the POS tag of the current word is preposition or article.
- **isSingleChar:** This feature returns true if the current word contains only a single character. Most of the times named entities are more than a character long; however, in some cases, a single character can appear as an individual entity as well. *For example, \dot{T} he R Company is currently the biggest threat to this nation.∴ In this sentence, the letter R is a part of an ORGANIZATION-type named entity- ‘The R Company’*
- **Article or Noun Predecessor based feature:** To emphasize on the previous words POS, we check if the previous word is an article or if it is a noun. Researches show

that in these cases, these two features can play an important role in identifying the possible entities.

- **Word Length:** We keep the length of the current word as a feature.
- **Term Frequency:** This is basically the count of how many times a particular term/word has appeared in the total dataset. As the datasets vary in length, to keep this feature value unbiased, we normalize it by dividing with the total word count of the dataset.

4.1.2 Domain-Dependent Features

These domain-dependent feature sets are used along with the ones discussed above to identify the named entities in tweets, blogs, reviews, or other user-generated Web contents. It is a common practice among twitter or other social media users to use symbols or special characters like hashtags while generating content. To increase the recall of our system for these datasets, we have to include these features in our classifier. However, feature overfitting can lead to performance degradation in other domains.

- **Includes a Dot in the word:** This Boolean feature returns true if the word we are considering contains a dot. Usage of ‘dot’ is pretty common in case of account names or mail ids of social media users.
- **Includes Hyphen:** Similar to the previous one, this feature returns true if the word we are considering contains a hyphen.
- **URL based feature:** Presence of URLs is common in data generated for the Web, and hence, to address these values, we have introduced this feature which returns true if the current token is an URL.
- **isInitSymbol:** If the first character of the current word is a symbol, this feature returns true. We can call this a special feature for domains like twitter where usage of @ is common for profile names and ‘#’ for trending topics.
- **containsDigit:** This feature returns true if the current token contains digits in any position of the word.

4.2 Classification Module

After the feature extraction and feature file building phase, we sent our training files to a CRF-based classifier. Conditional random fields (CRF) are a class of statistical modeling method, often used for labeling or parsing of sequential data, such as natural language text or biological sequences. We have used CRF++ (version 0.58),³ a simple, customizable, and open-source implementation of conditional random fields (CRFs) for segmenting/labeling sequential data. It is designed for generic purpose and can

³<https://taku910.github.io/crfpp/#download>.

be applied to a variety of NLP tasks, such as named entity recognition, information extraction, and text chunking.

We have evaluated our system for domain-independent feature set once and then including all the features. However, the evaluation results did not appear to be satisfactory. Hence, an additional ‘tag rectifier’ module is introduced into the system after the classifier.

4.3 *Tag Rectifier Module*

The output from the classifier is fed to a noise cleansing module which works in two phases: First, it checks whether a stopword or a non-noun term has been tagged as an entity. If yes, our system negates the decision made by the classifier and tags it as a non-entity. This is done to increase the precision of the system by avoiding non-entities being tagged as entities.

In the second phase, a checking is done for all the noun words against our gazetteer entries. The role of the gazetteer is to identify entity names in the text based on lists of already known names of specific entity types. We have used the ANNIE Gazetteer lists that come with the GATE,⁴ which is open-source software used for text processing. Though there are default methods available in ANNIE for gazetteer matching, we developed our own matching algorithm against our own gazetteer list for each of the entity types by merging some of the lists from ANNIE.

For each word with a noun POS tag, we do the following:

- i. Check whether the word matches with names from any of the lists.
- ii. If no match is found, keep the classifier outcome as final and end the process. Otherwise, go to the next step.
- iii. If the word matches with only one list (i.e., a single entity type) and the word starts with a capitalized character, tag that particular word with the corresponding entity type of the list it was matched against, and end the process. Otherwise, go to the next step.
- iv. When the system finds a match from multiple lists (i.e., more than one entity type) for a single word, it relies on the classifier used in the previous step of entity extraction for entity type disambiguation and keeps the tag given by the classifier as the correct tag.
- v. End.

⁴<https://gate.ac.uk/>.

	Twitter	Wikipedia	Review	Blog	NEWS	Literature	Mythology
Twitter	39.26	30.63	19.58	15.01	22.79	21.13	30.38
Wikipedia	20.75	86.41	18.35	56.46	69.34	63.73	78.85
Review	14.65	38.16	43.87	17.02	1.49	17.02	86.58
Blog	27.45	58.93	-	57.62	67.43	53.42	67.23
NEWS	3.21	24.98	4.08	26.6	84.99	41.78	10.46
Literature	12.54	26.22	16.69	21.76	16.1	39.74	21.21
Mythology	13.71	54.6	-	60.38	32.3	58.94	88.64

Fig. 1 Achieved F-scores taking entire feature set for all testsets with single-genre training set

5 Results and Observations

Here, we discuss the results achieved by our system considering both supervised and semi-supervised hypothesis. The outcomes for each experimental scenario were evaluated in terms of precision, recall, F-measure, and accuracy.

5.1 Hypothesis 1: Single-Genre Training

For our first hypothesis based on semi-supervised approach, we considered that tagged data is available only from one domain and using that as the training set, we need to build a system that will work for all the other domains. Hence, we built seven different classifiers using seven different datasets as training separately. The datasets were split into a 70–30% ratio for training and testing purpose, respectively. Then, using each of these classifier models one by one, test sets for all the datasets were tagged and evaluated. Initially, we built our feature files for all the datasets including both domain-dependent and domain-independent features and fed them to a CRF-based classifier [1](#).

The instances where the training and testing data are from the same domain are highlighted. As we can see, models built from Wikipedia article data, News data, and mythology data perform very well for test sets from the same domain, while the review model performs poorly for all the datasets. It fails to detect even a single entity in two domains. Also, all the models perform poorly in terms of recall against the review test set, which means the classifiers are not able to detect enough number of valid entities from the review texts. We noticed that the recall for twitter test set is maximum for the model which was built using Twitter as the training set.

As there are some features which work well only for a specific genre of text and we are trying to build a system that can yield a good amount of Named Entities from any genre of text given, we tried to note its performance based on only the domain-independent feature set. Though the evaluation scores reduced for a few particular instances, on a whole, the differences among different entity sets were more balanced [2](#).

	Twitter	Wikipedia	Review	Blog	NEWS	Literature	Mythology
Twitter	38.82	30.47	16.25	15.01	23.64	20.96	29.99
Wikipedia	20.94	86.39	-	56.77	61.98	63.32	78.71
Review	14.44	37.38	43.87	17.02	14.44	17.02	86.58
Blog	27.19	58.31	-	56.93	61.31	53.08	67.23
NEWS	3.08	28.41	2.43	18.26	85.37	36.31	11.86
Literature	12.36	26.35	33.69	21.4	15.98	54.91	23.2
Mythology	9.62	53.65	-	60.27	24.54	58.31	87.57

Fig. 2 Achieved F-scores taking only domain-independent feature set for all testsets with single-genre training set

	Twitter	Wikipedia	Review	Blog	NEWS	Literature	Mythology
Twitter	72.23	61.59	51.76	57.51	60.72	59.01	62.64
Wikipedia	35.86	89.53	14.57	62.77	68.45	68.71	82.62
Review	18.23	17.48	53.04	19.35	15.15	19.08	17.12
Blog	29.48	63.65	15.5	63.49	57.35	53.33	65.78
NEWS	11.79	35.58	8.26	26.78	84.31	40.58	19.35
Literature	77.02	80.63	82.13	85.64	81.37	83.99	83.5
Mythology	13.62	71.26	8.55	59.97	51.58	61.21	88.19

Fig. 3 Achieved F-scores after implementing Gazetteer list for all testsets with single-genre training set

As it can be seen, the scores for a few in-domain (training and testing data are from the same domain) classifiers have deteriorated slightly (at most 1%), but the difference among the results for different sets of training data for a single classifier model has reduced significantly. Non-entities had the lion's share of words from all the datasets, which eventually made our system biased toward this particular class. As a result, the system showed a visible decline in recall score for all the domains as it had the tendency to tag most of the entities as non-entity. To remove this bias toward non-entities, we divided the class OTHERS into parts-of-speech-specific classes. This essentially means that an adverb which is non-entity will be assigned the tag OTHERS-ADVERB, while a noun which is not an entity will go to the class OTHERS-NOUN. This strategy helped to attain a greater balance in terms of words per class ratio and thus removed the bias that caused a poor recall. We followed this strategy in both of our hypothesis. As the results for some of the cross-domain (test and train data from different domains) classification are not satisfactory, we introduced a final noise clearance module using the gazetteer list prepared on the outcomes from domain-independent feature set. As discussed earlier, this module works in two phases improving precision and recall of entity identification respectively 3.

Significant improvement in terms of both precision and recall was observed after including this final module to our system. While the in-domain F-measure for the domains mythology, Wikipedia article, and literature are all above 80%, the results

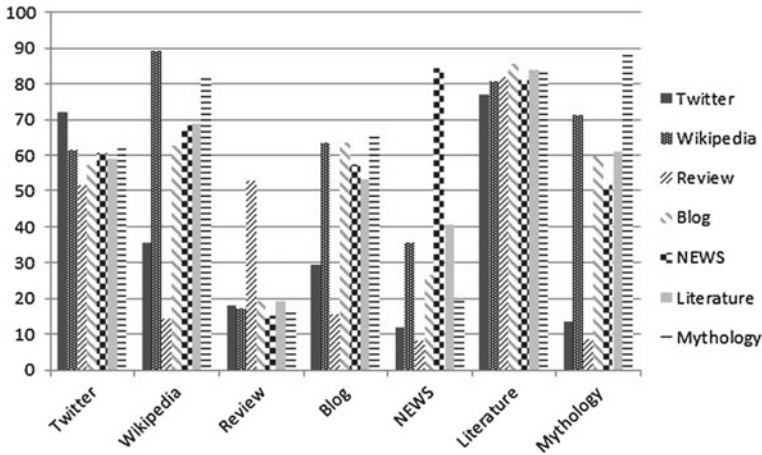


Fig. 4 Achieved F-scores after implementing Gazetteer list for all testsets with single-genre training set

for the user-generated contents, i.e., blog, review, and Twitter did not cross the 80% mark. Though, it can be safely said that the results have improved than the original outcome before implementing this module. In Fig. 4, we represent the final f-score results of our system following semi-supervised domain adaptation technique in a chart.

Another interesting observation that we could make from the evaluation results is that the accuracy of our system is almost always above 90%. This happens due to a large number of true negatives; i.e., our system correctly detected a large number of non-entities.

5.2 Hypothesis 2: Mixed-Genre Training

In the second hypothesis which follows the supervised approach of domain adaptation, we trained our model using data from a combination of two domains, and then, the model was used to tag data from those domains separately as well as mixed. For training, we have taken 20,000 instances from each domain, i.e., a total of 40,000 instances for training, and for testing, we have taken 5000 from each and merged them to have a test set of 10,000 instances. However, for testing using data from a single domain, we kept the same test set of Hypothesis 1.

In contrast to results obtained in Hypothesis 1, we can observe from Table 2 that all the results have improved if we have our classifier model trained with data from both the domains, instead of a single domain. Except for review, our system achieved a satisfactory result with above 80% F-measure in many test cases. The best result achieved was for the merged training sets of News and Wikipedia, on the mixed

Table 2 Achieved F-scores with mixed-genre training set

Train 1	Train 2	Test 1	Test 2	Mixed
News	Blog	82.96	66.54	86.29
News	Wikipedia	82.74	88.81	89.26
News	Twitter	82.83	69.28	78.76
News	Review	82.93	44.58	82.37
News	Literature	82.95	83.68	87.12
News	Mythology	83.21	86.90	87.01
Blog	Literature	64.81	83.22	85.76
Blog	Mythology	66.90	87.44	87.73
Blog	Wikipedia	65.72	83.78	87.02
Blog	Review	63.82	39.18	40.84
Blog	Twitter	54.19	66.61	65.81
Literature	Mythology	84.34	87.12	83.08
Literature	Wikipedia	83.84	78.44	81.03
Literature	Review	82.33	40.93	54.85
Literature	Twitter	80.17	65.96	65.67
Mythology	Wikipedia	86.78	83.50	86.32
Mythology	Review	86.90	30.65	80.39
Mythology	Twitter	85.78	65.73	75.45
Wikipedia	Review	87.79	38.88	81.37
Wikipedia	Twitter	80.36	68.98	74.71
Review	Twitter	40.41	66.42	65.23

testset with 89.26% precision. Both precision and recall have improved significantly for the user-generated Web contents like Twitter and blog entries. Best results were achieved for twitter when combined with a neutral dataset like News, where the influence of external factors on language is minimal. Wikipedia articles also follow a similar writing approach so that the texts are comprehensible to users globally. This characteristic makes these two datasets balanced and unbiased to train our classifier with. As an outcome, we notice that our system achieved the most prominent results (above 80% for almost all the mixed test sets) when we used at least one of these datasets during training.

Another interesting observation that one can make is that our system performs best for texts of closely related genres like mythology and literature or news and Wikipedia articles which share common language and literary characteristics.

Similar to Hypothesis 1, our system delivered excellent results in terms of accuracy. We can conclude that our system performs better if we have at least a few training instances from the domain we have as test set.

6 Conclusion

One of the main aims for this module was to build a model that will be able to extract named entities of types PERSON, LOCATION, ORGANIZATION from texts of different genres and domains (e.g., Twitter, blog, news, reviews, literature). Working on two hypotheses, single training set hypothesis and mixed training set hypothesis, we extracted domain-dependent and domain-independent feature values from texts and sent them to a CRF-based classifier and finally modifying the annotated tags using gazetteer lists. Though we achieved modest results in both the hypotheses, a few following measures can be taken for further improvements.

- Instances from the mixed training set that have a greater impact on the results with respect to a test set will be given greater weightage during its training so that the classifier may work better for that particular domain.
- The inclusion of entity-type specific features like trigger words or n-gram prefix-suffix to increase the recall of our system.
- We will try to devise methodologies for distinguishing closely related entity pairs such as CHARACTERS-PERSONS and THING-PRODUCT along with the identification of various other entity types such as EVENTS etc.

Acknowledgements The work is supported by Visvesvaraya Young Faculty Ph.D. Research Scheme, under Ministry of Electronics and Information Technology (MeitY), Media Lab Asia, Government of India.

References

1. H. DaumIII, D. Marcu, Domain adaptation for statistical classifiers. *J. Artif. Intell. Res.* **26** (2006)
2. J. Jiang, C. Zhai, Instance weighting for domain adaptation in NLP, in *Proceedings of the Annual Meeting of the Association of Computational Linguistics (ACL)* (2007)
3. H. DaumIII, Frustratingly easy domain adaptation, in *Proceedings of the Annual Meeting of the Association of Computational Linguistics (ACL)* (2007)
4. A. Kumar, A. Saha, H. Daume, Co-regularization based semi-supervised domain adaptation, in *Advances in Neural Information Processing Systems 23*, ed. by J.D. Lafferty, C.K.I. Williams, J. Shawe-Taylor, R.S. Zemel, A. Culotta. Curran Associates, Inc., (2010), pp. 478–486. <http://papers.nips.cc/paper/4009-co-regularization-based-semi-supervised-domain-adaptation.pdf>
5. A. Sun, R. Grishman, Cross-domain bootstrapping for named entity recognition, in *Proceedings of the SIGIR 2011 Workshop on Entity-Oriented Search* (2011)
6. G. Rizzo, M. van Erp, Making sense of microposts (#Microposts2016) named entity recognition and linking (NEEL) challenge, in *6th Workshop on Making Sense of Microposts (#Microposts2016)* (2016)
7. B. Pang, L. Lee, A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. in *Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics*, Association for Computational Linguistics, Stroudsburg, PA, USA, ACL '04 (2004). <http://dx.doi.org/10.3115/1218955.1218990>
8. J. Nothman, N. Ringland, W. Radford, T. Murphy, J.R. Curran, Learning multilingual named entity recognition from wikipedia. *Artif. Intell.* **194**, 151–175 (2013). [10.1016/j.artint.2012.03.006](http://dx.doi.org/10.1016/j.artint.2012.03.006), <http://dx.doi.org/10.1016/j.artint.2012.03.006>

9. E. Stamatatos, W. Daelemans, B. Verhoeven, P. Juola, A. López-López, M. Potthast, B. Stein, Overview of the author identification task at pan, in *CLEF 2015 Evaluation Labs and Workshop – Working Notes Papers* (CEUR, CEUR, Toulouse, France, 2015)
10. J. Blitzer, R. McDonald, F. Pereira, Domain adaptation with structural correspondence learning, in *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing*, Association for Computational Linguistics, Stroudsburg, PA, USA, EMNLP '06 (2006), pp 120–128. <http://dl.acm.org/citation.cfm?id=1610075.1610094>
11. S. Ben-david, J. Blitzer, K. Crammer, O. Pereira, Analysis of representations for domain adaptation, in *In NIPS*, (MIT Press, 2007)
12. S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, J.W. Vaughan, A theory of learning from different domains. *Mach. Learn.* **79**(1–2), 151–175 (2010). <http://dx.doi.org/10.1007/s10994-009-5152-4>

Relation Extraction from Cross-Genre Unstructured Text



Promita Maitra and Dipankar Das

Abstract In this work, we present our approach to find relationships, both taxonomic and non-taxonomic, among the named entities extracted from texts of different genres. Instead of trying to assess the taxonomic relationships among named entities, we have adopted a context word-matching technique to assign separate scores to a different pair of entities, for different taxonomic relations. For non-taxonomic relations, we have mostly focused on verb-based relations and have also proposed a simple system to assign possible sentiment polarity labels among entity-pairs. This is an attempt to build a single unified system which can scan through texts of any genre and provide a fair idea of how the possible named entities are related among each other, assessed on three distinct space-intrinsic property, action, and sentiment.

1 Introduction

While extracting relations among entities, we need to understand the various factors associated with the meaning of the unstructured text. The difficulty of such a task lies in identifying those factors, designing suitable algorithms, and implementing methods to handle those factors effectively in order to achieve a consensus upon retrieved relations among entities. This process can be supervised, unsupervised, or hybrid. Relations can be of two major types—taxonomic and non-taxonomic relations. To describe them briefly, taxonomic relations are the ones which have a hierarchical or tree structure while non-taxonomic relations do not display such natural structures. The atomic nature of non-taxonomic relations makes them increasingly difficult to identify. In this work, the task of identifying meaningful relations was further complicated by the cross-genre nature of the text, which involves different documents belonging to a wide range of topics, and displaying unique writing styles.

P. Maitra (✉) · D. Das
Jadavpur University, Kolkata, India
e-mail: promita.maitra@gmail.com

D. Das
e-mail: dipankar.dipnil05@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_39

We considered different types of named entities and it is infeasible to relate them in a taxonomic manner. Hence, we assign scores to each entity pair for different taxonomic relations based on their context words. For non-taxonomic relations, we rely on verbs of sentences in which the entity pair co-occurs and then gain more insight by further filtering and clustering with VerbNet.¹

While Sect. 2 is dedicated to discuss the works already done in this area, Sect. 3 briefly describes the dataset that we used in this work. Section 4 contains the challenge, the detailed approach, and the sample results of taxonomic relation extraction. In Sect. 5, our work on extracting relations which are non-taxonomic in nature has been explained. We conclude our present work and discuss the scope of future work in Sect. 6.

2 Related Works

Keeping in mind that not much work has been done in this particular field, we will try to discuss some methods already available to address taxonomic and non-taxonomic relation extraction tasks separately on different domains. To extract taxonomic relation, at first [3] introduced lexico-syntactic patterns in the form of regex. Pattern-based heuristic approaches have also been used widely for information extraction by researchers like [4]. The main idea behind these lexico-syntactic approach is very simple: The entire text is scanned for instances of a defined pattern, which can be mapped with each other as concepts having taxonomic relations. The PROMETHEE system, introduced by Morin [9] demonstrates semi-automatic extraction of semantic relations, as well as refinement of lexico-syntactic patterns. Snow et al. [12] constructed classifiers to identify the hypernym relationship between terms from dependency trees of large corpora. Terms with recognized hypernym relation are extracted and incorporated into a man-made lexical database, WordNet [2], resulting in the extended WordNet, which has been augmented with over 400,000 synsets. Ponzetto and Strube [11] and Suchanek et al. [13] mined Wikipedia and came up with hierarchical structures of entities and relations.

For non-taxonomic relation extraction, previous researches have made extensive use of verbs. There are two approaches: top-down approach, where the verbs appearing frequently in texts are extracted first and considered as relation labels, and then the concepts associated with those verbs are extracted; and bottom-up approach where the domain concepts are identified and extracted first and then the verbs appearing frequently in their context are identified as relations between concept pairs.

Co-occurrence distribution of words help discover their distributional properties in statistical approach. Semantic distances of the words can be used to determine the correlated concept pairs, which can be considered for non-taxonomic relations. In case of non-taxonomic relations, the lexico-syntactic approach is primarily based on string matching patterns on tokens of text and syntactic structure in unstructured texts.

¹<https://verbs.colorado.edu/~mpalmer/projects/verbnet.html>.

The research work in non-taxonomic relation discovery was first initiated by Maedche and Staab [6–8] and Text2Onto [1], the same association rule mining algorithm with a confidence score was taken to find out correlated concept pairs which were discovered using shallow text processing [15]. The ASIUM ontology learning tool [10] extracts syntactic frames of verbs from given text documents. OntoLearn [14] trains an available machine learning algorithm, TiMBL, an open-source software package implementing several memory-based learning algorithms with the help of a reduced set of FrameNet relations, to extract relations in a particular domain. When it comes to non-taxonomic relation labeling, most of the existing applications use verbs or verb phrases frequently occurring in the context of the pair being considered. As verbs play a major role in communicating the sense of a valid sentence, such associations are considered as labels of the unnamed relations. Earlier in Text-to-Onto, non-taxonomic relations were labeled manually by humans. Kavalec et al. [5] have built an extension to Text-to-Onto early version to include the automation of relation labeling.

3 Dataset Description

For the proposed research, we needed unstructured textual data from different genres of writing such as social media texts, mythological texts, etc. As we are focusing on cross-genre text, it was essential to collect texts from multiple topics in a single dataset. The complexity of our domain adaptation problem is hence twofold. First, it does not revolve around a particular theme or topic; and secondly, it does not deal with texts having some regular structure. We noticed that the social media texts can provide some important insights as they are vastly irregular in structure, follow no standard grammatical rules, incorporate various foreign words, symbols, and emoticons. These unusual characteristics make it quite difficult to extract relevant information from these kinds of noisy texts. For our work, we have mainly considered data from three domains: Twitter, Blog, and Review. As each of these social media texts serves different purposes, there is a distinct difference in the style of writing and the structure of the text. On the other hand, the mythological texts that we have considered diverge manifold from the contemporary style of writing that we are acquainted with. Hence, we have considered both of these genres of writing to evaluate the performance of our system. In addition to these, a Wikipedia article corpus and one NEWS corpus have also been included in our experiment as News and Wikipedia articles follow a neutral reporting manner of writing without many regional or cultural biases.

4 Taxonomic Relation Extraction

The system we are considering in present work deals with data from different genres, which can be on any random domain. As the domain is not fixed, it is practically impossible to define domain-specific entity types or concepts for ontology building. Hence, we have only considered different types of Named Entities as top-level entities. It can be safely stated that names, i.e., proper nouns do not generally hold a taxonomic relation (hypernym–hyponym or synonym–antonym) with other proper nouns. Hence, it is not possible to get direct taxonomic relations among the entities that we are considering. Moreover, the linguistic pattern-based approach which has been unanimously adopted to find out instances of these relations from the text is not going to work in our task either. The Hearst patterns or similar other patterns that other researchers have added over time are very much corpus specific and adhere to a particular writing style. On the contrary, we need to build a system that will work effectively on texts from various genres, which follow a different literary style. These patterns will not be able to reach satisfactory recall score on datasets like Twitter, Blog, or Review where the authors follow informal writing approach while adding content. These texts do not strictly follow standard grammar rules and include various foreign words and additional symbols. As direct taxonomic relations like hypernym–hyponym or meronym–holonym are not possible among named entities, we try to assign a score to each of these relations and for entity pairs based on the context which they frequently appear in. The relations which we have considered for calculating scores are hypernym–hyponym, meronym–holonym, and synonym–antonym. We have used relations from WordNet in our algorithm to keep our system abstract enough to work in various genres and effective enough to reach a satisfactory accuracy level.

4.1 Proposed Approach

To build a system that can be applied to all the domains present in the dataset, we need to keep the algorithm independent of the linguistic pattern or the contents of any particular dataset type (Fig. 1).

The steps of the algorithm are as follows:

- A. For each dataset, create an entity bag of each type: PERSON, LOCATION, ORGANIZATION (and THING, EVENT for twitter).
- B. For each entity, repeat step C to K.
- C. Crawl the entire dataset and extract all the words that appear in this entity's context (+/-3 word window) with their frequency.
- D. Filter out top frequent ten context words per entity. Do this for all entities.
- E. Take another entity from same or different entity type that co-occurs at least once with that entity and compares them based on six relations from WordNet: hypernym, hyponym, meronym, holonym, synonym, and antonym.

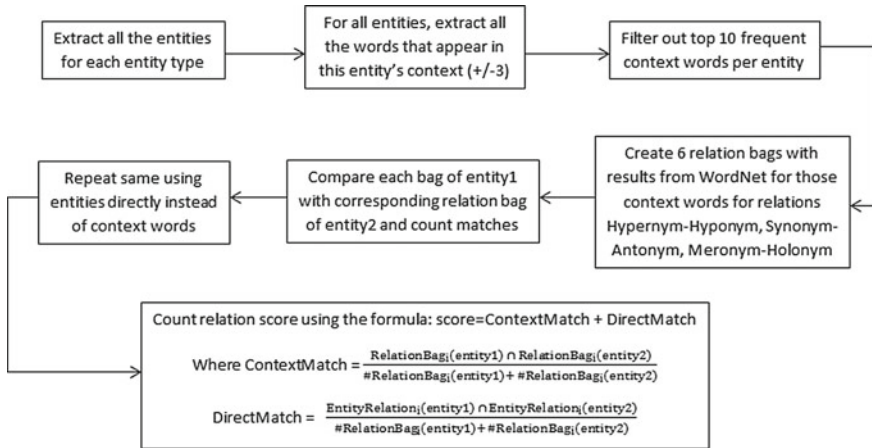


Fig. 1 Proposed model for taxonomic relation extraction

- F. Take all the ten context words of an entity and add all the words obtained from WordNet for a particular relation (from the six mentioned above) to a bag, e.g., Put all the words that we can extract from WordNet for synonym relation of those context words to ‘Synonym’ bag.
- G. Construct six different bags of words for the above-mentioned relation types for each of these two entities.
- H. Compare each relation bag of one entity to corresponding relation bag of the other entity and count number of overlaps.
- I. Calculate relation score for this particular entity pair using the formula:

$$CW\text{-Relation-Score}_i = \frac{\text{RelationBag}_i(\text{entity1}) \cap \text{RelationBag}_i(\text{entity2})}{\#\text{RelationBag}_i(\text{entity1}) + \#\text{RelationBag}_i(\text{entity2})}$$
 CW: Context Word; *i*: hypernym, hyponym, synonym, antonym, holonym, meronym
- J. Also, along with the context word relation score calculations, do a similar calculation for entities themselves, e.g., compare among the hypernyms of entity1 and entity2 from WordNet and count the matches. Then divide the count with the total number of hypernyms found for entity1 and entity2 combined.

$$DM\text{-Relation-Score}_i = \frac{\text{EntityRelation}_i(\text{entity1}) \cap \text{EntityRelation}_i(\text{entity2})}{\#\text{RelationBag}_i(\text{entity1}) + \#\text{RelationBag}_i(\text{entity2})}$$
 DM: Direct Match; *i*: hypernym, hyponym, synonym, antonym, holonym, meronym
- K. In the last stage, we add these two scores CW-Relation-Score(*i*) and DM-Relation-Score(*i*) to obtain the final score of an entity-pair for a relation-type. Continue for other entity-type combinations as well. End.

Table 1 Snapshot of taxonomic relation score for blog entities

Entity-pair	Hypernym	Hyponym	Antonym	Synonym	Meronym	Holonym
Berlin - Karina	0.0813	0.0295	0	0.2439	0.1463	0.1071
Germany - Starbucks	0.0683	0.0091	0	0.2626	0.2727	0
Sam - New York	0.4860	0.2082	0.2222	0.4961	0.5	0
NFL - League	0.1738	0.5026	0	0.2306	0	0

4.2 Sample Results and Observations

For each entity pair that co-occurs, we have assigned a score to each of the six relations. The score will vary from 0 to 1, where 0 means the entities are not related at all and 1 means that they are completely related for that particular relation. In Table 1, we present a sample snapshot of these relations for a few blog entities.

Similarly, scores for all other possible entity-pairs and for all the seven datasets were obtained. We have analyzed the possible inferences from the results obtained:

1. Karina is a railway station in San Jose and Berlin is a city. So it can be easily assumed that they will probably have a meronym–holonym relation as a city may have a railway station. However, city and railway station are a top-level abstraction of the entities that we have here. Hence, we do not get a very high score in these fields, but the score is not 0 either.
2. For the last entity-pair mentioned, we get two entities NFL and League. As we know that the NFL is a type of League, we observe a high score for the relations hypernym–hyponym, whereas no score for meronym–holonym. Also, the context words they appear in is similar and so we get a positive score for synonym, whereas no score for antonym.

5 Non-taxonomic Relations

Non-taxonomic relations are those for which a sensible and reusable taxonomy cannot be created using the relations, i.e., entities cannot be represented in a hierarchical fashion based on the relations among them. Non-taxonomic relation extraction is argued to be as one of the most difficult tasks and often skipped in ontology learning mechanism. It is this kind of relations which reveal more about a particular domain as taxonomic relations are restricted to some specific relations only, causing a hindrance to explore the domain in an exhaustive manner. As these relations vary immensely owing to the diverse nature of domains to extract an ontology from, it is very difficult

to figure out how many types of relations are there to be extracted. The main challenge at present, as already discussed, is that we are not dealing with a particular domain in focus. We are more concerned with different forms of unstructured texts, which in turn, can be based on multiple random topics. For example, one of our datasets is twitter data, which can consist of tweets on any random topic on earth, following to particular syntactic structure. In such cases, it is not possible for a particular domain expert to come up with a few fixed relations which will be able to cover all kind of relations that can possibly be present in the domain. To address this issue at hand, we fully depend on the verbs occurring between two concepts. The advantage of using verbs in relation mining is that it works for both of the subtasks, relation discovery and relation labeling. There can broadly be two approaches to implement this method:

- Bottom-up approach - In this approach, we first extract in-domain concepts appearing in the text at hand. Second, verbs co-occurring with those concepts are extracted for labeling of the relation.
- Top-down approach - Here, verbs frequently appearing in text are extracted first for relation labels. Then we extract concept pairs occurring along with these verbs.

In our proposed method, we mainly adopted the Bottom-up syntactic approach where we use the already extracted Named Entities from the previous step as concepts and try to extract relations among them based on co-occurrence using the freely available verb lexicon VerbNet.

5.1 Overview of VerbNet

VerbNet is a verb lexicon with syntactic and semantic information for English verbs, adhering to Levin verb classes (Levin 1993) for systematic construction of lexical entries. It consists of approximately 5800 English verbs and groups them into 274 first-level verb classes according to shared syntactic behaviors, exploring the generalizations of verb behavior. Even though the nature of VerbNet classification is syntactic, the verbs of a given class share semantic regularities as well because, according to Levin's hypothesis, the syntactic behavior of a verb is largely influenced by its meaning. This domain-independent lexicon takes full advantage of the systematic link between syntax and semantics that motivates these classes and thus provides a clear and regular association between syntactic and semantic properties of verbs and verb classes (Kipper et al. 2000; Dang et al. 2000). To make this association explicit, a set of thematic roles is assigned to each syntactic argument in a given verb class as well as some selectional restrictions to each of these theme roles.

5.2 Proposed Approach

As we have this major disadvantage of not knowing the domain theme beforehand, and texts from a single genre can be on any arbitrary topic and even multiple topics as well, there is no way we can design a flexible set of possible relations that needs to be extracted, even after rigorous discussion of a domain expert and an ontology developer. Hence, we make use of the verbs present at hand; as it has been seen that verb is an essential part of making sense out of a randomly drawn unstructured text. But it won't be enough to just identify a lot of relations from the text. We needed to find a way to cluster them, in due course which will be providing us with more insight about those entities. The logic being, there is a high chance of concepts or entities being similar in some dimension, which may or may not be very definitive one, if they belong to the same cluster. VerbNet, being such an efficient and free-to-use resource, which was not used for this kind of task before, was chosen as the primary tool to explore the dataset for verb based non-taxonomic relation clustering (Fig. 2).

The actual method goes through a series of preprocessing and evaluation tests, which are briefly described below:

- The sentences in which two entities of the same or different types occur together are extracted from each domain.
- Only the sentences in which the entities occur in a window size of maximum ten are filtered out, i.e., there will be at most ten words in between the entities. This is done to restrict the number of irrelevant entity-relationships that are extracted.
- We tag the sentences with Stanford CoreNLP POS tagger and only the verbs are extracted from those filtered sentences.
- Using Stanford CoreNLP lemmatizer, extracted verbs are transformed into the corresponding lemma, so that they can be used to search VerbNet to acquire a greater

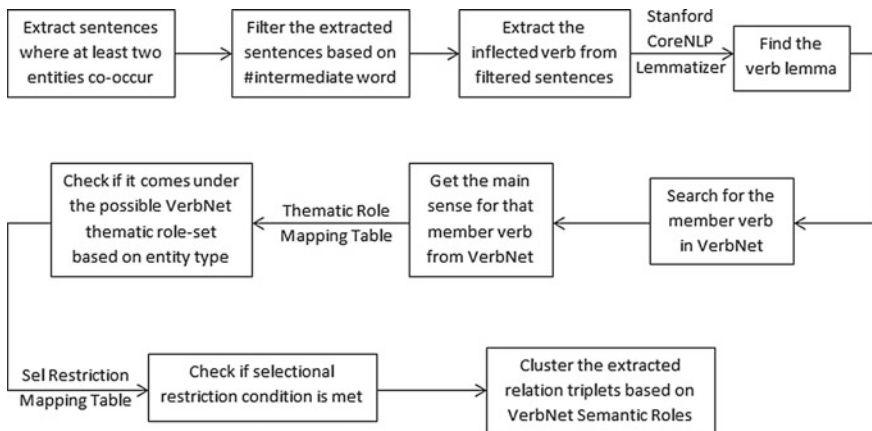


Fig. 2 Proposed model for verb based non-taxonomic relation extraction

understanding. Lemmatization usually uses a vocabulary and does a morphological analysis of words, normally aiming to remove inflectional endings only and to return the base or dictionary form of a word, which is known as the lemma. If confronted with the token *saw*, stemming might return just *s*, whereas lemmatization would attempt to return either *see* or *saw* depending on whether the use of the token was as a verb or a noun.

- Next, the lemmatized form of the verb is searched for in VerbNet and the main sense, of which this verb form is a member verb is extracted as the relation sense.
- After we get the main sense of that particular verb which appeared in our dataset, we need to check if it belongs to the set of thematic roles which are allowed for those particular types of entities this verb is appearing in the context of. The thematic role mapping table is mentioned below (Table 2).
- We try to further filter the verbs using selectional restrictions associated with each thematic role. We did a mapping for entity type and possible selectional restriction in the same manner as of the thematic roles. The mapping is provided below (Table 3).
- In a similar manner, all possible verb-based relations among all possible entity pairs of any two types are extracted. As the latest version of VerbNet has 274 first level classes in which all the other verbs are categorized into based on the sense they are communicating, we would have got 274 different clusters for ‘entity-relation-entity’ triplets. But it’s practically impossible to visualize that many classes for extracted relations. Hence, we incorporated the concept of thematic role in our clustering algorithm. As there are at most 23 thematic roles, any relation extracted from any of the domain will come under one or more of those 23 classes only. Compared to the previous 274 classes, it is much easier for users to understand the sense of those relations and use or extend them effectively for future use.

5.3 Sample Results and Observation

The number of relations extracted per entity-pair varies with the dataset and also with entity-pair types. A PERSON-type entity-rich ‘mythology’ dataset extracts a

Table 2 Thematic role mapping

Entity-type	Allowed thematic roles from verbNet
Person	Actor, agent, co-agent, beneficiary, experiencer, patient, co-patient, recipient
Location	Location, source, destination, initial_location
Organization	Agent, co-agent, beneficiary, recipient
Thing	Instrument, material, product
Event	Time, topic

Table 3 Selectional restriction mapping

Entity-type	Allowed selectional restriction chains from verbNet
Person	Concrete->natural->animate->human
Location	Location->regionPP
Organization	Organization
Thing	Concrete->phys-obj->artifact->tool
	Concrete->solid->rigid,
	Concrete->substance,
	Concrete->int-control->machine
Event	Time

large number of relations for all the entity-pairs that include PERSON. In the ‘contemporary literature’ domain, it is seen that we get relations only for the pairs which include at least one PERSON type entities. Similarly, for other domains as well, it can be witnessed that our system failed to extract any relation for a few entity types. For example, the count of entities for THING-EVENT or EVENT-EVENT entity pairs in twitter, ORGANIZATION-ORGANIZATION in the blog, and LOCATION-LOCATION in the review is nil. This can be explained by the notion that our system only detects relation when two different entities of an entity-pair co-occur in a single sentence. Moreover, one of these entities must be the subject of the verb and also satisfy the thematic role and selectional restriction mapping in VerbNet done for its corresponding entity type. So we can conclude from our observations, that the entity-pairs having nil relation counts in all the domains either never appear together in a single sentence, or the sentences they co-occur in, do not contain a verb which is appropriate for the entity types they belong to, in terms of thematic role assigned in VerbNet or selectional restrictions for that role. If we try to note the variations in terms of relation counts, we will notice a striking difference among the datasets like twitter and contemporary literature or mythology. While the relation counts for Twitter data is very less or nil, we find the same count to be very high for mythology dataset. One possible reason for this outcome is that we deal with short texts in twitter, and hence, the probability of two entities co-occurring together is low. A similar trend can be witnessed in Blog dataset as well. But in case of literature and especially mythological texts, our system acquires more instances of these, even in the review dataset as well; because they deal with characters and so chances of entities appearing together is higher. We have seen some noisy outcomes as well. To list a few:

- Harry Potter <celebrate> Harry_P

One can easily notice the problem here: Both the entities actually refer to same character Harry Potter, but as the forms are different, the system will treat them as different entities and try to find relations among them. This kind of issues can be removed if we implement an additional entity linking module before relation

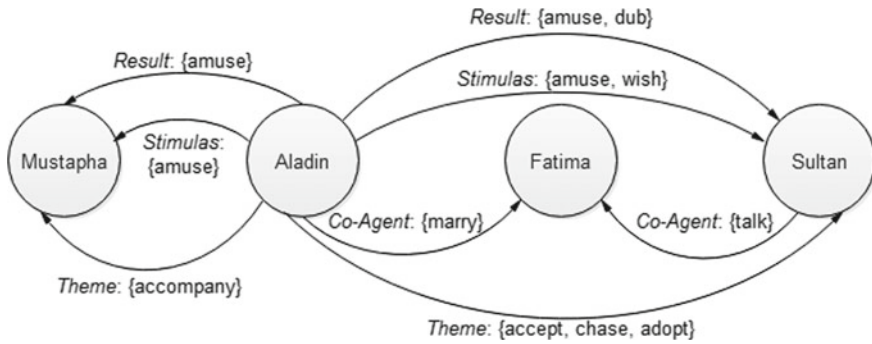


Fig. 3 Snapshot of extracted relations and clusters

extraction phase and merge all forms of the same entity together to indicate a single actual reference.

- Aladin <marry> Sultan
 Going by the famous short story, it can be said that this particular relationship is not correct. However, according to our system output, these two entities are connected via this verb because they appear in each other’s context along with the verb. This particular problem can be avoided by adding an additional check to keep only the relations if two participating entities are related to the subject-verb-object frame. But this will significantly reduce the recall for another type of entity-pairs other than the ones which contains at least one PERSON type entity, as entities like LOCATION or ORGANIZATION hardly appears as a subject of the verb in a sentence.
- Consider the sentence: Jason was not ready to agree with Julia this time. For this particular sentence, the relation Jason Julia <agree> will be extracted which is directly opposite to the sense here. For these cases, a negation checker needs to be included in the system.

A snapshot containing a few of such verb relations among a subset of entities from contemporary literature domain is presented in Fig. 3. Here, the entity set is [Aladdin, Mustafa, Fatima, Sultan]. The thematic-role-based clusters that we see here are Theme, Result, and Co-Agent. And the relations they contain are *amuse*, *dub*, *marry*, *accompany*, etc.

5.4 Sentiment Polarity Scores

In addition to the verb-based relation set, we also try to check if there exists any sentiment relation between two entities based on the polarity score among them, calculated over top frequent context words. We have used SentiWordNet 3.0.0 to build a module that will calculate the effective sentiment score of the context words

associated with any entity-pair. The primary disadvantage that we have faced while trying to examine this kind of relations among entities is that we had no annotated data ready to extract features from and train a model. Hence, the module we employed solely depends on SentiWordNet scores. We have diverse datasets containing text from the web such as tweets, blog data, IMDB review data, as well as wiki article corpus, contemporary literature corpus, and mythological text. We needed to build a system which will be simple enough that it can be applied to all formats of texts without user interference or domain-wise modification reducing any kind of bias as much as possible.

The proposed scheme works as described below.

- A. At first, we take an entity-pair of same or different type and extract all the sentences from the dataset where they co-occur.
- B. Excluding the entities themselves and any other entity that might appear in the context, all other words that appear in those sentences are collected in their lemmatized form (using Stanford CoreNLPlemmatizer) to build a bag of words for that particular entity pair.
- C. We do a SO (Subject-Object) analysis on that bag of words that were collected parsing the entire dataset. We only keep the words that can have a significant contribution (with a score >0.25 or <-0.25) on polarity score.
- D. We are left with only a handful of words that have negative/positive polarity scores associated with them. We extract scores from SentiWordNet for each of the words separately mentioning the POS they appear within that sentence.
- E. We take the average over all the context word sentiment scores for that particular entity pair. Comparing the score with prefixed mapping, we assign a sentiment polarity relation between those entities.
- F. Finally, after labeling each entity pair that co-occurs, we put them into three distinct classes: Positive, Negative, and Neutral.

Some of the sample results this module has produced are:

- Entity-Pair: America(LOCATION)-Lincoln(PERSON)
Context Words: debate, generally, consider, most, famous, political, carefully

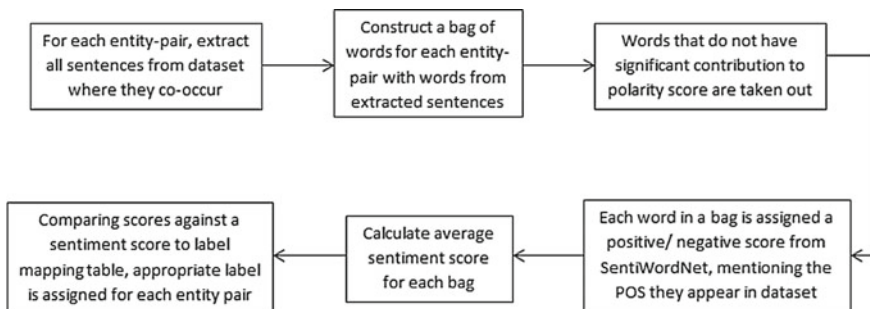


Fig. 4 Proposed model for sentiment polarity scoring of entity pairs

Polarity: POSITIVE

[Entity-pair extracted from blog dataset]

- Entity-Pair: Duryodhana(PERSON)-Army(ORGANIZATION)

Context Words: solicit, thee, offer, leadership, deprive, run, drag, arrow, battle, keen, shaft, slay, mighty, slaughter, killing, destroyer

Polarity: NEGATIVE

[Entity-pair extracted from mythology dataset] (Fig. 4).

6 Conclusion

The main challenge of the present work was twofold, as we needed to extract relations from multiple datasets of different genres having diverse characteristics in terms of text structure, vocabulary used, the influence of external agents like region, time, culture, platform, etc. In addition to this, the datasets we are considering can be on any random topic or theme, making identification of domain-specific relation-type almost impossible.

The primary aim of present work was to build a system that will work equally in all the domains in determining the degree and type of entity-pair closeness, and hence we kept our approach as generalized as possible. As we have only considered named entities, we proposed a module that assigns scores to each pair of entities for six different taxonomic relations (Synonym, Antonym, Hypernym, Hyponym, Holonym, Meronym) based on context word analysis. We have proposed two different non-taxonomic relation extraction schemes: One is verb-based, and another is entity-pair sentiment based. In the verb-based relation extraction module, we extract relations based on entity co-occurrence and the verb present, further filtering and clustering the extractions based on thematic role and selectional restrictions from VerbNet. In sentiment-based relation module, we try to predict a polarity label for each pair on the entity that co-occurs, based on the context words of co-occurrence. Further improvement of the present recall factor is possible by introducing a co-reference resolution model or weighted context word scoring based on occurrence. Refinement of relations by introducing additional modules like negation detection or further filtering of verb clusters.

Acknowledgements The work is supported by Visvesvaraya Young Faculty Ph.D. Research Scheme, under Ministry of Electronics & Information Technology (MeitY), Media Lab Asia, Government of India.

References

1. P. Cimiano, J. Völker, A framework for ontology learning and data-driven change discovery, in *Proceedings of the 10th International Conference on Applications of Natural Language to Information Systems (NLDB)*. Lecture Notes in Computer Science, vol. 3513 Springer, (Springer, 2005), pp 227–238
2. C. Fellbaum, *WordNet* (Wiley Online Library, 1998)
3. M.A. Hearst, Automatic acquisition of hyponyms from large text corpora, in *Proceedings of the 14th Conference on Computational Linguistics-Volume 2, Association for Computational Linguistics* (1992) pp 539–545
4. J.R. Hobbs et al., The generic information extraction system, in *MUC* (1993) pp 87–91
5. M. Kavalec, A. Maedche, V. Svátek, Discovery of lexical entries for non-taxonomic relations in ontology learning, in *SOFSEM*, vol. 4 (Springer, 2004), pp. 249–256
6. A. Maedche, S. Staab, Discovering conceptual relations from text, in *Proceedings of the 14th European Conference on Artificial Intelligence* (IOS Press, 2000a), pp 321–325
7. A. Maedche, S. Staab, Mining ontologies from text, in *Knowledge Engineering and Knowledge Management Methods, Models, and Tools* (200b), pp 169–189
8. A. Maedche, S. Staab, *Semi-automatic engineering of ontologies from text*, in *Proceedings of the 12th International Conference on Software Engineering and Knowledge Engineering* (IL, USA, Chicago, 2000c), pp. 231–239
9. E. Morin, Extraction de liens sémantiques entre termes à partir de corpus de textes techniques. Ph.D. thesis (1999)
10. C. Nedellec, Corpus-based learning of semantic relations by the ilp system, asium. Learning language in logic (2000), pp. 461–491
11. S.P. Ponzetto, M. Strube, Deriving a large scale taxonomy from wikipedia. *AAAI* 7, 1440–1445 (2007)
12. R. Snow, D. Jurafsky, A.Y. Ng, Learning syntactic patterns for automatic hypernym discovery, in *Advances in Neural Information Processing Systems* (2005) pp 1297–1304
13. F.M. Suchanek, G. Kasneci, G. Weikum, Yago: a core of semantic knowledge, in *Proceedings of the 16th International Conference on World Wide Web, ACM* (2007), pp 697–706
14. P. Velardi, R. Navigli, A. Cuchiarrelli, R. Neri, Evaluation of ontolearn, a methodology for automatic learning of domain ontologies. *Ontol. Learn. Text: Methods Eval. Appl.* **123**, 92 (2005)
15. M.K. Wong, S.S.R. Abidi, I.D. Jonsen, A multi-phase correlation search framework for mining non-taxonomic relations from unstructured text. *Knowl. Inf. Syst.* **38**(3), 641–667 (2014)

An Efficient Algorithm for Detecting and Measure the Properties of Pothole



Amitava Choudhury, Rohit Ramchandani, Mohammad Shamoan,
Ankit Khare and Keshav Kaushik

Abstract In road transportation, computational measurement is widely used. To ensure safe driving, pothole detection is one of the challenging tasks. Very few tools or applications have been proposed to identify any cracks or pothole while driving a car. This paper focused on an efficient methodology for pothole detection. Input images captured by camera (placed on the top of car), which captures the side view or front view of the road, then by converting the same image into its bird's-eye view to see the image from top. After image preprocessing, region of interest has been computed and then blob detection is applied to detect the actual pothole if exists on the road.

Keywords Pothole detection · Computer vision · Image segmentation · Image processing

1 Introduction

In recent years, application of computational intelligence is growing day by day. Emerging countries like India is performing on digital-based operation in every field of civilization. A computational approach is not only bounded within high com-

A. Choudhury (✉) · R. Ramchandani · M. Shamoan · A. Khare · K. Kaushik
School of Computer Science,
University of Petroleum and Energy Studies, Dehradun, India
e-mail: a.choudhury2013@gmail.com

R. Ramchandani
e-mail: ramchandani.rohit16@gmail.com

M. Shamoan
e-mail: shamoanmohd14@gmail.com

A. Khare
e-mail: akhare@ddn.upes.ac.in

K. Kaushik
e-mail: keshavkaushik@ddn.upes.ac.in

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_40

puting performance, but also the application will provide positive impact on nation building. Nowadays, the concept of computational intelligence used in agriculture, cultivation, weather prediction, transport, etc. When it comes to the point of nation development, country government should have more focus on transport system. It has been observed that a good transportation system can create a better developed nation, and transportation can be secured only when the roads are safe and well established. However, it is also been observed that unwanted speed breaker, potholes, and cracks on road create a difficult journey experience which results in many accidents. In this regard, many computational-based works are already been proposed by many researchers to detect speed breakers, cracks on roads but out of them, very few have discussed about pre-alert system before reaching a pothole. Potholes are small, mainly bowl-shaped depression in the road surface.

Pothole detection and identification is a very challenging task using computer vision approach. Sometimes, it may require higher cost. However, earlier pothole was detected by several tools but to ensure the same, human technician is required to verify. When image is captured from vehicle's top position, side view of actual road is also captured, which may lead to low accuracy while processing. In this paper, it has been described that how the side view of an image is converted into bird's-eye view, which is more effective to locate the region of interest from the whole image, and then using blob detection algorithm, pothole of the image is identified. This paper is organized as follows: In Sect. 2, various literature has been described. Proposed methodology is discussed in Sect. 3. In Sect. 4, experimental result and analysis are presented and finally Sect. 5 concludes with future scope of this paper.

2 Background Study

In recent practice, pothole detection is one of the very popular research areas using computer vision. Vigneshwar and Hemakumar discuss various image processing techniques [1] to detect potholes. *K*-means clustering-based segmentation approach was described in their paper having 83% overall accuracy. Histogram and canny edge-based segmentation techniques were described in [2, 3]. Buza et al. proposed spectral cluster-based identification techniques to detect actual region in [4]. Blob-based pothole detection technique is described in [5]. Bird's-eye view transformation technique is described in [6]. An embedded system is described in [7] to detect pothole from road images and send the location information using GPS to road maintenance authority. Z. Zhang, X. Ai, C. K. Chan, and N. Dahnoun proposed an efficient algorithm for pothole detection using stereo vision [8]. Another image processing technique is described to detect lanes, potholes, and recognize road signs in Indian roads in [9]. Buza [10], Seung-KiRyu [11], Koch [12], Mandal [13] used histogram and canny edge detection-based segmentation for detection of potholes. Many researchers most commonly use edge detection technique and thresholding technique to detect potholes.

3 Proposed Methodology

Several image processing algorithms are already described in Sect. 2. From the literature survey, it is clear that, detecting pothole from moving vehicle capturing image is very difficult task. Hence, the focus of this paper is to detect pothole efficiently. In Fig. 1, step-by-step process is shown.

3.1 Input Image

Road images were captured (video) using camera placed on the vehicle. Extract frames from video to apply further process. Figure 2 describes the input steps.

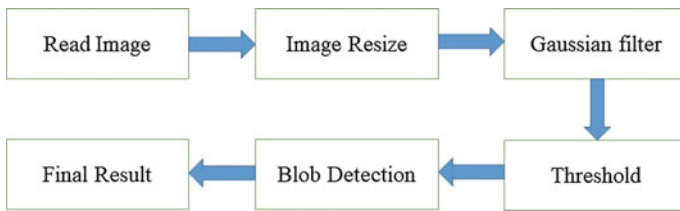


Fig. 1 Progress flow

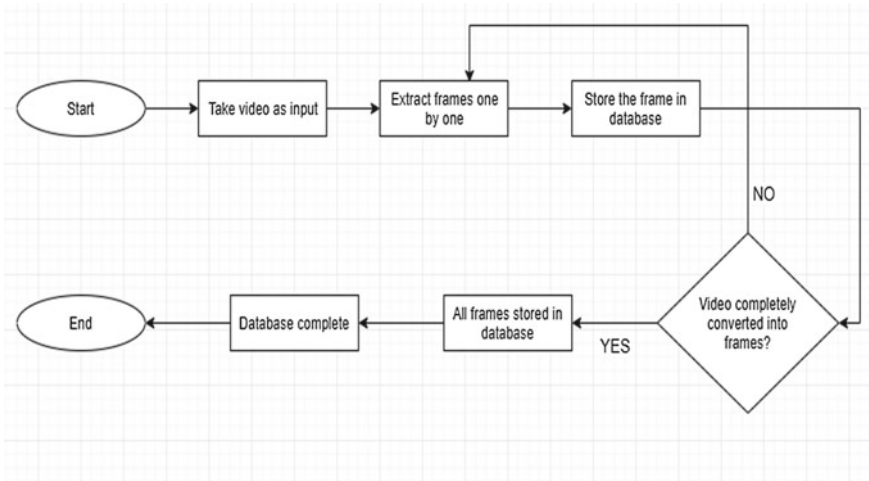


Fig. 2 Process to capture input image

3.2 Preprocessing Steps

1. Resize the input image.
2. Grayscale conversion and applying Gaussian filter.
3. Convert into binary image.
4. If pixel frequency is greater than the threshold value, then a particular value is assigned to that pixel.

3.3 Resizing

The size of captured image is very high which will increase process computing time. Hence, it is required to resize the original image size. It is done without removing the important feature of the image, i.e., the region of interest in which the pothole lies and thus it cannot be removed in this process (Fig. 3).



Fig. 3 a–d are sample input images after resize

3.4 Steps to Convert Captured Image into Bird's-Eye View (Top View)

Captured image consists of a trapezoidal region of interest. To calculate the result accurately, transformation is applied on trapezoidal region into the rectangular region. For proper vision of the region, projective transformation is used which gives the system accurate information. This process will remove unnecessary information by wrapping the transformed image into the size of the frame, which is the region of interest. It is H more efficient than Affin Transformation is that it takes only three points and our concern are the vertices of the trapezium.

Projective transformations can be represented as below:

$$M_2 = Hm_1,$$

where m_1 represents homogeneous pixel coordinate before transformation and m_2 represents homogeneous pixel coordinate after transformation.

$$H = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & 1 \end{bmatrix}$$

where H is homography matrix [14].

Projective transformation using homography matrix is written as the multiplication of affine transformation H_a and proper projective transformation H_p [15].

$$H = H_a H_p$$

where $H_p = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ h_{31} & h_{32} & 1 \end{bmatrix}$ and $H_a = H_t H_r H_s$

where

$$H_t = \begin{bmatrix} Su & 0 & 0 \\ 0 & Sv & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad H_r = \begin{bmatrix} c & -s & 0 \\ s & c & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad H_s = \begin{bmatrix} 1 & 0 & \delta u \\ 0 & 1 & \delta v \\ 0 & 0 & 1 \end{bmatrix},$$

such matrices describe the transformation of scaling, rotation, and translation. Where $c = \cos \Phi$, $s = \sin \Phi$, where Φ is the rotation angle in counterclockwise direction; Su and Sv , δu , δv are the scale factors and translation values at the horizontal and vertical directions, respectively.

From Fig. 4, it is clearly shown that using top view, more clear information of pothole is identified, which is very helpful to detect feature.

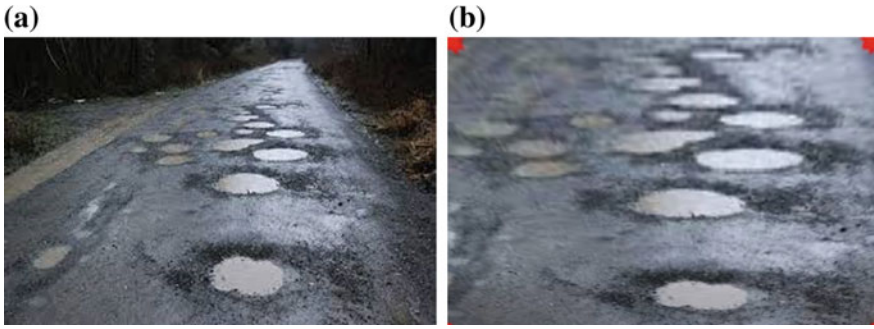


Fig. 4 a Original image before transformation and b image after transformation

3.5 *GrayScale Conversion, Filtering and, Binarization*

Median Filtering

The median filter is used to remove random noise in grayscale image and gives a smoothed output image. In addition, median filter maintains the integrity of image regions and boundaries.

Gaussian Filtering

The difference of Gaussian filter (DoG) was connected as a pre-preparing channel to unique pothole picture for better edge discovery with lessened clamor. It finds the distinction between two sigma estimations of two Gaussian profiles and discovers the edges in the grayscale picture (Figs. 5 and 6).

$$\text{DoG}_\sigma(x, y) \stackrel{\text{def}}{=} \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2 + y^2}{2\sigma^2}\right) - \frac{1}{2\pi(0.5\sigma)^2} \exp\left(-\frac{x^2 + y^2}{2(0.5\sigma)^2}\right) \quad (1)$$

Fig. 5 After filtering and gray scale applied on Fig. 4b



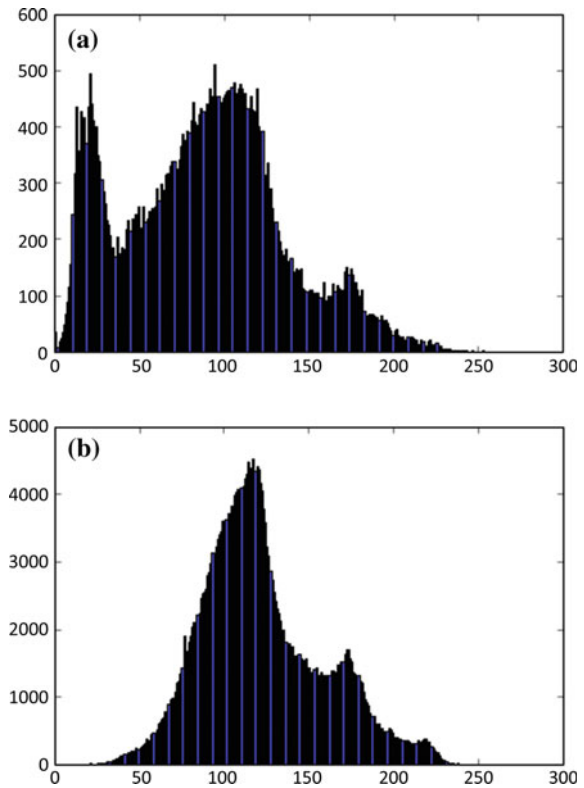


Fig. 6 After filtering and gray scale applied on Fig. 4a

3.6 Histogram Analysis

Histogram is the pen sketch of graphs or plot, which considers the frequency of grayscale image. From the histogram analysis, the effect of smoothness is clearly visible (Fig. 7).

Fig. 7 **a** Histogram analysis of Fig. 4a before filtering.
b Histogram analysis of Fig. 4a after filtering



3.7 Pothole Detection (Blob Detection)

A blob is a group of connected pixels in an image that share some common properties. In blob detection algorithm, we make use of three parameters to find blob, namely thresholding, filter by size, and filter by shape.

Thresholding: Converting image into binary where filter by size is blobs-based filter size by setting parameters (Fig. 8).

- 1 A blob is a group of connected pixels in an image that share some common property.
- 2 In blob detection algorithm, we make use of three parameters to find blob namely thresholding, filter by size, and filter by shape.
- 3 **Thresholding:** Converting image into binary form.
- 4 **Filter By Size:** Blobs based on size by setting parameters.

To display the candidate region, we use blob which is a common property shared by a group of connected pixels in an image; it will make use of three parameters, namely thresholding, filter by size, and filter by shape.

- (1) **Thresholding:** To convert the image into binary format.
- (2) **Filter By Size:** It is used to filter the blob based on size by setting the parameters e. g. **filter By Area = 1** and setting **minimum area = 100** will filter out all the blobs that have less then 100 pixels.
 - **Color.** In this parameter, comparison is done between the intensity of the binary image formed at the center of a blob to the blob color. If the difference

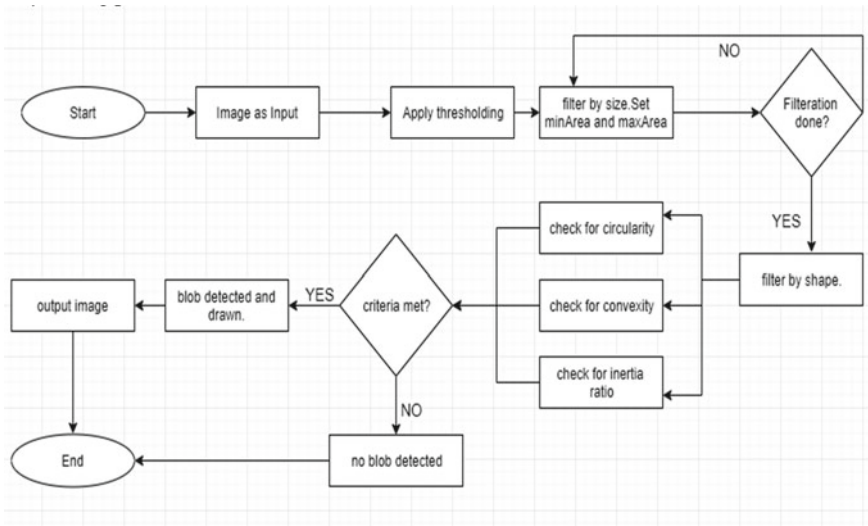


Fig. 8 Step-by-step region selection process using blob detection

occurs, then the blob is filtered out. **e. g.,** Use **blob Color = 0** to extract dark blobs and **blob Color = 255** to extract light blobs.

- **Area.** This is the bounded range of area in which blob is drawn. The minimum area (inclusive) and the maximum area (exclusive) are specified for this filter.

(3) **Filter By Shape:** Consists of further three parameters:

- **Convexity.** The range of convexity between minimum convexity and maximum convexity is specified to extract the blobs.

$$\text{Convexity} = \frac{\text{Area of the blob}}{\text{Area of Convex Hull}} \tag{2}$$

To filter by convexity, set **filter By Convexity = 1**, followed by setting $0 \leq$ **minimum Convexity** ≤ 1 and **maximum Convexity** (≤ 1).

- **Inertia Ratio:** this measures how elongated a shape is. For example, for a circle, this value is 1; for an ellipse, it is between 0 and 1; and for a line, it is 0. To filter by inertia ratio, set **filter By Inertia = 1**, and set $0 \leq$ **minimum Inertia Ratio** ≤ 1 and **maximum Inertia Ratio** (≤ 1) appropriately.
- **Circularity.** Extracted blobs have circularity between minCircularity (inclusive) and maxCircularity (exclusive).

$$\text{Circularity} = \frac{4 * \pi * \text{area}}{\text{perimeter} * \text{perimeter}} \tag{3}$$

4 Result and Analysis

Proposed algorithm is tested on 120 images. Such input images were extracted from several video frames. To measure the performance of the proposed model, *F1* score is calculated. *F1* score is obtained by the following expression

$$F1 - \text{Score} = \frac{2 * \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where Precision = $\frac{TP}{TP+FP}$ and Recall = $\frac{TP}{TP+FN}$

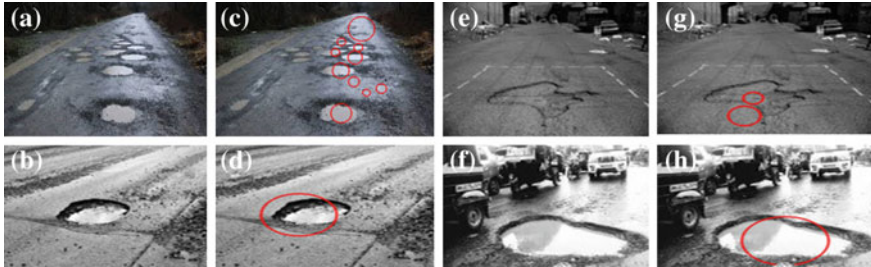
TP = True positive; FP = False Positive, FN = False Negative.

Below table describes the *F1* score for proposed method.

From the comparison Table 1, it has been shown that proposed method gives better result than existing work (Fig. 9).

Table 1 Result comparison

No. of samples	Method follows	Area measure	<i>FI</i> -score	Accuracy %
120	Proposed method	Yes	0.8433	84.83
120	Described as [1]	No	0.8060	80.60

**Fig. 9** a, b, e, f original images; c, d, g, h successful detection

5 Conclusion and Future Scope

In this paper, a pothole detection system is proposed. While detection of pothole is focused on the research, it is obvious to ensure the process with accurate result within very less computing time. Other properties like area, perimeters of hole are also identified with the help of proposed method. Proposed method tested on 120 different test samples and nearly 85% of accuracy has been achieved. Initially, the size of dataset is small, but in future, our aim is to test the same algorithm on large dataset.

References

1. K. Vigneshwar, B. Hema Kumar, Detection and counting of Pothol using image processing techniques, in *Conference Proceeding of the IEEE International Conference on Computational Intelligence and Computing Research*, 2017, India
2. C. Koch, I. Brilakis, Pothole detection in asphalt pavement images, in *Advanced Engineering Informatics*, vol 25 (Elsevier, 2011), pp. 507–515
3. M. Mandal, M. Katageri, M. Gandhi et al., Automated management of pothole related disasters using image processing and geotagging. *Int. J. Comput. Sci. Inf. Technol.* **7**(6), 97–106 (2015)
4. E. Buza, S. Omanovic, A. Huseinovic, Pothole detection with image processing and spectral clustering, in *Recent Advances in Computer Science and Networking* (Springer, 2013), pp. 48–53
5. D. Djamaluddin, A. Achmad, R. Parung, Prototype of vehicles potholes detection based blob detection method. *J. Theor. Appl. Inf. Technol.* **95**, 2509–2515 (2017)
6. I.S. Kholopov, Bird's eye view transformation technique in photogrammetric problem of object size measuring at low-altitude photography. *Adv. Eng. Res.* **133**, 318–324 (2017)
7. H. Sawalakhe, R. Prakash, Development of Roads Pothole Detection System Using Image Processing. *Lecture Notes in Electrical Engineering* (Springer, 2018), pp. 187–195

8. Z. Zhang, X. Ai, C.K. Chan, N. Dahnoun, An efficient algorithm for pothole detection using stereo vision, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '14)*, May 2014, Italy, pp. 564–568
9. A. Danti, J.Y. Kulkarni, P.S. Hiremath, An image processing approach to detect lanes, Pot holes and recognize road signs in Indian roads. *Int. J. Model. Optim.* **2**(6), 658–662 (2012)
10. E. Buza, S. Omanovic, A. Huseinovic, A. Pothole detection with image processing and spectral clustering, in *2nd International Conference on Information Technology and Computer Networks (ITCN '13)*, Antalya, Turkey. 2013; pp. 48–53
11. S.-k. Ryu, T. Kim, Y.-R. Kim, Image-based pothole detection system for ITS service and road management system, in *Mathematical Problems in Engineering*, Article ID 968361, vol 2015 (Hindawi Publishing Corporation, 2015)
12. C. Koch, I. Brilakis, Pothole detection in asphalt pavement images. *Adv. Eng. Inform.* **25**, 507–515 (2011)
13. M. Mandal, M. Katageri, M. Gandhi et al., Automated management of pothole related disasters using image processing and geotagging. *Int. J. Comput. Sci. Inform. Technol. (IJCSIT)* **7**(6), 97–106 (2015)
14. R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. (Cambridge University Press, Cambridge, 2003), pp. 87–117
15. G. Xu, Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition. A Unified Approach* (Kluwer Academic Publishers, Dordrecht, 1996), pp. 26–41

Quantitative Interpretation of Cryptographic Algorithms



Aditi Jha and Shilpi Sharma

Abstract Encryption is the method of sending messages via code. It essentially jumbles up the content of a message so that it would not make sense to a third party trying to gain access to the material. Both the sender and the recipient have the right key which encrypts and decrypts the message. It is the most effective method of securing communication. Encryption algorithms are divided into two categories: symmetric and asymmetric. In this paper, some of the most widely used encryption algorithms 3DES, AES and RSA have been implemented using C# programming language and .NET framework. After the implementation of these algorithms, they have been compared on the following parameters—encryption time, decryption time and throughput.

Keywords 3DES · AES · Cryptography · DES · RSA

1 Introduction

Security is of the paramount importance in today's technological world. Sending emails and files that contain confidential information is dangerous because these are sent in an unsecured form. Encryption algorithms are of two types: symmetric and asymmetric. Symmetric cryptography uses a single key for both encryption and decryption. In this system, both parties need to agree on a key, i.e., it has to be shared between them. Asymmetric cryptography uses two keys—a public key and a private key. Symmetric and asymmetric encryption methods are used together to combine speed with efficiency. Such a cryptosystem is known as a *hybrid cryptosystem*. A brief description about the algorithms is given ahead:

A. Jha (✉) · S. Sharma

Amity School of Engineering and Technology (ASET), Amity University, Noida, India
e-mail: aditi.jha20@gmail.com

S. Sharma

e-mail: ssharma22@amity.edu

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_41

459

1.1 3DES

DES is a 64-bits block cipher, i.e., it encrypts data 64-bits at a time. There are two inputs in DES—a plaintext (64-bits) which is to be encrypted and a key (56-bits). The output is a 64-bits cipher. *3DES* was developed in order to enhance DES without actually having to come up with a new algorithm altogether. *3DES* has a key length of $3 \times 56 = 168$ bits which increases security. The output is a 64-bits ciphertext.

1.2 AES

AES is a symmetric block cipher algorithm which means that the input plaintext is used in chunks of 128-bits. There are three types of AES, namely AES-128, AES-192, and AES-256 having key sizes of 128, 192 and 256-bits, respectively. The number of “rounds” differs according to the type of AES.

1.3 RSA

RSA was designed by Ron Rivest, Adi Shamir and Leonard Adleman of MIT. It is an asymmetric public cryptography cipher. It is the first algorithm to use two keys—a public key and a private key. A major advantage of *RSA* is that any one of the keys can be used for encryption and the opposite key is used for decryption. The public and private keys are linked using a mathematical function.

Table 1 gives a comparison between different symmetric and asymmetric algorithms on to use two keys—a public key and a private key. A major advantage of *RSA* is that any one of the keys can be used for encryption and the opposite key is used for decryption. The public and private keys are linked using a mathematical function.

Table 1 gives a comparison between different symmetric and asymmetric algorithms on the basis of structure, modification and known attacks which have been performed on them.

This paper draws a comparison between three commonly used algorithms, namely *3DES*, AES and *RSA*. The performance of these algorithms is measured on the basis of encryption time, decryption time and throughput with varying key lengths. The aim is to find the most suited algorithm for encrypting and decrypting data.

Tools and classes are provided in the .NET setup are used for simulating the algorithms and calculating time used for plaintext encryption and ciphertext decryption. This paper is organized as follows.

Literature review is described in Sect. 2 which contains observations and conclusions drawn from related work. Section 3 contains methodology which describes the system specifications on which the simulation has been performed. It also contains

Table 1 Comparison of structure, modifications and attacks on algorithms

Algorithm	Structure	Modification	Known attacks
DES	Feistel	No	Brute force attack
3DES	Feistel	From 56 to 168 bits	Brute force, chosen plaintext, known plaintext
AES	Substitution—permutation	256 key length (in multiples of 64)	Side channel attack
CAST 128	Feistel	128,256 bits	Chosen plaintext attack
BLOWFISH	Feistel	64–448 key length in multiples of 32	Dictionary attack
TWOFISH	Feistel	Key of any length up to 256 bits	Side channel attack
RC6	Feistel	124–2048 key length in multiples of 32	Brute force attack, analytical attack
IDEA	Feistel	No. 128-bits key	Man-in-the-middle attack
RSA	Factorization	Multi prime RSA, multi power RSA	Factoring the public key
ECC	Elliptic curve theory	Yes, curve dependent	Side channel attacks

a brief description about the parameters used for evaluation. Results are described in Sect. 4 which have been graphically represented. The conclusion of the paper is drawn in Sect. 5. Finally, the future scope of cryptography is contained in Sect. 6.

2 Literature Review

To gain more perspective, the comparison results obtained from other sources have been discussed here.

In [1] a comparison between DES, AES and RSA is performed on four text files. It is concluded that the encryption speed for AES is the fastest, i.e., it consumes least encryption time. On the other hand, RSA takes the longest time to encrypt the same text file. In case of decryption, AES is again the fastest and RSA the slowest. However, the encryption/decryption time difference between AES and DES is very minor.

In [2] a comparative analysis has been done on DES, 3DES, AES and RSA. The cryptographic algorithms have been simulated using VB.NET. The evaluation was performed on the following parameters: input size of plaintext, throughput and encryption time. It has been concluded that AES is better in performance as its encryption time is less than that of other algorithms. It was also concluded that power consumption of AES is on the higher side but still lower than 3DES and RSA.

The power consumed by DES is less than AES but in terms of security DES is much more vulnerable so it is not preferred.

In [3] the energy consumption of different symmetric encryption algorithms has been carried out on a hand-held device. The findings are as follows: “after only 600 encryption of a 5 MB file using triple—DES the remaining battery power is 45% and subsequent encryption are not possible as the battery dies rapidly.”

In [4] the computation time and memory usage analysis has been performed for DES, AES and RSA. The experiment was performed on five text files of different sizes. Based on the results of the experiment, it was concluded that AES and DES consume least encryption time as compared to RSA. In case of memory usage, DES consumes least memory. The time difference in encryption is negligible in DES and AES.

In [5] five encryption algorithms have been evaluated, namely DES, 3DES, AES, RSA and Blowfish (symmetric block cipher algorithm). They were implemented in Java using Eclipse IDE with the help of Java security and java crypto packages. Text files of different sizes have been used to evaluate the above-mentioned algorithms on the following parameters—encryption/decryption time, entropy—it is a measure of the degree of randomness of information. High entropy indicates better performance of cryptosystem or cryptographic algorithm. The algorithms are also evaluated on avalanche effect—the property by which the output or ciphertext of an algorithm changes significantly due to a very minor change in input (desirable attribute of cryptographic algorithms), number of bits required for encoding optimally and the amount of memory used.

It is concluded that RSA takes the highest encryption and decryption time and blowfish takes the least time.

In case of entropy, AES and Blowfish produce a very high rate of entropy. This is because AES uses numerous s-boxes and p-boxes and in Blowfish round function is used on s-array and p-array.

In terms of the avalanche effect, it is observed that AES shows the greatest avalanche effect and RSA shows the least. Thus, in cases where integrity and confidentiality need to be maintained, AES can be used.

Overall, in this paper it was observed that Blowfish takes up the least time and memory for encryption/decryption. With regard to cryptographic strength, however, AES is the most suitable algorithm.

3 Approach

For simulation purposes, 3DES, AES and RSA have been used as mentioned earlier. The simulation was performed using C# in .NET framework on HP Pavilion 15-cc5xx with the system configuration of i7-7500U @ 2.70 GHz and 12 GB RAM. Operating system used was Windows 10 (32-bits).

Windows Forms Application in .NET framework has been used for creating the GUI. The form created has the following components:

Key

A field for entering a user defined key (in case of symmetric algorithms). For RSA, the key is automatically generated by RSACryptoServiceProvider class.

Plaintext

A field which requires the user to enter the plaintext. It can contain characters, numbers and/or special characters.

Encrypt Button

The encrypt button when clicked handles the event for encrypting the plaintext entered by the user.

Encrypted Text

This field contains the ciphertext which has been derived after applying the algorithm when the encrypt button is clicked.

Decrypt Button

The decrypt button similarly handles the event for converting the ciphertext back into the plaintext.

Decrypted Text

This field contains the original plaintext which has been obtained after decrypting the ciphertext after clicking the decrypt button.

As can be seen the right side of the form contains the results are obtained in various fields.

Encryption Time

This field displays the encryption time taken by an algorithm in seconds and milliseconds.

Decryption Time

This field displays the decryption time taken by an algorithm in seconds and milliseconds.

Plaintext Size

This field displays the length of the plaintext entered by the user in bytes.

Ciphertext Size

This field displays the size of the ciphertext in bytes.

Key Size

This field displays the key size in bits.

The performance of algorithms has been carried out on the following parameters:

- Encryption time
- Decryption time
- Throughput.

Encryption time is the time taken by an algorithm to encrypt a plaintext entered by the user. Decryption time similarly is the time taken to convert ciphertext back into the plaintext.

To calculate the time in milliseconds functions provided in the system. Diagnostics namespace provided in C# (Figs. 1 and 2).

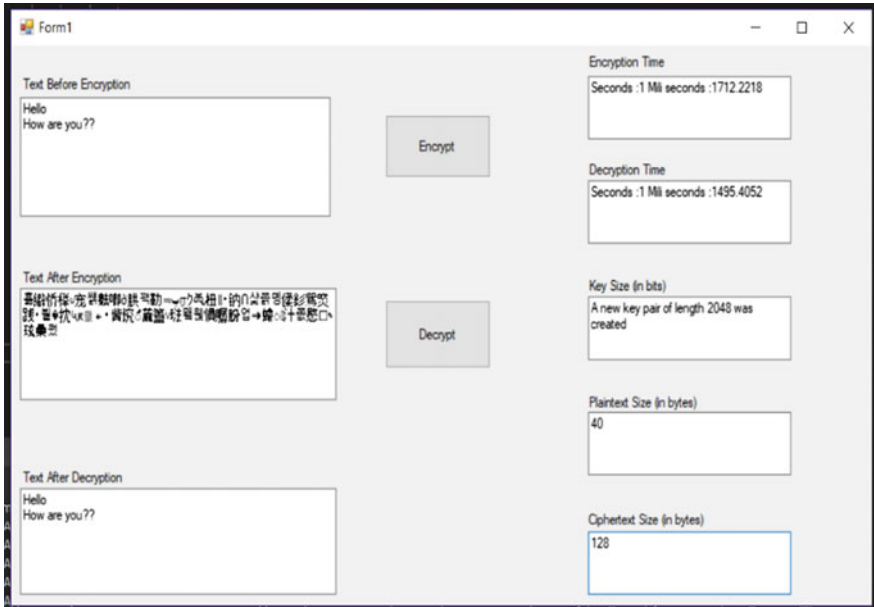


Fig. 1 Screenshot of simulation GUI where key generation is automatic

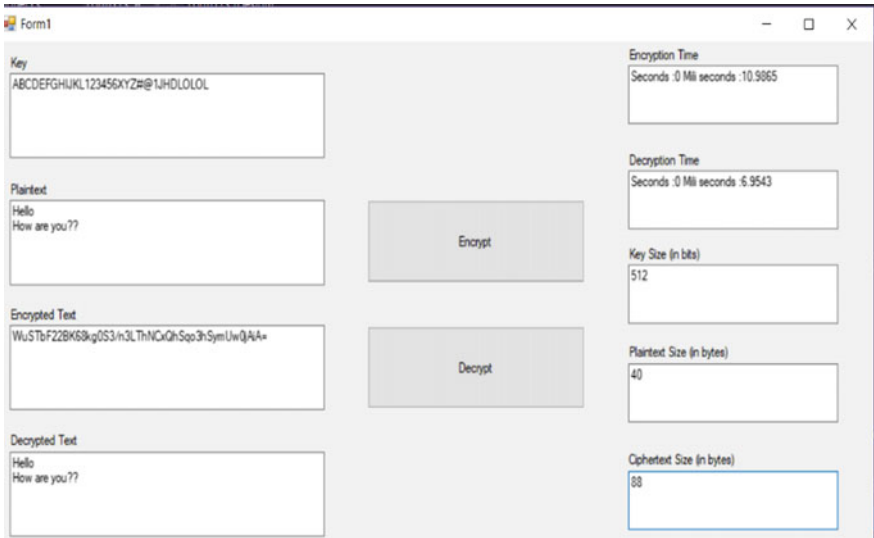


Fig. 2 Screenshot of simulation GUI where key is given by user

Fig. 3 Comparison of encryption and decryption time for key size 512 bits

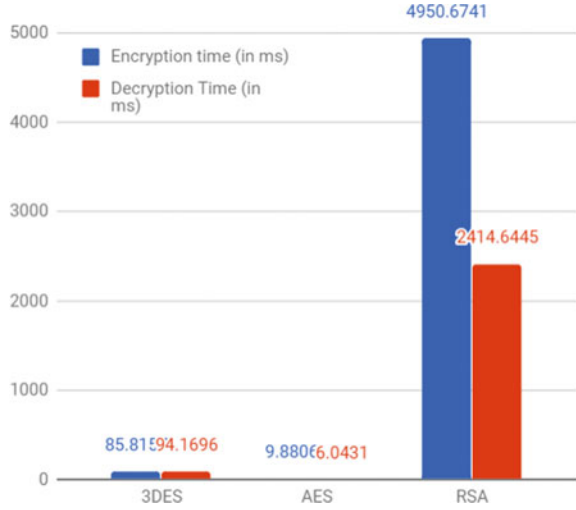
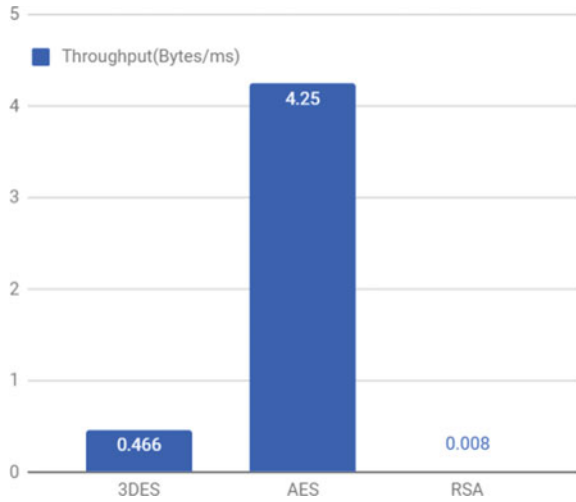


Fig. 4 Throughput (bytes/ms) for key size 512 bits



4 Implementation and Results

For a given size of plaintext, encryption time and decryption time taken by the three algorithms is depicted by the graphs shown in Fig. 3.

For different key sizes, the graphs for comparison in encryption and decryption time as well as the throughput are shown. The throughput is calculated as length of the plaintext divided by the time taken for encryption.

The following graph shows throughput in bytes/ms of DES, AES and RSA for key size 512 bits (Fig. 4).

Fig. 5 Comparison of encryption and decryption time for key size 1024 bits

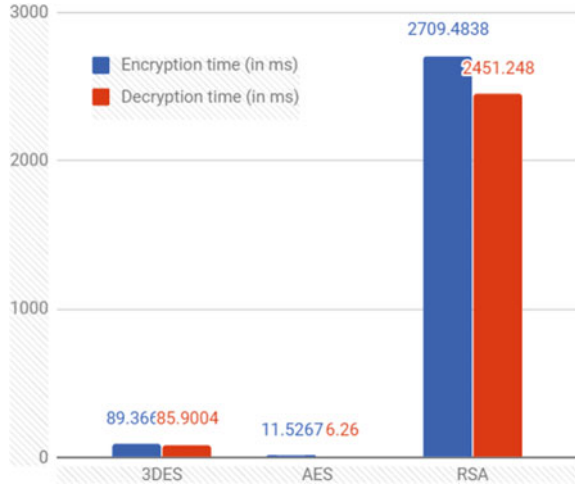
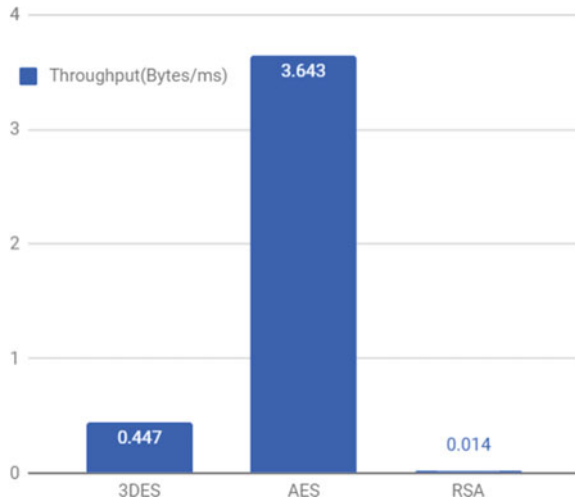


Fig. 6 Throughput (bytes/ms) for key size 1024 bits



For key size 1024 bits, the graphs comparing evaluation parameters are shown in Figs. 5 and 6.

For key size 2048 bits, the graphs comparing evaluation parameters are shown in Figs. 7 and 8.

5 Conclusion and Future Work

In the given paper, we have conducted a comparative analysis of various symmetric and asymmetric algorithms. An introduction about various symmetric and asym-

Fig. 7 Comparison of encryption and decryption time for key size 2048 bits

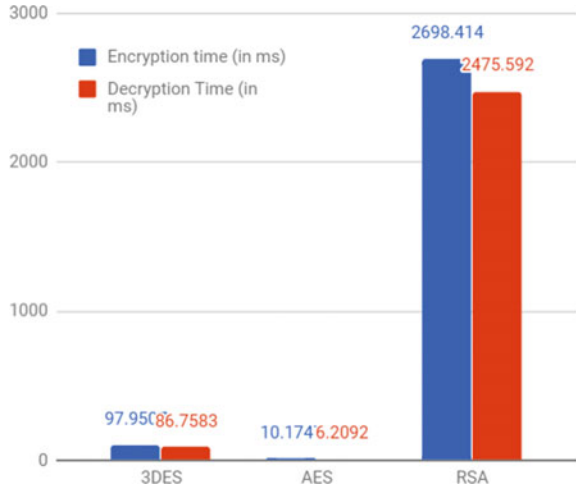
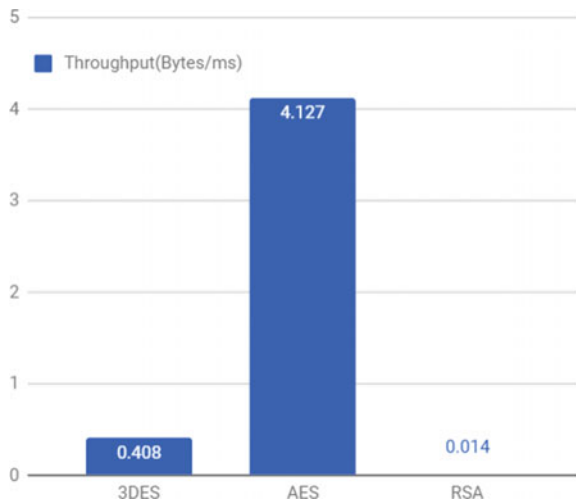


Fig. 8 Throughput (bytes/ms) for key size 2048 bits



metric algorithms along with their working has been provided. From the simulation performed above, we can see that RSA is a slower cipher than AES.

With varying key sizes, the encryption and decryption time for AES remains practically the same. In case of RSA, the encryption time reduces with increasing key size.

Practically, a cryptosystem uses both the RSA and the AES for effective encryption. Usually, RSA is used to encrypt a key (128 or 256 bits) to send over the network (such as the Internet). This key is then used to encrypt a larger data file using AES because it takes up less computing power.

In other words, RSA is used to encrypt the symmetric key which is then used in the encryption of the larger data file with AES.

The throughput is nominally affected with the change in key size. In case of AES, the throughput very marginally decreases with an increase in key size, so a larger key size decreases the speed of the algorithm. For 3DES also the decrease in throughput is very nominal.

In RSA, the throughput very marginally increases with the increase in key size meaning that the performance of RSA is enhanced with a greater key length.

The throughput for RSA is the least and that of AES is the greatest. Higher throughput indicates that AES consumes least power among the evaluated algorithms.

6 Future Scope

Homomorphic Encryption

Homomorphic encryption is a new approach in the field of cryptography. Fully homomorphic encryption would allow your data to be processed without ever having to give away access to it. For example, a particular web application could process one's tax return using encrypted financial information without really seeing it.

Honey Encryption

Another approach is honey encryption. It is particularly useful against brute force attacks. For every wrong guess, a phony result is generated. Since each guess generates a plausible result, it will be difficult for the attacker to tell when they have made a correct guess.

Quantum Cryptography

Quantum cryptography is a new and emerging field. It combines quantum physics with cryptography instead of using traditional mathematical functions to encrypt/decrypt data. The idea is to use quantum properties of photons to develop a cryptosystem which cannot be breached. The operation of photons is relatively well known and they can be used to carry in data over fiber optic cables. In practice, quantum cryptography has been used in IBM labs over short distances. The maximum distance over which this has been applied to date is 60 km using fiber optic cables. Beyond this distance, BERs or bit error rates cause the system to become unsuitable for transfer. So far transmission has only been achieved in ideal weather conditions. Further work and technological advancements are needed to make quantum cryptography a viable approach over long distances.

References

1. P. Mahajan, A. Sachdeva, A study of encryption algorithms AES, DES and RSA for security. *Global J. Comput. Sci. Technol.* (2013)
2. S. Amritpal et al. Comparative study of DES, 3DES, AES and RSA. *Int. J. Comput. Technol.* **9**(3), 1164–1170 (2013)

3. N. Ruangchaijatupon, P. Krishnamurthy, encryption and power consumption in wireless LANs-N, in *The Third IEEE workshop on wireless LANS* (2001)
4. P. Prajapati et al., Comparative analysis of DES, AES, RSA encryption algorithms. *Int. J. Eng. Manag. Res.* **4**(1), 292–294 (2014)
5. P. Patil et al., A comprehensive evaluation of cryptographic algorithms: DES, 3DES, AES, RSA and Blowfish. *Procedia Comput. Sci.* **78**, 617–624 (2016)
6. J. Thakur, N. Kumar, DES, AES and blowfish: symmetric key cryptography algorithms simulation based performance analysis. *Int. J. Emerg. Technol. Adv. Eng.* **1**(2), 6–12 (2011)
7. M. Marwaha et al., Comparative analysis of cryptographic algorithms. *Int. J. Adv. Eng. Tech/IV/III/July-Sept* 16, 18 (2013)
8. M. Vekariya, Comparative analysis of cryptographic algorithms and advanced cryptographic algorithms. *Int. J. Comput. Eng. Sci.* **1**(1), 1–8 (2015)
9. Z. Hercigonja, Comparative analysis of cryptographic algorithms. *Int. J. Digit. Technol. Econ.* **1**(2), 127–134 (2016)
10. A. Abdel-Karim, Performance analysis of data encryption algorithms (2006)
11. S.M. Seth, R. Mishra. Comparative analysis of encryption algorithms for data communication 1 (2011)
12. A.L. Jeeva, Dr V. Palanisamy, K. Kanagaram, Comparative analysis of performance efficiency and security measures of some encryption algorithms. *Int. J. Eng. Res. Appl. (IJERA)* **2**(3), 3033–3037 (2012)

Ransomware Attack: India Issues Red Alert



Simran Sabharwal and Shilpi Sharma

Abstract WannaCry ransomware attack is the latest global cyberattack which usually strikes Microsoft Windows Operating Systems, and the payment is stipulated in Bitcoin cryptocurrency. Business and public institutions are the primary targets but the private individuals are not unharmed now. The illegal activities come under cybercrime. Ransomware regularly lands as a phishing email, or spam, or a fake software upgrade—which taints the PC once the recipient clicks a link or attachment and holds the computer captive by encrypting the data and demanding a ransom payment for decrypting everything. Ransomware attack is a breach of right to personal liberty under the Indian Constitution. It is an infringement of our fundamental right to privacy in the Constitution of India. Information Technology Rules provide protection to personal information. India is a green field for cybercrime as ransomware does not come under the IT Act. There is no national cybersecurity legislation that can elaborate the responsibilities of stakeholders. As the threat has originated from beyond India, the investigations will end up at a dead end. This paper has been devised with a motive to generate awareness among people so that they can take proper preventions against attack.

Keywords Ransomware attack · Cybercrime · Cybersecurity · Crypto · Bitcoins · Malware attack · WannaCry

1 Introduction

In today's enterprises along with a keen peer competition in the business societies, there are also an increasing number of sophisticated information security threats in the cyber world since business contracts and online presence are considered as an

S. Sabharwal (✉) · S. Sharma
Amity School of Engineering and Technology (ASET), Amity University, Noida, India
e-mail: simranksabharwal11@gmail.com

S. Sharma
e-mail: ssharma22@amity.edu

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_42

essential profit-driven alley and moreover as a necessary means for global competency.

With growing Internet facilities and its competence to support businesses with great ease, organizations have started storing their critical business information on systems. These systems are constantly linked to the Internet. While singleton users of personal computers are also no exclusions. Individuals and users are inspired to store their private information on mobile devices and laptops, due to the availability of high-speed Internet facilities. Unaware of the fact that these devices might not be possessing enough security measures. This across the board and effectively available web frameworks have examined that the malware projects, for example, Trojan Horse, worms, and spyware shape a downpour of logical perspectives, and distinctive compatible procedures have been proposed to reduce and erase the digital threats [1].

Symantec specialists examined a particular attack in more detail for 4 weeks and found that 2.9% of the compromised clients paid payment, enabling the offenders to almost certainly acquire \$33,600 everyday. It implies the hoodlums would have effortlessly made \$394,000 in a month [2]. Accordingly, the quantity of ransomware assaults has expanded extensively.

Ransomware is essentially a kind of malware that makes the reports on a victim's PC remote and after that requests the victim to pay a remuneration (for the most part as Bitcoins) to achieve access to the lost records [3]. This class of malware has existed for decades, but in recent years there has been a sharp hike in the number of such cases.

Across the board use in the ongoing years has caused loss of countless millions in client misfortunes each year [4]. Exacerbating to this issue, an expanding number of law prosecution organizations have likewise been the victims of ransomware [5, 6], losing significant documents and constraining such establishments to disregard their own guidelines and pay the attackers.

The conduct of malware incorporates stealing information, sending certifications to attackers, and sending prime short message services (SMSs) [7]. To direct huge-scale attack and make benefit from it, the malware designers tend to assault the working framework with high share of the market. As indicated by the article by the International Data Corporation (IDC) [8], Android OS has monopolized the worldwide smartphone Operating Framework Market with a 78% stake at the primary quarter. Trojans stay as the fundamental portable malware compose. The TrendLabs Research reports anticipated that in 2016, the quantity of dangers will be more than twice of 2015.

2 Literature Review

The heinous AIDS Trojan [9] attack in 1990 was the first-ever crypto-ransomware attack. It was distributed through a disk handed over to the visitors at an international symposium about AIDS. The software first encrypted the file names from the system

and then displayed a demand for ransom to a place in Panama. The perpetrator's motive was more of a reprisal on the conference organizers rather than monetary gains, but in any case the attack was indecisive since a program to restore the file names was soon published. But over the years the Internet has made malware distribution uncomplicated. A few programs were available which used encryption to render a system counterproductive and claim a ransom but were easily undone since they all used the same key for all encryptions and did not shield their keys well. Then in 1996, Adam Young and Moti Yung presented a paper [10] illustrating how to set up an exclusive encryption key for each system and then lodging it with a master public key enclosed in the virus software. The climax of their mechanism was that the infected machine did not need to interact with the enforcer until the ransom was paid.

Reference [11] concludes that the remediation speed by security teams around the world has been quite remarkable. Microsoft, at the same time, released a patch which no longer backed Windows Server 2003, Windows 8, and Windows XP. Meanwhile, global security units hastened to patch the susceptible systems and shut the unprotected ports. While also the Recorded Future Intel Cards for WannaCry are being reviewed so that we can promptly identify any associated IP addresses and hashes.

According to Ref. [12], the most elementary way to cut down cyber risk is patching because of its ability to minimize the attack surface.

Anyway, the organizations ought to fix this update and numerous others as fast as possible, irrespective of its repercussion at this point. With the effect organizations are seeing, in the event that they do get infected, it is smarter to update, plug the gap, and fix a broken application than to have your frameworks held for ransom.

According to the Kaspersky Lab's Global Research and Analysis Team, their System Watcher component has been quite essential in stopping these attacks. The System Watcher component has the dexterity to rollback the variations done by ransomware in the event that a pernicious sample manages to bypass the other defense setups. This is immensely helpful in case a ransomware sample slips past the defense and tries to encrypt the data on the disk [13].

Reference [14] states that despite its capacity to proliferate so rapidly, the ransomware activities taken by the malware are not particularly novel. As discovered by security researcher MalwareTech and Talos, this malware was programmed to bail out upon a lucrative connection to that server, which would stop the malware altogether. We should all be obliged to MalwareTech for setting up the concavity, which caused this outbreak to slow down sooner than it otherwise would have. This malware is easily modifiable. As mentioned above, other analysts are already finding variants in the wild. If you're using Windows and have not patched yet, now is the high time to do it. And while you are at it, make sure to test your backups to build some confidence that you won't be forced at later stages to choose between paying up a ransom and losing data lest the worst happens to you or your organization.

According to Ref. [15], the actions and the existence of two hardcoded IP addresses (192.168.56.20, 172.16.99.5) can be well used to identify the escapade by employing network encroachment prohibition systems. Server message block (SMB) packages also store a ciphered payload, which dwells to exploit the shellcode and the file

launcher.dll. In the interim of the study, it was found that the malware is ciphered through a 4-byte XOR key (“0x45BF6313”). As announced by numerous sources, the malware dropper contains code to check to two particular areas before executing its ransomware or the system exploited codes. Further more the WannaCry employs three Bitcoin wallets to obtain payoff from its preys. An estimate can be made of approximately how much money the attackers would have made as of now by looking at the payment statements for these wallets.

It has been estimated that the attackers have earned a little over BTC 15.44 (US\$27,724.22). In spite of the fact that this is not much considering about the quantity of contaminated machines; however, these numbers are expanding and may turn out to be considerably higher in the coming days. As indicated by the reports, various associations crosswise over in more than 90 nations have been affected.

Analyst Ian Johnston stressed it was “very important” to realize that the actions taken so far only stopped one sample of the ransomware. But there is nothing stopping them from erasing the domain check and attempting again, so it is especially crucial that any unpatched system is patched as swiftly as possible [16].

According to Ref. [17], cyber-extortion methods can be etched back to the late 1980s. However, a modern wave of ransomware attacks began in 2005, and it is a kind of malicious software (malware) that attackers are spreading with the intention of not just wrecking data as was done by the traditional attacks; but encrypting and charging ransom for the services to recover the data. Ransomware is a form of scareware; that is, the user is forced to pay the ransom in response to fear of losing their data. The manifestation of this type of malware is on a hike. It is a fruitful business model for the criminal organizations that set up these outbreaks. Payment is generally done via Bitcoin, with appeals in the zone of 500–1000 USD. This supposedly “nominal” price is set to hike as time goes on, making it look alluring to pay at the earlier stages. This is a remunerative business for the cybercriminals, certain computations suggest that an approximate figure of 200 million USD annually is extorted by the criminal groups. Opinion is often given not to pay the ransom, as this supports the criminal biz model; however, it may be the only way to retrieve the lost data.

Ransomware is one of the blossoming digital threats which has accomplished consideration and thinking in the ongoing years. It is a malware that swindlers endeavor to introduce on your PC and utilize an assortment of lockout systems to counteract you access to your PC data. It at that point compels you to pay certain measure of money to reestablish the usefulness of your system. In some cases, the lockout message is shown to be from the nearby law authorities expressing that you have carried out a wrongdoing like child pornography or unapproved access to copyrighted data, compelling you to pay fine or else you are undermined to detainment.

Symantec specialists have contemplated a particular attack in detail for 4 weeks and found that 2.9% of the traded off clients paid ransom, empowering the culprits to possibly gain \$33,600 on a solitary day. It implies lawbreakers could have effectively made \$394,000 in multimonth. Due to this, the quantity of Ransomware attacks has expanded extensively since the criminals find this as a mean to earn easy and fast money. In spite of the fact that with advances made in the field of digital security, the

Ransomware assaults are observed to decrease this year, but its genuine harm cannot be dismissed [1].

According to Ref. [18], malware is a threat that will always surround technology. What started out as software to just brake or make a system stop working, which had caused an inconvenience for the end users. Now it has become a lucrative business for criminals to collect personal information and hold that information at stake for money. It comes down to a cat-and-mouse game to try to stop your personal data from becoming compromised and being stolen. While there is no 100% effective way to stop ransomware, steps can be taken to deter cybercriminals from deploying ransomware attacks. Malware can be prevented if the end users have a good understanding and take steps to protect themselves while online. Educating individuals in cybersecurity can be one of the best investments in stopping the spread of malware within an organization. Setting up security policies that can detect and stop ransomware from running and ruining their data and developing new techniques on analyzing the potential threats that come from malwares can give the users the confidence that the integrity of their data is not going to be compromised. Malware is like the flu it would always be around, but with education and steps to ensure network security, damages from malware can become more as an inconvenience rather than a catastrophe.

While exact subtle elements of how these ransomware flare-ups started are not outstanding, they generally begin when a client is misdirected into clicking a link or opening an attachment of some destructive email. A freeware that is contemplated to harm or wreck the PC is then downloaded to the client's PC aphoristically, and it quickly encodes the majority of the information on that framework and likely spreads out over the web to encrypt information on other machines also, in this way rendering all information inaccessible. The victim is then introduced to a message detailing that every one of the documents have been encrypted, and in the event that they cannot pay the requested payoff sum inside a limited period of time, all of the information would be lost.

Once an invasion has been launched, users have three primary options:

1. try to recover their lost data from any alternate,
2. pay the ransom amount, and
3. surrender their data to the attackers.

Many users pay out the ransom amount due to fear of losing their data but yet are not easily handed over their data back. Thus, organizations need to make efforts and generate awareness to prevent these attacks and recover quickly in case such a thing happens [19].

According to Krishna Chinthapalli, a neurology registrar [20] the number of ransomware attacks has quadrupled from 2015 to 2016, and so has the ransom amount been paid to hackers. In the UK, one-third of NHS institutions have reported a malware attack. Hospitals are the optimal bait for these associations. These hospitals have incorrigible medicolegal records and data for an expanding number of diurnal functions, ranging from patients' appointments to analyzing imaging and reports. Hospitals pose a higher probability than any other organizations to pay for agile

restoration of their data records. Hospitals and their employees need to keep up digital hygiene—that is, keeping hardware and programming as secure as could reasonably be expected. This incorporates representatives ending up less “click happy” when reading emails and continuous reinforcements are additionally similarly essential. They can likewise utilize tape drives, which are secured from digital hacking.

In other words, ransomware is a classification of malware that possesses some digital assets from its victims and asks for restitution in the form of Bitcoins for the release of their assets. Ransomware attacks were first witnessed in Russia in 2005–2006 and since then have largely changed their tricks and spots. The most recent upsurge of ransomware attacks is coveting targets in a very peculiar way—tracking their geographic locations and petrifying them with a hoax that fakes their respective countries’ police forces while holding their whole system and data incommunicado. These attacks are known as the “Police Trojan” attacks. The social networking scam in this case is purporting as their local police force and informing the victim that his computer is suspicious of illegal activities and thus need to pay a ransom in order to reuse it.

This phenomenon is becoming a potent landscape rather than a single detached malware incident. As the business model of ransomware attacks improves and that of fake antiviruses worsens, more criminal groups are rising on board.

Once the victim pool becomes too cognizant of this trap, the cybercriminals would most probably switch to a different trick.

Although technically the Trojan is not very progressive, it has certain evident and interesting features. Most essentially, it has been drafted to be onerous to remove. Some variants cannot be handily uninstalled even on safe booting the computer [21].

Reference [22] traces a range of maturing challenges that the controllers and law imposition organizations would need to remember. Key regions singled out incorporate structure risks, the utilization of remote and portable technologies, more refined malware, new recognizable proof and installment frameworks, PC-aided frauds, exploitation of more younger people, cerebral property intrusion, and industrial trailing. Effective pursuit and important retributions for these wrongdoings would require well-suited policing and ongoing jurisdictional revisions.

No single solution can be developed to respond to such cybercrimes.

Retaliating these perils is a multifaceted threat and demands competent strategy and synergic endeavors on the part of a large realm of government, lawmakers and the private sector bodies. Conceivable briefing for action can incorporate alluring the ICT security industry in the scheme of safe software and hardware and endowing project forces committed to the investigation and pursuit of technology-enabled scandal cases and furthermore upgrading the training and educational efficiency of police, prosecutors, and IT professionals.

As the ICT sector continues to boost, there will be an upsurge in the moments for offenders to operate illicitly. Severe and legitimate concerns are present about the modes in which modern technologies are prone to be exploited in the coming years. Technology-empowered outrages as of now go over a colossal range of exercises. These incorporate violations that concerning breaches of individual or corporate desolation, wrongdoings carried out by people that deliberately modify the information

within corporations or government firms driven by the goal of attaining revenue, individual or political intentions, and wrongdoings that draw in endeavors to rattle the working of the Internet.

Warnings have been given by Kaspersky, one of the prominent antivirus company that ransomware is a severe menace. This is because there exists no secure way to restore the encrypted data.

Ransomware can thus be defined as a fragment of noxious software that escapades a user's PC shortcomings to creep into the victim's system and cypher all his/her files; then, the attacker retains the files sealed nisi the victim consents to pay the ransom. During a quintessential attack, the attacker gets into the affected computer by probing his uncovered system amenabilities. It is easier for an attacker to get into the weakly configured computer; in case, the system had been previously attacked by a Worm/Trojan.

Attackers then look for distinct types of decisive files with extension names such as .txt, .doc, .rft, .ppt, .chm, .cpp, .asm, .db, .db1, .dbx, .cgi, .dsw, .gzip, .zip, .jpg, .key, .mdb, .pgp, .pdf. Well aware of the fact that these files are of pivotal relevance to the victims, he then cyphers these files, making them difficult for the victim to approach. Later the victim receives an e-mail or sees a pop-up window, which makes demand for the ransom that would decipher the inaccessible files.

Testing framework vulnerabilities, ransomware unflinchingly endeavors to take command over the victim's records or PC till the victim accepts the attacker's requests, generally by transferring assets as Bitcoins to the dedicated online currency accounts, for example, eGold or Webmoney stores. The future attacks would presumably be a result of combining strong cryptanalysis with malware to raid and hack information systems [23].

3 Observations

The research objectives have been postulated by analyzing the literature, the outcomes of various surveys, and their corresponding results. The literature review focused on the evolution, transmittal, and moderation methods of ransomware and the analysis of latest drifts in criminology mindsets. Further, the survey was approved with research papers written by renowned professionals.

The methodology followed in the same can be divided into three steps:

A. Survey and Interview Conduct

The foundation round of interview presented insights and inputs, which were used to settle the review questions to be interrogated. Both victims and non-victim groups were anonymously asked to fill out surveys by using the social idea as a validating file. The judgments and views of different volunteers were taped, along with their alertness and inceptive reactions to ransomware. In order to maximize the pace of the survey, a number of different instruments were used.

B. Maintaining the Cohesion of the Specifications

The vision obtained from the relevant literature details about ransomware shipment and remission modes was united with the exploration of varied moderation strategies. Explicit inspection of analytical like age, gender, and level of education was done to fathom their reliance on the losses foreseen or provoked, kind of mitigation schemes, and user awareness. The investigation was so designed to maintain the confidentiality of the information entered. The interviewees assessed the review, and on the basis of their feedback, a modified assessment was devised.

C. Scrutiny of the Survey Result

Statistical reasoning of all the variables affecting “LOSSES INCURRED” was used to inspect the survey results. The codependence of the autonomous factors (for instance, “the sort of delivery strategy” utilized and the “kind of organization”) is utilized to prescribe a few measures to control ransomware delivery and misfortunes acquired amid the same. These socio-socioeconomics are picked based on literature commendations [24]. This plans to help in proposal for alleviation of ransomware or lessening misfortunes by controlling or influencing the parameters with higher reliance.

4 Analysis

Following are the three categories in which the Malware Analysis and Detection Methods can be classified:

- Static Analysis Method
- Dynamic Analysis Method
- Hybrid Approaches.

Since the vast majority of the strategies just recognize ordinary malware which do not utilize variation systems, recent investigates have proposed a few propelled few procedures to enhance the precision in malware detection. Following are the detection algorithms for malware variants:

1. Static Analysis

This methodology incorporates the analysis of the examination based on signature, permit, and element. The signature-based component removes the semantic plans or highlights [25] and makes an elite signature relating to a particular malware. In this way, it neglects to identify the variation or obscure malware. While the permission-based method identifies dangerous permission requests to detect malware, the element-based method dismantles the APP excerpt and analyzes the definition and bytecode synergy elements to identify the vulnerabilities. Basic static analysis is straightforward and can be done quickly, but in case of sophisticated malware it is largely unaffected, since it is highly prone to miss important behaviors [26–28].

2. Dynamic Analysis

Here, malicious activities are diagnosed by setting up applications on an adversary or by using control systems [29].

This manner models users behavior and provides human-like inputs. Using this method directly in the devices may cause resource consumption. Current malware can show a wide assortment of sly systems intended to vanquish dynamic investigation including testing for virtual conditions or dynamic debuggers, postponing execution of pernicious payloads, or requiring some type of intuitive client input.

3. Hybrid Approaches

They average the advantage of the static and dynamic analysis schemes both.

5 Stages of a Ransomware Attack

The Ransomware threats have a variety of approaches to attack systems. However, there are common phases of ransomware in most of the processes. Ransomware is automated rather than receiving instructions from the host machine, infecting the system in a stealth mode. The major steps of a ransomware attacks have been broken down into major phases such as:

1. Campaign and Distribution

This is the first stage where ransomware attempts to deceive the victims to download and run attachments by using social engineering or pushing the users to visit weaponized Web sites, which lead the process to infection process.

2. Infection and Staging Process

In this step, the file initializes an installation process in the system by itself. However, the executable files set key functions in the windows registry files in order to be efficient after the system reboot and file recovery. Also, the ransomware establishes connection with a random server, or C2 server from TOR or Darknet to communicate with the hacker's infrastructure, which is quite useful in sending back the information about infected machines. Last but not the least, these files attempt to delete the shadow copy files from the windows systems.

3. Scanning and Searching for Contents

In this stage, the ransomware has already been installed and it starts to look for files and documents both locally and from the network. However, many ransomware attacks prioritize network shares over local drivers. During the scanning process, the ransomware codes leave some sort of notes from the files and directories. Moreover, it searches both mapped and unmapped network accessible systems over the networked areas for documents and shared files.

4. Encryption Process

This step is considered as one of the most challenging parts, wherein ransomware begins to encrypt all the files, which have been discovered during scanning process by using the encryption methods such as AES and RSA. In the initial stages, the ransomware files check the PHP proxy server to start the encrypting process, and in some cases, the ransomware encrypts both the extensions and content of the file. However, it deletes copies of the original files immediately. During this process, the ransomware starts to establish new connections in C2 server on the TOR network in order to get more information from the hackers and send back encryption keys for the damaged files. In addition, the connections can be used for some other purposes such as sending the instructions for the victim and navigating them as to how to get access back to their encrypted files.

5. Payday

After the attacked machine has been infected and the data has been encrypted, hackers then force the victims to pay the ransom within a stipulated period of time to restore their data. In most of the attacks, victims are usually provided with instructions to pay the ransom by sending links or locking the screens in order to get the decryption keys. The digital currency used to pay ransom is called “BITCOINS,” where each Bitcoin costs about 150 USD.

6 Result

The above graph (Fig. 1) is a depiction of cases the number of threats between 2012 and 2015. While there were observed 35% of ransomware attack in 2012, it splurged over the years and rose to 800% by 2015. The number of threats doubled from the year 2014 to 2015 [30].

The pie in Fig. 2 indicates the country most affected by ransomware attacks in the year 2016. The USA, which is the utmost advanced country in terms of computer infrastructures and is a home to most of the high-end companies, has been the most affected country giving people losses of data worth billions of dollars. Next most affected country is the Asian Country, Japan, while on the other hand countries like Russia, Italy, and Taiwan have been the least affected among the major countries [30].

Fig. 1 Graph depicting the number of threats between 2012 and 2015

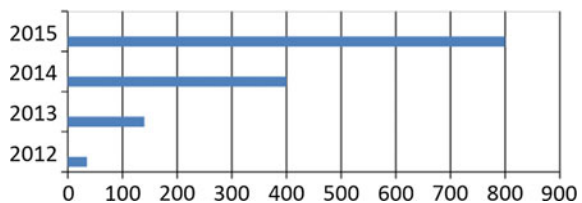
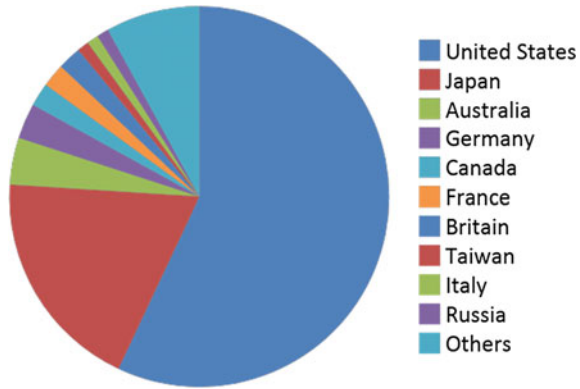


Fig. 2 Pie chart depicting the ratio of countries affected by ransomware



Malware examination is offered through a cloud service. It has built-in advantages of using pay-as-you-use services functioning over virtual platforms over the Internet with worldwide reach anywhere subordinately and not having to fear about anything in house breach over the regional network attacking users and IT servers alike.

Cloud Infrastructure offers an upper hand of not being narrowed by the hardware or computing power, thus enabling highly scalable setup. This type of framework-related composition helps present anti-malware services when enforced over periods of time, indicating and evaluating huge database and malware logs.

In addition to this, the service can be made customizable for the end users by equipping them the capacity to upload and amend logs/executable and even capture images of the infected systems. The geek users can also be offered a virtual test bed to perform their own laboratory analysis.

Another advantage of this system is its capability to apprise each user as soon as a new malicious payload is encountered.

7 Conclusion

The paper has been designed as an assessment model based on social engineering methods to restrict ransoms. In order to protect one’s data from a ransomware attack, it is required to make pre-detection, rather than post-detection.

Nowadays, cyberattacks have been going global and have affected variety of organizations and endpoint users. While, hackers use different approaches and tools including ransomware threats to take control over the targeted systems which would eventually lead to a huge damage such as in business, healthcare system, industry sector, and other fields. “Ransomware is on track to be a \$1 Billion crime in 2016.” Also, over twenty-five variants of ransomware families have been identified and over four-thousand ransomware attacks happen on a daily basis since January 1, 2016. While ransomware is not going away any time soon, you can surely defend yourself

and your organization against it—if you are well prepared. “Under the light of this idea, we work to understand the common forms of ransomware threats: identifying the root and types of ransomware and also diagnosing effective types over different platforms.” Also, how ransomware works in the systems and possible changes which can be made by ransomware.

Combating ransomware is a challenge, and all of us have to play a part in it. While creating new technologies and products we need to consider that the normal use cases are not sufficient any longer. The primary challenge for product designers is to enhance security and by taking malicious activities and scenarios into deliberation. We need to train ourselves in the basic security practices to protect our data, such as avoiding to click on malicious links or attachments and patching credulous software vulnerabilities. We need to attain more knowledge about the threats of ransomware and correspondingly take steps to prepare for and curtail hazards from these attacks.

In today’s world, we are online and always connected through the Internet. In the past, threat from malware attacks was specifically limited to desktops and laptops. But now with the advent of mobile phones there is an urgent need that we start looking at these devices and how they might pose a threat to the security of our data. Watches too can now easily link to our smartphones over WiFi and Bluetooth networks and homes are becoming more integrated with technology, while most of these smart devices hold very less data and a factory restore would easily fix the infected part. But smartphones are becoming our all-in-one devices which store all our data from our pictures to credit card information and bank details. A highlight for the future studies would be to analyze how vulnerable are these devices to grant an attacker a back door into the system. The future that technology holds would enable us to get any information at the blink of an eye, but there is always a threat of the system getting infected and the data getting lost [31].

On the one hand, ransomware can be extremely unnerving—the encrypted records can basically be viewed as irreparable damages. Be that as it may, by appropriately setting up the frameworks, it is just a mere nuisance. A few tips that can help keep ransomware from wrecking one’s day are backing up your data regularly, using a trustable site, updating your software regularly, using System Restore to retrieve a known-clean state, unplugging from suspected networks immediately. And there are many such small ways which can be implemented in daily life to save the users from the risk of a ransomware attack.

Acknowledgements This research would not have been successful without the help and support of many people around us.

This research was supported by Amity School of Engineering and Technology, Amity University, Noida. We thank our colleagues and friends who provided insight and expertise that greatly assisted the research.

We thank Kaspersky Lab for assistance with the pie chart depicting ratio of countries affected by ransomware and the graph depicting the number of threats between 2012 and 2015.

References

1. S. Thakkar, Ransomware-exploring the electronic form of extortion
2. A. Kharraz et al., Cutting the gordian knot: A look under the hood of ransomware attacks, in *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment* (Springer, Cham, 2015)
3. L. Kelion, Cryptolocker ransomware has infected about 250,000 pcs. BBC, 12/2013 (2013)
4. G. O’Gorman, G. McDonald, *Ransomware: A Growing Menace* (Symantec Corporation, 2012)
5. N. Scaife et al., Cryptolock (and drop it): stopping ransomware attacks on user data, in *2016 IEEE 36th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 2016
6. B. Fraga, Swansea police pay \$750 “ransom” after computer virus strikes. (En línea) 15 de noviembre de 2013. (Citado el: 15 de diciembre de 2013.)
7. S. Ramu, Mobile malware evolution, detection and defense. EECE 571B, term survey paper (2012)
8. S. Song, B. Kim, S. Lee, The effective ransomware prevention technique using process monitoring on android platform. *Mobile Information Systems* 2016 (2016)
9. K. Laffan, A Brief History of Ransomware (Varonis, 2015)
10. A. Young, M. Yung, Cryptovirology: extortion-based security threats and countermeasures, in *Proceedings of 1996 IEEE Symposium on Security and Privacy*. IEEE 1996
11. M. McCartney, Margaret McCartney: the NHS needs big, firm IT pants. *BMJ* **357**, j2352 (2017)
12. D.F. Sittig, H. Singh, A socio-technical approach to preventing, mitigating, and recovering from ransomware attacks. *Appl. Clin. Inf.* **7**(2), 624–632 (2016) (*PMC*. Web. 28 July 2017)
13. P. Seshagiri, A. Vazhayil, P. Sriram, AMA: static code analysis of web page for the detection of malicious scripts. *Procedia Comput. Sci.* **93**, 768–773 (2016)
14. A. Hass, C. Moloney, W.J. Chambliss, *Criminology: Connecting Theory, Research and Practice* (Taylor & Francis, 2016)
15. M. Patyal et al., Multi-layered defense architecture against ransomware
16. S. Bradshaw, *Combating Cyber Threats: CSIRTs and Fostering International Cooperation on Cybersecurity* (2015)
17. C. Moore, Detecting ransomware with honeypot techniques, in *Cybersecurity and Cyberforensics Conference (CCC), 2016*. IEEE, 2016
18. J. Wilson, and Follow Following Unfollow James Wilson, *Ransomware: Extortion Through Malware*
19. D.F. Sittig, H. Singh, A socio-technical approach to preventing, mitigating, and recovering from ransomware attacks. *Appl. Clin. Inform.* **7**(2), 624 (2016)
20. K. Chinthapalli, The hackers holding hospitals to ransom. *BMJ* **357**, j2214 (2017)
21. D. Sancho, *Police Ransomware Update*. Technical report, Trend Micro Incorporated (2012)
22. K.K.R. Choo et al., *Future Directions in Technology-Enabled Crime: 2007–09*. (Australian Institute of Criminology, Canberra 2007)
23. X. Luo, Q. Liao, Awareness education as the key to ransomware prevention. *Inf. Syst. Secur.* **16**(4), 195–202 (2007)
24. B.N. Giri, N. Jyoti, M. Avert, The emergence of ransomware, in *9th Annual Association of anti-Virus Asia Researchers (AVAR) International Conference–Digital Security: Prevention to Prosecution*, Auckland, NZ 2006
25. Y. Feng et al., Apposcopy: semantics-based detection of android malware through static analysis, in *Proceedings of the 22nd ACM SIGSOFT International Symposium on Foundations of Software Engineering*. ACM 2014
26. A.P. Fuchs, A. Chaudhuri, J.S. Foster, *Scandroid: Automated Security Certification of Android* (2009)
27. L. Lu et al., Chex: statically vetting android apps for component hijacking vulnerabilities, in *Proceedings of the 2012 ACM conference on Computer and communications security*. ACM, 2012
28. E. Chin et al., Analyzing inter-application communication in Android, in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*. ACM, 2011

29. G. Suarez-Tangil et al., Evolution, detection and analysis of malware for smart devices. *IEEE Commun. Surv. Tutor.* **16**(2), 961–987 (2014)
30. W.C. Hsieh, C.-C. Wu, Y.-W. Kao, A study of android malware detection technology evolution, in *2015 International Carnahan Conference on Security Technology (ICCST)*. IEEE, 2015
31. J. Wilson, James, and Follow Following Unfollow James Wilson, *Ransomware: Extortion Through Malware*

An Automated Dual Threshold Band-Based Approach for Malaria Parasite Segmentation from Thick Blood Smear



Debapriya Paul, Nilanjan Daw, Nilanjana Dutta Roy and Arindam Biswas

Abstract Thick blood smear examination essentially plays an important role in rapid screening of malaria parasite. In this paper, an automatic malaria parasite detector is proposed to perceive the malaria-infected blood cells in a stained thick blood smear image. The detector hence can verify presence of malarial parasites in blood cells and count the number against each white blood corpuscle. It could precisely accelerate the process of diagnosing malaria in human blood and thus reduces the chances of human errors. Here, an automated image segmentation technique has been developed that proposes to efficiently improve the accuracy over manual techniques. The experimental results show that the proposed method can provide impressive performance in segmenting the malarial parasites from a thick blood smear image on some published malaria dataset and also taken under an origami-based paper microscope. The algorithm makes use of certain preprocessing methods for comparatively better results in the proposed intensity-based segmentation technique, followed by a dual threshold band-based technique. The experimental results are comparable to other existing and traditional malaria segmentation processes.

Keywords Segmentation · Malaria diagnosis · Parasites · Biomedical imaging · Dual Canny thresholding

D. Paul (✉)

Cognizant Technology Solutions, Saltlake, Kolkata, India

e-mail: paul.debapriya20@gmail.com

N. Daw

Department of Computer Science, IIT, Bombay, India

e-mail: nilanjandaw@gmail.com

N. D. Roy · A. Biswas

Department of Information Technology,

Indian Institute of Engineering Science and Technology, Shibpur, India

e-mail: nilanjanaduttaroy@gmail.com

A. Biswas

e-mail: barindam@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling*

and Graphics, Advances in Intelligent Systems and Computing 937,

https://doi.org/10.1007/978-981-13-7403-6_43

1 Introduction

Malaria is considered as one of the most common protozoan infestations in human beings, spread over almost 91 countries. India is third among 15 countries having the highest cases of malaria and deaths due to the disease [1]. The industrialized world has mostly been free of malaria-related fatalities; it remains a major source of fatalities in developing countries where healthcare facilities are lacking. A reason for this high mortality rate is delay in the diseases diagnosis due to lack of infrastructure. The test results are also dependent upon the competence of the examiner and are thus prone to human errors. Lack of qualified technical support in rural areas only aggravates these situations. Primary diagnosis of malaria by thick smear examination is cheap and highly sensitive technique that speeds up the examination process also, advocated by World Health Organization (WHO). For such a rapid examination in a research setup, man power, human errors, and time could be reduced by using automated computational techniques. This system primarily aims at removing the technical and logistical difficulties involved in bringing healthcare facilities to the rural areas. It would speed up diagnosis, lower chances of human errors, and reduce the cost factor involved in manual testing processes prevalent in rural areas. All tests performed by the system are also automatically logged on a remote server which serves as a dataset repository. This research, which has a huge commercial impact also, was performed on around 2000 images of a Nigerian database [2]. Experiments are also performed on few real blood samples captured by foldscope [3], an origami-based lightweight pocket microscope, which is attached with mobile phone. Its remote testing facility will give accurate results within fraction of a minute. Image segmentation process plays a challenging role here which aims to assign a label for every pixel such that, in a digital image, there are pixels that can share certain common characteristics based on color, texture, or intensity. It partitions an image into separate segments that ideally correlates to distinct real-world objects, making segmentation critically important for image understanding and content analysis [4]. A particular pixel is then distinguished from the adjacent pixels based on the same computed properties. It is of primary importance in automated image analysis such that the portions of interest can be suitably extracted for a specific purpose. A wide variety of segmentation techniques has been developed over the decades which continually increases every year. A particular segmentation technique cannot be performed for all applications; hence, a detailed survey of all previous works is indispensable [5]. Further, there has been always the need for a state-of-the-art, generalized algorithm that can cater to all types of image analysis. Such image segmentation techniques use either fuzzy set theoretic approaches, some use spatial details while others use gray level histogram approach according to the paper by Shi et al. [6]. Applications based on computer vision necessitate the use of image segmentation to extract the required portions of an image. This is the primary step required to extract meaningful information from an image as required by the system under consideration. Almost all commonly used computer vision applications use effective image segmentation techniques to function properly. Some of them include content-based image retrieval systems used

by search engines to look up images based on their semantics, automated biomedical imaging systems [7] used in tumor [8] and other pathological detection [9], PET and MRI scans, tissue analysis, etc. Another field where image segmentation is widely used is in the field of automated object detection [10] and tracking like in pedestrian detection for self-driving cars and in recognition software like facial [11], iris, or fingerprint [12] recognition. Image segmentation techniques are also used in automated surveillance software like in automated traffic monitoring and video monitoring systems. Image analysis involves effective segmentation techniques viz edge detection, region-clustering, thresholding, compression-based methods, and neural networks. Fuzzy-C-Means (FCM) [13], Gaussian Mixture Model (GMM) [14], k-means clustering [15], and Active Contour Model [16] are various algorithms developed later, to improve the efficiency of the segmentation process. Some of the most common image segmentation processes merit discussion in this context.

1.1 Thresholding

The simplest method by far in the field of image segmentation is using the process of thresholding. In this method, an image is simply segmented based on the individual pixel densities. One of the most important works in this field is by Otsu [17] who proposed an algorithm which uses an intensity-based maximum variance method. Newer fuzzy logic-based methods [18] have also been developed which provides better results as compared to traditional systems.

1.2 Clustering Methods

Clustering is done in images with the aim to reduce the amount of information by clubbing or categorizing similar data items together. Graph theoretical-based methods comprise of a search for particular combinatorial structures in the similarity graph wherein clustering is a classic approach that reduces to a search for a complete subgraph, known as a clique [19]. A simple clustering method is k-means clustering which has low computational complexity than Fuzzy-C-Means (FCM). The clusters thus formed by k-means clustering do not overlap to produce better results [20].

1.3 Edge Detection Methods

Edge detection [21] is another well-known method for image segmentation. Typical digital images often consist of multiple objects separated from each other by well-marked boundaries. In such cases, the boundaries are characterized by a sharp

change in image intensity values. Many common image processing software relies on this feature to properly segment images. Two well-known edge detection-based segmentation techniques are Canny and Sobel [22].

1.4 Histogram-Based Methods

Histogram-based method is a simple segmentation technique, often used in various computer vision applications. For grouping the pixels into regions, this method uses a histogram for the selection of gray levels in an image. Histogram-based segmentation methods hugely rely on the finding of a good threshold, not taking into account the concept of connectivity, thus the coordinates of the pixels are not required for the same [23].

1.5 Graph-Based Partitioning

Graph-based partitioning is another well-known method for image segmentation. It depends on the feature that for mostly homogeneous images, pixels characteristics like intensity and spectrum values tend to have a clustering effect. Thus in this method, these clusters are modeled as nodes of a weighted undirected graph. The graph is then partitioned according to the cluster dissimilarities. Common techniques include random walks [24], minimum spanning tree-based segmentation [25], and min-cut algorithms [26].

1.6 Dual Clustering Methods

Dual clustering is a method that works on two domains, where one domain is the optimization domain while the other one caters to the constraint domain [27]. It offers object clustering in spatial and non-spatial domains that cannot be suitably detected through traditional clustering methods [28].

The proposed algorithm is a departure from the previous image segmentation techniques discussed as above. This novel method initially acquires a three-channel color image, which is sharpened using unsharp masking followed by denoising using mean shift filtering. Finally, the image is segmented using our proposed dual Canny-based algorithm coupled with a hierarchical contour detector.

2 Related Works

Sio et al. [29] show enhancement of image by adaptive histogram equalization for infected erythrocyte detection where edge contours were connected together to form closed boundaries around the erythrocytes. Then to split clumps of two or more erythrocytes into constituent cells of interest, they have used clump splitting method. Within the large boundary edges of erythrocytes, these parasites were identified by certain regions.

An image segmentation method to separate infected erythrocytes and parasites from a blood smear based on various color models and k-means clustering was proposed by Abdul-Nasir et al. [30].

To improve the efficiency of microscopic examination, an automated malaria parasite and infected erythrocyte segmentation (MPIE) method was proposed by the authors of [31]. They have used an weighted Sobel operation to compute the image gradient.

A combined algorithm for malaria parasite detection from thick smear has been proposed by Chakraborty et al. [32] where focus is given on morphological operations and color-based pixel discrimination technique.

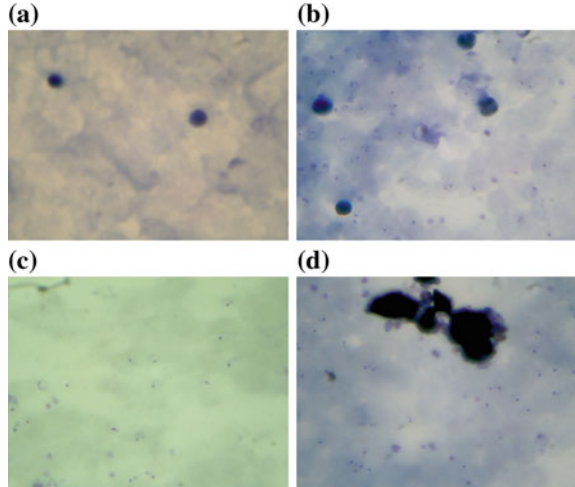
3 Proposed Algorithm

The image segmentation algorithm initially involves the acquisition of a three-channel Red-Green-Blue (RGB) color image ($I \in (m * n * 3)$). In this study, the malaria images have been captured from the thick blood smear prepared by University of Uganda [refer]. Some images are also captured by foldscope-mounted mobile phone and stained by using JSB or Giemsa staining. The acquired image is then sharpened using multiple passes of unsharp masking [33], followed by denoification using mean shift filtering [34]. The result obtained is grayscaled to a single channel ($I \in (m * n * 1)$). It is then morphologically transformed and run through a dual Canny method as will be described below. The results obtained are finally put through a segmentor to detect multiple generations of overlapped image segments (Figs. 1 and 2).

3.1 Unsharp Masking

The first phase in this multi-pass process is blurring phase. A deep copy(I') of the digital image(I) is obtained and blurred by convoluting the image pixels with a kernel ($f(x, y)$) given by the function in Eq. (2). This results in smoothing of edges and

Fig. 1 Samples of the captured malaria images



removal of pixel-level aberrations. For cells (n) in the flattened image matrix, the convolution function is mathematically represented as follows:

$$f(x, y) * I'[n] = \sum_{m=-\infty}^{\infty} f[m] \cdot g[n - m] \tag{1}$$

The Gaussian kernel can be generated as below:

$$f(x, y) = I' \cdot e^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)} \tag{2}$$

where amplitude is denoted by the pixel intensity, the center is (x_0, y_0) , and the standard deviation is defined by (σ_x, σ_y) in the x - and y -axes, respectively. The original and the blurred images are then passed through a linear blend operator with linear weights α and β and a constant argument γ to obtain a partially sharpened image I_2 . The original image is then replaced with the image thus obtained.

$$I_2 = \alpha * I + \beta * I' + \gamma \tag{3}$$

$$I = I_2 \tag{4}$$

The above process is repeated until a satisfactorily sharpened image has been obtained. The final image is then passed to the mean shift filter.

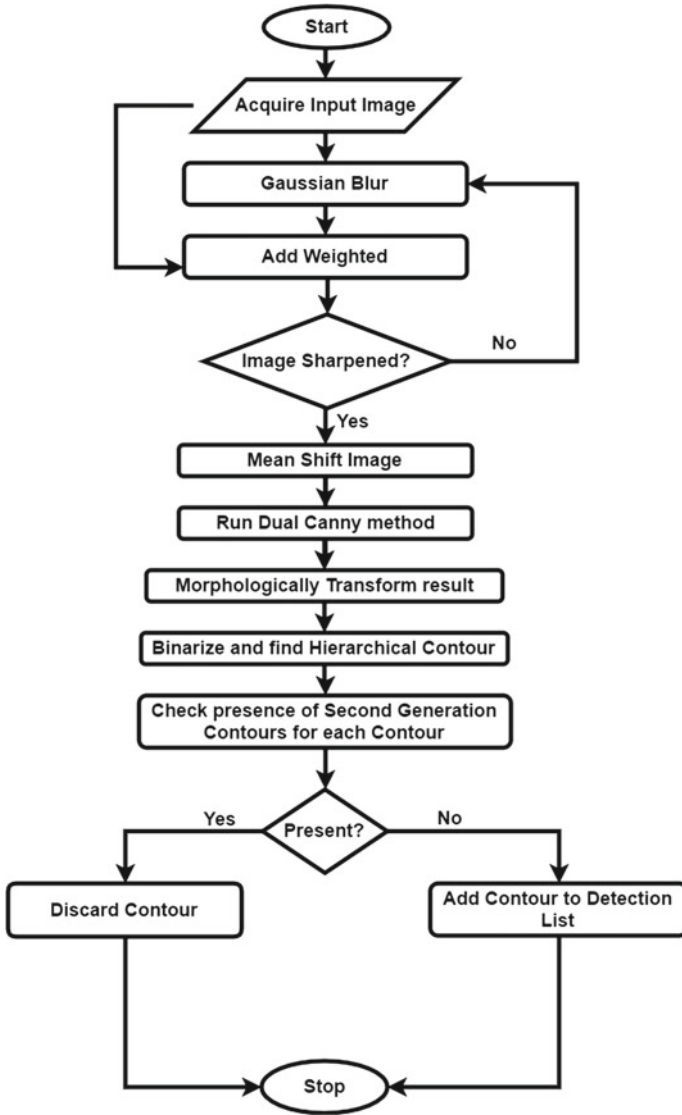


Fig. 2 The flowchart of the proposed algorithm is given as above

3.2 Mean Shift Filtering

During this phase, every pixel $(x, y) \in I(x, y)$ in the joint color hyperspace is passed through a mean shift iteration filter given by the function

$$(x', y'): x - sp \leq x \leq x + sp, y - sp \leq y \leq y + sp \quad (5)$$

where sp is the spatial window radius. After the filtering process, the initial pixel values are reassigned to the normalized values as obtained by the above procedure.

$$\forall_{i,j} I(x_i, y_j) \leftarrow (x'_i, y'_j) \quad (6)$$

The output of this filter is an image with its fine-grain texture and color gradients flattened.

3.3 Dual Canny Segmentation

Canny edge detection is a multistage algorithm used to detect edges with varying intensities. The smoothed image is run through a fast Sobel kernel both in the vertical and the horizontal directions to get the first order derivatives (G_y) and (G_x), respectively. This is then used to find the gradient and direction of each edge detected by the Sobel kernel.

$$\nabla_G = \sqrt{G_x^2 + G_y^2} \quad (7)$$

$$\theta_G = \tan^{-1} \frac{G_y}{G_x} \quad (8)$$

The gradients are then used to discard non-edge portions of the image by using a technique known as non-maximum suppression where local minimums are suppressed. Finally, a hysteresis-based algorithm is used to find connected edge locations. In our algorithm, we run this Canny segmentor twice, hence the terminology given as dual Canny segmentation. First, a low threshold pass is run on the input image to obtain an edge matrix E_{low} . This causes the system to detect even partially visible low-intensity edges. However, the low threshold band means that the system falsely extrapolates multiple disjoint segments into a single blob. This in most other studies is the main cause of failure. Another cause of error is that the low threshold band may cause the formation of multiple generations of overlapped segments in images with high Gaussian noise. Following the low threshold band run the Canny segmentor is run again with the threshold set to a very high limit to obtain the high threshold edge matrix E_{high} . This causes most of the low-intensity edges to be ignored with a high amount of fragmentation happening even within singular segments formed due

to under-extrapolation. These edge matrices are then passed on to a morphological transformer for furthering processing.

3.4 Morphological Transformation

Morphological transformation is a type of non-linear signal processing which uses mathematical set theory for the analysis of images. The edge matrices E_{high} and E_{low} , here denoted only as E , from the above module are passed through a series of morphological transformers. E is first convolved with a kernel B , such that the minimal pixel value overlapped by B is assigned as the new value to that anchor point. The following equation represents this mathematically.

$$E \ominus B = \{z \in E | B_z \subseteq I\} \tag{9}$$

where E is the obtained image and B is the kernel element. B_z is the vector Z 's translation of the kernel B .

$$B_z = \{b + z | b \in B\}, \forall z \in E \tag{10}$$

This operation is followed by a similar transformation where the maximal pixel value overlapped by kernel B is assigned as a new value of the image anchor point. This can again be formulated as:

$$E \oplus B = \{z \in E | (B^s)_z \cap I \neq \emptyset\} \tag{11}$$

Here B^s is the symmetric of B , given by:

$$B^s = \{x \in E | -x \in B\} \tag{12}$$

This causes the removal of multi-pixel aberrations or specks caused due to Gaussian noise. The image obtained at the end of this pipeline is again passed through the transformer, however, this time in the opposite direction, meaning the maximal pixel transformer acting first followed by the minimal pixel transformer. The image obtained at the end of this pipeline is mostly free of noises.

3.5 Binarization

The edge matrices are then binarized using an unsupervised method which considers the 0th and the 1st order cumulative moments of the matrices. This converts the image pixels from 8-bits to 1-bit binary values, optimizing both spatial and temporal requirements for future calculations. The threshold range for image binarization is

calculated by minimizing the weighted within-class deviation ($\sigma_w^2(t)$), where t is the threshold value calculated experimentally. The weighted, within-class deviation is calculated according to the following equation

$$\sigma_w^2(t) = q_1(t)\sigma_1^2(t) + q_2(t)\sigma_2^2(t) \quad (13)$$

where

$$q_1(t) = \sum_{i=1}^t P(i) \quad \& \quad q_2(t) = \sum_{i=t+1}^I P(i) \quad (14)$$

given, P_i is the probability of background class occurrence

Let $\mu_1(t)$ and $\mu_2(t)$ be defined as

$$\mu_1(t) = \sum_{i=1}^t \frac{iP(i)}{q_1(t)} \quad \& \quad \mu_2(t) = \sum_{i=t+1}^I \frac{iP(i)}{q_2(t)} \quad (15)$$

Then,

$$\sigma_1^2(t) = \sum_{i=1}^t [i - \mu_1(t)]^2 \frac{P(i)}{q_1(t)} \quad \& \quad \sigma_2^2(t) = \sum_{i=t+1}^I [i - \mu_2(t)]^2 \frac{P(i)}{q_2(t)} \quad (16)$$

The threshold thus obtained is applied on the matrix E which converts it into a binarized image E' . This operation is mathematically shown in Eq. 17.

$$E'(f)m * n = E'(f)_{m*n*p} * \sigma_w^2(t). \quad (17)$$

Here onwards, $E'(f)m * n$ is denoted as E for ease of representation.

3.6 Contour Detection and Segmentation

In the last and final step of the segmentation algorithm, we pass the binarized edge matrices E_{low} and E_{high} through a hierarchical contour detector to obtain the contour maps C_{low} and C_{high} , respectively. We discard all second generation contours from the C_{low} , as they have a high probability of being formed due to dual thresholding of the same boundary, as given below

$$\begin{aligned} C_{low} &\leftarrow C_{low} - \forall_i(i) \\ &\in C_{low} \mid \text{parent}(C_{low}[i]) \neq \emptyset, \text{parent}(\text{parent}(C_{low}[i])) \\ &= \emptyset \end{aligned} \quad (18)$$

Finally as the last step, a contour map C is consisting of the union of the contour maps C_{low} and C_{high}

$$C = C_{low} \cup C_{high} \tag{19}$$

4 Results and Discussion

This paper is concluded by presenting certain numerical benchmarks obtained by running our algorithm on multiple sets of objects with varied background and foreground objects of interest. We show the stepwise working of our algorithm starting with the original image, followed by other intermediate stages of the same image. Finally, we show the segmented image obtained at the end of the pipeline. In presenting our results, we will be citing standard benchmark values of specificity and sensitivity. Here, in the present context, we have worked on all the images of a standard dataset (refer makerere) published by University of Uganda. Also, we have used our own database which contains 20 images collected from JSB (Jaswant Singh Bhattacharya) stained thick blood smear slides. The database was created by capturing images by foldscope-mounted cell phone camera and staining has been done by heterogeneous group of technicians. All the results cited here were obtained by running our algorithm on a Windows-based system running on a 2.2 GHz processor with 8 GB of RAM.

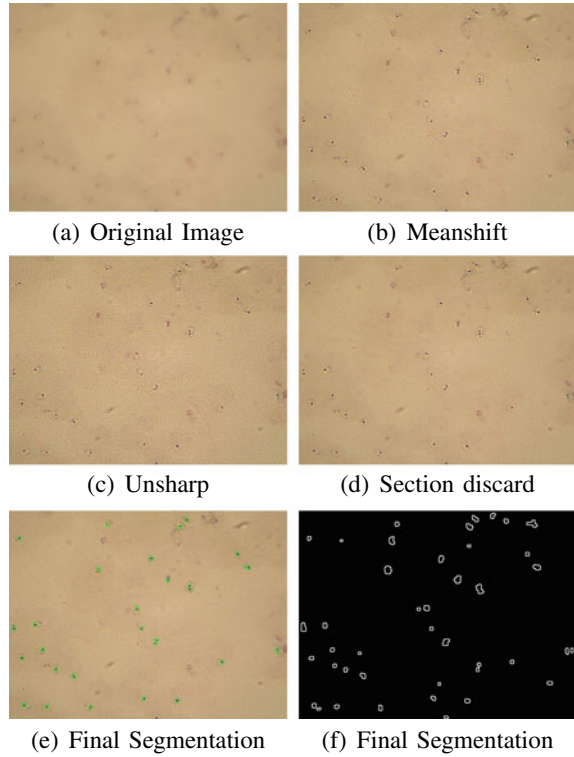
The results given in Table 1 show certain benchmark values. Column 2 gives the true positive (TP) values, Column 3 gives the false positive (FP) values, Column 4 gives the false negative values for the images indexed in Column 1 and shown in Figs. 3 and 4. The table also shows the sensitivity (S), which is calculated as the ratio between the true positives and the sum of true positives and false negatives, and the accuracy (Acc) which is calculated as ratio between the sum of true positives and true negatives, and the sum of true positives, true negatives, false negatives, and false positives.

We can see from table that for each of the cases, the accuracy and the sensitivity of the system are well above 80%. If we take into consideration the blood smear denoted in Fig. 3a, a visual investigation will immediately confirm that the enhanced image shown in Fig. 3b is more distinct with a higher overall contrast. A manual count of the cells and other stained bodies shown in the image 3f, when compared

Table 1 Table depicting the benchmark values

Image No.	TP	FP	TN	FN	S (%)	Acc (%)
2c	38	4	-	1	97.43	88.37
2f	26	4	-	-	100	86.67
2i	42	3	-	6	87.5	87.35

Fig. 3 Segmentation results obtained from the original blood smear



with the results obtained on running our segmentation algorithm, depicted in Fig. 3c provides 38 instances of true positives. However, it mis-segments four non-existent objects missing one at another place. This yields a sensitivity of 97% and an accuracy of 88%. After testing the algorithm on the whole dataset as mentioned above, similar results were obtained, of which four are shown in Table 1. The mean sensitivity obtained during the whole exercise was approximately 94.98% while the accuracy was approximately 87.46%. Similar results were obtained from the proposed algorithm when tested on the fundus photographs of the retina as mentioned in Fig. 3.

A comparative analysis of the proposed algorithm with Sobel image segmentation technique [35] and Canny image segmentation technique [36] has been given in Figs. 5 and 6, respectively (Figs. 5 and 6).

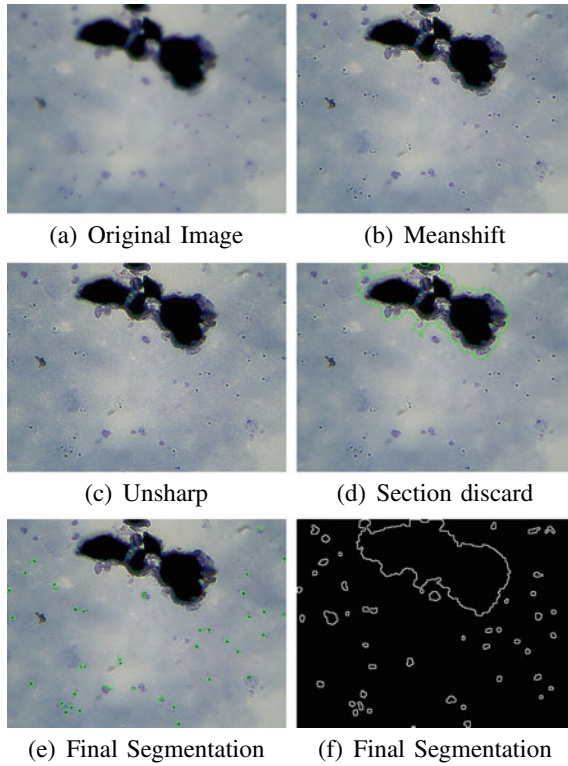


Fig. 4 Segmentation results obtained from the original blood smear

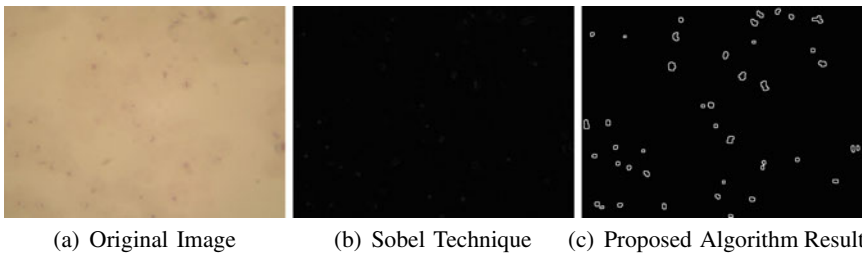


Fig. 5 Comparison with Sobel image segmentation technique by testing on the plasmodium dataset

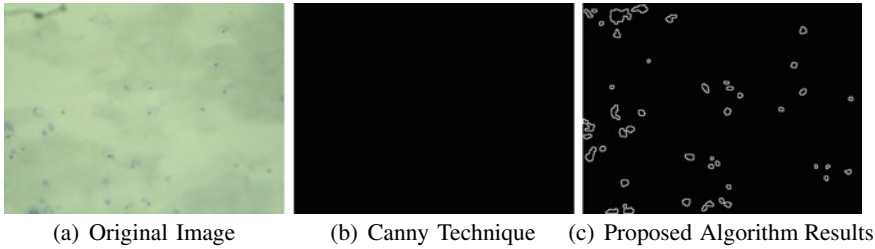


Fig. 6 Comparison with Canny image segmentation technique by testing on the plasmodium dataset

5 Conclusion

The proposed method defined in the present context is a novel generalized way of segmenting images using a dual segmentation-based technique. An image is first enhanced and then passed through a pipeline consisting of two Canny-based segmentors with varying threshold levels to successfully segment any image without any detectable loss of information. The algorithm was tested on a dataset of 2703 hand labeled malaria plasmodium infected blood smears and also on our own dataset. In each of the cases, there was a considerable improvement of the results obtained for detecting boundaries in natural images.

Acknowledgements The authors would like to acknowledge the Department of Biotechnology, Government of India, for financial support vide [ref. number No. BT/IN/Indo-US/Foldscope/39/2015] to carry out this work. We also thank NVIDIA Corporation for gifting us with Titan Xp GPU for faster implementation and testing of the proposed approach.

References

1. *The Malaria Control Programme, World Malaria Report* (2015)
2. K. National Treatment Guidelines, 2 edn. (Uganda Ministry of Health (MOH) 2003)
3. M. Prakash, *Foldscope* (Department of Bioengineering, Stanford University, Prakash Lab, 2015)
4. H. Zhang, J.E. Fritts, S.A. Goldman, Image segmentation evaluation: a survey of unsupervised methods. *Computer Vision and Image Understanding* **110**(2), 260–280 (2008)
5. Y.J. Zhang, A survey on evaluation methods for image segmentation. *Pattern Recogn.* **29**(8), 1335–1346 (1996)
6. J. Shi, J. Malik, Normalized cuts and image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **22**(8), 888–905 (2000)
7. D.L. Pham, C. Xu, J.L. Prince, Current methods in medical image segmentation 1. *Ann. Rev. Biomed. Eng.* **2**(1), 315–337 (2000)
8. W. Wu, A.Y. Chen, L. Zhao, J.J. Corso, Brain tumor detection and segmentation in a crf (conditional random fields) framework with pixel-pairwise affinity and superpixel-level features. *Int. J. Comput. Assist. Radiol. Surg.* **9**(2), 241–253 (2014)

9. S. Kamalakannan, A. Gururajan, H. Sari-Sarraf, R. Long, S. Antani, Double-edge detection of radiographic lumbar vertebrae images using pressurized open dgvf snakes. *IEEE Trans. Biomed. Eng.* **57**(6), 1325–1334 (2010)
10. J.A. Delmerico, P. David, J.J. Corso, Building facade detection, segmentation, and parameter estimation for mobile robot localization and guidance, in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1632–1639 (2011)
11. X. Chen, L. Xu, W. Wang, X. Li, Y. Sun, C. Politis, Computer-aided design and manufacturing of surgical templates and their clinical applications: a review. *Expert Rev. Med. Dev.* **13**(9), 853–864 (2016)
12. H.B. Kekre, V.A. Bharadi, Finger-knuckle-print region of interest segmentation using gradient field orientation & coherence, in *2010 3rd International Conference on Emerging Trends in Engineering and Technology (ICETET)*. IEEE, pp. 130–133 (2010)
13. O. Jamshidi, A.H. Pilevar, Automatic segmentation of medical images using fuzzy c-means and the genetic algorithm. *J. Comput. Med.* (2013)
14. Y.-C. Wu, Gaussian mixture model (2005)
15. J.A. Hartigan, M.A. Wong, Algorithm as 136: a k-means clustering algorithm. *J. R. Stat. Soc. Ser. C (Appl. Stat.)* **28**(1), 100–108 (1979)
16. F. Leymarie, M.D. Levine, Tracking deformable objects in the plane using an active contour model. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(6), 617–634 (1993)
17. N. Otsu, A threshold selection method from gray-level histograms. *Automatica* **11**(285–296), 23–27 (1975)
18. M.E. Yüksel, M. Borlu, Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic. *IEEE Trans. Fuzzy Syst.* **17**(4), 976–982 (2009)
19. M. Pavan, M. Pelillo, A new graph-theoretic approach to clustering and segmentation, in *2003 IEEE Computer Society Conference on Proceedings Computer Vision and Pattern Recognition*, vol. 1. IEEE, pp. I–I (2003)
20. H. Ng, S. Ong, K. Foong, P. Goh, W. Nowinski, Medical image segmentation using k-means clustering and improved watershed algorithm, in *IEEE Southwest Symposium on Image Analysis and Interpretation*. IEEE, (2006), pp. 61–65
21. H. WilliamThomas, S.P. Kumar, A review of segmentation and edge detection methods for real time image processing used to detect brain tumour, in *2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCCIC)*. IEEE (2015), pp. 1–4
22. M. Sharifi, M. Fathy, M.T. Mahmoudi, A classified and comparative study of edge detection algorithms, in *Proceedings of International Conference on Information Technology: Coding and Computing*. IEEE (2002), pp. 117–120
23. N. Bonnet, J. Cutrona, M. Herbin, A no-threshold histogram-based image segmentation method. *Pattern Recogn.* **35**(10), 2319–2322 (2002)
24. L. Grady, Random walks for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **28**(11), 1768–1783 (2006)
25. C.T. Zahn, Graph-theoretical methods for detecting and describing gestalt clusters. *IEEE Trans. Comput.* **100**(1), 68–86 (1971)
26. Z. Wu, R. Leahy, An optimal graph theoretic approach to data clustering: theory and its application to image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(11), 1101–1113 (1993)
27. C.-R. Lin, K.-H. Liu, M.-S. Chen, Dual clustering: integrating data clustering over optimization and constraint domains. *IEEE Trans. Knowl. Data Eng.* **17**(5), 628–637 (2005)
28. L. Jiao, Y. Liu, B. Zou, Self-organizing dual clustering considering spatial analysis and hybrid distance measures. *Sci. China Earth Sci.* **54**(8), 1268–1278 (2011)
29. S.W.S. Sio, W. Sun, S. Kumar, Z.W. Bin, S.S. Tan, H.S. Ong, H. Kikuchi, Y. Ohima, W.S.K. Tan, Malaria count: an image analysis based program for the accurate determination of parasitemia. *J. Microbiol. Methods* **68**(1), 11–18 (2007)
30. A.S. Abdul-Nasir, M.Y. Mashor, Z. Mohamed, Colour image segmentation approach for detection of malaria parasites using various colour models and k-means clustering. *WSEAS Trans. Biol. Biomed.* **10**(8), 888–905 (2013)

31. M.-H. Tsai, S.-S. Yu, Y.-K. Chan, C.-C. Jen, Blood smear image based malaria parasite and infected-erythrocyte detection and segmentation. *J. Med. Syst.* **39**(10), 118 (2015)
32. K. Chakraborty, A. Chattopadhyay, A. Chakraborti, T. Acharya, A.K. Dasgupta, A combined algorithm for malaria detection from thick smear blood slides. *Health Med. Inform.* **6**(1) (2015)
33. T. Luft, C. Colditz, O. Deussen, Image enhancement by unsharp masking the depth buffer. *ACM* **25**(3) (2006)
34. D. Comaniciu, P. Meer, Mean shift: a robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 603–619 (2002)
35. N.R. Pal, S.K. Pal, A review on image segmentation techniques. *Pattern Recogn.* **26**(9), 1277–1294 (1993)
36. S.C. Zhu, A. Yuille, Region competition: unifying snakes, region growing, and Bayes/mdl for multiband image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(9), 884–900 (1996)

A Secured Biometric-Based Authentication Scheme in IoT-Based Patient Monitoring System



Sushanta Sengupta

Abstract In this research work, the biometric-based authentication scheme for IoT-based patient monitoring system has been studied. IoT-based patient monitoring system helps the patients to enjoy the healthcare-related services sitting at remote location in their homes. Patient's privacy, safety, and security in this case are very much essential. Authentication technique in this regard is the unique selling point for establishing the safe and secure communication between the patient and the medical server. Jiang et al. have proposed and analyzed a biometric-based authentication scheme in 2017. It was seen that Jiang et al.'s scheme fails to protect the communication system against some of the vulnerable attacks like denial-of-service attack, replay attack, man-in-the-middle attack, offline password guessing attack, smart card stolen attack, forward secrecy attack, user anonymity attack, mutual authentication attack, etc. In order to prevent these security attacks, an enhanced biometric-based authentication scheme using biometric hash function and time stamping in the proposed cryptographic algorithm has been proposed. An informal security analysis of the proposed scheme is also done here. The authentication proof using BAN (Burrows-Abadi-Needham) logic is also done in this paper. For hospitals, a close and prompt monitoring might be required for the critical patients who are admitted in respective critical care units. In order to monitor and diagnose the health of the ailing patients more effectively and efficiently, the smart health application using IoT-based infrastructure with biometric-based authentication system would be required. Considering the poor physiological condition of the patients, some wearable wireless sensor devices would be a better option. The system would monitor patient's condition throughout the day and would send the different health-related parameters of the patients to the relevant doctors through sensors and human interfacing system like mobile app. The doctors sitting at the remote location would see and monitor every data of the patients through their app and can suggest relevant medicines to the patients immediately. The inherent nature of the system would reduce the lead time for the doctors to come and visit physically to the patients and diagnose. The whole

S. Sengupta (✉)

Tata Consultancy Services Ltd., Gitanjali Park, Action Area-II, Rajarhat,
New Town 700156, West Bengal, India
e-mail: Sushantasengupta199@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_44

501

architecture of IoT in this case would be requiring some biometric-based authentication scheme in order to maximize the security and safety features of the system. This also in turn will reduce any kind of vulnerabilities or risks in the system. As the patient monitoring system deals with the life and death of a patient and which is very critical and sensitive for the patient, hence a robust and secured system is very much required for this architecture to work. In this thesis, a secured and robust biometric-based authentication system for IoT-based patient monitoring system has been discussed. This scheme is suitable, secure, and effective for healthcare applications.

Keywords Internet of things · Biometric-based authentication · Burrows-Abadi-Needham · Elliptic curve cryptosystem · Attacks

1 Introduction

In today's Internet-enabled world, Internet of Things (IoT) provides a spectacular advantage of the patients at home and the medical service providers at hospitals to interact between them more efficiently and effectively in a quicker possible way. The ailing or critical patients who are physically unfit to move around smoothly and are unable to reach to the doctors at hospitals, IoT gives a tremendous facility to them to communicate with the doctors and other medical facilitators and get the necessary prescriptions and suggestions very fast so as to mitigate the emergency situations efficiently. The very basic problem lying with this system is the privacy preservation of the information and the reliability of the information flowing in the communication channel. To ensure confidentiality, integrity, and authenticity in the communication channel, a robust and realistic authentication scheme is required.

There are many two-factor and three-factor authentication schemes studied in biometric-based authentication field. Jiang et al. in 2017 studied and analyzed bio-hashing-based three-factor authentication schemes for telecare medical information system using Elliptic Curve Cryptosystem. In this thesis, Jiang et al.'s scheme was studied, and it was found that this scheme is vulnerable to denial-of-service attack, replay attack, man-in-the-middle attack, offline password guessing attack, smart card stolen attack, forward secrecy attack, user anonymity attack, and mutual authentication attack. An enhanced biometric-based authentication scheme to cater these attacks and did some validation of the scheme using BAN (Burrows-Abadi-Needham) logic is proposed in this paper. Informal security analysis of the scheme is also done in the paper.

2 Related Work

In the context of remote user authentication scheme, in 2015, Li et al. [1] in their paper have reviewed the Arshad et al.'s authentication scheme and have come up with some weaknesses of the scheme. Li et al. talked about the password guessing attack and user impersonation attacks which Arshad et al. are likely to not withstand. Biometric along with bio-hashing has been discussed in the paper in order to resist offline password guessing attack and denial-of-service attack. In July 2015, Amin et al. [2] have talked about the schemes of Lu et al.'s scheme and have come up with some improvement over the authentication scheme on telecare medical information system. Amin and Biswas have shown that Lu et al.'s scheme is prone to user anonymity, new smart card issue, patient and medical server impersonation attack and they have used an enhanced smart card-based Elliptic curve cryptographical scheme. In 2017, Jiang et al. [3] have provided few more improvements over Lu et al.'s scheme. In this paper, Jiang et al. have shown that Lu et al.'s paper fails to protect user privacy and is susceptible to identity revelation attack, identity guessing attack and tracking attack, offline password guessing attack, user and server impersonation attack. In 2016, Xie et al. [4] reviewed Wu et al.'s authentication scheme for mobile network and identified that Wu et al.'s scheme is susceptible to impersonation attacks. Xie et al. suggested a three-factor authentication scheme to fix the weakness of Wu et al.'s scheme. In July 2016, Kumari et al. [5] discussed and reviewed the authentication scheme of Amin–Biswas and identified that the scheme is vulnerable to some attacks like privileged insider attack, strong replay attack, stolen smart card attack, offline password guessing attack, and user impersonation attack. Saru Kumari et al. proposed a new RSA-based remote user authentication scheme for TMIS which overcome all security drawbacks of Amin–Biswas. In February 2017, Limbasiya et al. [6] presented a brief survey concerning various remote user authentication scheme and their vulnerabilities. Authors have also discussed possible attacks against the remote user authentication scheme.

In the context of IoT-based patient monitoring system, Rallapalli et al. [7] proposed an IoT-based healthcare system using sensors, RFID tags, and WLAN using ZigBee protocol. In 2015, He et al. [8] have discussed the system architecture and security requirements of RFID authentication system in the context of healthcare environment using ECC. Several ECC-based RFID authentication schemes have also been discussed in their paper. In 2016, Saadeh et al. [9] in their Cybersecurity and Cyber Forensics Conference paper discussed the main security issues in IoT-layered architecture. The paper also discussed some authentication techniques of IoT and classified and compared between those. In 2017, Ouaddah et al. [10] discussed on several access control in the Internet of things and the big challenges lying behind the same. The author proposed an OM-AM (objective, model, architecture, and mechanism) authorization reference model for IoT. In 2016, Quian et al. [11] have discussed a novel secure architecture for the Internet of Things. A layered architecture has been proposed for IoT and challenges of each layer have been dis-

cussed and studied. Several existing technologies have been discussed in this paper by which the proposed architecture can be secured.

3 Objective of the Research Paper

The main objective of the research paper is to devise a biometric-based secured remote user authentication scheme which provides all kind of securities against possible vulnerabilities in the IoT-based patient monitoring system in hospitals. In literature, it can be seen that biometric-based authentication system for remote user authentication is one of the best authentication techniques from security perspective. In our IoT-based patient monitoring system, similar kind of authentication process has been implemented in order to secure patient-related sensitive information from outside world.

4 Proposed Biometric-Based Authentication System

4.1 Notations

The meaning of the different notations used in this paper is described as follows:

U_i	i th remote user
ID_i	Identity of U_i
PW_i	Password of U_i
$H(B_i)$	Biometric feature of U_i
S	The server
p	Large prime number
x	The private key of server
$h(\cdot)$	Hash function
TR_i	Registration date and time of user U_i obtained from server's clock
ΔT	A legal time interval expected for transmission delay
d_u	Random nonce chosen by user
\oplus	Bitwise XOR operation
\parallel	Concatenation operation
TR	Time when user receives the message
TS	Time when server receives the message

4.2 Proposed Scheme

The proposed scheme of biometric-based authentication system consists of three phases:

- (a) Initialization Phase.
- (b) Registration Phase.
- (c) Login Phase.
- (d) Authentication Phase.
- (e) Password Change Phase.

4.2.1 Initialization Phase

The authorized server performs some computation based on some algorithmic approach to initialize the system and enable the service. In our proposed scheme, the server used elliptic curve-based cryptosystem. Server chooses a group of finite fields.

4.2.2 Registration Phase

In registration phase, user U_i wants to access the service from server. U_i will choose his own identity as ID_i and a password PW_i and input his biometric B_i . To secure his password, user chooses a random number b and performs $h(b \oplus PW_i)$ and sends a registration request to s with parameter $\{ID_i; h(b \oplus PW_i); H(B_i)\}$.

After receiving this request from user, server checks whether ID_i matches with any other ID already stored in his database if not then S stores ID_i and registration time TR_i in his database for new user entry. S computes $K_i = h(ID_i || h(x || TR_i))$,

$$A_i = h(x || TR_i) \oplus h(ID_i || h(b \oplus PW_i) || H(B_i)),$$

$$V_i = h(K_i || ID_i || h(b \oplus PW_i)), K_{pub} = xP$$

S issues a smart card containing parameter $\{K_{pub}; A_i; V_i; l; h(\cdot); H(\cdot)\}$ to U_i . When U_i receives smart card he/she stores his/her secrete parameter (b) in it (Fig. 1).

4.2.3 Login Phase

Whenever U_i wants to access the service from S , he/she sends a login request message. For this, U_i inserts his smart card in system's terminal and enters ID_i^* , PW_i^* and $H(B_i^*)$ then smart card computes the following parameter:

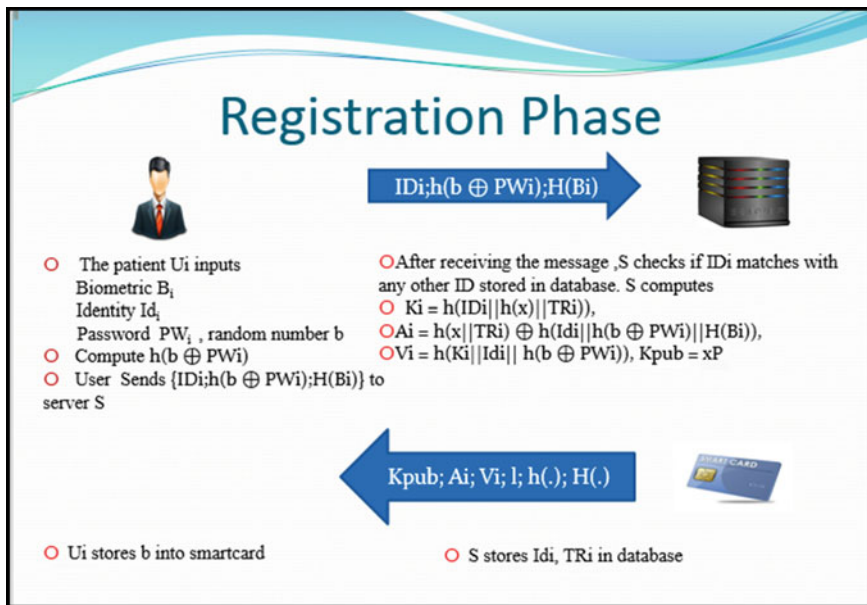


Fig. 1 Activities in registration phase

$$\begin{aligned} (H(x||TR_i))^* &= A_i \oplus h(ID_i^*||h(b \oplus PW_i^*)||H(B_i^*)), \\ K_i^* &= h(ID_i^*|(h(x||TR_i))^*), \\ V_i^* &= h(K_i^*||ID_i^*||h(b \oplus PW_i^*)) \text{ and compares} \end{aligned}$$

$V_i^* = V_i$, if successful then SC_i chooses random nonce d_u and generates login request message as follows:

$$\begin{aligned} M_1 &= d_u P \text{ and } AID_i = ID_i^* \oplus d_u K_{pub}, \\ M_2 &= h(ID_i^*||K_i^*||d_u K_{pub}||T_1) \end{aligned}$$

Smart card sends $[AID_i; M_1; M_2; T_1]$ to S at T_1 (Fig. 2).

4.2.4 Authentication Phase

After receiving the login request at time T_S , S first computes $T_S - T_1 \leq \Delta T$. If this is true, S believes that the messages is fresh and proceed to compute:

$$\begin{aligned} M_1^s &= x M_1 = x d_u P \\ ID_i &= AID_i \oplus M_1^s = ID_i \oplus d_u K_{pub} \oplus x d_u P = ID_i \oplus d_u x P \oplus x d_u P \end{aligned}$$

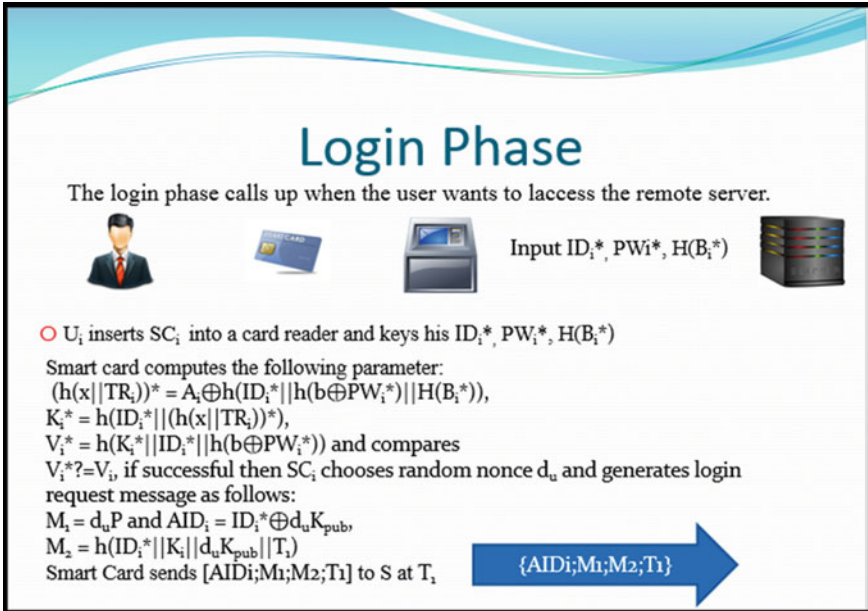


Fig. 2 Activities in login phase

The server searches the database to find ID_i which matches with ID_i . If any match found then obtain TR_i from database and compute:

$$h(x||TR_i) \text{ and } K_i = h(ID_i || h(x||TR_i))$$

$$M_2^* = h(ID_i^* || K_i || M_1^* || T_1) \text{ and checks if } M_2^* = M_2.$$

If both are same, the S believes that U_i is authentic with valid identity provided by U_i .

S chooses a random nonce d_s and computes

$$M_3 = d_s P$$

$$Z = d_s M_1 = d_s d_u P$$

$$M_4 = h(Id_i || K_i || Z || T_2)$$

S sends response $\{M_3; M_4; T_2\}$ to U_i

U_i receives the message at time T_r , first checks the freshness of the response by computing $T_r - T_2 \leq \Delta T$. If this holds good, U_i believes the response message is new and computes:

$$Z^* = d_u M_3 = d_u d_s P$$

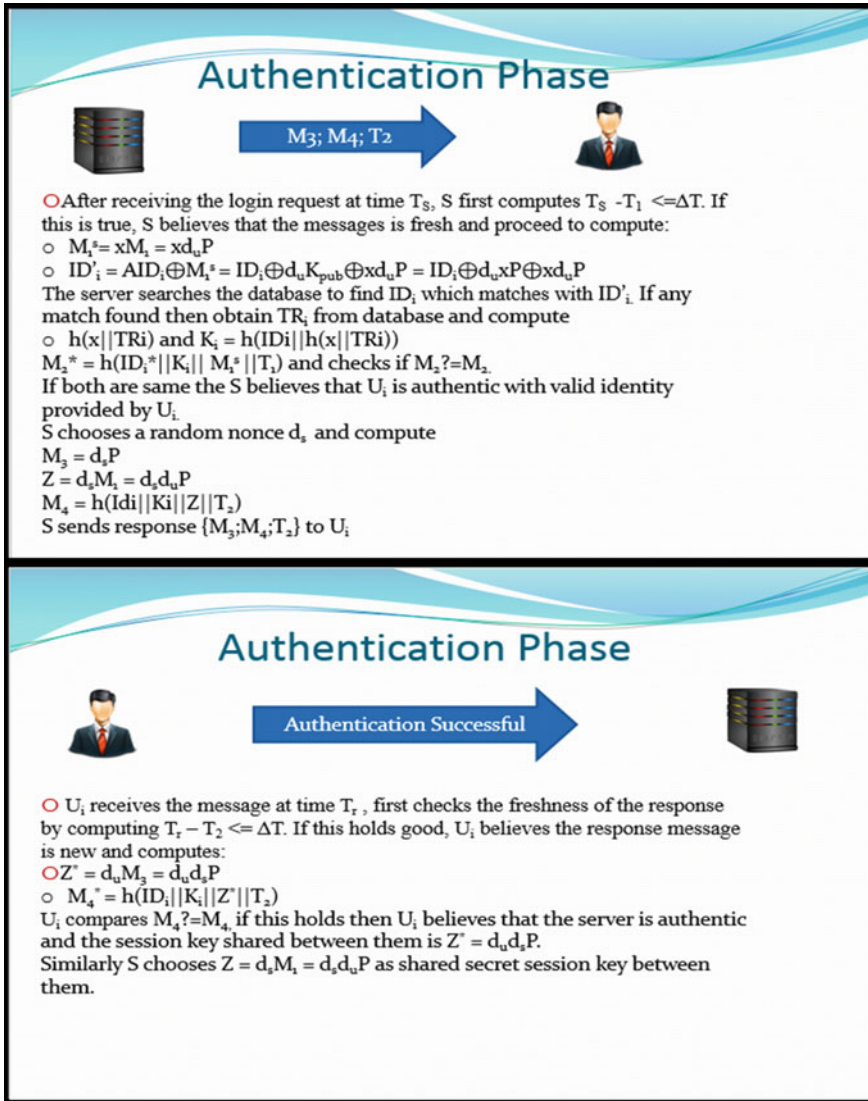


Fig. 3 Activities in authentication phase

$$M_4^* = h(ID_i || K_i || Z^* || T_2)$$

U_i compares $M_4^* = M_4$, if this holds then U_i believes that the server is authentic and the session key shared between them is $Z^* = d_u d_sP$.

Similarly S chooses $Z = d_sM_1 = d_s d_uP$ as shared secret session key between them (Fig. 3).

4.2.5 Password Change Phase

If U_i wants to change his password, he has to send some request to change it. U_i enters SC_i into the system and enters current ID_i , PW_i , and $H(B_i)$. Then SC_i will compute:

$$\begin{aligned}(h(x||TR_i))^* &= A_i \oplus h(ID_i^*||h(b \oplus PW_i^*)||H(B_i^*)), \\ K_i^* &= h(ID_i^*||h(x||TR_i))^*, \\ V_i^* &= h(K_i^*||ID_i^*||h(b \oplus PW_i^*)) \text{ and compares } V_i^*? = V_i\end{aligned}$$

If this comparison holds true, the system validates U_i , login into local system and U_i is allowed to change his password.

U_i enters new password PW_{i-new} .

SC_i computes the following:

$$\begin{aligned}A_{i-new} &= A_i \oplus h(ID_i||h(b \oplus PW_i)||H(B_i)) \oplus h(ID_i||h(b \oplus PW_{i-new})||H(B_i)) \\ A_{i-new} &= h(x||TR_i)h(ID_i||h(b \oplus PW_i)||H(B_i))h(ID_i|| \\ &\quad h(b \oplus PW_i)||H(B_i)) \oplus h(ID_i||h(b \oplus PW_{i-new})||H(B_i)) \\ A_{i-new} &= h(x||TR_i) \oplus h(ID_i||h(b \oplus PW_{i-new})||H(B_i)) \\ V_i &= h(K_i^*||ID_i||h(b \oplus PW_{i-new}))\end{aligned}$$

System stores $\{A_{i-new}; V_{i-new}\}$ in the SC_i (Fig. 4)

5 Authentication Proof Using BAN Logic

5.1 BAN Logic

The BAN logic was named after its inventors Mike Burrows, Martin Abadi, and Roger Needham. BAN logic is one of the methods for the analysis of cryptographic protocols. BAN logic makes a simple way of representing the authentication protocol verification against formal way of cryptographic protocol representation. The basis of the logic is the belief of the party in the truth of a formula. A verification in BAN logic does not necessarily imply that no attacks on the protocol are possible. A proof with the BAN logic is a good proof of correctness based on the assumptions.

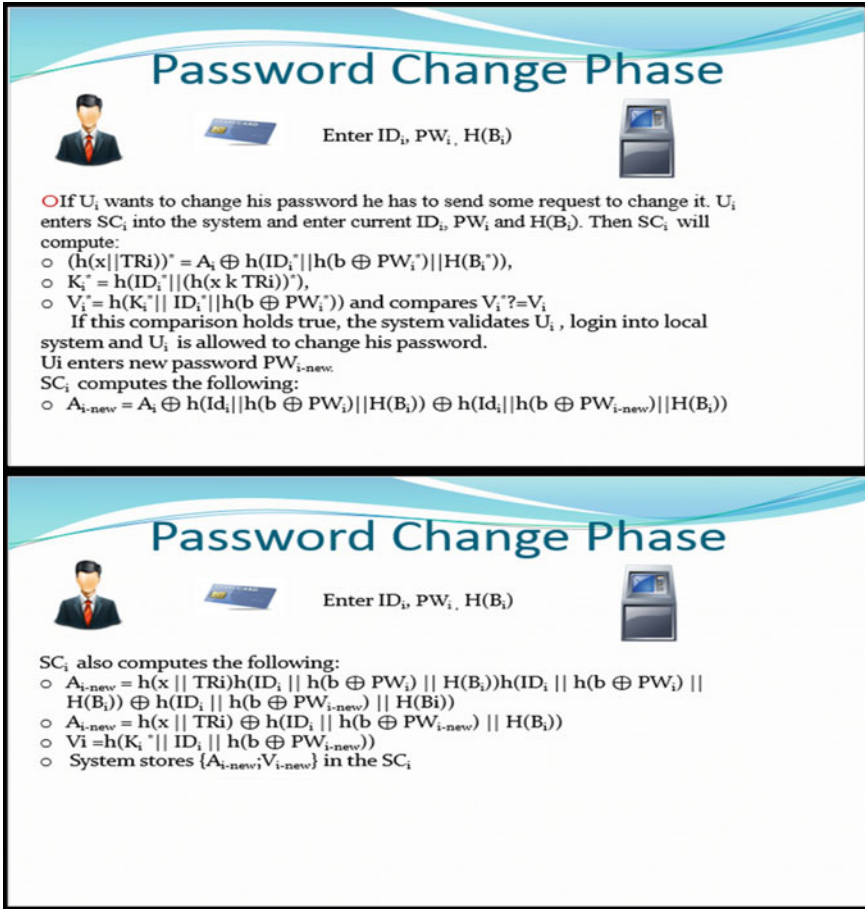


Fig. 4 Activities in password change phase

5.2 Notation Summary

- P ≡ X P *believes* X
- P < X P *sees* X
- P ~ X P *once said* X
- #(P) P is *fresh*
- P ⇒ X P has *jurisdiction over* X
- P \xrightarrow{K} Q K is a *good key* for communicating between P and Q
- $\overset{K}{\leftarrow}$ P P has K as a *public key*
- P $\overset{X}{\leftarrow}$ Q X is a *secret* known only to P and Q
- ⊕(X)_Y X *combined with* (secret) Y

5.3 Different Steps on Authentication Proof Using BAN Logic

Step 1. Goals of proposed scheme

1. $U_i \models (U_i \xleftrightarrow{SK} S)$
2. $U_i \models S \models (U_i \xleftrightarrow{SK} S)$
3. $S \models (U_i \xleftrightarrow{SK} S)$
4. $S \models U_i \models (U_i \xleftrightarrow{SK} S)$

Step 2. Authors have consider certain assumption about proposed scheme:

1. $A_1: U_i \models z\{T_1, T_2, d_u\}$
2. $A_2: S \models z\{T_1, T_2, d_u\}$
3. $A_3: U_i \models U_i \xleftrightarrow{K_i} S$
4. $A_4: S \models U_i \xleftrightarrow{K_i} S$
5. $U_i \models \{ID_i, d_u, T_1\}$
6. $S \models U_i \models \{ID_i, d_u, T_1\}$
7. $S \models \{d_u, T_2\}$
8. $U_i \models S \models \{d_u, T_2\}$

Step 3. Idealized form of proposed scheme is as follows:

1. *Message1*: $U_i \rightarrow S: \{AID_i: \langle ID_i \rangle_{K_{pub}}, M_1: \langle d_u \rangle_p, M_2: (ID_i, d_u)_{K_i}, T_1\}$
2. *Message2*: $S \rightarrow U_i: \{M_3: \langle d_u \rangle_p, M_4: (ID_i, M_1, d_u)_{K_i}, T_2\}$

Step 4. The proof of proposed scheme using BAN logic:

U_i sends login request *message1* to S .

1. $S_1: S \triangleleft (AID_i: \langle ID_i \rangle_{K_{pub}}, M_1: \langle d_u \rangle_p, M_2: (ID_i, d_u)_{K_i}, T_1)$
According to assumption A_4 , S_1 & message meaning rule.
2. $S_2: S \models U_i \mid \sim M_2$
According to assumption A_2 & freshness conjuncatenation rule.
3. $S_3: S \models \#M_2$
According to assumption S_2 , S_3 & nonce verification rule.
4. $S_4: S \models U_i \models M_2$
According to assumption A_6 , S_4 & jurisdiction rule
5. $S_5: S \models M_2$
According to statements S_3 , S_4 & Session key introduction rule
6. $S_6: S \models (U_i \xleftrightarrow{SK} S)$ **Goal 3** is achieved
According to statement S_2 , S_3 , S_4 & nonce verification rule
7. $S_7: S \models U_i \models (U_i \xleftrightarrow{SK} S)$ **Goal 4** is achieved
After successful authentication of U_i , S sends response *message2* to U_i .

8. $S_8 : U_i \triangleleft (M_3 : \langle d_t \rangle p, M_4 : (ID_i, d_s)_K, T_2)$
According to assumption A_3 , S_8 & message meaning rule
9. $S_9 : U_i \models S \mid \sim M_4$
According to assumption A_1 & freshness concatenation rule
10. $S_{10} : U_i \models \#M_4$
According to Statement S_9 , S_{10} & nonce verification rule
11. $S_{11} : U_i \models S \equiv M_4$
According to assumption A_8 , S_{11} & Jurisdiction rule
12. $S_{12} : U_i \models M_4$
After this verification smartcard authenticates S and according to assumption S_{10} , S_{11} & session key introduction rule
13. $S_{13} : U_i \models (U_i \xrightarrow{SK} S)$ **Goal 1 is achieved**
According to statement S_9 , S_{10} , S_{11} & nonce verification rule
14. $S_{14} : U_i \models S \models (U_i \xrightarrow{SK} S)$ **Goal 2 is achieved**

6 Informal Security Analysis

The informal security analysis ensures that the proposed biometric-based authentication scheme has the ability to ensure different security attacks including the attacks found in Jiang et al.'s scheme. The following attacks are based on assumptions that a malicious attacker completely controls the whole communication channel connecting the patient and the telecare server in login and authentication phase. So, attacker can eavesdrop, modify, insert, or delete any message transmitted via public channel.

6.1 Denial-of-Service Attack

Whenever U_i inserts his/her smart card into local system to generate login request message, to resist against DOS attack proposed scheme has used verification of ID_i , PW_i , and $H(B_i)$. If user enters any of these parameters incorrectly, the result $h(x \parallel TR_i)$ will be wrong. Hence, next computation of K_i will result in a wrong output and V_i verification will not match. During a certain period, if U_i is not able to enter his/her parameter correctly then local system suspends the session.

6.2 Replay Attack

In proposed scheme to thwart replay attack, global synchronized time stamp is used. Also use of random nonce provides more security from any kind of modification in the message. Suppose attacker tries to modify login request message $\{AID_i; M_1;$

$M_2; T_1\}$ by replacing time T_1 as T_a and send $\{AID_i; M_1; M_2; T_a\}$ to S . When S computes $T_s - T_a \leq \Delta T$ successfully passes. Next server computes $M_2^s = h(\text{ID}_i || K_i || M_1^s || T_a)$. Then compare $M_2^{*?} = M_2$, this comparison will fail because $M_2 = h(\text{ID}_i^* || K_i^* || d_u x P || T_1)$. Therefore, proposed scheme resists against replay attack.

6.3 Man-in-the-Middle Attack

If attacker tries to impersonate as legal user U_i using a false ID_a and a random nonce d_a , computes $M_1^a = d_a P$, $AID_i^a = \text{ID}_a \oplus d_a K_{\text{pub}}$, to compute $M_2^a = h(\text{ID}_a || K_i^a || d_a K_{\text{pub}} || T_a)$ attacker has to compute K_i^a which requires x and TR_i , where x is known only to the server and TR_i is unique for each user. Therefore, attacker will not be able to compute K_i^a . Somehow if attacker compute false K_i^a and sends login request message. When server retrieves ID_a from AID_i^a and if matches with any user stored in database then also due to uniqueness of TR_i for each user K_i will differ, hence $M_2^{*?} = M_2^a$ will fail.

If the attacker tries to impersonate as a legal server by sending response message $\langle M_3; M_4; T_2 \rangle$ by modifying time stamp T_2 as T_a and random nonce d_s as T_a attacker will compute $M_3^a = d_a P$, $Z_a = d_a M_1 = d_a d_u P$ now to compute $M_4^a = h(\text{ID}_i || K_i || Z_a || T_a)$ attacker requires to know ID_i and K_i which is infeasible to compute without knowing x and TR_i . Therefore, attacker cannot impersonate as a legal server.

6.4 Offline Password Guessing Attack

Assume, U_i loses smart card or stolen by the attacker. Attacker retrieves all the parameters stored into smart card, i.e., $\{A_i; V_i; K_{\text{pub}}; l; h(\cdot); H(\cdot)\}$

where

$$A_i = h(x || \text{TR}_i) \oplus h(\text{ID}_i || h(b \oplus \text{PW}_i) || H(B_i)),$$

$$V_i = h(K_i || \text{ID}_i || h(b \oplus \text{PW}_i)), K_{\text{pub}} = xP, l, h(\cdot), H(\cdot)$$

Now attacker guesses ID_i, PW_i but cannot enter biometric which matches with the biometric feature of original U_i and also does not know x and TR_i therefore attacker will not be able to compute A_i^* to match with A_i and without retrieving K_i attacker will not be able to compute V_i and compare with V_i to be assure that he/she has guessed correct ID_i and PW_i .

Suppose attacker has intercept login request message $\{AID_i; M_1; M_2; T_1\}$ to S at time T_1 and server's response message $\{M_3; M_4; T_2\}$ and try to guess ID_i from $AID_i = \text{ID}_i \oplus d_u K_{\text{pub}}$ but attacker does not know the value of random nonce d_u .

6.5 *Smart Card Stolen Attack*

Suppose attacker steals smart card of U_i and retrieves all the parameters stored into smart card, i.e., $\{A_i; V_i; K_{\text{pub}}; l; h(\cdot); H(\cdot)\}$, then he tries to launch login request message $\{AID_i; M_1; M_2; T_1\}$. However, without knowing x , TR_i he/she will not be able to generate K_i , therefore, it is impossible to compute M_2 . So, smart card loss attack is not possible in our scheme.

6.6 *Forward Secrecy Attack*

Assume that server's long-term secret key x disclosed accidentally. Since in proposed scheme shared secret is computed using random nonce chosen by both parties, i.e., $SK = d_u d_s P$ therefore, it is not possible to reveal all previous session keys.

6.7 *User Anonymity Attack*

A system must provide user anonymity to provide user privacy. In proposed scheme, PW_i and $H(B_i)$ is not sent over public channel during login and authentication phase. ID_i , is not sent in direct form over the public channel. Any personal data like ID_i , PW_i , and $H(B_i)$ is not stored in the smart card in their direct form. Therefore, our scheme provides user anonymity.

6.8 *Mutual Authentication Attack*

The proposed scheme provides mutual authentication between two parties, i.e., U_i and S using challenge and response method. U_i sends a random number d_u as a challenge to S along with $K_i = h(ID_i || h(x || TR_i))$. When the server retrieves K_i is true then confirms that user is authentic. Then S sends response d_u back to U_i along with a challenge d_s . Only legal server can compute K_i and retrieve d_u using secret key of S therefore, when U_i finds that response is correct then he/she authenticates the S . Server knows that if the user is legal then only he/she will be able to retrieve challenge d_s and compute session key successfully.

7 Proposed IoT-Based Patient Monitoring System

Functions of a smart system consist of three parts—sensing, actuation, and controlling. The smart system builds the decision depending on the current obtainable information. The system does some predictive analysis of the data by executing some clever movements. In the healthcare segment, the IoT-based canny technologies help in diagnosis, prevention, and patient monitoring system very simple. The system provides the eminence in facility and also diminishes the healthcare system's cost as a whole.

The smart RFID tags are attached to the body of the patients. These smart tags consist of several biosensors which sense the temperature, blood pressure, blood sugar, heartbeat, etc. of the patient. In Figure-8, flowchart of smart RFID tag is pictorially depicted. Once the patient arrives at the hospital, the RFID sensors sense the functional signals and subsequently these signals are transformed to electrical signals. The equivalent electrical signal is then transformed into digital signal. The digital information is then kept in RFID.

From the local server, the information that is stored is transferred to the medical server through WLAN. While transferring the information to the medical server, the system identifies if the patient has any past health records or not. If the server does not have any old health record then a new individuality is created in the server and it stores the data in its backend tables. This data is then moved to the medical practitioner.

After the doctor obtains the detail health information of patients from the medical server, he tries to diagnose the problem. If the condition of the patient is stable, then the doctor prescribes next course of action, which is available in the local server. However, if the patient is in critical condition, doctor prescribes the patient surgery or relevant tests accordingly and monitors his/her health.

When patient's condition is not normal, the doctor suggests operation. Here patient's treatment is performed in two levels-

- Pre-surgery monitoring.
- Post-surgery monitoring.

7.1 Pre-surgery Monitoring

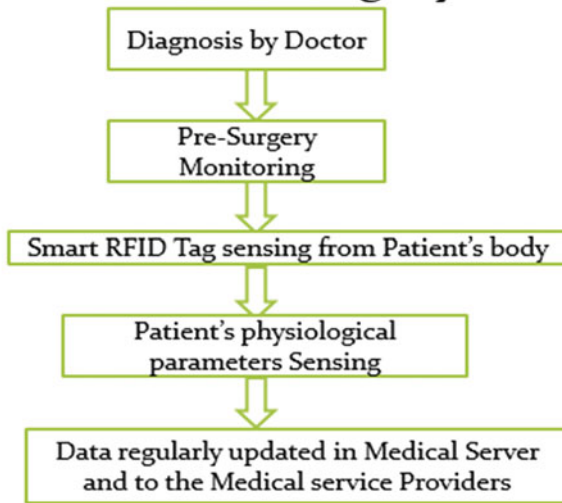
With the use of RFID tag, the patient is monitored continuously in pre-surgery monitoring phase. Smart RFID tags consist of different sensors to track blood pressure, ECG, sugar level of blood, temperature of human body, etc.

- **ECG Sensor:** ECG sensor is one of the sensors in smart RFID tag. Using some disposable electrodes on the chest of the patient, this sensor is usually attached to the body of the patient. This sensor senses the biological signals which are generated during human muscle contraction and expansion. The physiological

signal received from the human body is transformed into an electrical signal. This electrical signal is passed through analog-to-digital converter (ADC), filter and different amplifier and ultimately is sent to the local server.

- **Blood Glucose Sensor:** One of the major biosensors is the blood glucose sensor. Glucose sensors are tiny electrodes which are inserted under patient’s skin. The glucose sensor senses the physiological signal and converts it to digital signal which is the digital data. This digital data is diffused to the server where medical information is stored. Continuous glucose monitoring (CGM) is a blood sugar sensor used to monitor the glucose level throughout the day.

Flowchart of Pre-Surgery Monitoring



7.2 Post-surgery Monitoring

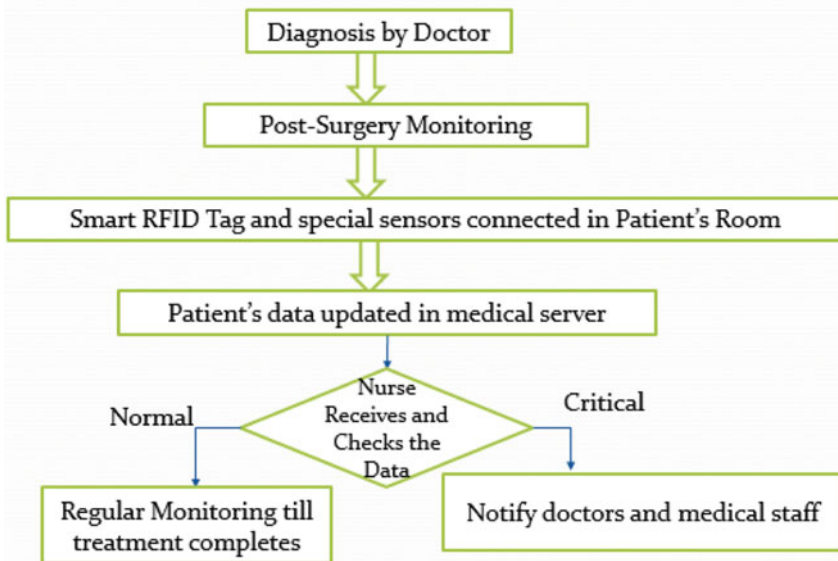
Post-surgery monitoring of a patient is a vital part. Recovery of the patient is highly dependent on post-surgery care and the intensity of anesthesia provided to the patient during operation. The health condition of the patient will very much depend on how closely the post-surgery monitoring process is dealt. Post-surgery monitoring rooms are smartly connected with sensors connected everywhere. Every patient is attached to RFID tags and sensors which not only sense temperature, blood pressure, etc., but they also have special sensors as follows:

- **Dehydration Sensors:** Dehydration sensor is a wearable sensing device which senses when dehydration of the patient body happens above a threshold level. The sensor is attached as wristband in the hand of the patient and it consists of very small and micro-needles which are attached to the skin to measure the amount of fluid inside the body.

- **Pulse Rate Sensor:** The pulse rate sensor which is also equivalent to a heart-bit sensor is used to measure the number of pulses produced by the heart. The basic principle behind this sensor is the change in luminous intensity which is actually proportional to the rate of change in flow of the blood.
- **Flow Sensor:** The flow sensor actually senses the rate of fluid flow inside the body. Infrared sensors are used to identify the droplets generating from the saline and it in turn warns the medical staff when it exceeds some edge level.

Sensors sense the physiological signals and they are converted into electrical signal. The analog electrical signal is then converted into digital signal. This digital signal or digital record is stored in RFID tag. Medical staff receives the sensed information from the RFID tag uninterruptedly without any bodily observation.

Flowchart of Post-Surgery Monitoring



8 Conclusions and Future Scope

In this paper, the security of Jiang et al.'s biometric-based authentication scheme was discussed, and it was discovered that the scheme is vulnerable to some of the security attacks. After this, the researcher analyzed and proposed the scheme and described how this scheme will protect the communication channel from denial-of-service attack, replay attack along with other attacks like man-in-the-middle attack, offline password guessing attack, smart card stolen attack, forward secrecy attack, and user anonymity attack. It was also discussed how the proposed scheme does the mutual authentication. The proposed scheme also facilitates the password change

phase of the registered user. Some authentication proof using BAN (Burrows-Abadi-Needham) logic is also done here. The biometric-based authentication scheme in an IoT-based patient monitoring system has been discussed here, which enhances the security of the whole system as a whole. This paper also discussed one of the proposed IoT-based patient monitoring schemes using RFID technology having biometric-based authentication for the communication in the system. This scheme is able to continuously monitor patient's blood sugar level, pressure level, temperature, heart-bit, and other critical parameters and provides feedback to the concerned medical practitioners in order to mitigate the high-risk factors involved with a patient's health conditions and provide fast recovery.

For future, the cryptanalysis and verification of the enhanced scheme need to be conducted using the other techniques like AVISPA and random oracle methods. The scope of research can be extended to cryptanalysis in multi-server and multi-user environment having more complexities in securities. Biometric-based authentication scheme can also be extended to other Internet-based applications like cloud-based ERP application securities and other business areas like Supply-Chain and CRM (Customer Relationship Management).

References

1. L. Yanrong, L. Li, H. Peng, Y. Yang, An enhanced biometric based authentication scheme for telecare medical information systems using elliptic curve cryptosystem. *J. Med. Syst.* **39**, 32 (2015)
2. R. Amin, S.K.H. Islam, G.P. Biswas, Design and analysis of an enhanced patient-server mutual authentication protocol for telecare medical information system. *J. Med. Syst.* **39**:137 (2015)
3. Q. Jiang, Z. Chen, B. Li, J. Shen, L. Yang, J. Ma, Security analysis and improvement of bio-hashing based three-factor authentication scheme for telecare medical information systems. *J. Ambient. Intell. Hum. Compute* (2017)
4. Q. Xie, Z. Tang, K. Chen, Cryptanalysis and improvement on anonymous three-factor authentication scheme for mobile networks. *Comput. Electr. Eng.* (2016)
5. A.K. Sutrala, A.K. Das, V. Odelu, M. Wazid, S. Kumari, Secure anonymity-preserving password-based user authentication and session key agreement scheme for telecare medical information systems
6. T. Limbasiya, N. Doshi, An analytical study of biometric based remote user authentication schemes using smart cards. *Comput. Electr. Eng.* **59**, 305–321 (2017)
7. H. Rallapalli, P. Bethelli, IoT based patient monitoring system. *Int. J. Comput. Commun. Instrum. Engg* (2017)
8. D. He, S. Zeadally, An analysis of RFID authentication schemes for internet of things in healthcare environment using elliptic curve cryptography. *IEEE Internet Things J.* (2015)
9. M. Saadeh, A. Sleit, M. Qatawneh, W. Almobaideen, Authentication techniques for the internet of things: a survey. *cybersecurity and cyberformsics conference* (2016)
10. A. Ouaddah, H. Mousannif, A.A. Elkalam, A.A. Ouahman, Access control in the internet of things: big challenges and new opportunities. *Comput. Netw.* (2016)
11. J. Quian, H. Xu, A novel secure architecture for the internet of things. *Int. Conf. Intell. Netw. Collab. Syst.* (2016)

Automatic Facial Expression Recognition Using Geometrical Features



Tanmoy Banerjee, Sayantan De, Sampriti Das, Susmit Sarkar
and Spandan Swarnakar

Abstract Every person expresses himself or herself with different facial expressions. Nowadays, recognition of facial expression is a popular topic for researchers due to the practical function in a wide variety of areas. The learning of human facial expressions is a tough and tricky domain in pattern research society. Each individual facial expression is produced by non-rigid object contortions, and these contortions are man-dependent. The main motto of this paper was to invent and apply a structure for the automatic recognition of human facial expression.

1 Introduction

In every automatic expression analysis system, researchers have used lots of different approaches for recognizing some small set of expressions: happiness, surprise, anger, fear and disgust (Fig. 1) [1, 2]. In our system, we also tried to recognize this set of expressions by using some geometrical features. Expressions are changed for every different emotion by changing two or three discrete facial features. Sometimes this emotion change can be determined by changing one discrete feature, such as anger can be represented by showing tightening lips, surprise can be represented by raising eyebrow [3] and so on.

T. Banerjee · S. De (✉) · S. Das · S. Sarkar · S. Swarnakar
Department of Computer Science and Engineering, Institute of Engineering
and Management, Saltlake, Kolkata, India
e-mail: sayantande31@gmail.com

T. Banerjee
e-mail: tanmb1991@gmail.com

S. Das
e-mail: sampriti.das06@gmail.com

S. Sarkar
e-mail: susmit.sarkar97@yahoo.com

S. Swarnakar
e-mail: swarnakarspandan1997@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling
and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_45



Fig. 1 Six universal expressions

The expression is of great value in the domain of human–computer interface as probable input, and it is especially valid in the case of voice-stimulated control system. It is found when humans are talking, 55% of communication takes place via expression, whereas merely 7% occurs via spoken words.

The method of deriving a facial expression from original pictures is utilized to create animate synthetic characters, and this is also employed in video telephony with narrow bandwidth. If we transmit the facial expression sequence in the place of transmitting the video, the real video can be derived. Computer animation can also be implemented in one more use of this.

Automatic recognition of expressions is very challenging experiment, as there are many underlying features that can have an effect on the mode of facial expressions. One factor is the presence of subject differences like the texture of the skin, hairstyle, age, gender and ethnicity. All these factors have a huge influence on the expression of the face. In addition to the differences in appearance, there might be differences in expressiveness; that is, individuals express their emotion differently from each other. Expressiveness specifically refers to the degree of facial plasticity, morphology, a frequency of intense expression and the overall rate of expression. Each difference is a very important aspect of identity, and characteristic facial actions may be used as a biometric property to improve the accuracy of the facial recognition algorithms.

The research work mainly aspires at cracking the dilemma in a proficient approach. Section 2 portrays the related work. Section 3 features the proposed methodology. Section 4 shows the experimental results, and the last section concludes the research paper.

2 Related Work

Suwa et al. established the preliminary automatic facial expression recognition system which tries to explore facial expressions by tracking the movement of 20 identified marks in a picture series in 1978 [4]. After that many researchers have done many works and proposed different computer systems to realize and employ this usual form of human communication.

In the sphere of psychology, the facial action coding system [5] is used to portray minor alterations in the facial descriptions. Figure 2 [6] shows some of the FACS action units from the total of 44. These are associated with the contraction of some specific set of facial muscles. Researchers have given automation priority to run with the modern world, and some efforts have been made in that field [7].

We can point out the FER as an exceptional face recognition system or just a component of a face recognition system. A quick look at the common architecture of face recognition should be enlightening.































Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
					
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
					
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink
Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
					
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
					
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
					
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Fig. 2 FACS action unit

The initial action in face recognition is the identification of the face on which the performance of the whole arrangement depends [6].

Principal component analysis (PCA), which was proposed by Turk and Pentland, is another detection system that relies on the eigendecomposition. Eigenfaces are the face images. In this proposition, the picture is characterized by a regular face with a set of “eigenfaces”. Considering the distribution of non-face pictures, Sung and Poggio [8] applied Bayes’ rule to get a likelihood assessment. Rowley et al. [9] applied neural networks, and Osuna et al. [10] made a kernel support vector machine that can distinguish between face and non-face images. A bootstrap algorithm is applied in these systems which works iteratively and gathers significant example. Schneiderman and Kanade [11] made use of AdaBoost learning to create a classifier-based wavelet illustration of image, but this is computationally pricey. Viola and Jones [12] substitute wavelets with Haar features [13, 14]. This system is the earliest real-time frontal-view face detector [15].

Ample of work have been done on these, and there are upgradations proposed by many researchers. Lenhart et al. [16] utilized rotated Haar features. Li et al. [17, 18, 16] recommend a multi-view face detection system.

3 Proposed Methodology

The main motto of the paper is to create a framework and make a facial recognition system. Face action representation and appropriate algorithms are chosen for every individual step of a system for various approaches in this regard.

In this paper, we have used three basic systems to build the algorithm: Sect. 3.1 face detection and tracking, Sect. 3.2 feature extraction and Sect. 3.3 expression recognition.

3.1 *Face Detection and Tracking*

The very initial component of this structure is the detection of the face and the localization of the landmark in the picture. There are basically two ways we applied for the detection of face and its parts; those are face detection using Viola and Jones framework and template matching algorithm.

3.1.1 **Viola and Jones Framework for Face Detection**

The initial object detection framework to equip competitive object detection was proposed in 2001 by Paul Viola and Michael Jones renowned as Viola and Jones [4] framework. The picture is depicted by a set of Haar-like features. A chief attribute

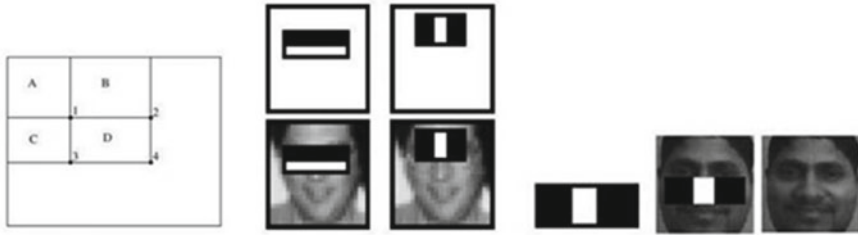


Fig. 3 Haar-like feature detection by Viola and Jones



Fig. 4 Eye detection using template matching

of Haar wavelet-like aspects is that they can be worked out resourcefully by means of an integral picture [11].

To get the feature value, we deduct pixel sum acquired by white rectangle from the pixel sum of grey rectangle. These two rectangles give us the contrast between two vertically or horizontally neighbouring areas. There are many approaches as three rectangles are applicable to find the contrasted area put among two alike regions and four rectangles to get the diagonally placed areas (Fig. 3).

3.1.2 Template Matching

In digital image processing, the artistry to find small expressions of a picture that resembles a template image is known as template matching. The main utilization of this process is manufacturing as a unit of quality control, a manner to navigate a mobile robot or as a manner to discover the edges in picture.

In our proposed system, we used the template matching algorithm for executing eye detection. At first, we took an image of the eye (both left eye and right eye) as shown in Fig. 4. It was utilized to discover the eye part in the visage that was discovered by the Viola–Jones algorithm.

3.2 *Feature Extraction*

There are two concepts applied to elaborate the expression geometric or appearance attributes which can be applied individually or collectively. For geometrical feature-supported systems, the face is characterized by a group of observed facial points. Distortions among the neutral state and the current state are the factors of facial action. This process needs trustworthy methods for point's exposure and observing which is troublesome to extract [19]. Appearance-based processes calibrate the changes in the impression, which primarily depends on texture analysis. Though Gabor filters are verified to be stronger in face reflection analysis, they take time as well as memory.

Wrinkling of facial muscles manufactures alterations for both the direction and magnitude of the movement on the skin boundary and in the impression of permanent and transient facial attributes. Illustrations of the permanent attributes are the lips and eyes which have become stable with age. The observed outcomes show that our procedure is strong for observing facial features [20].

3.2.1 *Multistate of Facial Components*

A multistate facial component model was developed after the detection of facial components in near frontal images. In Table 1 [1], the models which include the fixed facial component (i.e. eye, eyebrows, mouth and lips) of frontal images are illustrated.

The face which is traced is illustrated by a combination of facial points in geometric feature-based systems. Distortion among the neutral state and current frame is measured as the factor of facial action. Feasible methods are required for point detection and tracing in this way which is not easy to put into action [19]. The changes in appearance are examined by the appearance-based methods which are texture analysis-based mainly. Gabor filters give good result for scrutinizing the facial expression, but they are both time and memory taking.

Due to deflationary in facial muscles yields changes on the skin, and as a result, permanent and temporary facial features are changed. Only some permanent features (e.g. lips, eyes) do not change with the age. We consider the first frame in a neutral expression. The templates of the permanent features are initialized in the first frame. The experimental outcomes demonstrate that our course of action is well-built for observing facial features [20].

After tracking all the facial component, we implement some methods on the components. From this Table 1, we see that all components have a different state for a different expression.

Table 1 Different component of facial expression


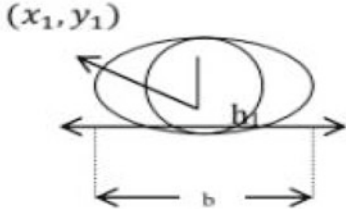
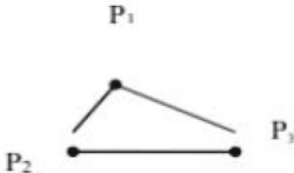


Component	State	Description/Feature
Eye	Closed	
	Open	
Eyebrow	Present	
Lip	Open	
	Close	



Fig. 5 Eye and eyeball detection using template matching and circle Hough transform

3.2.2 Eye and Eyeball Detection

For the study of facial expressions, eye is a valuable part of the face. Eye viewers mostly monitor the eye when it is open and navigate the eye regions. For discriminating facial action units, we required to discriminate the state of eyes whether it is open state or closed state. It measures the height of the open eye, the radius of the iris and the corners. In Table 1, the eye model basically includes “open” and “closed”.

In both cases of eye, some feature is the same, but not all. In case of opened eyes, we find the circle of the eyeball, but in case of closed eyes, we could not find any circle for the eyeball. Closed eyes can only be defined as a straight line (as shown in Table 3). In our proposed system, when eyes are opened we first try to find the circle for the eyeball using circle detection algorithm as mentioned below. Then we consider the circle with the maximum radius. As a result, we find the eyeball as shown in Fig. 5.

In our proposed system when eyes are opened, we first try to find the circle for the eyeball using a circle detection algorithm. Then we consider the circle with the maximum radius. As a result, we find the eyeball. Whenever we get the eye ball, we can detect the centre easily. This centre is used later to calculate the distance between eyebrow and eyeball for the different expression of face due to the scarcity of space, and we could not show the eyebrow and lip detection.

3.3 Expression Recognition

Facial expression is the micro-expressions of the face that helps to convey and show the emotions and other hints about the physical and psychological condition of a human. These subtle hints can be identified, and this is the form of non-verbal communication.

In our system after detecting several parts of a facial component, we tried to recognize the expression. In the previous section, we described six basic facial expressions. Now we tried to recognize this using several methods. When we used the feature extraction method, we used edge detection algorithm and also further experiments. As an example, when we get the eyeball we also detect the edge of the eyeball. After

that when we check for maximum intensity change vertically for eyebrow detection, this edge detection is more useful for the correct result. The basic theory of edge detection used is given below.

3.3.1 Edge Detection Algorithm

The process of locating the points of an image with the help of discontinuity of brightness is known as edge detection, and the set of curved line part on which the sharply changed brightness points are organized is known as edge. For one-dimensional approach, it is recognized as step detection and for signal discontinuities over time is change detection. Edge detection is an elementary tool for image processing, machine vision and computer vision. We identify the sharp changes in the brightness to detain chief events and alter in property.

The effect of pertaining an edge detector to a picture may guide to a set of associated curves that point out the margins of matter, surface markings as well as curves that correspond to disruptions in surface orientation. Thus, pertaining an edge detection algorithm to a picture may radically shrink the quantity of information to be processed and may, therefore, strain out information that may be considered as a lesser amount of significant, while defending the vital structural phenomenon of a picture. If the edge detection step is flourishing, the successive mission of interpreting the data contents in the real picture may, therefore, be substantially cut down. Edge detection is one of the elementary steps in image processing, image analysis, image pattern recognition and computer vision techniques. There are numerous procedures used for edge detection, like Roberts, Prewitt, Sobel, Canny and so on. In this paper, we used Prewitt and Canny in most of the cases for edge detections. Illustration of an example of this technique is given in Fig. 6.

After detecting the edge, we calculate the distance from the centre of the eyeball to eyebrow, between the upper lip and lower lips, etc. For calculating this distance, we used Euclidean distance.

4 Experimental Results

In our system, we used two common databases for our algorithm—the extended Cohn–Kanade (CK+) dataset and Japanese female facial expression (JAFFE) database.

4.1 *The Extended Cohn-Kanade (CK+) Dataset*

In this record, 210 grown human beings of 18–50 years of age were recorded with the help of two hardware synchronized Panasonic AG-7500 cameras (Fig. 7). Out

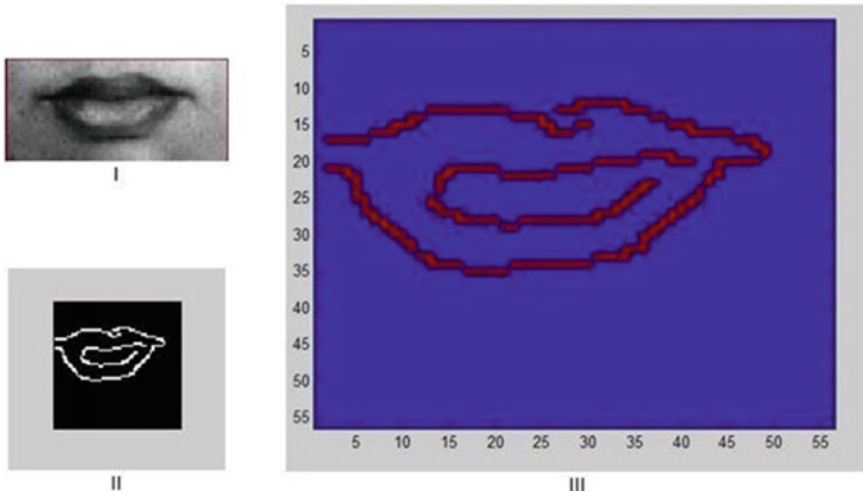


Fig. 6 Example of edge detection I, normal lips II and III: edge detection by using canny



Fig. 7 Different expressions used in CK+ dataset

of those 210 people participated, 69% are female, 81% Euro-American, 13% Afro-American and 6% other groups. The researcher instructed them to present 23 facial displays which were comprised of single action units and combinations of action unit. Every individual display initialized with a neutral face and ended with a neutral face too. The images were frontal or 30 degree viewed that with either 640×490 or 640×480 pixel arrays with 8-bit greyscale or 24-bit colour values. The full information of the record is in [19].

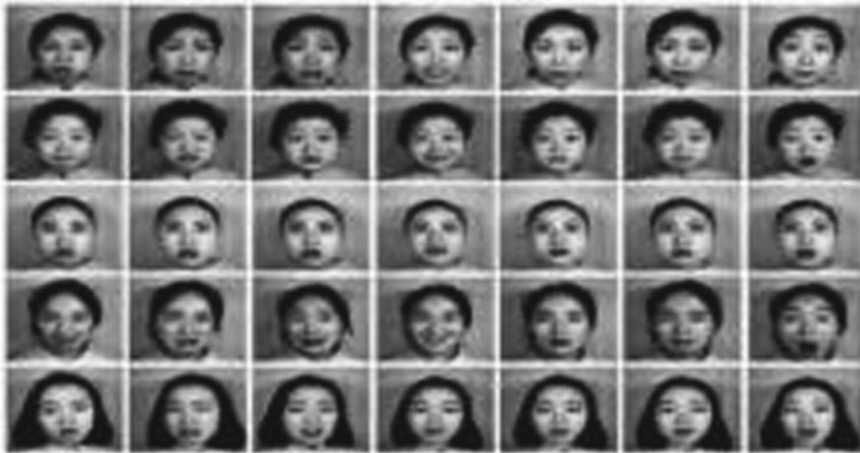


Fig. 8 Different expression used in JAFFE dataset

Table 2 Confusion matrix for facial expression detection in CK+ dataset

	Natural	Anger	Disgust	Surprise	Sadness	Happy	Fear
Natural	92	2	1	0	2	0	3
Anger	1	48	20	2	17	1	11
Disgust	0	6	89	0	3	0	2
Surprise	0	0	1	96	1	2	0
Sadness	6	21	20	3	33	1	16
Happy	1	13	8	17	6	51	4
Fear	3	13	14	4	23	3	40

4.2 Japanese Female Facial Expression (JAFFE) Database

The actual record includes 213 photographs of 7 different facial expressions (6 fundamental and 1 neutral) with the help of 10 Japanese ladies (Fig. 8). Every single picture has been evaluated with respect to the six basic emotion adjectives by 60 Japanese subjects. The record was designed and brought together by Michael Lyons, Miyuki, Kamachi and JiroGyoba.

With the help of previously discussed methods and its results, we have given some examples in this part and according to this, we create our whole system. The results are shown in this figure shown below:

The performance of our system for features for emotion detection is given in Tables 2 and 3 for two different datasets used in our system, and it is noted that three basic emotions, named disgust, neutral and surprise are carried out and gave better results with respect to the other emotions. This outcome is very much useful as they make many changes in the face because they are distinctive emotions.

Table 3 Confusion matrix for facial expression detection in JAFFE dataset

	Natural	Anger	Disgust	Surprise	Sadness	Happy	Fear
Natural	96	1	1	0	2	0	3
Anger	6	38	22	2	19	1	12
Disgust	0	1	93	0	2	0	4
Surprise	1	0	0	97	0	2	0
Sadness	1	23	17	0	39	1	19
Happy	0	19	6	14	11	47	3
Fear	7	23	18	2	14	8	28

Performance of our system is described by using confusion matrix for two different datasets (CK+ and JAFFE), respectively. In this matrix, we used six universal poses of the face (anger, disgust, surprise, sadness, happy and fear), and one is for natural. So in this matrix, seven rows and seven columns are there to represent the result. From FACS manual, we can see that the action unit always differs for a different expression. By comparing this action unit, we get the result.

From these two Tables 2 and 3, we see that we get almost the accurate result for natural, disgust and surprise. For example, when we try to find out the expression of surprise for CK+ dataset, we get 96 correct results for 100 subjects. So accuracy is 96% for that expression.

5 Conclusion and Future Work

The main motive of this paper was to idealize the field of facial expression recognition. To start with, the psychological principles for facial actions examination suggest that this area of science has been broadly considered in terms of application and computerization. Computer software substituted the manual face study used by psychology.

Moreover, this area of research has still many troubles in such systems, and a lot of research work is still required in the sphere of performance and practical application improvement. We have provided the plan and execution techniques of facial expression recognition system in this paper. This basically studies the six basic emotions by making a video of facial behaviour. Full automation and environment independence are the two special features of this paper.

Applying the approach, all expressions were not predictable. Before this, we see in outcome part that our system was giving a good outcome merely in three expressions. Now for a superior outcome in the future, we crave to utilize lots of others facial component like nose wrinkling, eyebrow distance and so on. We anticipate that using this module we will get a more precise outcome in future.

References

1. Y. Tian, T. Kanade J.F. Cohn. Recognizing action units for facial expression analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, **23**(2), (2001)
2. M.F. Valstar, M. Pantic. Fully automatic recognition of the temporal phases of facial actions. *IEEE Trans. Syst., Man Cybern.-Part B: Cybern.* **42**(1), (2012)
3. L. Florea, R. Boia. Eyebrow localization for expression analysis. In: *Intelligent Computer Communication and Processing (ICCP), 2011 IEEE International Conference Publications*, 2011
4. M. Suwa, N. Sugie, K. Fujimori. A preliminary note on pattern recognition of human emotional expression (1978)
5. P. Ekman, W. Friesen, The facial action coding system: a technique for the measurement of facial movement (1978)
6. A.K. Jain, S.Z. Li, *Handbook of Face Recognition* (Springer-Verlag, New York Inc, Se caucus, NJ, USA, 2005)
7. M. Pantic, I. Patras, Dynamics of facial expression: recognition of facial actions and their temporal segments from face profile image sequences. *IEEE Trans. Syst., Man, Cybern., Part B* **36**(2), 433–449 (2006)
8. K.-K. Sung, T. Poggio. Example-based learning for view-based human face detection (1998)
9. H. A. Rowley, S. Baluja, T. Kanade, Neural network-based face detection (1998)
10. E. Osuna, R. Freund, F. Girosi. Training support vector machines: an application to face detection (1997)
11. H. Schneiderman, T. Kanade. A statistical method for 3d object detection applied to faces and cars (2000)
12. P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features (2001)
13. F. Crow, Summed-area tables for texture mapping (1984)
14. P.Y. Simard, L. Bottou, P. Haffner, Y.L. Cun, Boxlets: a fast convolution algorithm for signal processing and neural networks (1998)
15. P. Viola, M. Jones. Robust real-time object detection (2001)
16. S.Z. Li, L. Zhu, Z.Q. Zhang, A. Blake, H. Zhang, H. Shum, Statistical learning of multi-view face detection (2002)
17. S.Z. Li, Z. Zhang. Floatboost learning and statistical face detection (2004)
18. S. Li, Z.Q. Zhang, H.-Y. Shum, H. Zhang. Floatboost learning for classification (2002)
19. K. Fukui, O. Yamaguchi, Facial feature point extraction method based on a combination of shape extraction and pattern matching. *Syst. Comput. Jpn.* **29**(6), 49–58 (1998)
20. G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, T.J. Sejnowski, Classifying facial actions. *Int. J. Pattern Anal. Mach. Intell.* **21**(10), 974–989 (1999)

A Review on Different Image De-hazing Methods



Sweta Shaw, Rajarshi Gupta and Somshubhra Roy

Abstract This paper is written to present a review on the different haze removal techniques. Image de-hazing is the most important field of research in image processing and analysis. The process of de-hazing is important to make human understanding easier in low visibility. Haze is a state of poor air quality characterized by a pale scent appearance of atmosphere and reduces visibility. It is caused by high concentration of air pollutants suspended in the atmosphere that scatter and absorb sunlight. Because of this, the picture gets whitened and its contrast is reduced. The de-hazing techniques are very much in use in the field of numerous applications in the field of computer vision such as video surveillance and detection of object. The thorough purpose of this paper is to delve different procedures of efficaciously expel haze from any image and checks the drawbacks of the existing methods.

Keywords De-hazing · Depth map-based · Polarization-based · Dark channel prior

1 Introduction

The main aim of image processing is to visualize, sharpen, restore, retrieval, and recognize an image. Haze, fog, smog, etc. vitiate the image condition of outside scene. Haze present in an image is an abrasive issue as it quells the contrast and changes the color. Many applications do not work reliably because of it. The presence of haze also cutbacks the clearness of underwater pictures. As a result, removal of these bad weather situations is critical and unavoidable area of computer graphics.

Many issues have been created by the haze present in atmosphere in the field of terrestrial photography, as here for image distant subjects, the diffusion of enormous quantity of heavy atmosphere is necessary. Because of the consequences of scatter-

S. Shaw (✉) · R. Gupta · S. Roy
Department of Computer Science and Engineering, Institute of Engineering and Management,
Kolkata, India
e-mail: swetaka1988@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_46

533

ing of light by the haze particles, the contrast and hence the visibility get reduced. The prodigy of image de-hazing debauches picture quality, which adversely affects many applications which uses image processing. The scattering is basically due to attenuation and air light. Haze attenuates the light reflected from the objects which is then mixed with extra light present in atmosphere. The main motive of this technique is to boost the reflected light from the mixed light. An image's resolution and strength can be boosted by usage of the de-hazing technique. This can be achieved by using many image de-hazing techniques such as DCP, ICA, and depth map-based method.

2 Image De-hazing Methods

Following are the two major classifications of methods of image de-hazing.

2.1 *Multiple Image De-hazing Method*

In this de-hazing method, more than one image of the same scene is taken. There are many variables in an image. This method obtains the variables which are known and averts the variables which are unknown. The following methods fall under this category.

Method Based on Various Weather Conditions

Here many pictures are collected at different weather conditions. As the primitive idea, the differences of various pictures of the same object are taken. All the pictures have different prospects of devoted medium. It has some advantages as well as disadvantages. This method can surely improve the visibility, but one need to wait until the properties or rather the weather changes. So, it is time-consuming (Fig. 1).

Polarization-Based Method

Here, multiple pictures are taken at different polarization filters. In this method, the underlying principle is to take many pictures of similar view at various polarization degrees. To achieve this, the polarizing filter that is affixed with the camera is rotated. But this method is not so accurate for dynamic scenes (Fig. 2).

Depth Map-Based Method

In this procedure, the information of the depth is used for de-hazing. Here, a single image is used and it has been assumed that three-dimensional geometrical prototype of this view is imparted by any existing database and also assumed that the texture of the scene is given. This model is then aligned with hazy image to provide the depth of the scene. This method gives truthful results. But the disadvantage of this method is that it needs user interaction and cannot be applied automatically. This method needs many resources which are sometimes very difficult to obtain (Fig. 3).



Fig. 1 Hazy and De-hazed images

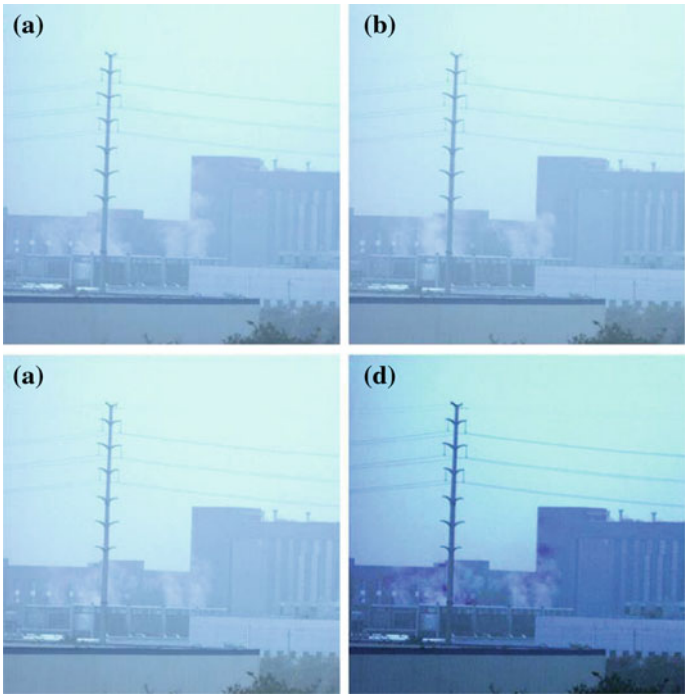


Fig. 2 **a** Image taken at polarization degree 0, **b** image taken at polarization degree 45, **c** image taken at polarization degree 90, and **d** the de-hazed image

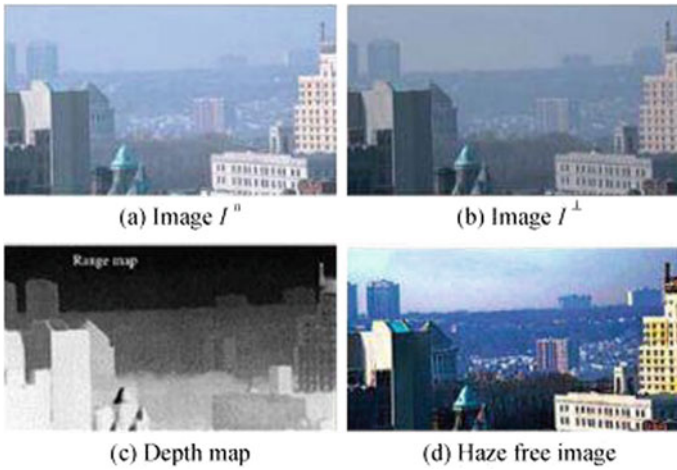


Fig. 3 De-hazing using depth-based method



Fig. 4 Images at different contrasts, de-hazed image

2.2 Single Image De-hazing Method

Unlike the multiple image de-hazing method, in this procedure, only one picture is taken. In this procedure, all the required information of the scene is taken from a single image, taken as the input. Researchers are more interested toward this method than the other one. The following methods fall under this category.

Method of Contrast Maximization

Haze abates the contrast. So, removing haze means embellishing the contrast. In this method, the contrast is enhanced, but there are many constraints. In this method, the brightness and depth are actually not improved. Only the visibility is enhanced by enhancing the contrast. Because of this, the saturation in the output image is very large, which is a great disadvantage of this method (Fig. 4).



Fig. 5 Hazy and de-hazed images

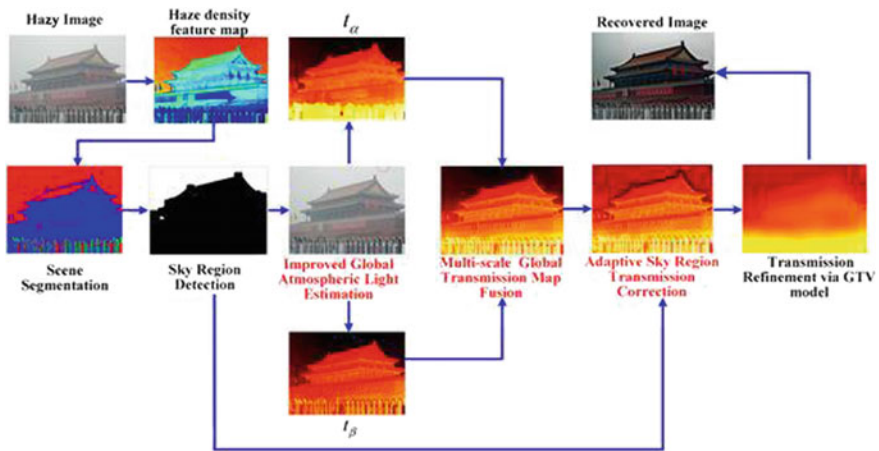


Fig. 6 De-hazing using DCP

Independent Component Analysis

This is a statistical procedure where two additional elements of the input image are divided. Here it is assumed that the transmission and shading of surface are statistically orthogonal in local patch. This procedure gives good result in sparse haze as it produces good result but in dense haze, it does not work properly (Fig. 5).

Dark Channel Prior Method

This is one of the most effective methods of image de-hazing. If we take all the lowest intensity value, then the minimum of it is the dark channel. In this method, the atmospheric light and transmission are calculated first of all for the purpose of de-hazing an image. Although this method gives successful output in both sparse and dense haze, it does not give good result when the scene image is same to the air light, as here the atmospheric light cannot be calculated accurately (Fig. 6).



Fig. 7 Hazy and de-hazed images

Antistrophic Diffusion

In this method, the haze is removed without removing the important parts of the image such as edges and lines so that image understanding is not disturbed. In this method the antistrophic diffusion for refining air light map from DCP. The benefit of this method is that it is very flexible and works well in dense haze (Fig. 7).

3 Literature Survey

Tarel et al. [13] have given an algorithm for single image de-hazing, which depends on filtering approach. In this method, there are few linear operations that require many frameworks for alteration. The advantage of this algorithm is its speed which allows application of visibility restoration in real-time application of de-hazing of image.

He et al. [16] gave a very uncomplicated but adequate method (DCP) to de-haze an input image. It is a kind of stats of outside non-hazy pictures. Here, the estimation of haze transmission is achieved. This method doesn't effectively when air-light & surface object are same.

Xu et al. (2012) gave an enhanced DCP. The DCP was improved by rehabilitating the tedious soft matting part with the quick bilateral filter. It improved the previous dark channel prior algorithm by having fast execution speed and greater efficiency.

Ullah et al. [19] gave another improved DCP. In the new version of dark channel prior, both achromatic and chromatic features of input picture are considered. Here, the contrast and color vibrancy of restored images are further improved. Here, the least of saturation and intensity is considered instead of RGB components.

Arora et al. (2014) have proposed a new integrated dark channel prior (IDCP) method by combining dark channel prior (DCP), contrast limited adaptive histogram equalization (CLAHE), and gamma correction methods. Main motive of this algorithm is to boost the certainty of intelligent transportation system (ITS).

Ansia et al. (2014) gave a method to cynosure on contrast-based single image de-hazing. A white balancing is used to obliterate the color. The estimation of saliency map is done for the determination of the discernible region of the image. Then morphological operator is used to abolish the specularities. Then intensified CLAHE approach is used to augment the color contrast. This method produces better result for homogenous haze images.

JunMaro et al. (2014) developed a function for reckoning the haze degree for automatic detection of foggy image with different haze degree value. This method is advantageous in unusual weather conditions for computer vision-based applications such as video surveillance.

4 Gaps in Literature

A haze-free image possesses evident contrasts compared to a hazy image. There are many vision applications for which image de-hazing is important. After studying different methods/algorithms, it has been revealed that the researchers have overlooked many aspects. Using the literature survey, many research gaps are achieved which are listed below.

- The methods which are mentioned above neglect the noise removal.
- The methods also neglect the issue of irregular illumination which degrades the algorithm.
- No efforts have been taken to integrate the methods of DCP and CLAHE.
- Researchers have not made enough approach toward underwater images.
- No measure is taken toward remote sensing images.

5 Conclusion

Image de-hazing has been a prior research field in image processing. It has been a very important research as it is used to make the visibility possible in bad weather. There are many methods and algorithms proposed till now. The main two categories of these methods are single image de-hazing and multiple image de-hazing. Out of which, single image de-hazing is found to be more accurate and efficient than multiple image de-hazing as this method is less time-consuming and requires less resources. But none of these methods is suitable for all the circumstances. Some method lacks in one circumstance while the other lacks in any other circumstance. Hence, the methods which are already present need to be modified in such a way that they overcome all the problems simultaneously.

References

1. S.K. Nayar, S.G. Narasimhan, Vision in bad weather. Proc. Seventh IEEE Int. Conf. Comput. Vis. **2**, 820–827 (1999)
2. S.G. Narasimhan, S.K. Nayar, Chromatic framework for vision in bad weather. Proc. IEEE Conf. Comput. Vis. Pattern Recognit. **1**, 598–605 (2000)
3. Y.Y. Schechner, S.G. Narasimhan, S.K. Nayar, Instant de-hazing of images using polarization, in *The Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. I-325. 2001
4. S.G. Narasimhan, S.K. Nayar, Vision and the atmosphere. Int. J. Comput. Vis. **48**(3), 233–254 (2002)
5. S.G. Narasimhan, S.K. Nayar, Contrast restoration of weather degraded images. IEEE Trans. Pattern Anal. Mach. Intell. **25**(6), 713–724 (2003)
6. S.G. Narasimhan, S.K. Nayar, Interactive (de) weathering of an image using physical models, in *IEEE Workshop on Color and Photometric Methods in Computer Vision*, vol. 6, no. 6.4, p. 1., (France 2003)
7. S. Shwartz, E. Namer, Y.Y. Schechner, Blind haze separation. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. **2**, 1984–1991 (2006)
8. N. Hautière, J-P. Tarel, D. Aubert, Towards fog-free in-vehicle vision systems through contrast restoration. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), 1–8 (2007)
9. J. Kopf, B. Neubert, B. Chen, M. Cohen, D. Cohen-Or, O. Deussen, M. Uyttendaele, D. Lischinski, Deep photo: model-based photograph enhancement and viewing. ACM Trans. Graph. (TOG) **27**(5), 116 (2008)
10. R. Fattal, Single image dehazing. ACM Trans. Graph. (TOG) **27**(3), 72 (2008)
11. Tan, R.T.: Visibility in bad weather from a single image, in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, pp. 1–8, (2008)
12. Z. Xu, X. Liu, N. Ji, Fog removal from color images using contrast limited adaptive histogram equalization, in *2nd International Congress on Image and Signal Processing (CISP)*, pp. 1–5 (2009)
13. J-P. Tarel, N. Hautiere, Fast visibility restoration from a single color or gray level image, in *12th International Conference on Computer Vision*, pp. 2201–2208 (2009)
14. J. Yu, C. Xiao, D. Li, Physics-based fast single image fog removal, in *10th IEEE International Conference on Signal Processing (ICSP)*, pp. 1048–1052 (2010)
15. F. Fang, F. Li, X. Yang, C. Shen, G. Zhang, Single image dehazing and denoising with variational method, in *IEEE International Conference on Image Analysis and Signal Processing (IASP)*, pp. 219–222 (2010)
16. K. He, J. Sun, X. Tang, Single image haze removal using dark channel prior. IEEE Trans. Pattern Anal. Mach. Intell. **33**(12), 2341–2353 (2011)
17. J. Long, Z. Shi, W. Tang, Fast haze removal for a single remote sensing image using dark channel prior. in *International Conference on Computer Vision in Remote Sensing (CVRS)*, pp. 132–135 (2012)
18. Y.-Q. Zhang, Y. Ding, J.-S. Xiao, J. Liu, Z. Guo, Visibility enhancement using an image filtering approach. EURASIP J. Adv. Signal Process. **1**, 1–6 (2012)
19. E. Ullah, R. Nawaz, J. Iqbal, Single image haze removal using improved dark channel prior, in *Proceedings of International Conference on Modelling, Identification & Control (ICMIC)*, pp. 245–248 (2013)
20. M.S. Hitam, W.N.J.H.W. Yussof, E.A. Awalludin, Z. Bachok, Mixture contrast limited adaptive histogram equalization for underwater image enhancement, in *IEEE International Conference on Computer Applications Technology (ICCAT)*, pp. 1–5 (2013)
21. A.W. Setiawan et al.: Color retinal image enhancement using CLAHE, in *International Conference on ICT for Smart Society (ICISS)* 2013
22. Y. Xu, J. Wen, L. Fei, Z. Zhang, Implementation code of clahe
23. E.H. Kaur, R. Mahajan, Modified haze removal using dark channel prior, gabor filter and clahe on remote sensing images (IJCEA) 2014

An OverView of Different Image Algorithms and Filtering Techniques



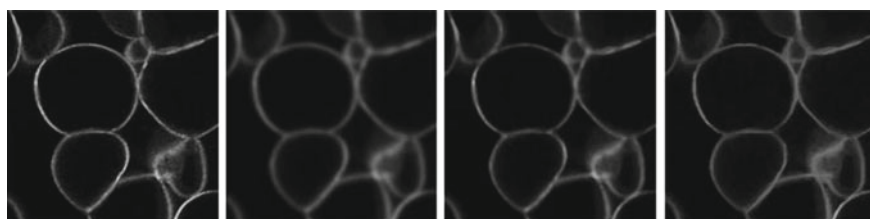
Sumit Prakash, Abhas Somya and Ayush Kumar Rai

Abstract Image filtering is a very important part of image processing techniques. This is the most popular technique which comes under image processing. It is used to get maximum information from original picture. This is becoming popular in the market as most of the people are using different filters in their cell phones. Actually, the basic idea behind the filters is to recognize the edges and reduce the difference between the pixels of edges. Different types of algorithms are there to do such type of functions.

Keywords Gaussian filtering · Median filtering · Linear filtering · Spatial filtering · Temporal filtering · Mean filtering

1 Introduction

Image processing is a method to perform some operations on an image, in order to get maximum output or information from it. A variety of algorithms linear and nonlinear algorithms are used for filtering the images. Image filtering makes possible several useful tasks in image processing. It is used for smoothing of images.



(a) Original image (b) Gaussian filtering (c) Median filtering (d) Directional filtering

S. Prakash (✉) · A. Somya · A. K. Rai
Department of Computer Science and Engineering, Institute of Engineering and Management,
Kolkata, India
e-mail: sumitprakash1988@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_47

This paper consists of different types of filtering algorithms which includes:

- Gaussian filtering.
- Median filtering.
- Linear filtering.
- Spatial filtering.
- Temporal filtering.
- Mean filtering.

These algorithms are useful in nullifying the effect of noise from the image. It includes smoothing, sharpening, and edge enhancement. Filtering is a neighborhood option in which the value of any given pixel is customized by using the values of the pixels in the neighborhood of the corresponding input pixel. Another type of filter can be used to reverse the effects of blurring on a particular picture.

Nonlinear filters have quite different behavior compared to linear filters. For nonlinear, the filter output or response of the filter does not obey the principles outlined earlier, particularly scaling and shift invariance. Moreover, a nonlinear can produce results that vary in a non-intuitive manner.

2 Different Image Filters and Algorithms

2.1 Mean Filter

Mean filtering is a simple and easy to implement method of smoothing images, i.e., reducing the amount of intensity variation between one pixel and the next. It is often used to reduce noise in images.

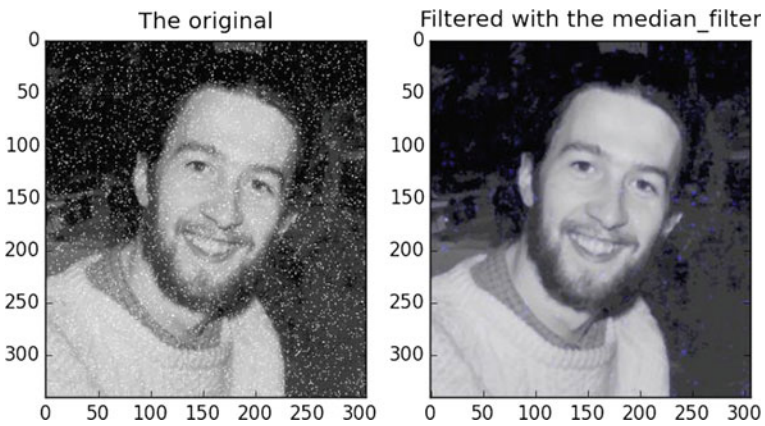
In mean filter, we basically replace the value of each input pixel is updated the mean value of all its neighboring pixel including the pixel taken as input pixel. So, when we replace the pixel value by its surrounding mean value, we make that pixel compatible with the surrounding. In other words, we eliminate the odd one out by replacing it with a value which is expected by its surrounding pixels. Mean filtering is indirectly convolution filter. A kernel matrix is used to represent the shape size and values of the neighboring pixels. The larger is the kernel size, more severe is the smoothing effect (Fig. 1).

2.2 Median Filter

The median filter is also used to remove noise from the image. But, it is observed that median filter often outperforms mean filter in several aspects. It is observed that median filter does better at preserving the useful details of the input image during elimination of noise. Also, median filter eliminates noise to a greater extent.

$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$
$\frac{1}{9}$	$\frac{1}{9}$	$\frac{1}{9}$

Fig. 1 A sample kernel for mean filtering



Like the mean filter, the median filter also focuses on a single pixel at a time. But, instead of simply replacing the pixel value with the mean pixel value of its surrounding pixels, it updates the input pixel value with the median of its surrounding pixel values. In order to calculate the median first, all the pixel values from the surrounding neighborhood are sorted in ascending order and then the input pixel is replaced with the pixel value in middle of the sorted sequence. If there are two middle values, we use the basic method used while calculating median of a set of numbers.

123	125	126	130	140
122	124	126	127	135
118	120	150	125	134
119	115	119	123	133
111	116	110	120	130

Neighbourhood values:

115, 119, 120, 123, 124,
125, 126, 127, 150

Median value: 124

2.3 Gaussian Blur

The Gaussian operator is a 2-D operator that is used to “blur” images and remove noises. We might think that it is similar to mean filter, but it uses a different type of kernel matrix that is similar to the shape of a Gaussian hump.

Gaussian blur is a filter that is often used to smooth or blur a digital picture. It replaces pixels with the average value of weights of neighboring pixels. The weights are the results of Gaussian probability distribution, so the pixels nearer to the input pixels are given higher priority. We can apply the value of the Gaussian function at the center of a pixel in the mask, but this is not exact because the value of the Gaussian varies nonlinearly across the pixel. So, we calculate the value of Gaussian by integrating it over the entire pixel. The value 273 in the fraction is the sum of all the values in the mask.

1	4	7	4	1
4	16	26	16	4
7	26	41	26	7
4	16	26	16	4
1	4	7	4	1

$\frac{1}{273}$

2.4 Spatial Filtering

An image $g(x, y)$ is formed by passing the image $f(x, y)$ through a 2-D convolution product system in presence of an impulse response function $h(x, y)$.

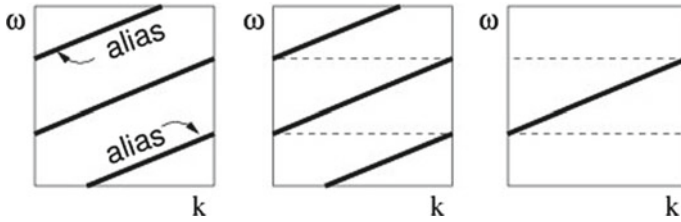
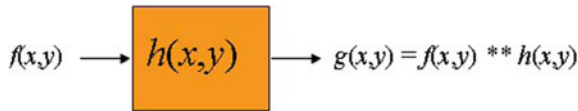


Fig. 2 A low-pass anti-aliasing (triangle filters)



The above-mentioned function $g(x, y)$ is defined as

$$g(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x - x', y - y')h(x', y')dx'dy'$$

The impulse response function $h(x, y)$ is defined as-

$$h(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \delta(x - x', y - y')h(x', y')dx'dy'$$

The serial fashion of association of these systems is equivalent to a unique linear and spatially invariant system having an impulse response equal to the convolution product of individual impulse responses in the space.

The parallel fashion of association of these systems is equivalent to a unique linear and invariant system whose impulse response is equal to the sum of the individual impulse responses unlike the serial fashion which is product of individual impulse responses.

2.5 Temporal Filter

The temporal filtering is a useful tool in designing stacking operators of various types. But when there is rapid change in the length and direction of the travel time gradient, the estimation of filter cutoff becomes erroneous at the curved parts of the operator. Therefore, there is a possibility that the high-frequency part of the output becomes erroneous and distorted leading to a loss in image detailing and resolution. So, it is not an ideal filter when the output contains high-frequency components (Fig. 2).

The method of low-pass filtering is less prominent in the case of a 3-D integral operator. We can take the length of a triangle filter proportional to the of the time gradient, the maximum of the gradient components in the two directions of the operator space, or the sum of these components.

$$\delta t = \Delta x |\partial t / \partial x| + \Delta y |\partial t / \partial y|$$

3 Conclusion

We have done research on different filtering algorithms. From the results and implementation of all filtering methods, we can conclude that the median filtering is the best among all of the other filters. Since median filter is easy to implement and also gives more accurate results as compared to mean filter, it is also used in day to day life activity. The median filter is the most efficient of the lot in removing noise and smoothing the image. Median filter is the best filter when we consider removing of salt and pepper noise. The results from the research can be very useful in analyzing the image much better.

Bibliography

1. M. Brookes, Algorithms for max and min filters with improved worst case performance. IEEE Trans. Circuits Syst., **47**(9), 930–935 (2000). GraphiCon'2007 accuracy is good
2. Website: <http://homepages.inf.ed.ac.uk/>
3. M. Freeman, *The Digital SLR Handbook*. published in 23 October, 2005
4. Website: www.thorlabs.com
5. A. Harrison (2012) *NZ Arab Annual Conference 2012* (Wellington, New Zealand, 2012), pp. 25–26
6. E. Dubois, S. Sabri, Noise reduction in images sequences using motion- compensated temporal filtering. IEEE Trans. On Commun. COM-**32**(7), 826–231 (1984)
7. Website: StackOverflow.com, 'Source of the identity revealing human image'

Design of a Quantum One-Way Trapdoor Function



Partha Sarathi Goswami and Tamal Chakraborty

Abstract Of late security has become a key concern of data transmission mechanism over a communications channel. In an asymmetric cryptographic system, a public key is shared across an insecure medium. This makes the data exchange vulnerable to potential threat from various attackers. This paper proposes the design of a one-way trapdoor function built upon a quantum public key and a classical private key-based encryption–decryption technique of the secret message. The mapping between numbers used in the classical paradigm and their corresponding quantum states is established through the proposed quantum one-way trapdoor function.

Keywords Quantum cryptography · One-way trapdoor function · Qubit

1 Introduction

One of the foremost challenges of data transmission in today’s world is security. An electronic system that is to be made reliable requires sufficient safety mechanisms in order to function properly. The security of data depends on a good encryption algorithm. In case of a symmetric key system, an identical key is needed for encryption and decryption. In case of asymmetric key encryption, a pair of keys, viz. a public and a private key, are used for data encryption and decryption, respectively. Therefore, generation of the pair of keys (public and private) is critically important in any cryptosystem.

Numerical one-way trapdoor functions, which are “easy” to calculate while “hard” to inverse in the absence of some extra (the trapdoor) information, are commonly

P. S. Goswami (✉)

Department of Computer Application, A. J. C. Bose Polytechnic, North 24 Parganas, India

e-mail: goswamipsg@rediffmail.com

Department of Technical Education & Training, Government of West Bengal, Kolkata, India

T. Chakraborty

Mrinalini Dutta Mahavidyapath, West Bengal State University, Kolkata, India

e-mail: tamal@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling*

and Graphics, Advances in Intelligent Systems and Computing 937,

https://doi.org/10.1007/978-981-13-7403-6_48

used in asymmetric cryptography. The main feature of these functions is that they offer an efficiently solvable problem to the genuine users, whereas any illicit user or adversary is faced with a problem that is computationally hard. This barrier of complexity between the valid users and the unauthorized users is the main concept behind a lot of modern public-key cryptosystems.

This paper introduces a one-way trapdoor function based on the axioms of quantum theory. The unit of quantum information is a quantum bit or qubit, vis-à-vis a bit (which takes a value of either 0 or 1) in a classical system. A qubit can exist as the superposition of the quantum $|0\rangle$ and $|1\rangle$ states, which depict the ground and excited states of a single electron. The proposed algorithm presents a quantum one-way trapdoor function for the generation of the pair of keys by rotation of qubits.

The rest of this paper is organized as follows: Section “Related Work” discusses various related works in this domain. Section “Application of Quantum Computing on Cryptographic Systems” provides an overview of quantum cryptography. Section “One Way Trapdoor Function” gives a mathematical background of one-way trapdoor functions. Section “Rotations of Qubit Using Quantum Trapdoor Function” focuses on the rotation of qubits based on quantum trapdoor functions. Section “Encryption and Decryption of the Pain Text” introduces the proposed algorithms. Section “Security” discusses various security aspects of the given procedure, and finally Section “Conclusions” concludes the discussions.

2 Related Works

Bennett et al. [1] first proposed the protocol on quantum cryptography named as “BB84” where BB stands for Bennett and Brassard, respectively. The algorithm was based on Heisenberg’s Uncertainty Principle. Bennett [2] suggested a more basic version of “BB84” called “B92.” Yang and Kuo [3] suggested an improved quantum key exchange protocol using “BB84” and “B92.” One of them was to achieve a quadratic time complexity on average; another was to increase to an efficiency of 42.9% and thereby has the average complexity $O(n^{2.86})$. Houshmand et al. [4] proposed “An entanglement-based quantum key distribution protocol” in which an updated version of Cabello’s [5] definition of efficiency of quantum key distribution rules was used to compare between their procedure and “BB84.” Aldhaheri et al. [6] proposed “A novel secure quantum key distribution algorithm,” which overcame the weaknesses in “BB84” and “B92.” In their work, they exchanged the session key over the quantum channel. Moreover the users’ authentication and privacy were preserved by trading the arbitrary basis and nonce. Odeh et al. [7] presented a new model for quantum key distribution (QKD) [8] between three parties in which there was a trusted center that provided clients with the requisite secret information to securely communicate between each other. The performance of the algorithm was enhanced by removing redundant rounds of checking the quantum bases and verification.

3 Application of Quantum Computing on Cryptographic Systems

Quantum cryptography [9] applies principles of physics to develop a secure cryptosystem. The word quantum refers to the smallest particles of matter and energy. Quantum theory [10] explains everything that exists and nothing can be in violation of it. Quantum key distribution [11] is built on the fundamentals of quantum physics and information theory. The key to be distributed must be secret. Quantum key distribution assures concealment and confidentiality of data transmission. An essential and fairly important characteristic of quantum mechanics is quantum entanglement [12], a physical phenomenon that happens when quantum particles perform in such a way that the quantum state of each particle cannot be defined separately. This quantum phenomenon is often used in quantum cryptographic systems.

4 One-Way Trapdoor Function

Let A be a set of numbers and Q be the set of quantum states [13] of a system. Then a one-way function can be depicted as a mapping $M: A \rightarrow Q$ which is simple to evaluate but difficult to invert. A quantum one-way trapdoor function [14] is one, which is reversible by means of some a priori (trapdoor) information (Fig. 1).

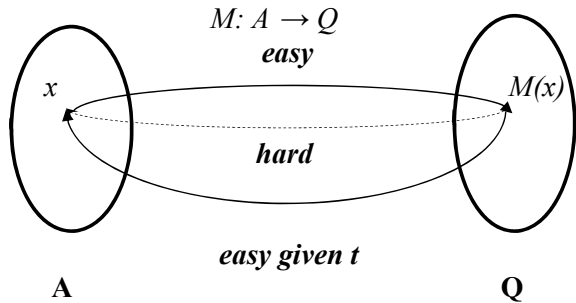
Assuming $A \in Z_n = \{0, 1, 2, \dots, n-1 | n \in N\}$ as the input set of integers for the one-way trapdoor function and the quantum state $|\varphi_a\rangle$ be the corresponding output, with the initial state being $|0\rangle$ in the Hilbert space H ; then for any randomly

chosen $A \in Z_n$, a rotation operator $\widehat{O}: H \rightarrow H$ may be defined as follows: $|0\rangle \xrightarrow{\widehat{O}} |\varphi_a\rangle$

Hence all possible sets of output states of the one-way trapdoor function will be

$$Q \equiv \{|\varphi_a\rangle | A \in Z_n\}, Q \in H.$$

Fig. 1 One-way trapdoor function



So, if $M: Z_n \rightarrow Q$ is a one-to-one mapping, then there exists a unique $A \in Z_n$ such that $|0\rangle \rightarrow |\varphi_a\rangle$ and the cardinality of Z_n is same as that of Q .

5 Rotations of Qubit Using Quantum Trapdoor Function

An arbitrary single-qubit state on the $x - z$ plane is as follows:

$$|\Phi(\theta)\rangle = \cos \frac{\theta}{2} |0_z\rangle + e^{i\varphi} \sin \frac{\theta}{2} |1_z\rangle$$

where $0 \leq \theta \leq \pi$ and $0 \leq \varphi \leq 2\pi$ define a point on a unit sphere, known as the Bloch sphere [15], as illustrated (Fig. 2).

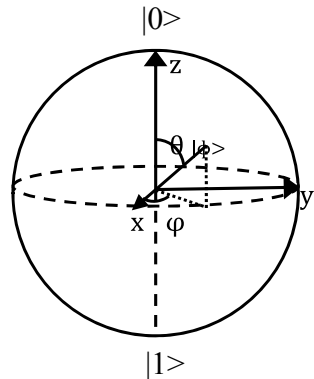
Therefore, a quantum bit can denote a range of states on the $x - z$ Bloch plane as against the conventional way of storing a discrete variable by means of two real values, given by “0” and “1.” Now a rotation about the y -axis $\hat{R}(\theta) = e^{-i\hat{y}\frac{\theta}{2}}$ is the Pauli [16] Y operator yields

$$|\Phi(\theta)\rangle = \hat{R}(\theta) |0_z\rangle, \text{ where } \hat{y} = i(|1_z\rangle\langle 0_z| - |0_z\rangle\langle 1_z|)$$

Since the Pauli operators are both unitary and Hermitian, the rotation vector can be expanded as:

$$\begin{aligned} \hat{R}(\theta) &= e^{-i\hat{y}\frac{\theta}{2}} \\ &= \cos \frac{\theta}{2} I - i \sin \frac{\theta}{2} \hat{y} \\ &= \begin{pmatrix} \cos \frac{\theta}{2} & -\sin \frac{\theta}{2} \\ \sin \frac{\theta}{2} & \cos \frac{\theta}{2} \end{pmatrix} \end{aligned}$$

Fig. 2 Bloch sphere



where I is the identity matrix.

Since the input to the quantum trapdoor function is a random integer A , which is evenly spread throughout the set of all integers modulo 2^n , where n is a natural number and a qubit is initialized as $|0_z\rangle$, a string of length n is adequate to find the input A for a fixed n .

Performing a y -rotation of the qubit by $A(\theta_n)$ with $\theta_n = \frac{\pi}{2^{n-1}}$ for any constant natural number n , yields:

$$\Omega_n = \left\{ |\Phi_{A(\theta_n)}\rangle \mid A \in Z_{2^n}, \theta_n = \frac{\pi}{2^{n-1}} \right\},$$

Where $|\Phi_{A(\theta_n)}\rangle$ can be expressed as:

$$|\Phi_{A(\theta_n)}\rangle \equiv \widehat{R}(A(\theta_n))|0_z\rangle = \cos \frac{A(\theta_n)}{2} |0_z\rangle + \sin \frac{A(\theta_n)}{2} |1_z\rangle \quad (1)$$

Hence, in the absence of prior knowledge of both Z_{2^n} and Ω_n will be secret.

Thus for any arbitrarily chosen value of n and A , it is easy to map $A \rightarrow |\Phi_{A(\theta_n)}\rangle$ as it rotates only a single qubit.

Inverse of the mapping $A \rightarrow |\Phi_{A(\theta_n)}\rangle$ is to recover A from the given qubit $|\Phi_{A(\theta_n)}\rangle$ chosen randomly from the known set Ω_n . The inverse of the function is to identify the different non-orthogonal states [17] chosen randomly from Ω_n for a specified n . As n increases, the number of non-orthogonal states also increases and for $n \gg 1$, the nearest overlapping so obtained as the following:

$$\langle \Phi_{A(\theta_n)} | \Phi_{A+1(\theta_n)} \rangle = \cos \frac{\theta_n}{2} \rightarrow 1$$

Therefore, for a sufficiently large value of n it is computationally improbable to differentiate all the possible states [18, 19].

6 Encryption and Decryption of the Pain Text

6.1 Key Generation

In the cryptosystem, every user will generate a pair of keys,

- (i) Private Key (K_{pv}).
- (ii) Public Key (K_{pb}).

The following algorithm illustrates the key generation process:

Algorithm KeyGen

1. Let $n \gg 1$ be a random positive integer.
2. Let A be a set of random integer strings of length N where $A = \{A_1, A_2, A_3, \dots, A_N\}$ with A_j chosen independently from Z_{2^n} and $A \in Z_{2^n}$, and let X be a random binary string of length N where $X = \{X_1, X_2, X_3, \dots, X_N\}$ with $X_i \in \{0, 1\}$.
3. Obtain N qubits in the state $|0_z\rangle = 0^{\otimes N}$ by applying N parallel Hadamard operations.
4. For each qubit A_j in A
 - 4a. Rotate the qubit $\widehat{R}^{(j)}(A_j\theta_n)$ by angle $\theta_n = \pi/2^{n-1}$ if and only if $X_j = 1$, such that the j th qubit $|\Phi_{A_j(\theta_n)}\rangle_j = \widehat{R}^{(j)}(A_j\theta_n)|0_z\rangle$ takes the form

$$|\Phi_{A(\theta_n)}\rangle \equiv \widehat{R}(A(\theta_n))|0_z\rangle = \cos \frac{A(\theta_n)}{2} |0_z\rangle + \sin \frac{A(\theta_n)}{2} |1_z\rangle$$

where $0 \leq \theta < 2\pi$.

5. Private Key $K_{pv} = \{n, A\}$ and Public Key $K_{pb} = \{N, |\Phi_{(\theta n)}^{(pk)}\rangle\}$ with the N qubits states $|\Phi_{(\theta n)}^{(pk)}\rangle \equiv 0_j^{\otimes N} |\Phi_{A_j(\theta n)}\rangle_j$.

6.2 Encryption

Let S be the sender and R be the receiver. Now S wants to send R a κ -bit message $\omega = (\omega_1, \omega_2, \dots, \omega_\kappa)$, with $\omega_j \in \{0, 1\}$ and $\kappa \leq N$. To encrypt the plain text without changing the sequence of the public-key qubits, the following algorithm is employed:

Algorithm Encrypt

1. Obtain R 's authentic public key K_{pb} .
If $\kappa > N$, S requests R to increase the length of her public key.
2. Encrypt the j th bit of S 's message ω_j by the rotation $\widehat{R}^j(\omega_j \pi)$ on the qubits of the public key, which becomes

$$|\Phi_{A_j, \omega_j(\theta n)}\rangle_j = \widehat{R}(\omega_j \pi) |\Phi_{A_j(\theta n)}\rangle_j$$

3. Obtain the quantum encrypted message as:

$$|\Phi_{A_j, \omega(\theta n)}^{(Kpy)}\rangle = \otimes N_j = 1 |\Phi_{A_j, \omega_j(\theta n)}\rangle_j \text{ is sent to } R.$$

The encoding of the message is in the first κ qubits of the ciphertext so that, in the decipher process R only concentrates on this part of the ciphertext, neglecting the rest of the $N - \kappa$ qubits, that does not contain any extra information.

6.3 Decryption

The following algorithm is used by R to decrypt the ciphertext $|\Phi_{A, \omega(\theta_n)}^{(K_{py})}\rangle$ to obtain the message m .

Algorithm Decrypt

1. Apply $\hat{R}^{(j)}(A_j \theta_n) - 1$ on the j th qubit of the encrypted text.
2. Calculate each qubit of the ciphertext in the basis $\{|0z\rangle, |1z\rangle\}$.

7 Security

The foremost intention of an attacker E is to decrypt the original message from the encrypted message to R . There is always an additional determined objective concerning the retrieval of the private key from R 's public key [20]. A secure function $|\Phi_{A_j, \omega_j(\theta_n)}\rangle_j$ is said to be broken if any of the two purposes is satisfied provided that the opponent E has access to all of the texts intended for the receiver R .

Without doing authentication which is an integral measure of quantum cryptography any secure communication is susceptible to a mimic attack [20, 21]. To focus on its importance, it is described in the algorithm Encrypt that the sender S should acquire a replica of R 's public key.

The private key (K_{pv}) of each entity is $K_{pv} = \{n, A\}$ where $n \gg 1$ is a random positive integer and A is a set of random integer strings of length N where $A = (A_1, A_2, A_3, \dots, A_N)$ with A_j chosen independently from Z_{2^n} and $A \in Z_{2^n}$.

The entropy for n is $H(n) = \log_2 |\tilde{N}|$, where $|\tilde{N}|$ denotes the number of elements in \tilde{N} .

Again the entropy for A is $H(\frac{A}{n}) = N_n$

Hence, we can infer that the entropy of the private key K_{pv} is,

$$H(K_{pv}) = H(n) + H(\frac{A}{n}) = \log_2(|\tilde{N}|) + N_n \gg N$$

To satisfy the above criteria, it is sufficient to either have a sufficiently large value of n or set $\log_2(|\tilde{N}|)$ to be very much larger than N .

Thus E 's information gain is very less as compared to $H(K_{pv})$, which therefore remains basically unknown to him.

8 Conclusions

This paper proposes the design of a one-way trapdoor function using rotation of quantum bits. In Encrypt algorithm, the user creates a pair of asymmetric keys: (private and public). The sender enciphers the message with the receiver's public key through the rotation of qubits. An eavesdropper is unable to decode the secret text deprived of the knowledge of receiver's private key. The focus of the work is to illustrate how the axioms of quantum computing can establish a correct theoretical framework of designing a one-way trapdoor function. The proposed framework establishes a security barrier between authentic users and eavesdroppers, using the foundations of quantum public-key encryption.

References

1. C.H. Bennett, G. Brassard, Quantum cryptography: public key distribution and coin tossing, in *Proceedings of IEEE International Conference on Computers Systems and Signal Processing* (Bangalore, India, 1984), pp. 175–179
2. C. Bennett, Quantum cryptography using any two non-orthogonal states. *Phys. Rev. Lett.* **68**, 3121–3124 (1992)
3. C.-N. Yang, C.-C. Kuo, Enhanced quantum. Key distribution protocols using BB84 and B92 (2002)
4. M. Houshmand, S. Hosseini-Khayat, An entanglement- base quantum key distribution protocol, in *8th International ISC Conference Information Security and Cryptology (ISCISC)*, IEEE (2011), pp. 45–48
5. A. Cabello, Experimentally testable state-independent quantum contextuality. *Phys. Rev. Lett.* **101**, 210401 (2008)
6. A. Aldaheri, K. Elleithy, M. Alshammari, H. Ghunaim, *A Novel Secure Quantum Key Distribution Algorithm* (University of Bridgeport, 2014)
7. A. Odeh, K. Elleithy, M. Alshowkan, E. Abdelfattah, Quantum key distribution by using public key algorithm (RSA), in *Third International Conference on Innovative Computing Technology (INTECH)*, IEEE (London, United Kingdom, 2013)
8. M.L. Adleman, Molecular computation of solutions to combinatorial problems. *Science* **266**, 1021–1024 (1994)
9. M.A. Nielsen, I.L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, London, 2000)
10. A. Menezes, P. Van Oorschot, S. Vanstone, *Handbook of Applied Cryptography* (CRC Press, 1996)
11. C.-H. F. Fung, K. Tamaki, H.-K. Lo, Performance of two quantum key- distribution Protocols. *Phys. Rev. A* vol. **73** (2006)
12. G.J. Simmon, Symmetric and asymmetric encryption. *ACM Comput. Surv.* **11**(4), 305–330 (1979)
13. P.S. Goswami, P.K. Dey, A.C. Mandal, Quantum Cryptography: Security through Uncertainty, in *National Conference on Computing & Systems 2010* (2010), pp 86–89
14. P.S. Goswami, T. Chakraborty, H. Chatterjee, A novel encryption technique using DNA encoding and single qubit rotations. *Int. J. Comput. Sci. Eng.*, **6**(3), 364–369 (2018)
15. F. Bloch, Nuclear induction. *Phys. Rev.* **70**, 460–474 (1946)
16. Pauli matrices. Planetmath website. 28 March 2008
17. W.K. Wothers, W.H. Zurek, A single quantum cannot be cloned. *Nature* **299**, 802 (1982)

18. A. Young, The future of cryptography: practice and theory. *IEEE IT Prof. J.* 62–64 (2003)
19. V. Teja, P. Banerjee, N.N. Sharma, R.K. Mittal, Quantum cryptography: state-of-art, challenges and future perspectives, in *7th IEEE International Conference on Nanotechnology (2007)*, pp. 1296–1301
20. C. Elliott, D. Pearson, G. Troxel, Quantum cryptography in practice. Preprint of SIGCOMM (2003)
21. M.S. Sharbaf, Quantum cryptography: an emerging technology in network security. *IEEE*, (2011)

Relation Estimation of Packets Dropped by Wormhole Attack to Packets Sent Using Regression Analysis



Sayan Majumder and Debika Bhattacharyya

Abstract MANET is exposed to the possibility of being attacked or harmed by so many attacks like Black hole, Wormhole, Jellyfish, etc. The attacking nodes can easily set Wormhole attack by duplicating a route which is shorter than the original within network. In this paper, we estimate the pattern of relation between packets dropped to packets sent, that is packet delivery fraction (PDF) by Wormhole attack using regression analysis, a machine learning approach, by method of least square (MLS), and least absolute deviation (LAD). A comparative analysis between these two techniques is done. As there is high degree of linearity between packets sent and packets dropped by Wormhole attack, the pattern of dropped packet is proved by MLS regression and LAD regression. Accuracy is also tested. MLS and LAD regression algorithms are also developed to estimate this pattern for Wormhole attack. Wormhole attackers create a duplicate tunnel (wired link or high frequency) from source to destination. An illusion is created by this, that the distance through this tunnel from source to destination is minimal and must take less time. Comparing to this fake tunnel, the original route takes more time. So it is necessary to calculate the time taken to estimate the relation of packets dropped to packets sent by Wormhole attack. Better performance by linear regression analysis for pattern estimation of these dropped packets is proved by simulation, done by MATLAB simulator for Wormhole attack.

Keywords MANET · Absolute deviation · Regression · LAD · Wormhole attack · MLS

S. Majumder (✉)
Dream Institute of Technology, Kolkata, India
e-mail: Sayanmajumder90@gmail.com

D. Bhattacharyya
Institute of Engineering and Management, Kolkata, India
e-mail: bdebika@iemcal.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharyya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_49

1 Introduction

The unique characteristics to make data flow within the network securely by mobile ad hoc networks (MANETs) are

- (a) shared wireless medium
- (b) peer-to-peer architecture which is opened to all
- (c) high progressive network topology.

MANET is a decentralized network, which is wireless. It achieves routing by forwarding data packets to other nodes. It also supports multihop trend. MANET is exposed to the possibility of being attacked or harmed by so many attacks like Black hole, Wormhole [1], Jellyfish[2], etc. among which Wormhole attack is the most risky or hazardous attack[3].

Wormhole attack [4] is a dangerous attack. By duplicating a route within network, one can easily set this attack. This type of attack strikes on network layer. Wormhole attack creates confusion between original path and duplicated path created by Wormhole nodes. Wormhole link is the high-speed link by which two attackers are connected in case of Wormhole attack. We can establish link with the help of wired system and network cable.

Statistical analysis is an effective way to detect and prevent Wormhole attack. SAM [5] detects the multipath routing, which is exposed to Wormhole attack. SAM did not use any security systems. A path which is able to be relied on as honest or truthful is also discovered by cross-correlation [6] method. Capacity and delay of mobile ad hoc network is also discussed using correlated mobility [7]. To make a trust vector to detect Wormhole attack, Khin Sandar Win used correlation [8] method. Regression analysis gives us idea about the relationship between variables. It is one kind of modeling technique by which we can get the total idea about the state of connectivity between variables.

Among so many regression techniques, in this paper we have chosen linear regression technique to estimate the relation between packets dropped by Wormhole attack to packets sent.

Linear regression is the best technique, where the variable, which is dependent to other, is continuous and variable, which is independent, can be continuous.

Method of least square (MLS) is more reliable way to find the line of best fit. Least absolute deviation (LAD) is another regression technique [9] which provides a robust solution when outliers are present.

R. Venkataraman, M. Pushpalatha, T. Rama Rao used regression to generate a trust model [10] to point out the harmful nodes which are responsible for severe attacks over AODV and OLSR routing protocols.

Fraser Cadger, Kevin Curran, Jose Santos, and Sandra Moffett used regression technique in their machine.

Learning algorithms to predict the location coordinates [11] of next move of a node, so that the time taken by the node may be minimal with no security attack.

LAD regression is a new technique which can be also used as a good machine learning approach to estimate the pattern between bivariate data. As previously

R. Venkataraman or Fraser Cadger used only MLS regression to predict location or attack in MANET, we have done a comparative study between MLS and LAD regression.

2 Proposed Methodology

Previously, it is already proved that the correlation value between packets dropped to packets sent by wormhole attack under AODV routing protocol is 0.95 or higher, which indicates high degree of linearity. In this paper, the relation is estimated and proved by means of suitable equations, derived on the basis of available bivariate data, i.e., packets dropped by Wormhole attack and packets sent to nodes. We have estimated the pattern or relation between dropped packets and sent packets here with a comparative study between two types of regression

- (a) MLS regression.
- (b) LAD regression.

MLS Regression Algorithm to estimate relation Between Packets dropped to Packets sent by Wormhole attack.

If bivariate data (number of packets dropped to packets sent) are plotted as points on a graph paper, it will be found that the concentration of the points follows a certain decoration establishing the state of being connected between the variables. We have found that the trend of points is to be linear, so we will determine the best fitting straight line by method of least square. MLS is a mode satisfactory process. To find the line of best fit for the set $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$, where

x_i = Number of packets dropped., y_i = Number of packets sent.

Algorithm 1

Step 1: Find the mean of the packets dropped (x) by Wormhole attack and packets sent (y).

$$\bar{X} = \sum^n x_i / n$$

$$\bar{Y} = \sum^n y_i / n$$

Step 2: Find the correlation coefficients between packets dropped and packets sent (slope of the line).

$$m = \frac{\sum^n (x_i - \bar{x})(y_i - \bar{y})}{\sum^n (x_i - \bar{x})^2}$$

Step 3: Find the packets sent (y) intercept.

$$b = \bar{y} - m \bar{x}$$

Step 4: Use the slope m and the y -intercept b to form the equation of the best fit line.

$$Y = \bar{y} + b(x_i - \bar{x})$$

Finish: Plot the graph accordingly.

LAD Regression Algorithm to estimate relation between packets dropped to packets sent by Wormhole attack

Only step 2 is different from previous one that is to calculate correlation coefficient we have to find, $E = \Sigma e^2 = \Sigma |(x - \bar{x})|$, that means divide the deviation by absolute value but not squared value.

To find the line of best fit for the set $(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)$, where, x_i = Number of packets dropped., y_i = Number of packets sent.

Algorithm 2

Step 1: Find the mean of the packets dropped (x) by Wormhole attack and packets sent (y).

$$\bar{X} = \sum^n x_i / n$$

$$\bar{Y} = \sum^n y_i / n$$

Step 2: Find the correlation coefficients between packets dropped and packets sent (slope of the line).

$$m = \sum^n (x_i - \bar{x})(y_i - \bar{y}) / \sum^n |(x_i - \bar{x})|$$

Step 3: Find the packets sent (y) intercept.

$$b = \bar{y} - m \bar{x}$$

Step 4: Use the slope m and the y -intercept b to form the equation of the best fit line.

$$Y = \bar{y} + b(x_i - \bar{x})$$

Finish: Plot the graph accordingly.

3 Simulation and Result

We have used MATLAB simulation technique to describe the results, here. MATLAB gives us right environment to set up the node distribution for Wormhole attack. Here, we have followed AODV routing algorithm.

Simulation time taken is 240 s by MLS and 190 s by LAD with simulation area of 255×255 pixels. Node distribution scenario is established in Fig. 1.

We can see Wormhole attack scenario in Fig. 2. Wormhole attack scenario is established in between red and blue colored nodes. The nodes are malicious. All the data packets between red and blue nodes will be forwarded directly, due to Wormhole attack. We have created Wormhole scenario with nine nodes, among which one Wormhole is featured with two malicious nodes. Wormhole follows multipath routing. So, starting from each node packets, we have taken into consideration the routing to destination.

As we have used MATLAB simulator, the parameters used to simulate the algorithms like simulation time, simulation area, number of Wormholes, protocol used, etc. are listed in Table 1.

Fig. 1 Node distribution scenario

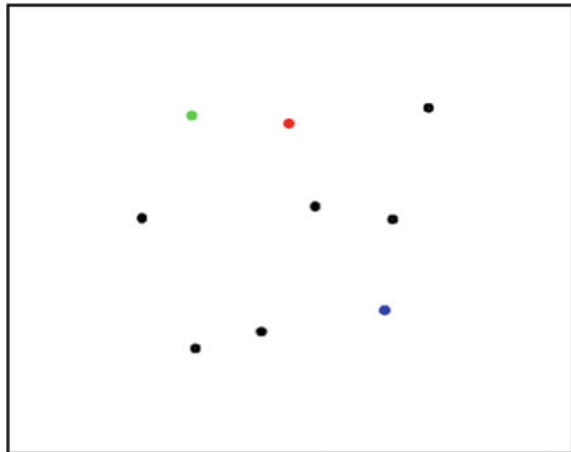


Table 1 Simulation parameters [12]

Parameters	Descriptions
Examined protocol	AODV
Simulation time	240 s (MLS), 190 s (LAD)
Simulation area	255×255 Pixel
Number of nodes	09
Malicious nodes	02
Number of wormholes	01

Fig. 2 Network affected by Wormhole

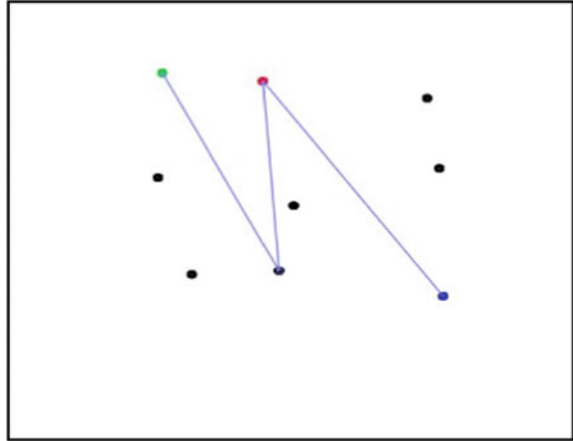
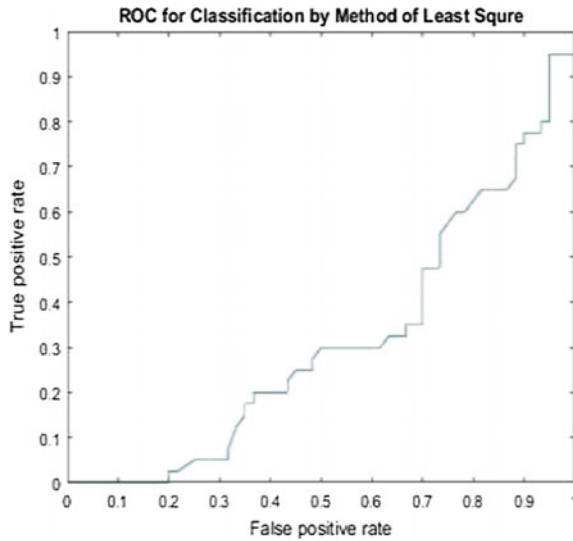


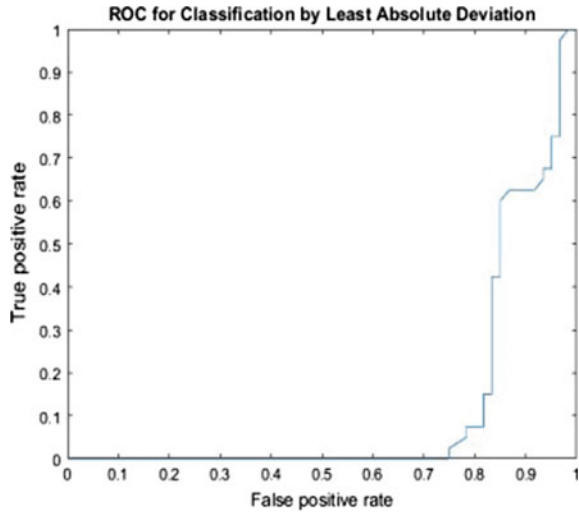
Fig. 3 Regression analysis by MLS



After calculating the number of packets dropped due to Wormhole attack to number of packets sent by nodes, we have calculated the correlation coefficients for each routing. Then, according to the Algorithm 1, MLS regression is depicted as plotted in Fig. 3. Along x -axis, false positive rate is taken, and along y -axis, true positive rate is taken.

From the above graph, we can observe that the true positivity is increasing accordingly up to 0.95. As we previously proved that correlation coefficient for Wormhole attack must be 0.95 or higher. From this graph, we also can draw the pattern of our line of best fit (best fitting curve) for the set of bivariate data.

Fig. 4 Regression analysis by LAD



After calculating the correlation coefficients for each routing, according to the Algorithm 2, LAD regression is depicted as plotted in Fig. 4. Along x-axis, false positive rate is taken, and along y-axis, true positive rate is taken.

Our observation from the above graph is that least absolute deviation technique is also increasing up to higher than 0.95. We can also draw the pattern of best fitting straight line from the bivariate data for LAD regression.

But the false positive rate is starting here from 0.75, which is much greater than the previous one that is 0.20 in case of method of least square.

We have taken nine nodes here, among which two nodes are malicious, because of which Wormhole attack takes into place. But anybody can take large number of nodes as a dataset and can prove this attack with two or more malicious nodes.

After doing regression analysis by MLS and LAD technique, it is estimated that there is a high degree of linearity between packets dropped and packet sent by Wormhole attack scenario. Line of best fit is measured and plotted in Fig. 5, which gives quite good result.

Accuracy is calculated for both the method of MLS and LAD regression analysis, and accuracy percentage is shown in Fig. 6.

$$\text{Accuracy} = \frac{\text{TN} + \text{TP}}{\text{TN} + \text{TP} + \text{FN} + \text{FP}},$$

where

- TN True negative,
- FN False negative,
- TP True positive.

Fig. 5 Line of best fit

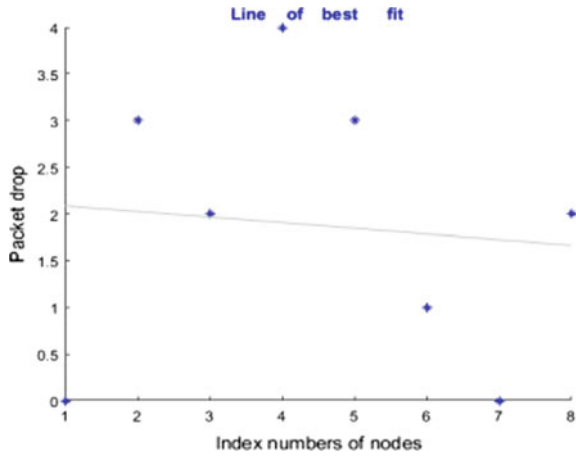
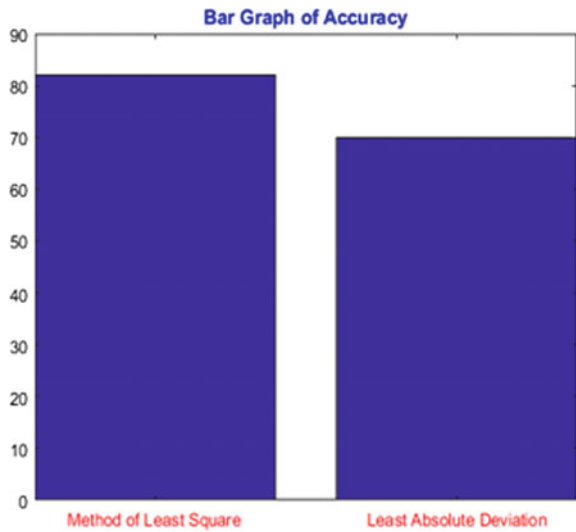


Fig. 6 Accuracy percentage



Using MLS, we have calculated accuracy as **82%** of our training data fitted well for best fit. Using LAD, we have found accuracy as only **70%** of our training data fitted well.

4 Conclusion

We have done here a comparative study between MLS regression and LAD regression to estimate the pattern of dropped packets to sent packets by Wormhole attack using AODV routing algorithm.

We can consider method of least square (MLS) as an effective statistical estimator to estimate the relation of packets dropped by Wormhole attack to the packets sent in MANET.

MLS gives more accuracy percentage than least absolute deviation (LAD) technique. MLS gives 82% accuracy, whereas LAD regression gives only 70% accuracy on training data of packets dropped to packets sent.

LAD regression takes less time than MLS technique, as the calculation overhead is minimal in case of LAD technique.

LAD regression can be used as a good statistical estimator where outliers are present because least absolute deviation technique is less affected by outliers.

So we can conclude that though accuracy percentage of MLS technique is 12% higher than LAD technique, one can use least absolute deviation technique to estimate the pattern more efficiently and in lesser time with large dataset, containing outliers also.

Three main advantages of LAD over MLS technique are,

1. Computational overhead is minimal in case of LAD regression as there is the only concept of absolute value but not squared value.
2. Execution time taken by LAD regression is also less than MLS regression.
3. Presence of outliers can be detected well by LAD technique.

Lastly, we can say that we have make a comparative study between MLS and LAD regression using AODV routing algorithm only for linear regression and we have found that LAD regression is quite better in case of large dataset, where the outliers are present.

References

1. S. Majumder, D. Bhattacharyya, Mitigating wormhole attack in MANET using absolute deviation statistical approach, in *2018 IEEE 8th Annual, on Computing and Communication Workshop and Conference* (2018), pp. 317–320
2. F. Xing, On the survivability of wireless ad-hoc networks with node misbehaviors and failures. *Secur. Comput., IEEE Trans. On* 7(3), 284–299 (2010)
3. M.N. Alslaim, H.A. Alaqel, S.S. Zaghoul. A comparative study of MANET routing protocols, in *2014 Third International Conference on e-Technologies and Networks for Development* (2014)
4. D.B. Roy, R. Chaki, N. Chaki, A new cluster-based wormhole intrusion detection algorithm for mobile ad-hoc networks. *IJNSA*, 44–52 (2009)
5. L. Qiana, N. Songa, X.F. Li, Detection of wormhole attacks in multi-path routed wireless ad hoc networks: a statistical analysis approach. *J. Netw. Comput. Appl.* 2005
6. G.K. Patnaik, M.M. Gore, Trustworthy path discovery in MANET—a message oriented cross-correlation approach, in *Workshops of International Conference on Advanced Information Net-working and Applications* (2011), pp. 170–177
7. R. Jia, F. Yang, S. Yao, X. Tian, X. Wang, W. Zhang, J. Xu, Optimal capacity–delay tradeoff in MANETs with correlation of node mobility. *IEEE Trans. Veh. Technol.*, 66(2), 1772–1785 (2017)

8. K. Sandar Win, Analysis of detecting wormhole attack in wireless networks, *Int. J. Electron. Commun. Eng.*, **2**(12), 2704–2710 (2008)
9. Y. Li, G.R. Arce, A maximum likelihood approach to least absolute deviation regression. *EURASIP J. Appl. Sig. Process.* **2004–12**, 1762–1769 (2004)
10. R. Venkataraman, M. Pushpalatha, T. Rama Rao, Regression- based trust model for mobile ad hoc networks. *IET Inf. Secur.*, **6**(3), 131–140 (2012)
11. F. Cadger, K. Curran, J. Santos, S. Moffett, MANET location prediction using machine learning algorithms, in *International Conference on Wired/Wireless Internet Communications*, (2012), pp. 174–185
12. S. L. Rosen, J. A. Stine and W. J. Weiland, A manet simulation tool to study algorithms for generating propagation maps, in *Proceedings of the 2006 Winter Simulation Conference*, Monterey, CA, (2006), pp. 2219–2224

A Review on Agricultural Advancement Based on Computer Vision and Machine Learning



Abriti Paul, Sourav Ghosh, Amit Kumar Das, Saptarsi Goswami, Sruti Das Choudhury and Soumya Sen

Abstract The importance of agriculture in modern society need not be overstated. In order to meet the huge requirements of food and to mitigate, the conventional problems of cropping smart and sustainable agriculture have emerged over the conventional agriculture. From computational perspective, computer vision and machine learning techniques have been applied in many aspects of human and social life, and agriculture is not also an exception. This review paper gives an overview of machine learning and computer vision techniques which are inherently associated with this domain. A summary of the works highlighting different seeds, crops, fruits with the country is also enclosed. The paper also tries to give an analysis, which can help researchers to look at some relevant problems in the context of India.

Keywords Computer vision · Machine learning · Agriculture · Plant phenotyping

A. Paul (✉) · A. K. Das · S. Goswami · S. Sen
A. K. Choudhury School of Information Technology, University of Calcutta,
Kolkata 700106, India
e-mail: paulabriti@gmail.com

A. K. Das
e-mail: amitkrdas.kol@gmail.com

S. Goswami
e-mail: saptarsi007@gmail.com

S. Sen
e-mail: iamsoumyasen@gmail.com

S. Ghosh
IBM India Pvt. Ltd., Kolkata, India
e-mail: souravghos@gmail.com

S. Das Choudhury
School of Natural Resources & Department of Computer Science and Engineering,
University of Nebraska-Lincoln, Lincoln, USA
e-mail: srutidc@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_50

1 Introduction

According to the most recent estimation by the United Nations (UN), growth of population will reach 9.7 billion and global demand of food is expected to increase 59–98% by 2050 [1]. In order to feed these huge populations, it is necessary to adopt advanced computing technologies like computer vision, machine learning, big data analytics, etc. in the different applications of agriculture. It has been found from the reports that the wastage of agricultural product is almost 40% due to the issues like lack of maintenance, non-identified like disease, etc.

There have been numerous research works on how computer vision and machine learning are used to improve conventional agriculture. Most of them are focused on early detection of disease or improvement of quality of foods or determination of color of affected area or identification of nutrient deficiencies in a particular crop based on visible phenotypes. They vary in terms of techniques as well as crops, type of data they work with, and the country on which where these have been applied. In our survey, we actually focused on the last decade's research efforts on the different techniques of computer vision and machine learning applied in agriculture.

The motivation of the paper is to structurally summarize the different components that are used in a typical smart agriculture pipeline starting from image acquisition to processing and then finally building the models. This also presents a country and crop-wise summary to highlight the most researched crops. The main aim of this research work is to perform a thorough survey by integrating the concepts of machine learning and computer vision for agriculture-based applications to help the researchers to have a deep insight into this area.

The paper is organized in the following form. Section 2 represents comparisons with other survey papers in this domain. In Sects. 3 and 4, different image acquisition techniques and computer vision techniques have been discussed. Machine learning models are outlined in Sect. 5. Section 6 shows country and crop-wise summary details along with India context are depicted. Section 7 contains relevant work summary in India, and Sect. 8 represents conclusion and future aspect of it.

2 Comparisons with Other Survey Papers

A number of review works have been done over the last few years on the applications of machine learning in the field of agriculture. Brosnan and Sun [2] provided an extensive survey on the applications of computer vision in food and agricultural products. It described how computer vision was used to inspect and grade fruits and vegetables.

A broad review on the advancement of learning techniques to evaluate food quality applying computer vision methodologies is presented by Du and Sun [3]. The review focused on how different machine learning techniques (artificial neural network, statistical learning, etc.) have been applied in assessing quality of cereals, fruits,

vegetables, and even meat. An extensive review of big data application and machine learning techniques in plants has been presented by Ma et al. [4]. The paper talks about big data infrastructure for plant science and summarizes different machine learning packages that are used in the plant science domain.

Gandhi and Armstrong [5] presented a comprehensive review of how different data mining techniques like artificial neural network (ANN), Bayesian network, support vector machine (SVM), and association rule mining (ARM) are used in the areas of agriculture. Image acquisition and image processing play a crucial role in acquiring necessary data in plant science. Perez-Sanz et al. [6] provide an overview of different image acquisition and processing techniques.

In this survey, we attempt to present an extensive summary on the research carried out in agricultural domain on machine learning application. Most of the machine learning research papers focus on early detection on disease of a particular crop based on visible phenotypes. They elaborate on the specific methods used to capture the images, image processing techniques to extract data, machine learning models used on the data for feature selection, detection and subsequently development of algorithm either for classification or prediction. This paper also focuses on the cross-validation of the model with the tested data.

In this paper, we added few more dimensions such as:

- A dedicated section to outline country and crop/seed-wise survey which has happened over a period of decades on this field.
- Summarizes the significant work done in India on machine learning application in agriculture.
- It highlights most frequently used image acquisition and processing techniques and machine learning algorithms.

3 Image Acquisitions

Image acquisition is a method where an image is generally captured by a camera and then converted into a manageable entity before processing it. This gives us the digital representation of a scene in terms of an image. Image is represented in terms of pixels [6]. The pixel values indicate light intensity in one or more spectral bands as well as related to the physical measures like depth, reflectance, absorption of sonic or electromagnetic waves, and nuclear magnetic resonance. A scene can be captured by digital camera, imaging sensors, or any digital device. By some special hardware, these image signals are amplified, filtered, transported, and enhanced. Based on the nature of the capturing devices, the output image data can be represented as ordinary 2D image or 3D volume or as the sequences of images.

3.1 *Image Acquisition Methods of Plant Phenotyping*

3.1.1 Digital Imaging

It is the creation of different photographic images such as the structure of an object or physical scene as captured by camera or similar devices. It is available from the analog mediums such as scanner and printed photographic paper/film.

3.1.2 Infrared (IR) Imaging

It is used for looking into the movements of the internal molecules. There are two popular infrared imaging techniques—NIR and far-IR (thermal IR), a combination of visible images and NIR images to detect the total health of plant.

3.1.3 Fluorescence Imaging

This methodology is deployed in different types of experimental setups. Ultraviolet light is reflected in a form of discrete wavelengths [6] through different plant components in the range of 340–360 nm. It measured with CCD camera along with sensitive fluorescence signals. These images have many usages like early detection of stress, metabolic status of a plant.

3.1.4 Thermographic Imaging

It is a method to improve the visibility of objects mainly in the dark environment. It helps to detect the object's infrared radiation, and image is created based on that data. It works without ambient light environment.

3.1.5 Spectroscopic Imaging

It is the measure of interaction of solar radiation with plants using remote sensors. This method helps to divide image into bands, moreover capable to identify the existence of selected elements from the tree.

3.1.6 Mono-RGB Imaging

Mono-RGB imaging process is a composite collection of different devices like imaging sensors, lens, specific H/W, and IO interface. Mono-RGB vision devices include

Table 1 Frequently used image acquisition techniques along with references

Image capturing techniques	References
Digital imaging	Wang et al. [10], Lu et al. [11], Gilandeh et al. [12], Haug et al. [13], Bendary et al. [14], Blasco et al. [15], Park et al. [16], Qureshi et al. [17]
Spectroscopic imaging	Sankaran et al. [18], Chlingaryana et al. [19], Singh et al. [20]
Thermal imaging	Wang et al. [21]
RGB images	Casanova et al. [22], McKinnon et al. [23], Ramya et al. [24], Tellaechea et al. [25]
Hyperspectral imaging	Behmann et al. [26], Backhaus et al. [27], Blasco et al. [15]
Radar images	Coopersmith et al. [28]
Scanned images	Chan et al. [29], Zareiforoush et al. [30], Ebrahimi et al. [31], Kuo et al. [32]
Sensor imaging	Johnsone et al. [33], Pantazi et al. [34], Emmi et al. [35]
Greenhouse hyper spectral reflection images	Rumpf et al. [36]
Fluorescence imaging spectroscopy	Yao et al. [37], Atas et al. [38], Casanova et al. [39], Pena et al. [40] Perez-Bueno et al. [41]

different smart tools to predict and improve crop yield. It also incorporates automated phenotyping prototype of large plants in the greenhouse.

Most frequently used image acquisition techniques are mentioned in Table 1.

4 Image Processing Techniques

Image processing is an important tool/technique for enhancement of agricultural product. After getting the raw images information through different image acquisition techniques, we apply various processing method for improving quality and extracting some useful information [7].

The major purposes of using image processing [8, 9] in the field of agriculture are

- (a) Identification of disease.
- (b) Improvement of quality of foods.
- (c) Determination of color of affected area.
- (d) Automation in management of crop.
- (e) Identification of nutrient deficiencies.

After image acquisition, the following steps take place:

- Image preprocessing.
- Image segmentation.
- Image representation and description.
- Feature selection.

4.1 Image Preprocessing

The objective of image preprocessing is to make the images better than the original viewed one. It is of two types: (a) compression (reducing storage requirement and bandwidth) and (b) enhancement (changing brightness and color).

4.2 Image Segmentation

It is the process of partitioning any image into numbers of objects or constituting parts and extracting the required attributes from it.

4.3 Image Representation and Description

It is the raw pixel data, output of the basic segmentation part. Description basically identifies and differentiates one class object from another.

4.4 Feature Selection

It is an important method which identified the most accurate information such that the features of interest are highlighted (Fig. 1; Table 2).

5 Machine Learning Models

Short descriptions of the different machine learning models which have been commonly used are discussed below:

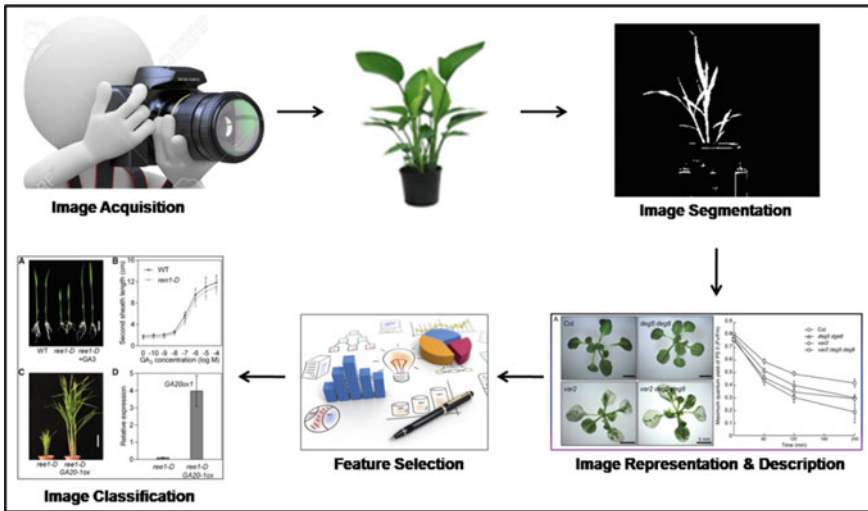


Fig. 1 Image processing pipeline

Table 2 Image processing techniques

Image processing tasks	Applied techniques	References
Preprocessing	Compression, background subtraction, noise removal, polygon shape creation	Wang et al. [10], Lu et al. [11], Haug et al. [13], Bendary et al. [14], Blasco et al. [15], Pantazi et al. [34], Kulkarni et al. [42], Kuo et al. [32]
Smoothing	Median filter, bilateral filter, gradient weighting filter	Blasco et al. [15], Singh et al. [43], Chaudhary et al. [44]
Segmentation	Super pixel, K-means clustering, fuzzy C-means, edge-based, fuzzy, nearest neighbors, and genetic algorithm	Wang et al. [10], Gilandeh et al. [12], Haug et al. [13], Bendary et al. [14], Wang et al. [21], Backhaus et al. [27], Blasco et al. [15], Arivazhagan et al. [45], Pantazi et al. [34], Zhang et al. [46], Namin et al. [47], Park et al. [16], Singh et al. [43], Bai et al. [48], Chaudhary et al. [44], Qureshi et al. [17], Park et al. [16], Kulkarni et al. [42], Ebrahimi et al. [31]
Feature selection	Principal component analysis, fast Fourier transform, outline extraction, histogram of wavelength, gray-level co-occurrence matrix	Wang et al. [10], Lu et al. [11], Bendary et al. [14], Wang et al. [21], Blasco et al. [15], Arivazhagan et al. [45], Zhang et al. [46], Namin et al. [47], Singh et al. [43], Bai et al. [48], Kulkarni et al. [42], Ebrahimi et al. [31], Phadikar et al. [49]

5.1 Naive Bayes Classifier

This is a classification technique developed using Bayes' theorem. This methodology works on the assumption that the presence of one of the feature of a class is unrelated to other features.

5.2 K-Means Clustering

This clustering methodology is a type of unsupervised learning which is applied on unlabeled data. The objective of this algorithm is to identify certain non-overlapping groups in the data, and K is a variable which represents the number of groups. The algorithm works in an iterative manner and assigns every data point to one of the K groups. Based on feature similarity data, points are clustered.

5.3 Support Vector Machines (SVM)

This is a well-known and very popular supervised algorithm which is used in both classification and regression problems. It applies nonlinear mapping methodology to transform the given training data to a higher dimension. In this new dimension, it looks for the linear optimal separating hyperplane. SVM is capable to identify hyperplane with the help of support vector and margin.

5.4 Artificial Neural Networks (ANN)

It is a theoretical model that depicts the relationship between input and output signals derived from biological neural networks where brain responds to stimuli from sensory inputs. In brain, interconnected cells are called neurons which are capable of huge parallel processing, and similarly in ANN, a network of artificial neurons or nodes help to solve the problems in different integrated methods.

5.5 Decision Trees (DT)

It is another supervised learning algorithm used in classification problems. This classification does not require parameter setting or domain knowledge and therefore suitable for exploratory knowledge discovery. It can handle multi-dimensional data. The learning and classification phases of decision tree is fast and simple

Table 3 Various machine learning model as used in existing research

Machine learning models	References
Naive Bayes classifier	Skrubej et al. [50], Blasco et al. [15]
<i>K</i> -means clustering	Kulkarni et al. [42], Hiary et al. [51], Wang H et al. [26], Qureshi et al. [17], Coopersmith et al. [28]
Support vector machines	Behmann et al. [26], Bendary et al. [14], Arivazhagan et al. [45], Yao et al. [37], Rumpf et al. [36]
Artificial neural networks	Sanz et al. [6], Lu Y et al. [11], Gilandeh et al. [12], Backhaus et al. [27], Johnsona et al. [33], Pantazi et al. [34]
Decision trees	Blasco et al. [15], Skrubej et al. [50]
Linear discriminate analysis	Bendary et al. [14], Blasco et al. [15]
Deep learning	Namin et al. [47], Ferentinos et al. [52], Kamilaris et al. [53], French et al. [54]

5.6 Linear Discriminate Analysis (LDA)

This is a well-established method to perform dimension reduction problems which is used in preprocessing for the applications like machine learning, pattern classification, etc. LDA separates two classes depending upon the concept of searching for a linear combination of variables.

5.7 Deep Learning

It is a fine-tuned subclass of broader machine learning family to learn and represent data. It is an algorithmic method based on artificial neural networks with multiple layers inspired by structure and function of human brain. Recently, deep learning techniques are applied in various challenges of the agriculture domain. This method has more predictive power.

Application of different machine learning and deep learning models in various research papers are shown below (Table 3).

6 Country and Crop/Seed-Wise Summary

In this section, a review of the crops, fruits, and vegetables has been discussed.

- In India, work has been done on sugarcane, rice, and grape.
- Some of the other top exported agricultural products of India are wheat and pulses, where some research effort can be undertaken (Table 4).

7 Summaries of Relevant Works in India

In this section, we explore the application of image processing and machine learning in agriculture domain in India. There have not been many instances of significant work in the past. However, on the back of conducive government initiatives and increased foreign direct investment (FDI), use of technology to improve yield has become a need of the hour. Consequently tech startups and companies have come up with products making lives easier for the peasants.

7.1 *Application on Image Processing in Agriculture*

A Bengaluru-based company Intello Labs uses deep learning to provide advanced image recognition technology. This is used in

- (a) Agricultural product grading: This is an automated quality analysis method of images to grade fresh vegetables, grains, fruits, cotton, etc. characterized by shape, color, and size. This application analyzes the photograph taken by the farmer and evaluates the quality of the agricultural produce online.
- (b) Alerts on crop infestation: Using their solution, any farmers can get an idea about the diseases, infections, and weeds grown by capturing an image of their crop. This approach exploits deep neural network and image processing techniques to recognize any diseases in the agricultural produce or pest infestation of pest in the crops. It also gives suitable direction on treatment as well as prevention mechanism.

7.2 *Predictive Analytics on Agriculture*

Both excess rainfall and drought are serious challenges to agriculture in India. Therefore, identifying the exact sowing time is a serious concern of the farmers in India. In order to address this problem, Microsoft jointly with International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) developed an Artificial Intelligence

Table 4 Detailed view of country-wise crops and its shortcoming is undertaken

Seed	Country	Problem	References
Apple	China	Plant disease leaf spot segmentation	Bai et al. [48]
Barley plant	Italy	Identification of early plant stress responses in hyper spectral images	Behmann et al. [26]
Carrot	Germany	Evaluating computer vision based precision agriculture tasks	Haug et al. [13]
Chili pepper	Turkey	Aflatoxin detection	Atas et al. [38]
Citrus	Spain	Development of a decision support system for effectively handling irrigation in agriculture	Torres-Sanchez et al. [55]
Cotton	UK	Automatic identifying plant disease visual symptoms	Camargo et al. [56]
Cucumber	China	Plant disease leaf spot segmentation	Bai et al. [48]
Cucumber	USA	Identification of leaf diseases and their subsequent classification.	Zhang et al. [46]
Grape	India	Application of neural networks for diagnosis and categorization of leaf diseases	Kulkarni et al. [42]
Mango	Mexico	Early detection of mechanical damage during postharvest	Blasco et al. [15]
Olive	Spain	Early detection and quantification of verticillium wilt disease	Tejada et al. [57]
Orchid	Taiwan	Applying color and texture feature for detection of disease	Huang et al. [58]
Rice	China	Applying color and texture feature for detection of disease	Yao et al. [37]
Rice	India	Categorization of diseases.	Phadikar et al. [49]
Rice	India	Designing of an incremental classifier for disease prediction	Sengupta et al. [59]
Rice	Taiwan	Detecting disease in rice seedlings by machine vision	Kuo et al. [32]
Sugar beet	Jordan	Fast and accurate detection diseases of leaves	Hiary et al. [51]
Sugar beet	Germany	Detection of the disease earl and differentiation of them	Rumpf et al. [36]

(continued)

Table 4 (continued)

Seed	Country	Problem	References
Sugarcane	Australia	Regional forecast for crop production	Everingham et al. [60]
Sugarcane	India	Categorizing the yield	Papageorgiou et al. [61]
Tobacco	Germany	Nutrition state classification	Backhaus et al. [27]
Tomato	Egypt	Evaluate ripeness	Bendary et al. [14]
Tomato	Slovenia	Automatic assessment of seed germination rate	Skrubej et al. [50]
Wheat	Iran	Automatic purity measurement device	Mollazade et al. [31]
Wheat	UK	Yield prediction	Pantazi et al. [34]

(AI)-based sowing app that used machine learning-based techniques as well as business intelligence.

Climate data (spanning over three decades) and AI techniques have successfully calculated the crop sowing period for Andhra Pradesh in the year 2015. Moisture Adequacy Index (MAI), the measure by which one can assess the degree of adequacy of soil moisture and rainfall to get the idea about the potential water requirement of crops, was applied to determine the optimal sowing period in that specific area in Andhra Pradesh.

Currently, the farmers across AP and Karnataka are getting a text intimation before sowing the seeds. During the stages of cropping, automated voice calls are alerting to a quite good number of farmers from Telangana, Maharashtra, and Madhya Pradesh if their agricultural products are at vulnerable to an attack of pest based on climatic conditions.

8 Conclusions

Applications of machine learning, deep learning, and computer vision techniques have advanced the field of agriculture significantly in recent times. In this survey, we have summarized important techniques and algorithms both from machine learning as well as computer vision which have been applied for improving conventional agriculture. A brief account of current adaptation in different states has also been discussed. Agricultural product and country-wise summary are done with a specific focus on India. This paper will give a vivid idea to the new researchers to understand the processes of machine learning and image processing associated with agriculture and motivate them to find out the relevant problems in this area.

References

1. World Population Prospects. The 2015 Revision, United Nations New York (2015)
2. T. Brosnan, D. Sun, Inspection and grading of agricultural and food products by computer vision systems. *Comput. Electron. Agric.* **36**, 193–293 (2002)
3. J.C. Du, D.W. Sun, Learning techniques used in computer vision for food quality evaluation a review. *J. Food Eng.* **72**, 39–55 (2005)
4. C. Ma, H.H. Zhang, X. Wang, Machine learning for big data analytics in plants. *Trends Plant Sci.* **19**(12), 798–808 (2014)
5. N. Gandhi, J.L. Armstrong, A review of the application of data mining techniques for decision making in agriculture, in *2nd International Conference on Contemporary Computing and Informatics* (2016)
6. F.P. Sanz, P.J. Navarro, M.E. Cortines, Plant phonemics: an overview of image acquisition technologies and image data analysis algorithms. *Giga Sci.* **6**(11), 1–18 (2017)
7. V.P. Kulalvaimozhi, G.M. Alex, J.S. Peter, Image processing in agriculture www.ijaetmas.com **04**(03), 142–151 (2017)
8. R. Chahar, P. Soni, A study of image processing in agriculture for detect the plant disease. *Int. J. Comput. Sci. Mob. Comput.* **4**(7), 581–587 (2015)
9. A.P. Janwale, S.L. Santosh, Digital image processing applications in agriculture: a survey (2015)
10. H. Wang, S. Zhang, W. Huang, Z. You, Plant diseased leaf segmentation and recognition by fusion of super pixel. K-means and PHOG. *Optik* **157**, 866–872 (2018)
11. Y. Lu, S. Yi, Z. Nianyin, Y. Liu, Y. Zhang, Identification of rice diseases using deep convolutional neural networks. *Neurocomputing* **267**, 378–384 (2017)
12. A.Y. Gilandeh, S. Sabzi, H. Javadikia, Machine vision system for the automatic segmentation of plants under different lighting conditions. *Biosys. Eng.* **161**, 157–173 (2017)
13. S. Haug, J. Ostermann, Crop/Weed field image dataset for the evaluation of computer vision based precision agriculture tasks, in *ECCV 2014 Workshops, Part IV, LNCS 8928*, (2015), pp. 105–116
14. N. Bendary-El, E. Hariri -El, E.A. Hassanien, Using machine learning techniques for evaluating tomato ripeness. *Expert. Syst. Appl.* **42**, 1892–1905 (2015)
15. V.N. Rivera, G.J. Sanchis, C.J. Perez, J.J. Carrasco, M.M. Giraldo, D. Lorente, S. Cubero, J. Blasco, Early detection of mechanical damage in mango using NIR hyper spectral images and machine learning. *Biosys. Eng.* **122**, 91–98 (2014)
16. S.D. Park, C. Kim, S. Yoon, A. Fuentes, A robust deep learning based detector for real-time tomato plant diseases and pests recognition. *Sensors* **17**, 2022 (2017)
17. S.W. Qureshi, A. Payne, K.B. Walsh, R. Linker, O. Cohen, M.N. Dailey, Machine vision for counting fruit on mango tree canopies. *Precision Agric.* **18**, 224–244 (2017)
18. S. Sankaran, A. Mishra, J.M. Maja, R. Ehsani, Visible-near infrared spectroscopy for detection of Huanglongbing in citrus orchards. *Comput. Electron. Agric.* **77**, 127–134 (2011)
19. A. Chlingaryana, S. Sukkarieha, B. Whelan, Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agric.* **151**, 61–69 (2018)
20. A. Singh, B. Ganapathysubramanian, K.A. Singh, S. Sarkar, Machine learning for high-throughput stress phenotyping in plants. *Trends Plant Sci.* **21**(2), 110–124 (2016)
21. F. Wang, L. Song, K. Omasa, J. Wang, Automatically diagnosing leaf scorching and disease symptoms in trees/shrubs by using RGB image computation with a common statistical algorithm. *Ecol. Inform.* **38**, 110–114 (2017)
22. C.B. Wetterich, R. Kumar, S. Sankaran, B.J. Junior, R. Ehsani, G.L. Marcassa, A comparative study on application of computer vision and fluorescence imaging spectroscopy for detection of Huanglongbing citrus disease in the USA and Brazil. *J. Spectrosc.* 1–6 (2013)
23. T. McKinnon, P. Hoff, Comparing RGB-based vegetation indices with NDVI for agricultural drone imagery. *AGBX* **02**, 1–17 (2017)

24. V. Ramya, A.M. Lydia, Leaf disease detection and classification using neural networks. *Int. J. Adv. Res. Comput. Commun. Eng.* **5**(11) (2016)
25. A. Tellaechea, G. Pajares, P.X. Burgos Artizzub, A. Ribeiro, A computer vision approach for weeds identification through support vector machines. *Appl. Soft Comput.* **11**, 908–915 (2011)
26. J. Behmann, J. Steinrucken, L. Plumer, Detection of early plant stress responses in hyperspectral images. *ISPRS J. Photogramm. Remote. Sens.* **93**, 98–111 (2014)
27. A. Backhaus, F. Bollenbeck, U. Seiffert, Robust classification of the nutrition state in crop plants by hyper spectral imaging and artificial neural networks, in *3rd Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)* (2011)
28. J.E. Coopersmith, S.B. Minsker, E.C. Wenzel, J.B. Gilmore, Machine learning assessments of soil drying for agricultural planning. *Comput. Electron. Agric.* **104**, 93–104 (2014)
29. D. Chen, K. Neumann, S. Friedel, B. Kilian, M. Chen, T. Altmann, C. Klukas, Dissecting the phenotypic components of crop plant growth and drought responses based on high-throughput image analysis. *Plant Cell* 4636–4655 (2014)
30. H. Zareiforoush, S. Minaei, R.M. Alizadeh, A. Banakar, Potential applications of computer vision in quality inspection of rice: a review. *Food Eng. Rev.* **7**, 321–345 (2015)
31. K. Mollazade, E. Ebrahimi, S. Babaei, Toward an automatic wheat purity measuring device: a machine vision-based neural networks-assisted imperialist competitive algorithm approach. *Measurement* **55**, 196–205 (2014)
32. F.Y. Kuo, L.C. Chung, J.K. Huang, Y.S. Chen, H.M. Lai, C.Y. Chen, Detecting Bakanae disease in rice seedlings by machine vision. *Comput. Electron. Agric.* **121**, 404–411 (2016)
33. D.M. Johnsona, W.W. Hsieha, J.A. Cannonb, A. Davidson, F. Bedardd, Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agric. For. Meteorol.* **218–219**, 74–84 (2016)
34. X.E. Pantazi, D. Moshou, T. Alexandridis, R.L. Whetton, A.M. Mouazen, Wheat yield prediction using machine learning and advanced sensing techniques. *Comput. Electron. Agric.* **121**, 57–65 (2016)
35. L. Emmi, M. Gonzalez de Soto, P. Gonzalez de Santos, Configuring a fleet of ground robots for agricultural tasks, in *ROBOT2013: First Iberian Robotics Conference* (2014), pp. 505–517
36. T. Rumpf, A.K. Mahlein, U. Steiner, E.C. Oerke, H.W. Dehne, L. Plumer, Early detection and classification of plant diseases with support vector machines based on hyperspectral reflectance. *Comput. Electron. Agric.* **74**, 91–99 (2010)
37. Q. Yao, Z. Guan, Y. Zhou, J. Tang, Y. Hu, B. Yang, Application of support vector machine for detecting rice diseases using shape and color texture features, in *ICEC '09 Proceedings of the 2009 International Conference on Engineering Computation* (2009), pp. 79–83
38. M. Atas, Y. Yardimci, A. Temizel, A new approach to aflatoxin detection in chili pepper by machine vision. *Comput. Electron. Agric.* **87**, 129–141 (2012)
39. J.J. Casanova, A.S. O'haughnessy, R.S. Evett, M.C. Rush, Development of a wireless computer vision instrument to detect biotic stress in wheat. *Sensors (Basel)* **14**, 17753–17769 (2014)
40. M.J. Pena, T.J. Sanchez, S.A. Prez, I.A. de Castro, L.F. Granados, Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors* **15**, 5609–5626 (2015)
41. M.L. Perez-Bueno, M. Pineda, F.M. Cabeza, M. Baron, Multicolor fluorescence imaging as a candidate for disease detection in plant phenotyping. *Front Plant Sci.* 1790 (2016)
42. S.S. Sannakki, S.V. Rajpurohit, B.V. Nargund, P. Kulkarni, Diagnosis and classification of grape leaf diseases using neural networks, in *Fourth International Conference on Computing Communications and Networking Technologies (ICCCNT)* (2013)
43. V. Singh, K.A. Misra, Detection of plant leaf diseases using image segmentation and soft computing techniques. *Inf. Process. Agric.* **4**, 41–49 (2017)
44. Chaudhary, P., Chaudhari, K.A., Cheeran, N.A., Godara, S.: Color transform based approach for disease spot detection on plant leaf. *Int. J. Comput. Sci. Telecommun.* **3**(6), (2012)
45. S. Arivazhagan, N.R. Shebiah, S. Ananthi, V.S. Varthini, Detection of unhealthy region of plant leaves and classification of plant leaf diseases using texture features. *Agric. Eng. Int: CIGR J.* **15**(1), 211–217 (2013)

46. S. Zhang, Z. Wu, Z. You, L. Zhang, Leaf image based cucumber disease recognition using sparse representation classification. *Comput. Electron. Agric.* **134**, 135–141 (2017)
47. T.S. Namin, M. Esmailzadeh, M. Najafi, B.T. Brown, O.J. Borevitz: Deep phenotyping: deep learning for temporal phenotype/genotype classification. CC-BY 4.0 International License (2017)
48. X. Bai, L. Zhang, X. Li, Z. Fu, X. Lv, A fuzzy clustering segmentation method based on neighborhood grayscale information for defining cucumber leaf spot disease images. *Comput. Electron. Agric.* **136**, 157–165 (2017)
49. S. Phadikar, J. Sil, K.A. Das, Rice diseases classification using feature selection and rule generation techniques. *Comput. Electron. Agric.* **90**, 76–85 (2013)
50. U. Skrubej, C. Rozman, D. Stajanko, Assessment of germination rate of the tomato seeds using image processing and machine learning. *Europ. J. Hort. Sci.* **80**(2), 68–75 (2015)
51. H. Al Hiary, S., Bani Ahmad, M. Reyalat, M. Braik, Z. Rahamneh, Fast and accurate detection and classification of plant diseases. *Int. J. Comput. Appl.* **17**, 8875–8887 (2011)
52. P.K. Ferentinos, Deep learning models for plant disease detection and diagnosis. *Comput. Electron. Agric.* **145**, 311–318 (2018)
53. A. Kamilaris, X.F. Prenafeta-Boldu, Deep learning in agriculture- a survey. *Comput. Electron. Agric.* **174**, 70–90 (2018)
54. P.A. French, P.M. Pound, A.J. Atkinson, J.A. Townsend, H.M. Wilson, M. Griffiths, S.A. Jackson, A. Bulat, G. Tzimiropoulos, M.D. Wells, H.E. Murchie, P.T. Pridmore, Deep machine learning provides state-of-the-art performance in image-based plant phenotyping. *Giga Sci.* **6**, 1–10 (2017)
55. N.H. Hellin, J. Rincon del Martínez, R. Miguel Domingo, F. Valles Soto, R. Sanchez Torres, A decision support system for managing irrigation in agriculture (2016)
56. A. Camargo, J.S. Smith, An image-processing based algorithm to automatically identify plant disease visual symptoms. *Biosys. Eng.* **102**, 9–21 (2009)
57. Z.J. Tejada, R. Calderon, A.J. Navas-Cortes, Early Detection and Quantification of Verticillium Wilt in Olive Using Hyperspectral and Thermal Imagery over Large Areas. *Remote Sens.* **7**, 5584 (2015)
58. K.Y. Huang, Application of artificial neural network for detecting Phalaenopsis seedling diseases using color and texture features. *Comput. Electron. Agric.* **57**, 3–11 (2007)
59. S. Sengupta, K.A. Das, Particle swarm optimization based incremental classifier design for rice disease prediction. *Comput. Electron. Agric.* **140**, 443–451 (2017)
60. L.Y. Everingham, W.C. Smyth, G.N. Inman Bamber, Ensemble data mining approaches to forecast regional crop production. *Agric. For. Meteorol.* **149**, 689–696 (2009)
61. R. Natarajan, J. Subramanian, I.E. Papageorgiou, Hybrid learning of fuzzy cognitive maps for sugarcane yield classification. *Comput. Electron. Agric.* **127**, 147–157 (2016)

Transformation of Supply Chain Provenance Using Blockchain—A Short Review



Sekhar Kumar Roy, Runa Ganguli and Saptarsi Goswami

Abstract Blockchain is a decentralized and distributed digital ledger that maintains a continuously growing list of time-sequenced records of transactions. Records are validated by the participating nodes and confirmed before it finally gets added to the immutable ledger. Information stored in blockchain is hardened against any sort of tampering or revision by any of the participating nodes. Verifying and transferring ownership by third-party intermediaries is redundant, as it is done by the decentralized protocol using complex math and cryptography. Inherent characteristics of blockchain are to enhance trust between the parties through transparency and traceability by lowering friction in transactions between the participants. Global supply chains are growing more complex and complicated. Current supply chain system operates under disparate methods between disconnected systems where there is no clarity of a single state of truth. Everyone in the ecosystem maintains their own current state in isolation, which leads to challenges like lack of transparency, traceability, accountability, and efficiency. In this paper, an attempt has been made to explore and analyze how blockchain or distributed ledger will aid efforts to improve supply chain provenance tracking.

Keywords Blockchain · Smart contracts · Distributed ledger · Ethereum · Provenance · Traceability · Supply chain provenance

S. K. Roy (✉) · R. Ganguli · S. Goswami
A K Choudhury School of IT, University of Calcutta, Kolkata, India
e-mail: sekhar.roy@gmail.com

R. Ganguli
e-mail: runa.ganguli@gmail.com

S. Goswami
e-mail: saptarsi007@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_51

1 Introduction

Supply chain is a crucial part of most businesses and is important to company success and customer satisfaction. Establishing a responsive and cost-effective supply chain is definitely a key challenge the world is facing today. One of the crucial factors is the lack of data management and integration. A supply chain is a network of retailers, distributors, storage facilities, transporters, and suppliers that participate in the production of goods, distribution, and sale of a product to the consumer. Almost all participants of that complex network stay hidden from consumers. Let us take an example. Imagine if a customer is served a burger at a restaurant, and something about it is not quite right. Is it incredibly hard to identify what went wrong? Each part of the sandwich—the meat, lettuce, tomato, bun, condiments, and cheese—came together on the plate from separate long and convoluted supply chains that go all the way back to a farm. Imagine if you can track the journey of the burger from “farm” to “fork.”

In this paper, an attempt has been made to discuss a potential solution to the problem using a disruptive technology called blockchain.

The organization of the rest of the paper is as follows: Other existing implementation and research in this domain have been discussed in Sect. 2. In Sect. 3, a brief description about blockchain and different aspects of the technology are mentioned. Supply chain ecosystem with or without blockchain has been covered in Sect. 4. An impartial evaluation of both the technologies is discussed in Sect. 5. Section 6 contains conclusion with direction for future work.

2 Related Works

In this section, an attempt has been made to briefly touch upon some of the areas/domain, which can be disrupted using the blockchain technology. Also in last year, initiatives in this domain have been conducted.

Blockchain can potentially bring value to many areas in financial service and beyond. Table 1 lists the domains where blockchain technology can be leveraged.

Various endeavors have already started in the supply chain domain. Below are the few initiatives, which have got enterprise attention.

- China has several recent food safety issues, including a government seizure in 2015 of 100,000 tons of smuggled pork, beef, and chicken. Motivation is to make Chinese pork supply chain system [1] safer in one of the meat's biggest global markets. This is an implementation where Walmart and Chinese pork industry joined hands together to develop a solution which was helped by IBM.
- Walmart struggles to identify and remove food that has been recalled. To identify the root cause of sickness, it takes days to trace the product, shipment, and vendor. This is an implementation done by IBM with a state Pilot-progress [2].

Table 1 Blockchain areas of application

Cryptocurrency exchange services	Trade finance	Supply chain provenance
<ul style="list-style-type: none"> • Exchange of cryptocurrency • Hosted wallet Solutions providing brokerage services, e.g., coinbase 	<ul style="list-style-type: none"> • Issuance and transfer of bill of lading and letter of credit • Supply chains management • Prevention of fraud during trade finance 	<ul style="list-style-type: none"> • Platform for transparently tracking materials/inventory throughout value chain
Capital markets	Trading platforms	Merchant services
<ul style="list-style-type: none"> • Securities transaction • Asset servicing • Derivative transaction • Applicability exist in pre-trade, post-trade, custody, and securities servicing 	<ul style="list-style-type: none"> • OTC trading • Platforms to trade a plethora of financial instruments, e.g., bonds, derivatives, stocks, etc. 	<ul style="list-style-type: none"> • Facilitation of cryptocurrency-based e-commerce gift cards/loyalty platforms on blockchain
Payments	Compliance	Investments and loans
<ul style="list-style-type: none"> • Cross-boarder–cross-currency payments. • Micro-transactions, e.g., tips • Both B2B and B2C solutions are potential areas • Real-time settlement of payments 	<ul style="list-style-type: none"> • Regulatory reporting • Proof of audit • AML for cryptocurrency 	<ul style="list-style-type: none"> • Escrow/custodian services • Streamline loans and settlements • Mortgage backed securities

- To aid robust proof of compliance to standards at the origin and along the chain, prevent the “**double-spend**” of certificates and explore the possibilities of using the blockchain technology to create an open system for bringing in transparency and traceability for food and other physical goods. Southeast Asia fishing industry has taken an initiative which is supported by provenance (a UK-based startup) [3].
- BHP Billiton (the mining giant) is leveraging the power of blockchain technology to track mineral analysis performed by outside vendors.
- The startup Everledger is focused on addressing the supply chain provenance for the diamond industry.

3 What Is Blockchain and Why It Matters?

A blockchain is a digital shared ledger of records that are arranged in chunks of data called blocks. Each block holds a list of transactions. These blocks then link with one another through a cryptographic validation technique known as a hashing function. Linked together, these blocks form an immutable, continuously growing unbroken chain—a blockchain which is secured from tampering and revision. The reason why

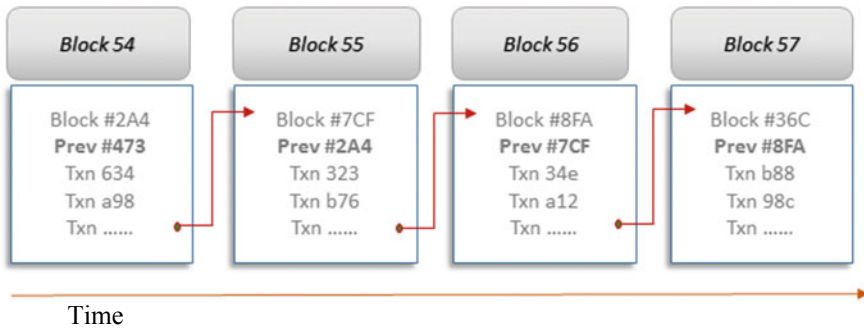


Fig. 1 Transactions, block, and chain

this type of data structure is used for things like cryptocurrencies and record-keeping is decentralization, meaning the records inside this chain are not stored in any single location, are accessible by everyone, and are immutable by any one party. A sample representation of the blockchain data structure is given in Fig. 1.

Blockchain is a type of database that takes a number of records and puts them in a block (as in Fig. 1). Each block which contains number of transactions is then “chained” to the next block using a cryptographic hash, thus allowing blockchain to be used like a ledger. This can be shared and corroborated by anyone with the appropriate permissions.

There is a common misconception that blockchain and bitcoin are synonymous; however, that is not quite true. Bitcoin is the first application of blockchain. In October 2008, Satoshi Nakamoto proposed a combined digital asset and peer-to-peer payments system in his paper [4]. The first bitcoin was minted on January 4, 2009; the first payment experiment was successfully performed on January 11. The open-source software was released on the January 15, enabling anyone with the required technical skills to get involved [5].

There are several broad categories of blockchain applications; however, five categories of uses for the blockchain can be broadly identified as—

- (a) **Currency**—Bitcoin began as a P2P electronic cash system works in the permissionless public network. Anyone can hold bitcoin and pay anyone without an intermediary. Examples: bitcoin [4], Litecoin.
- (b) **Payment Infrastructure**—You can use bitcoin to send money around the world. Merchants can accept bitcoin payments. This is slightly different from using bitcoin as a currency. Use cases include merchant processing and remittances. Examples: BitPay and Abra.
- (c) **Digital Assets**—The blockchain can be used to create digital assets such as stocks, bonds, land titles, and frequent flyer miles. These assets are created using protocols on top of the blockchain core protocol. Example protocols include colored coins and counterparty. Companies using this technology are Chain, NASDAQ, and Openchain.

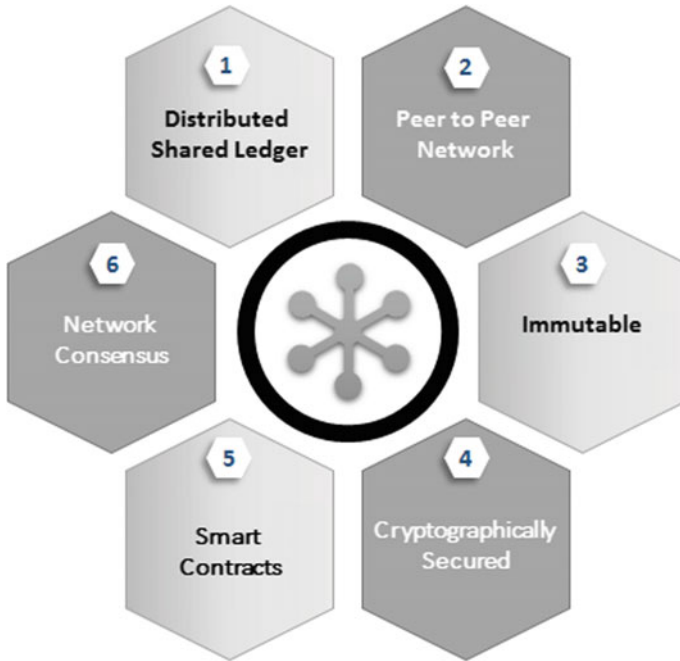


Fig. 2 Key constructs of blockchain

- (d) **Verifiable Data**—Create a verifiable record of any data, file, or business process on the blockchain. Examples: Proof of Existence, Factom, and BigchainDB.
- (e) **Smart Contracts**—Software programs that live on the blockchain and execute without the possibility of third-party interference. Examples: Ethereum [6], RootStock, Hyperledger, and Quorum [7].

In this paper, an attempt has been made to do a research of the verifiable data and smart contract aspect of blockchain and how that can be leveraged to reshape the loosely coupled supply chain ecosystem. Let us discuss blockchain bit more details (Fig. 2).

1. A time-sequenced store of transaction records in a distributed ledger shared across a business network.
2. Connected through P2P using each other.
3. Single source of truth across all/relevant participating entities kept inside an append-only ledger. Once written, no one can change it.
4. Transactions are secured, authenticated using public key infrastructure (PKI) and verified through cryptography provided by the framework.
5. Smart contract is contractual logic that is “Blockchain housed” and “executed” inside the decentralized application (DAPP).
6. This is a mechanism through which all the participants agree on the state of the ledger.

Table 2 Comparison between popular consensus

Proof of work (POW)	Proof of stake (POS)	Proof of burn (POB)
<p>The algorithm has to be solved by each node before a block can be published successfully. Whoever solves the algorithm first, becomes the owner of the block and gets the mining fee, if any</p> <ul style="list-style-type: none"> ✓ The difficulty of generating a block is adjusted to limit the rate at which new blocks can be created by the network—1 every 10 min ✓ For a block to be valid, it must hash to a value less than the current target; this means that each block indicates that work has been done generating it ✓ Each block contains the hash of the preceding block <p>Example “Bitcoin” [4]</p>	<p>Each node has to have coins to participate in the consensus process Who gets to mine the next block depends on weightage of the coins the node has, gets to create the block and associated fee if any</p> <ul style="list-style-type: none"> ✓ The node with highest weightage of coins has rights to create the next block ✓ Weightage depends upon the age of the coins. The network determines the minimum age of coins that can be put up for stake ✓ Coins that are put up for stake, come back to the wallet; however, as the age of these coins reset to 0, it cannot participate in the consensus process <p>Example “NXT” [11]</p>	<p>Method for distributed consensus across chains requires nodes to burn some coins in one chain, to get credits for some action in another</p> <ul style="list-style-type: none"> ✓ The user needs to burn (send cryptocurrency to an unspendable address) and prove that they have burnt the required amount before they can participate in a consensus in another chain ✓ This is expensive, just like proof of work, but more energy efficient way. It burns underlying asset <p>Example “Peercoin” [12]</p>

Every single blockchain constructs are important. However, to stay relevant with the theme of the paper, blockchain, consensus, and smart contract have discussed in a bit more details.

Blockchain Consensus?

Consensus algorithms exist to ensure deterrent for malicious nodes. Each node validates each transaction based on the copy of the ledger that they have locally. Successful validation of the transaction allows it to be eligible of addition to a block. The published blocks are also validated and confirmed by individual nodes. As long as the majority of nodes are true, valid transactions would go through and invalid transactions would be rejected. Table 2 is the comparison of most commonly used consensus protocols.

There are other protocols like proof of activity, proof of elapsed time, and proof of capacity, etc.

Blockchain and Smart Contract?

Smart contracts are computer protocols intended to facilitate, verify, or enforce the negotiation or performance of a contract [8]. Smart contract can also be considered as “Proof of Agreement.” Smart contracts have a wide range of possible use cases—for businesses, individuals, and for IoT devices themselves. A relevant example could be track supplier performance through smart contracts that can govern the compliance of supplier standards, metrics, and contractual obligations and ensure standardization.

Smart contracts are useful tool for reducing operational risk and can be thought of as an automated trustworthy workflow between participants without an intermediary.

4 Supply Chain Provenance Using Blockchain

Blockchain technology will have a profound impact on the way the supply chain works. However, to understand it better, let us discuss the traditional ecosystem challenges and how the proposed solution brings transformative impact.

(a) *Problems with Traditional Supply Chain ecosystem*

As described in Fig. 3, current supply chain system is done by disparate methods between disconnected enterprise resource planning (ERP) with problems of traceability and accountability between individuals and enterprises. There is no clarity of a single state of truth; everyone is maintaining their own current state of things, which leads to a number of challenges.

- Absence of trust and consensus between suppliers, manufacturers, distributors, and retailers, lack of collaboration among the different entities including raw material suppliers, production sites, warehouses, and retailers are the related problems. The problem is that the different decision makers do not have access to the information regarding the state of the entire SC network, and in addition, they usually operate under different objective functions.
- Lack of trusted information is available to the consumer. Counterfeit food and pharmaceutical products is a growing problem across the world; the contaminated study puts the value of imported fake goods worldwide at \$461 billion in 2013, compared with total imports in world trade of \$17.9 trillion [9]. Unsustainable and

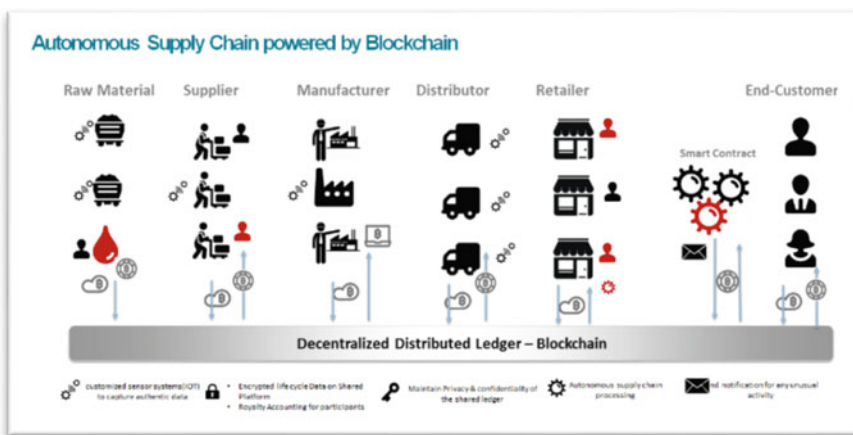


Fig. 3 Proposed supply chain ecosystem view

inefficient use of resources in the distribution of goods, higher operational costs, and lack of automation leads to inefficiency and delay in processing.

(b) ***Technology Advantage of Blockchain for Reshaping Supply Chain***

It is appealing when a supply chain needs three things: data to be tamperproof, data written by mutual consent, and absolute surety of who committed data. Blockchain's decentralized and secure environment for joint record-keeping provides the ideal foundation for supply chain making it democratize.

Interoperable—A modular, interoperable platform that eliminates the possibility of double spending.

Enhanced Transparency—Using customized sensor systems (IOT) with integrated blockchain technology -IoT integration means that non-invasive, non-destructive and immediate measurements may be made on very small devices—an example would be allowing embedded biosensors within food packaging material to track and trace its quality. As blockchain storage is immutable which creates a permanent, shareable, and actionable record of every moment in a product's trip through its supply chain, creating efficiencies throughout the global economy.

Autonomous—Smart contracts that reduce the need for trust parties by signing a digital agreement. To purchase a product, one of the parties may deposit a certain amount of funds in a smart contract backed escrow account. The funds will only be delivered to the provider if certain checkpoints are met (e.g., the supplier delivers by a certain timestamp, there is verifiable quality compliance and the supplying company is not violating environmental regulations).

Auditable—An auditable record that can be inspected and used by companies, standard organizations, regulators, and customers alike.

Cost-Effective—Creation of peer-to-peer marketplaces which decrease costs of marketing and distribution. A solution to drastically reduce costs by eliminating the friction between the parties.

Agile, Real time, Efficient—A fast and highly accessible supply chain infrastructure which leads to efficient and immediate processing of the transactions.

An attempt to show how the supply chain ecosystem looks which is powered by decentralized ledger/blockchain.

Figure 4 is proposed to be solution context using which most of the problems can be taken care of.

5 Impartial Evaluation: Traditional Versus Proposed Solution

In this section, we present an evaluation matrix (Table 3) by taking account of comparability with existing systems and comprehensiveness, and by examining the tradeoff between evaluation items [10].

Table 3 Evaluation matrix

#	Architectural aspects	Traditional/powered by centralized systems	Proposed/powered by blockchain/fully decentralized
1	Topology		
2	Transparency	Lack of open and trustworthy information across the supply chain	An ecosystem where information flows openly and securely between permissioned participants.
3	Interoperability	Lack of interaction between disconnected systems leads to inefficiency and obscureness. Internal and reconciliation required	Interoperability allowing a disconnected system to interact with each other lead to agile and real-time interaction making it efficient and cost-effective
4	Auditability and traceability	Difficult to audit and trace	Perfectly auditable and end-to-end traceable
5	Performance	Error-prone and loosely coupled/disconnected system lead to a lousy reconciliation process between the parties, which is slower and error-prone	Transaction processing in blockchain will always be slower than centralized databases, due to the below burdens: (a) signature verification, (b) consensus mechanisms, and (c) redundancy. However, since the systems are hyper-connected, there is no additional effort required for reconciliation
6	Security	Centralized disconnected ecosystem provides lesser security on the data	The combination of sequential hashing and cryptography along with its decentralized structure makes it very challenging for any party to tamper with it in contrast to a standard database
7	Scalability	Current supply chain process ecosystem maintaining their own current state of things in a silo which makes it difficult to scale	Blockchain promises a scalable system of supply chain visibility using interoperability mechanisms.
8	Integrity	Increasingly difficult to ensure the quality and integrity of raw materials and finished products along the chain, fight contamination and counterfeiting	Integrity achieved using the special crypto algorithm, storing of copies of the database on various nodes of the network, and absolute data synchronization on all nodes

(continued)

Table 3 (continued)

#	Architectural aspects	Traditional/powered by centralized systems	Proposed/powered by blockchain/fully decentralized
9	Immutability	No default support for immutability. There is always a distinct risk of loss due to potential human mishaps (intentional or not)	Once a transaction is verified, validated, and recorded in a block of data on the distributed ledger, it is irreversible and cannot be hidden or altered
10	Automation	Due to disconnected systems it is challenging to automate the ecosystem	Enforcing the agreements running and executing in a decentralized node using smart contracts, a system can make sure different suppliers are meeting their obligations. One can see deliveries at multiple locations, and track shipments based on the fulfillment of smart contract terms

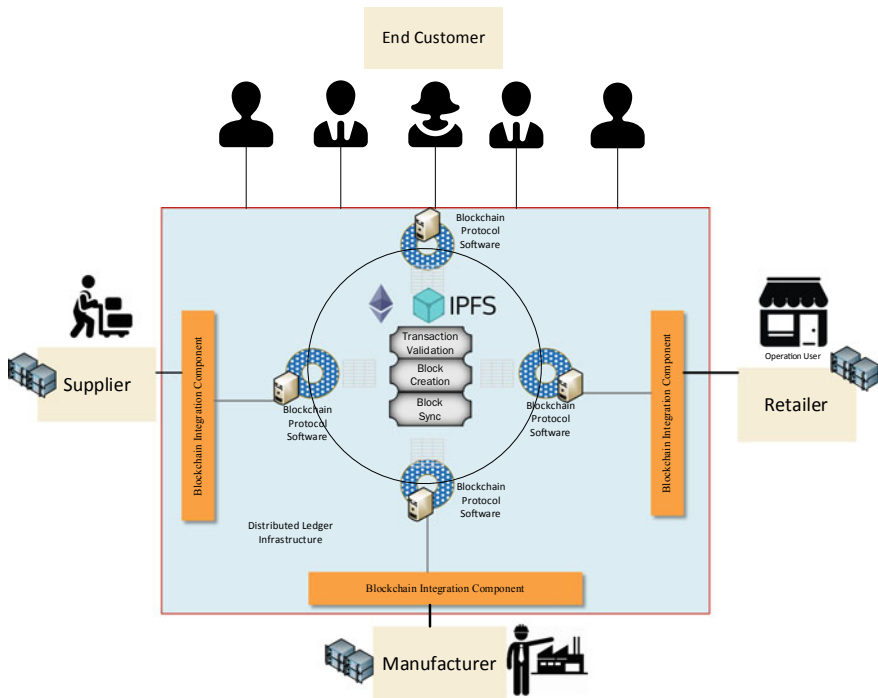


Fig. 4 Proposed blockchain-enabled solution context

6 Discussion and Conclusion

In the present study, an analysis has been performed highlighting various aspects of blockchain and supply chain provenance. It aims to provide a basis for discussion on the relevant uses of the transformative technology that could open up opportunities in the supply chain provenance space. Blockchain has been perceived mostly as digital currency; however, there is a great potential to use the power of the technology as smart contract and proof of audit. Once the life cycle data are persisted it can be considered “as a single source of trust” which can never be altered without the consensus of the majority of the decentralized actors. The author is fully aware of the fact that it is easier said than done as creating a consortium to bring all the parties under the same umbrella is a definite business challenge. There are quite a number of technology challenges also like the current state of the frameworks are not capable of handling the scale it’s required to operate in this space.

References

1. Chinese Pork on the Blockchain. [Online]. Available: <http://fortune.com/2016/10/19/walmart-ibm-blockchain-china-pork/>
2. Track and trace produce in U.S. stores [Online]. Available: <http://www.pymnts.com/blockchain/2016/walmart-launches-blockchain-technology-for-pork-and-produce/>
3. Track and trace yellowfin and skipjack tuna fish [Online]. Available: https://www.provenance.org/tracking_tuna_on_the_blockchain
4. S Nakamoto, *Bitcoin: A Peer-to-Peer Electronic Cash* (2008), [Online]. Available <https://bitcoin.org/bitcoin.pdf>
5. Bitcoin Wikipedia in [Online]. Available: <https://en.wikipedia.org/wiki/Bitcoin>
6. “Ethereum Whitepaper” [Online]. Available: <https://github.com/ethereum/wiki/wiki/White-Paper>
7. “JP Morgan Quorum” Whitepaper [Online]. Available: <https://github.com/jpmorganchase/quorum-docs/blob/master/Quorum%20Whitepaper%20v0.1.pdf>
8. “Smart Contract” Wikipedia [Online]. Available: https://en.wikipedia.org/wiki/Smart_contract
9. Economic Co-operation and development (OECD) report available in [Online]. Available: <http://www.oecd.org/industry/global-trade-in-fake-goods-worth-nearly-half-a-trillion-dollars-a-year.htm>
10. “Evaluation Forms for Blockchain-Based System ver. 1.0” by Information Economy Division [Online]. Available: http://www.meti.go.jp/english/press/2017/pdf/0329_004a.pdf
11. “NXT” Whitepaper [Online]. Available: <https://nxtplatform.org/get-started/developers/>
12. PeerCoin Whitepaper [Online]. Available: <https://peercoin.net/whitepaper>
13. “Provenancel Blockchain: the solution for supply chain transparency” [Online]. Available: <https://www.provenance.org/whitepaper>
14. Towards an Ontology-Driven Blockchain Design for Supply Chain Provenance Aug 28, 2016 Henry M. Kim, Marek Laskowski [Online]. Available: <https://arxiv.org/abs/1610.02922>
15. Ethereum, “Solidity documentation,” 2016. [Online]. Available: <https://ethereum.github.io/solidity/docs/home/>
16. A Supply Chain Traceability System for Food Safety Based on HACCP, Blockchain and Internet of Things 2017 [Online]. Available: <https://ieeexplore.ieee.org/document/7996119/>
17. IPFS—Content Addressed, Versioned, P2P File System by Juan Benet [Online]. Available: <https://ipfs.io/ipfs/QmR7GSQM93Cx5eAg6a6yRzNde1FQv7uL6X1o4k7zrJa3LX/ipfs.draft3.pdf>

A Framework for Predicting and Identifying Radicalization and Civil Unrest Oriented Threats from WhatsApp Group



Koushik Deb, Souptik Paul and Kaustav Das

Abstract Social media is an emerging area of research for data scientist. Lots of test dataset are available for Facebook, Twitter, and Instagram. Very few research has been done on WhatsApp group chat log due to scarcity of data openly available. WhatsApp is secured by encryption so WhatsApp is heavily used for terrorist activities, spreading riot, creating civil unrest, and disseminating radicalization. Data to conduct research on these issues are hardly available as these are very sensitive information. We are proposing here a framework to address this type of social issue. By NLP and semi-supervised machine learning, we can design a system which will predict the probability of WhatsApp group to be a threat for civilization. Success of this framework depends on how much actual data we can feed to system. Depending on training data, it will predict with more accuracy. This paper will show how this system can work properly in actual field.

Keywords Machine learning · Pattern mining · Natural language processing · Topic detection · Prediction system

1 Introduction

In today's world, social media plays a very important role in interpersonal communication. It has become a platform for users to share their opinions on various incidences and events around the world. Social media nowadays has a wide range of platforms for people share their views and ideas such as—social networks (Facebook, Twitter), image sharing (Instagram, Tumblr) and video sharing (YouTube, Dailymotion,

K. Deb · S. Paul (✉) · K. Das
IEM, Kolkata, India
e-mail: paul.souptik@yahoo.com

K. Deb
e-mail: koushikdeb2009@gmail.com

K. Das
e-mail: ddaskaustav66@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_52

Twitch), online messaging (WhatsApp, Viber, Telegram). Due to the wide-reach and availability of these media, many organizations and individuals use them for planning situations of civil unrest such as riots and rebellions. They share sensitive media on these platforms, which wound the sentiments of various communities, thereby inciting communal disharmony. Social media platforms are also used by many extremist and hate groups to promote online radicalization. Research has shown that extremist groups share their ideas, repulsive speeches, and violent images and videos to pollute the minds of people. The main objective of such organizations is to target the minds of the youths of today, who are the primary users of social media. This is one of the major issues faced by many countries nowadays and most countries are trying to find ways to counter this issue. Due to the high level of anonymity, many extremist and terrorist organizations also use online messaging applications such as WhatsApp for communicating among themselves. They use the various communication features (Video Calling, Messaging, and Voice Calling) of such online applications for the planning and implementation of their activities. The extended usage of social media for such activities provides a wide window for forecasting such activities in advance.

Online radicalization is a widespread problem faced by many countries around the globe. The availability of radical content and drafting of civil unrest situations are huge issues faced by the governments around the world. Law enforcement agencies are constantly trying to prevent the dispersion of such content and to apprehend the people involved in the dispersion of such content. In such a situation, automated analyzation and detection of extremist content can be very useful. As law enforcement agencies can be made aware of radicalization attempts instantly and they can perform the necessary actions. One of the major online messaging applications is WhatsApp. Our goal is to develop an analyzation framework for predicting radicalization attempts through WhatsApp messages. The framework will successfully predict whether radicalization is occurring through WhatsApp messages or not. This data can assist different law agencies to prevent online radicalization and to prepare for possible radical outbursts.

2 Related Work

The ongoing topic of predicting real-world events has received considerable interest from researchers around the globe. Researchers have concentrated their efforts on predicting real-world events by using a variety of prediction algorithms and tracking methods. This study is related to two different areas:- online radicalization, and predicting events and behavior using data from social media.

There is a profuse amount of research in the field of online radicalization, on its increasing rate and its effect in increasing civil unrest situations around the world. Studies performed on a micro-level by Mohammad Istiaque Rashid show the various methods and procedures used by extremist organizations to promote online radicalization. It has a step by step overview of the entire scheme of online radicalization. The study also highlights the impact, such online radicalization has on promoting

disruptive events [1]. A research performed by Dr Alex P. Schmid shows how radicalization can be assessed in accordance with mainstream political thought over a given period and also explores radicalization on a macro-, micro-, and meso-level. The research also examines various counter-radicalization and de-radicalization programs [2].

Prediction of events from data obtained from social being an intriguing topic has received the attention of many scholars from around the world. Rostyslav Korolov developed a method to predict protests by identifying mobilization in social media communication. The study used Twitter data obtain during the 2015 Baltimore events to predict future protests [3]. They correlated their predictions with the amount of Twitter tweets during that duration to obtain their conclusions. Nasser Alsaedi proposed a disruptive event detection framework which uses classification and clustering techniques on Twitter tweets [4]. They used ground-truth data based on the intelligence gather by the London Metropolitan police to test their framework. The study used MongoDB to store the data and stopword elimination and stemming to process the data. The data were classified for use by using a Naïve Bayes classifier. Pragan Debnath developed a framework, which would assess the need of different post-disaster assistance organizations using a string-wise keyword matching algorithm [5]. They obtained their data from the WhatsApp chat log of an NGO called Doctors for You. They used Python's built in natural language processing toolkit called WordNet and TextBlob's sentiment analysis tool for analyzing their data.

3 Collection of Data and Initial Analysis

WhatsApp group message of those people who are provoking riot is a tough job to collect. We have studied the report of north-east ethnic violence of 2012. We gathered data related to the incidents [6]. We have gathered and analyzed data of Assam Riot in 2012 [7]. We have taken help from the research "Blackberry riot" that happened in London, 2011. We have studied several sources like blog and newspaper report. Gathering WhatsApp chat data from civil unrest spreading group is not feasible for us due to security issue. We have made synthetic data based on dark web dataset. Though those dark web data are not WhatsApp chat but topic and discussion are same as we required in chat data [8].

We have also taken help Riot Related Journal (published by several journalists). These journals are cited in paper. We have studied several chat group log of common topic. With the help of these evidences, we have built our dataset otherwise research could not be done. That is why our focus in on framework not efficiency of methodology. We have come to know about some integral factors of riot conversation: People are frustrated with certain issue, some common type of enemy has been selected, lots of negative emotions have been surged up, promoting those negative emotion

25/08/2017, 17:48 - Rehman: is this news true?
 25/08/2017, 17:49 -Rijwan: yes brother i have also seen this news a few minutes before. Those jawans shoot an innocent suspecting terrorist.
 25/08/2017, 17:52 -Abdul: these can't continue..we have to reply them back. We are not cowards so we can't seat at home looking them killing us.

Fig. 1 Sample chat on WhatsApp

by others. Depending on these factors, we have created some synthetic data for riot conversation (Fig. 1).

Following are observation from dataset and available conversation:

As it is text data, it is highly unstructured.

There are multiple topics discussed but always shifted toward hatred.

Lots of regional terms have been used.

There is a mix of local languages and English.

Through the analysis of chat log, we have found there are some phrases and some sequential and non-sequential pattern of phrase which have been used multiple times. These phrases are common for most of the radicalization conversation. We can identify some frequently used phrase depending on the residential region of group. Not only phrases but also multimedia data play a good role in radicalization message detection. Video showing violent scene, audio containing hate speech, image provoking civil unrest are some integral part of these chat logs.

4 Design Approach for Threat Detection in WhatsApp Chat

4.1 Tool Used

To search pattern, we have mainly used natural language processing techniques. Data have been taken as input in text format. Nltk toolkit has been used to clean and preprocess data. Porter stemmer has been used to stem words into lexemes. We have extensively used FP-growth algorithm to mine pattern [9]. We have also used SPMining [10] and NSPMining [10] method to search sequential pattern and non-sequential pattern exhaustively. We have used SQL database to store all patterns.

4.2 Framework

This framework follows semi-supervised learning algorithm. More we will train this system, more efficient it becomes.

First, we will take input of chat log and preprocess data. We will only take nouns, nominal adjective, noun phrases, etc. Other parts of speech will be discarded. As other parts of speech change their meaning based on nouns. Words must be reduced into lexemes so that we can work with root of words. After this step, each sentence will remain as a collection of root words.

Next, we will execute the pattern mining algorithm to build sequential and non-sequential patterns. This pattern will be stored in database as variable length word collection. There must exist a threshold over frequency of pattern. If some pattern crosses that frequency level with respect to its occurrence in text, it will be stored in database else it will be ignored as trivial pattern.

While a patterns will be stored, its length L (number of element) and frequency F (number of occurrence in chat log) will be stored along with it. There will be one field for human intervention to assign T (0 or 1). Zero means that pattern is not concerned pattern to our required topic. One means this pattern is one of the required patterns.

Frequency Weight (FW) will be calculated for all patterns.

$$FW_p = \frac{F_p}{\sum_1^n F_p} * T$$

P is any pattern from total (n) pattern.

This frequency will be multiplied by length of pattern and pattern weight (PW) will be calculated.

$$PW_p = FW_p * L$$

There will be a single repository after collecting patterns and pattern weight (PW) from all chat log.

Whenever a test data will be feed, it will search for sequential and non-sequential pattern and store the one which passed the threshold frequency. Each pattern will be compared to the stored pattern.

Suppose for PAT p , r number of elements out of S_p elements are matched with r number elements out of L_q elements of PAT q in pattern database.

Then matching score will be

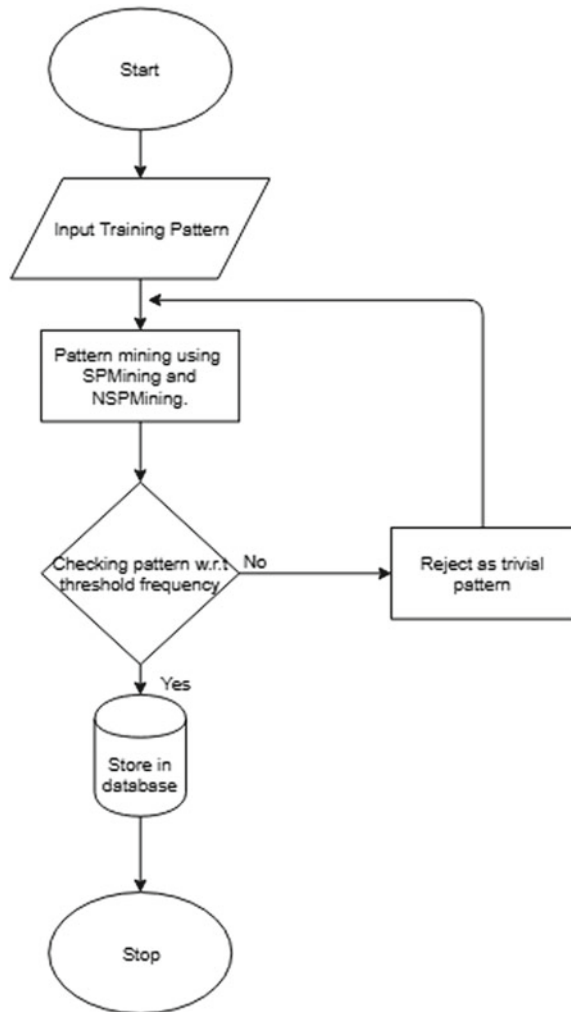
$$MScore = \frac{r}{S_p * S_q} * PW_p$$

If there is k number of matches for the same pattern, then score will be

$$MScore = \frac{r}{S_p * (S_q + S_r \dots kth \text{ Term})} * (PW_q + PW_r + \dots kth \text{ Term})$$

Every pattern will have its matching score. Total matching score will be the confidence of chat log.

Fig. 2 Flowchart for training data



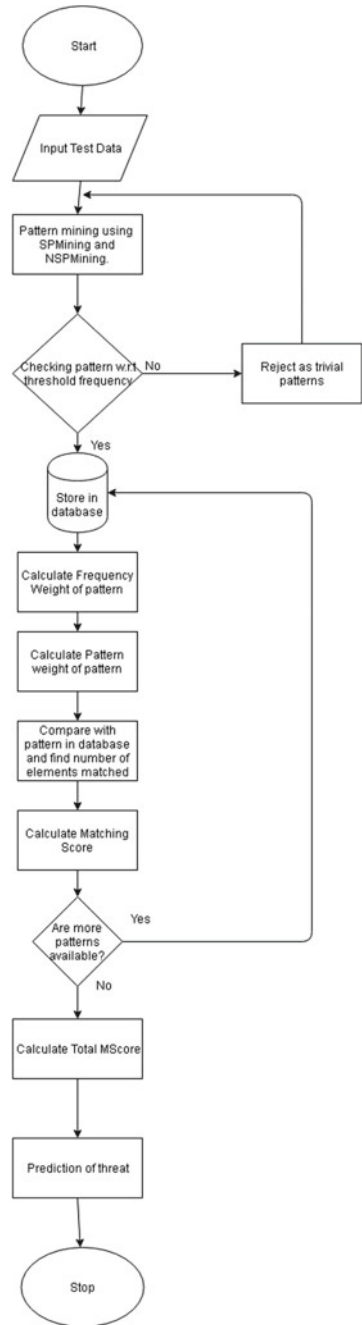
$$\text{TotMScore} = \frac{\text{MScore}_1 + \text{MScore}_1 + \dots + \text{MScore}_p}{p}$$

TotMScore will be a fraction that denotes probability of chat log being a threat spreading group.

Here, we have represented how training data work on this model (Fig. 2).

Here, we have represented how model will work on input data (Fig. 3).

Fig. 3 Flowchart for test data



5 Experimental Result

In the chat log, we can find fields such as name, phone number, and time. We have to remove these fields to get effective pattern.

Here, we are showing a case study on our synthetic data.

Sequential patterns we have received after processing a chat log containing riot-spreading conversation (Table 1).

Non-sequential patterns we have found much more than sequential. Here are some entries of NSP (Table 2).

After having this pattern database, we have to tag those malicious patterns 1 and remaining 0 (Tables 3 and 4).

Table 1 Sequential patterns from training data

Patterns	Frequency	Length
Good,Morning	32	2
Allah,Huakbar	26	2
indian,army	23	2
freedom,india	21	2
Azaad,kashmir,india	18	3
stone,pelting,bsf	15	3
hope,best	13	3
fund,villagers,welfare	9	3
money,destruction,india	5	3

Table 2 Non-sequential patterns from training data

Patterns	Frequency	Length
Good,Morning,guys	30	4
Allah,Huakbar,Allah,Hafiz	25	4
indian,army,notorious	21	3
bloody,freedom,india,torture	20	4
Road,track	19	2
Plan,attack	18	2
Azaad,kashmir,india,free	15	4
stone,pelting,bsf,jawans	13	4
hope,best,result	13	3
outrage,death,penalty	9	3
ultimatum,atrocities,attack	8	3
good,life,god	7	3
fund,villagers,welfare,goods	7	4
money,destruction,india,level	4	4

Table 3 Tagged sequential patterns

Patterns	Frequency	Length	T
Good,Morning	32	2	0
Allah,Huakbar	26	2	0
indian,army	23	2	0
freedom,india	21	2	1
Azaad,kashmir,india	18	3	1
stone,pelting,bsf	15	3	1
hope,best	13	3	0
fund,villagers,welfare	9	3	0
money,destruction,india	5	3	1

Table 4 Tagged non-sequential patterns

Patterns	Frequency	Length	Tag
Good,Morning,guys	30	4	0
Allah,Huakbar,Allah,Hafiz	25	4	0
indian,army,notorious	21	3	1
bloody,freedom,india,torture	20	4	1
Road,track	19	2	0
Plan,attack	18	2	1
Azaad,kashmir,india,free	15	4	1
stone,pelting,bsf,jawans	13	4	1
hope,best,result	13	3	0
outrage,death,penalty	9	3	1
ultimatum,atrocities,attack	8	3	1
good,life,god	7	3	0
fund,villagers,welfare,goods	7	4	0
money,destruction,india,level	4	4	1

This procedure will be continued for multiple chat logs which are already identified for spreading threats. So, our pattern database will grow and will contain various radicalization related patterns.

Next, we have to work with a test chat log. After processing test chat log, we will have new patterns from that.

Here is example of some patterns (sequential and non-sequential) (Table 5).

We can build our pattern database depending on collected data (Table 6).

In pattern hey, WhatsApp, the length is 2 and there is no match with stored pattern in the database. So MScore is 0.

For pattern hello, good, morning, we have a match in pattern database but tag is 0 so MScore will be 0.

Table 5 Sequential and non-sequential patterns from training data

S. No.	Patterns	Frequency	Length
1	Hey,whatsup	32	2
2	Hello,good,morning	29	3
3	Ideal,time	27	2
4	Plan,stone,attack	24	3
5	Retreat,gun	21	2
6	Bullet,life	18	2
7	Destruction	15	1
8	Hindu	13	1
9	Freedom,india	11	2

Table 6 Pattern database

S No.	Patterns	Frequency	Length	T	FW	PW
1	Good,Morning	32	2	0	0	0
2	Allah,Huakbar	26	2	0	0	0
3	indian,army	23	2	0	0	0
4	freedom,india	21	2	1	0.056603774	0.113208
5	Azaad,kashmir,india	18	3	1	0.04851752	0.145553
6	stone,pelting,bsf	15	3	1	0.040431267	0.121294
7	hope,best	13	3	0	0	0
8	fund,villagers,welfare	9	3	0	0	0
9	money,destruction,india	5	3	1	0.013477089	0.040431
10	Good,Morning,guys	30	4	0	0	0
11	Allah,Huakbar,Allah,Hafiz	25	4	0	0	0
12	indian,army,notorious	21	3	1	0.056603774	0.169811
13	bloody,freedom,india,torture	20	4	1	0.053908356	0.215633
14	Road,track	19	2	0	0	0
15	Plan,attack	18	2	1	0.04851752	0.097035
16	Azaad,kashmir,india,free	15	4	1	0.040431267	0.161725
17	stone,pelting,bsf,jawans	13	4	1	0.035040431	0.140162
18	hope,best,result	13	3	0	0	0
19	outrage,death,penalty	9	3	1	0.02425876	0.072776
20	ultimatum,atrocities,attack	8	3	1	0.021563342	0.06469
21	good,life,god	7	3	0	0	0
22	fund,villagers,welfare,goods	7	4	0	0	0
23	money,destruction,india,level	4	4	1	0.010781671	0.043127

For pattern 4, we have multiple matches. For 'plan', 'attack' we have a match in the 16th entry and for stone we have a match in the 6th entry.

For first part

$$\text{MScore} = (2/(3 * 2)) * 0.097 = 0.023$$

For second part

$$\text{MScore} = (1/(3 * 3)) * 0.121 = 0.0134$$

Total score for this particular pattern is 0.036.

In this way, we have to check for each pattern the score. The average score will be confidence of system for that chat log to thread-spreading group.

6 Conclusion and Future Scope

This is a proposed framework to handle growing radicalization issue. This can be used in multiple areas to detect mass sentiment. Most cumbersome job is tagging part which needs human intervention. We can build a topic-related corpus so that this tagging can be automated. Depending on this tagging, the topic will be detected. We can also find the member from that group who is the most verbal about that topic most frequently so the actual leader can be detected.

Here, computer vision can be used to understand meaning of the image. So we can predict depending on image data also. Mostly, images are used to provoke mass enragement. We can also use recorded voice file of WhatsApp group. Those recording can be converted to text and same procedure could be applied to understand intention of those recording. We have the temporal information with data. So, we can measure the frequency of speech by time. In the actual field, we can relate that frequency with incident timing and produce more effective patterns from their speech. This can be highly helpful to any government and security organization.

References

1. M.I. Rashid, *Online Radicalization: Bangladesh Perspective*, in Report of U.S. Army Command and General Staff College
2. A.P. Schmid, Radicalisation, de-radicalisation, counter-radicalisation: a conceptual discussion and literature review. *Int. Cent. Count.-Terror.- Hague Res. Pap.* (2013)
3. R. Korolov, D. Lu, J. Wang, G. Zhou, C. Bonial, C. Voss, L. Kaplan, W. Wallace, J. Han, H. Ji, On predicting social unrest using social media, in *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*

4. N. Alsaedi, P. Burnap, O. Rana, Can We Predict a Riot? Disruptive Event Detection Using Twitter. *ACM Trans. Internet Technol.*, **17**(2), Article 18, Publication date: March 2017
5. P. Debnath, S. Haque, S. Bandyopadhyay, S. Roy, Post-disaster situational analysis from whatsapp group chats of emergency response providers, in *Proceedings of the ISCRAM 2016 Conference—Rio de Janeiro, Brazil, May 2016* Tapia, Antunes, ed. by Bañuls, Moore and Porto
6. J.P. Pathak, Role of social media in reference to North-East ethnic violence (2012). *IOSR J. Hum. Ities Soc. Sci. (IOSR-JHSS)* **19**(4), Ver. V, pp. 59–66 (2014)
7. S. Roy, A.K. Shukla, Social media in the domain of communal violence: a study of assam riot 2012. *Int. J. Adv. Res. Innov. Ideas Educ.* (2016)
8. AZSecure-data.org Intelligence and Security Informatics Data Sets <https://www.azsecure-data.org/dark-web-forums.html>
9. A. Kaur, G. Jagdev, Analyzing working of FP-growth algorithm for frequent pattern mining. *Int. J. Res. Stud. Comput. Sci. Eng. (IJRSCSE)* **4**(4), 22–30 (2017)
10. S.-T. Wu, Y. Li, Pattern-based web mining using data mining techniques. *Int. J. E-Educ., E-Bus., E-Manag. E-Learn.* **3**(2) (2013)
11. G. Morrell, S. Scott, D. McNeish, S. Webster, The August riots in England understanding the involvement of young people, in *October 2011 Prepared for the Cabinet Office*

Flower Pollination Algorithm-Based FIR Filter Design for Image Denoising



Supriya Dhabal, Srija Chakraborty and Prosenjit Sikdar

Abstract This paper proposes flower pollination algorithm (FPA)-based two-dimensional FIR filter for the denoising of images. The efficacy of proposed FIR filter is tested on few standard images like Lena, Einstein, Cameraman, and Barbara and the noises include Gaussian, salt and pepper, and speckle. The results achieved by proposed filtering technique demonstrate a clear explanation of the performance improvement. Comparison with other available results in the literature indicates that the proposed method exhibits an average of 13% improvement in PSNR, 17% improvement in IQI, and 23% improvement in SSIM for Barbara image corrupted with Gaussian noise.

Keywords Denoising · FPA · PSNR · IQI · SSIM · Levy flight

1 Introduction

Noise is usually defined as an unwanted signal in image processing which corrupts the original information and reduces the overall quality of digital images [1–3]. It is a hindrance to smooth and easy image analysis or transmission through media. Thus, it must be removed from the image as much as possible to obtain nearly accurate information. Moreover, nowadays images are being transmitted over wireless systems extensively, which leads to deterioration in image quality at the receiver end. Different

S. Dhabal (✉) · P. Sikdar

Department of Electronics and Communication Engineering, Netaji Subhash Engineering College, Kolkata 700152, West Bengal, India
e-mail: supriya_dhabal@yahoo.co.in

P. Sikdar

e-mail: prosenjit.sikdar@yahoo.com

S. Chakraborty

Department of Biomedical Engineering, Netaji Subhash Engineering College, Kolkata 700152, West Bengal, India
e-mail: frauleinmouli@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_53

kind of noises can cause different form of degradation in image quality. Therefore, choice of the suitable filter for a certain image mostly depends on the kind of noise present in it. Hence, a number of approaches are introduced to suppress noise and consequently boost image quality over the past few decades [1–5].

In image denoising, filtering can be performed in both spatial and frequency domain. In the spatial domain denoising, arithmetic operations are used on neighborhood pixels to replace original pixel intensity with the new one. In contrast, the frequency domain filtering is performed by taking Fourier transform of the noisy image and it is faster than the spatial domain. In wavelet shrinkage method, noise is suppressed in low-frequency sub-bands and subsequently enhancement in pixel values takes place in high frequency band. Wavelets shrinkage methods are effective to reduce Gaussian noise; one such wavelet shrinkage method is known as SURE shrink [1, 2]. Similarly, matched bi-orthogonal wavelets are used to denoise medical and satellite images which are corrupted by Gaussian noise [3]. Due to the advantages of sparsity and multi-resolution property of wavelets, the authors in [4, 5] proposed dual-tree complex wavelet transform, and quincunx diamond filter bank to denoise images of medical importance. In the reduction of speckle noise, Lee and Kuan filters are the pioneers [6, 7]. Lee proposed a non-recursive algorithm where each pixel of the image is processed independently. This method is advantageous when it is used in real-time digital image processing applications. The filter proposed by Kuan is self-adaptive to the non-stationary local image statistics when it faces various kinds of noise which are signal-dependent. For multiplicative noise, this filter works, based on a systematic derivation of Lee's algorithm having few modifications which provide disparate estimators for the local image variance. Beside, the local statistics method, proposed by Frost, is applied to remove the noise from radar images [8]. The computation of local statistics for image denoising was also used by Yang et al. [9]. Chitroub [10] applied principal component analysis to denoise remote sensing images which are required in the compression and enhancement of images. To denoise images, Mohamadi et al. [11] used a combination of stochastic optimization algorithm and partial differential equation which exhibits more denoising ability considering the PSNR and visual quality. Bharathi and Govindan proposed a hybrid filtering technique for the reduction of impulse noise in medical images for better image quality [12]. In [13, 14], the authors applied bilateral and weighted bilateral filtering techniques for the denoising of both gray and color images. The bilateral filter is a type of nonlinear filters that implement spatial averaging and also preserve image edges.

With the advancement of machine learning paradigm, the researchers are using different metaheuristic optimization algorithms for the removal of noises from digital images. In most of the cases, two-dimensional (2D) FIR filters are designed using optimization algorithm and the filtering is performed in spatial domain or in frequency domain. Kochanat et al. [15] proposed artificial bee colony algorithm (ABC)-based 2D FIR filter for the removal of Gaussian noise. Similarly, the ABC algorithm is also used for the optimization of filter coefficients in despeckling of medical ultrasound images [16]. Hua et al. [17] utilized differential evolution-based particle swarm optimization for the denoising of images. Paiva et al. [18] worked with denoising

methods based on hybrid genetic algorithm where local search and mutation operators are used in conjunction with standard filtering. Huang presented a comparative study of anisotropic diffusion filter, Wiener filter, total variation filter; nonlocal means filter, and bilateral filter on the denoising of Chinese calligraphy images [19]. Kumar and Nagaraju used cluster-based Fuzzy Firefly Bayes filter to remove various kinds of noises [20]. For the further enhancement in image quality here an improved flower pollination algorithm is proposed for the design of 2D FIR filter which performs filtering in the spatial domain.

The aim of this research is to minimize the noise from corrupted images. The different type of noises like Gaussian, speckle, and salt and pepper are added in few standard images and the denoising is performed by minimizing the cost function using the proposed algorithm. The remaining parts of the paper are arranged in the following way: Sect. 2 presents the overview of the algorithm and the implementation. Section 3 highlights the improved algorithm. Section 4 confers about the results, discussions, and performance analysis. Finally in Sect. 5, the conclusion and further research scope are presented.

2 Overview of Flower Pollination Algorithm

Flower pollination algorithm is a bio-inspired metaheuristic optimization algorithm that mimics the pollination phenomenon of flowering plants [21]. In global pollination, pollen grains are carried by different pollinators (including insects) over long distances, ensuring pollination and reproduction of the fittest plants. The basic model of global pollination and flower constancy can be described as

$$x_i^{t+1} = x_i^t + L(x_i^t - g_*) \tag{1}$$

where x_i^t is the pollen i or solution vector x_i at t th iteration, g_* is the best solution among the current options selected from all solutions of the current generation or iteration, L is mainly a step size and the strength of pollination. Levy flight is used to mimic the insects' behavior of flying over long distances with different distance steps and taking $L > 0$ from a Levy distribution it can be written as

$$L \sim \frac{\lambda \tau(\lambda) \sin(\pi/2)}{\pi} \frac{1}{s^{1+\lambda}}, (s \gg 0) \tag{2}$$

$\tau(\lambda)$ being the standard gamma function, this type of distribution holds true for $s \gg 0$ or large steps. The local pollination and flower constancy is specified as

$$x_i^{t+1} = x_i^t + \epsilon (x_j^t - x_k^t) \tag{3}$$

where x_j^l and x_k^l represent pollens of different flowers from the same species of plant. Flowers in nearby neighborhoods and adjacent patches have a greater probability of undergoing pollination by local flower pollens compared to distant flowers.

3 FPA Algorithm for FIR Filter Design

To obtain FPA-based FIR filter, a random population of ‘ n ’ flowers consisting of filter order (N), pass-band frequency (ω_p) and stop-band frequency (ω_s) has to be initialized. Maximum number of iteration (max_iter) is defined along with the iteration (iter) which is initialized at 0. In the next step, the best solution (G_{best}) and the fitness of all flowers are evaluated. Now the switching probability (p) will be defined between global pollination and local pollination. A random number, say r , is selected between 0 and 1. If r is smaller than 0.5, then a vector L will be drawn using Levy distribution and global pollination will be applied, otherwise local pollination should be applied. After that 1D FIR filter is designed and applied transformation to obtain 2D circular filter. Now the fitness will be calculated and will be replaced if the fitness of the new solution is better than the old one and G_{best} will be updated. If ‘ iter ’ is greater than ‘ max_iter ’ then the best solution will be stored and 2D circular filter will be formulated. If ‘ iter ’ is less than ‘ max_iter ’ then again a random number will be selected and the same procedure will be followed till ‘ iter ’ is greater than ‘ max_iter ’. Here PSNR value is considered as the cost function in proposed algorithm. The flowchart of proposed algorithm is depicted in Fig. 1.

4 Results and Discussion

In this part, a comparison between the proposed method and other standard methods has been presented. To exhibit the performance of the proposed technique, three experiments are conducted on standard images like Cameraman (256×256), Barbara (256×256), Lena (512×512), and Einstein (512×512). In the simulation results, the corresponding images are assumed as Im 1, Im 2, Im 3, and Im 4, respectively. These images are adulterated with Gaussian, speckle, salt and pepper noises at different degradation levels. For quantitative analysis, following image quality metrics are considered to appraise the results of the proposed filter.

(a) Peak signal-to-noise ratio (PSNR):

$$\text{PSNR} = 10 \log \frac{\text{MAX}^2}{\text{MSE}} \quad (4)$$

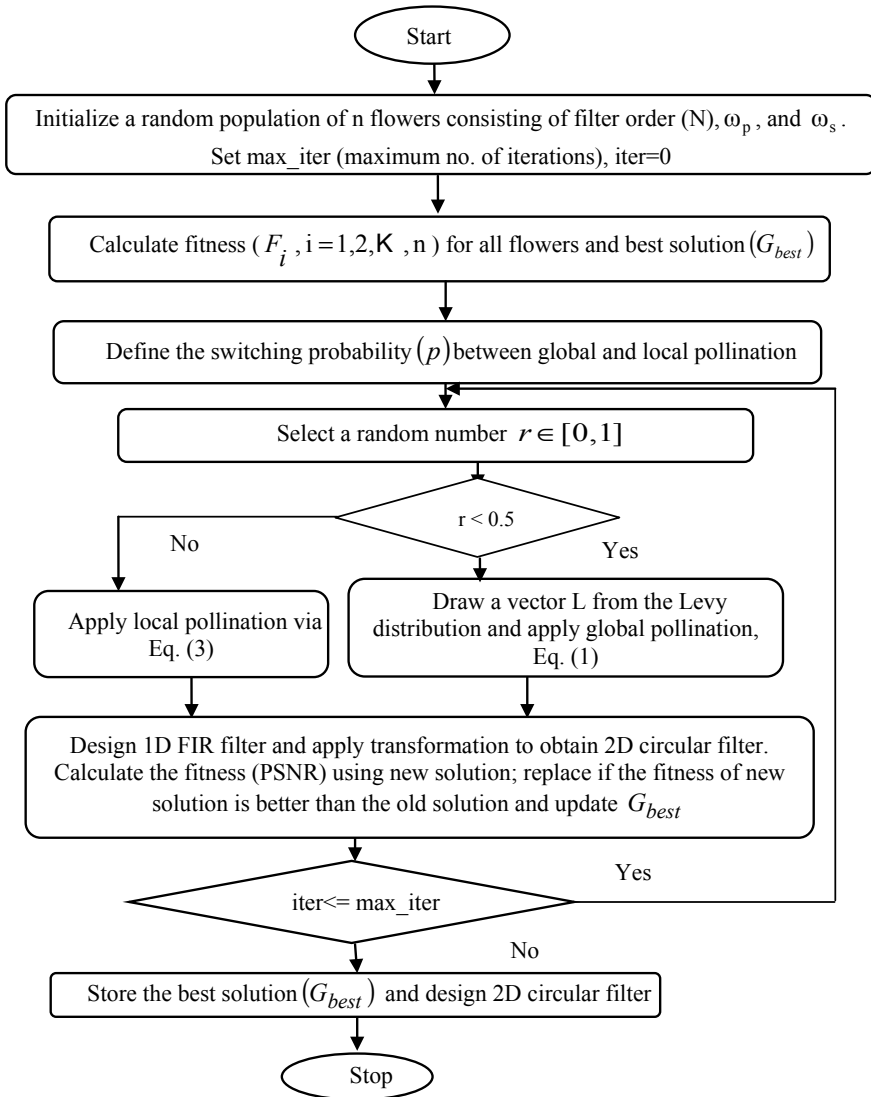


Fig. 1 Flowchart of proposed algorithm

where MAX denotes maximum possible pixel value of the image. MSE denotes mean squared error which is defined as the cumulative error between noisy and actual image. It can be mathematically expressed as:

$$\text{MSE} = \frac{1}{LM} \sum_{i=1}^L \sum_{j=1}^M (f(i, j) - g(i, j))^2 \quad (5)$$

where $f(i, j)$ is the original image, $g(i, j)$ is the image acquired by filtering, $L \times M$ is the dimension of the image.

(b) Image quality index (IQI):

$$\text{IQI} = \frac{\sigma_{ab}}{\sigma_a \cdot \sigma_b} \cdot \frac{2m_a m_b}{m_a^2 + m_b^2} \cdot \frac{2\sigma_a \sigma_b}{\sigma_a^2 + \sigma_b^2} \quad (6)$$

where m_a and m_b are the mean values of A and B images; σ_a^2 , σ_b^2 , σ_{ab} represent the variances of A , B , and covariance of $\{A, B\}$, respectively.

(c) Structural similarity index (SSIM):

$$\text{SSIM}(x, y) = \frac{(2\mu_x \mu_y + c_1)}{(\mu_x^2 + \mu_y^2 + c_1)} \cdot \frac{(2\sigma_{xy} + c_2)}{(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (7)$$

where μ_x , μ_y represent the average values of x and y ; σ_x^2 , σ_y^2 denote the variance of x and y ; σ_{xy}^2 = covariance of x and y ; $c_1 = (k_1 L)^2$ and $c_2 = (k_2 L)^2$. L is the dynamic range of pixel, $k_1 = 0.01$ and $k_2 = 0.03$.

The improvement in PSNR is calculated using Eq. (8) as follows:

$$\begin{aligned} \% \text{ improvement in PSNR} &= \frac{\text{PSNR (other method)} - \text{PSNR (proposed method)}}{\text{PSNR (other method)}} \\ &\times 100\% \end{aligned} \quad (8)$$

To compare the performance of proposed filter, the following filters are also implemented in MATLAB: Average filter, Median filter, Wiener filter, Lee filter, Kuan filter, Frost filter, standard bilateral filter (SBF), and robust bilateral filter (RBF). Three assessments are carried out to study the performance of proposed and above-mentioned denoising methods.

Assessment 1 In this experiment, Im 1, Im 2, and Im 4 are corrupted with Gaussian noise having variance (σ) = 0.4. These noisy images are denoised using FPA-based filter and eight other existing methods. PSNR, IQI, and SSIM values of different denoising methods are calculated for the performance evaluation and presented in Table 1. From Table 1, it is observed that proposed filter outperforms other conventional methods. Considering mentioned image quality metrics, the proposed approach

Table 1 PSNR, IQI, and SSIM values with Gaussian noise at variance = 0.4; shaded and bold cases indicate best results

Gaussian (0.4)	PSNR			IQI			SSIM		
	Im 1	Im 2	Im 4	Im 1	Im 2	Im 4	Im 1	Im 2	Im 4
Noisy	8.212	8.227	8.213	0.697	0.6229	0.6260	0.30137	0.1912	0.1137
Average	15.73	18.38	21.28	0.8950	0.9001	0.9116	0.64264	0.6517	0.7013
Median	14.31	14.19	14.52	0.84337	0.7756	0.7390	0.7250	0.5834	0.4712
Wiener	15.52	16.86	17.88	0.84369	0.7775	0.7259	0.68097	0.6095	0.5432
Lee	15.10	16.07	16.79	0.8324	0.7606	0.7122	0.6660	0.5717	0.4823
Kuan	16.17	17.94	19.55	0.90558	0.8514	0.7894	0.7051	0.6598	0.6320
Frost	15.95	17.31	18.35	0.88197	0.8104	0.7489	0.70513	0.6380	0.5728
SBF	8.291	8.294	8.282	0.69637	0.6216	0.6245	0.30839	0.1955	0.1171
RBF	15.70	16.89	17.91	0.86430	0.8045	0.7462	0.69567	0.6142	0.5461
Proposed	16.76	19.01	21.97	0.93129	0.9151	0.9315	0.73041	0.6919	0.7255

exhibits significant improvement for all images. For example, in Barbara image, the improvements in PSNR values are 3, 34, 13, 18, 6, 10, 100, and 12% as compared to Average, Median, Wiener, Lee, Kuan, Frost, SBF, and RBF filters, respectively. Similarly, the improvements in IQI are 2, 18, 18, 20, 7, 13, 47, and 14% and the improvements in SSIM are 6, 19, 14, 21, 5, 8, 100, and 13%. For visual analysis, the results of different denoising techniques are presented in Fig. 2a–d. Figure 2a represents the noisy image whereas Fig. 2b–d are the results of Average filter, SBF, and the proposed method. It is clearly observed that the execution of proposed denoising method is superior in quality than alternative methods.

Assessment 2 In this assessment, Im 1, Im 3, and Im 4 were contaminated with salt and pepper noise at variance (σ) = 0.6. In a similar way, PSNR, IQI, and SSIM values are investigated to appraise the performance of different methods and the results are presented in Table 2. From Table 2, it can be determined that for Einstein image the improvements in PSNR are 3, 24, 21, 27, 11, 17, 100, and 20% with respect to Average, Median, Wiener, Lee, Kuan, Frost, SBF, and RBF filtering methods. Also the improvements in IQI are 1, 0.2, 34, 37, 23, 31, 30, and 31% while for SSIM the improvements are seen to be 0.2, 3, 36, 53, 12, 27, 100 and 35%. Figure 3a–d reveal the results for visual differentiation. Figure 3a represents the noisy image whereas Fig. 3b–d are the results of Lee filter, Median filter, and the proposed method. It is quite evident that the performance of proposed denoising method is better and texture of the image is preserved effectively.

Assessment 3 In this experiment, Im 1, Im 2, and Im 4 are corrupted by speckle noise having variance (σ) = 0.4. The above-mentioned three image quality metrics of different denoising methods are calculated to analyze their operations and the results are summarized in Table 3. In this case also the proposed method shows appreciable improvement for all three images. For example, in Cameraman image

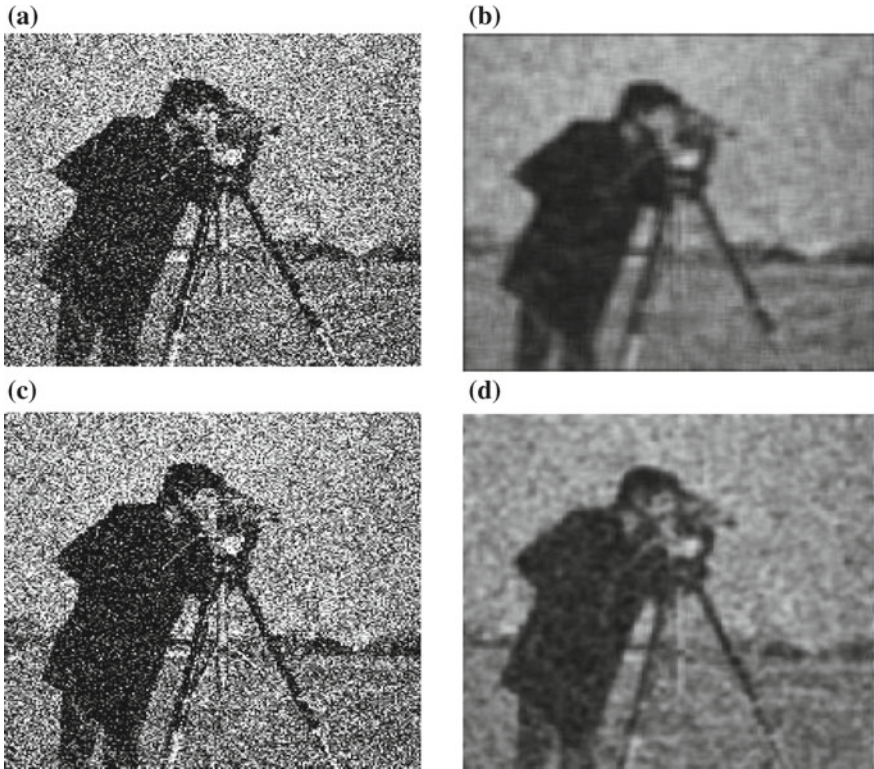


Fig. 2 **a** Noisy image, **b** average filter, **c** standard bilateral filter, **d** proposed

Table 2 PSNR, IQI, and SSIM values with salt and pepper noise at variance 0.6; shaded and bold cases indicate best results

Salt and pepper (0.6)	PSNR			IQI			SSIM		
	Im 1	Im 3	Im 4	Im 1	Im 3	Im 4	Im 1	Im 3	Im 4
Noisy	7.305	7.6444	7.8755	0.7565	0.7264	0.6981	0.2070	0.1382	0.0781
Average	14.70	17.388	19.725	0.9088	0.9363	0.8979	0.5119	0.5588	0.5516
Median	14.74	15.689	16.396	0.8975	0.9063	0.9008	0.7464	0.6893	0.5646
Wiener	14.09	15.671	16.832	0.7997	0.7426	0.6729	0.5328	0.4941	0.4052
Lee	13.72	14.996	15.962	0.7834	0.7232	0.6580	0.5200	0.4596	0.3590
Kuan	14.77	16.815	18.319	0.8692	0.8160	0.7352	0.5626	0.5559	0.4892
Frost	14.28	16.069	17.285	0.8207	0.7695	0.6904	0.5497	0.5187	0.4346
SBF	7.359	7.7076	7.9348	0.7551	0.7256	0.6939	0.2098	0.1383	0.0762
RBF	14.18	15.644	16.866	0.8163	0.7605	0.6868	0.5437	0.4935	0.4082
Proposed	15.34	17.760	20.305	0.9311	0.9273	0.9030	0.5887	0.5983	0.5502

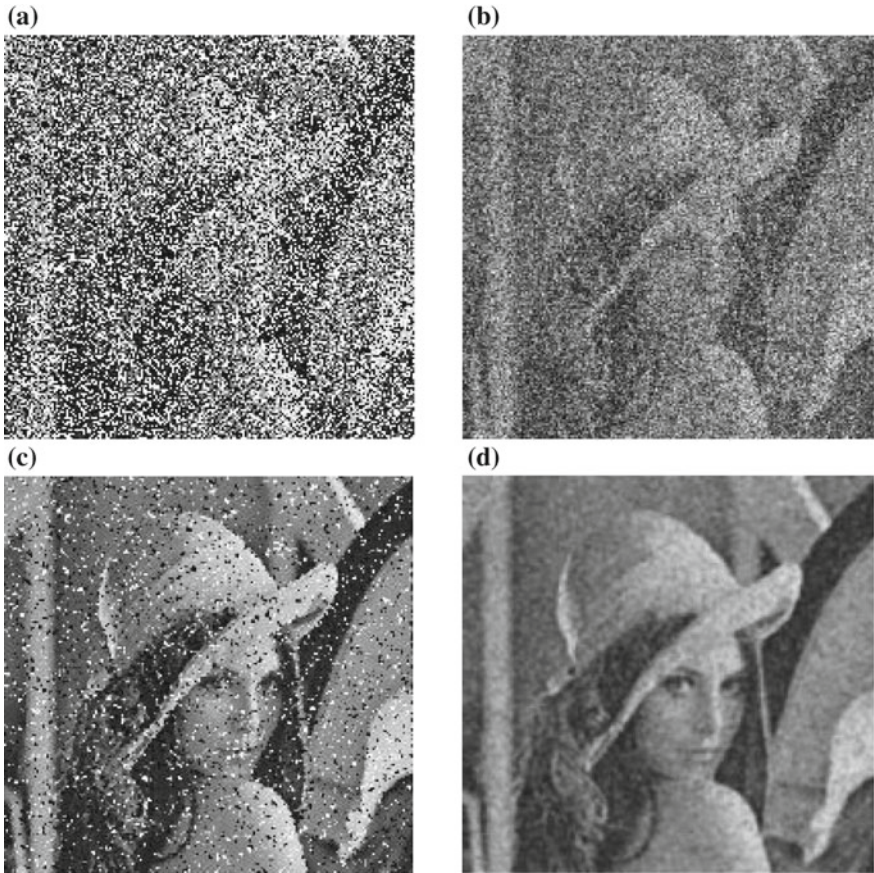


Fig. 3 a Noisy image, b Lee filter, c Median filter, d proposed

the improvements in PSNR are noted as 7, 27, 17, 11, 7, 25, 90, and 4% as compare to Average, Median, Wiener, Lee, Kuan, Frost, SBF, and RBF filtering methods. In case of IQI, the improvements are found to be 2, 7, 7, 3, 1, 0.3, 23, and 1% and for SSIM the improvements are 3, 12, 8, 4, 2, 0.3, 61, and 1%. For visual assessment, the results of different denoising techniques are mentioned in Fig. 4a–d. Figure 4a represents the noisy image whereas Fig. 4b–d are the results of Kuan filter, Frost filter, and the proposed method. It can be observed that the proposed technique is more efficient in terms of eliminating speckle noise than other existing methods.

Table 3 PSNR, IQI, and SSIM values with speckle noise at variance 0.4; shaded and bold cases indicate best results

Speckle (0.4)	PSNR			IQI			SSIM		
	Im 1	Im 2	Im 4	Im 1	Im 2	Im 4	Im 1	Im 2	Im 4
Noisy	10.73	11.171	11.25	0.771	0.6914	0.6837	0.56114	0.4147	0.2945
Average	19.33	20.369	24.07	0.930	0.9105	0.9341	0.89612	0.8246	0.8723
Median	16.28	17.518	18.57	0.888	0.8380	0.8005	0.82524	0.7585	0.7021
Wiener	17.69	18.413	19.69	0.885	0.8394	0.8021	0.85865	0.7758	0.7374
Lee	18.64	19.342	20.44	0.920	0.8727	0.8329	0.88501	0.8126	0.7715
Kuan	19.48	20.885	23.27	0.944	0.9228	0.8997	0.90125	0.8566	0.8616
Frost	20.36	21.315	23.21	0.947	0.9158	0.8844	0.92062	0.8713	0.8634
SBF	10.92	11.475	11.60	0.770	0.6926	0.6845	0.57235	0.4318	0.3118
RBF	19.89	20.531	22.59	0.941	0.9070	0.8774	0.91343	0.8502	0.8458
Proposed	20.74	22.341	26.13	0.950	0.9359	0.9487	0.92312	0.8902	0.9206

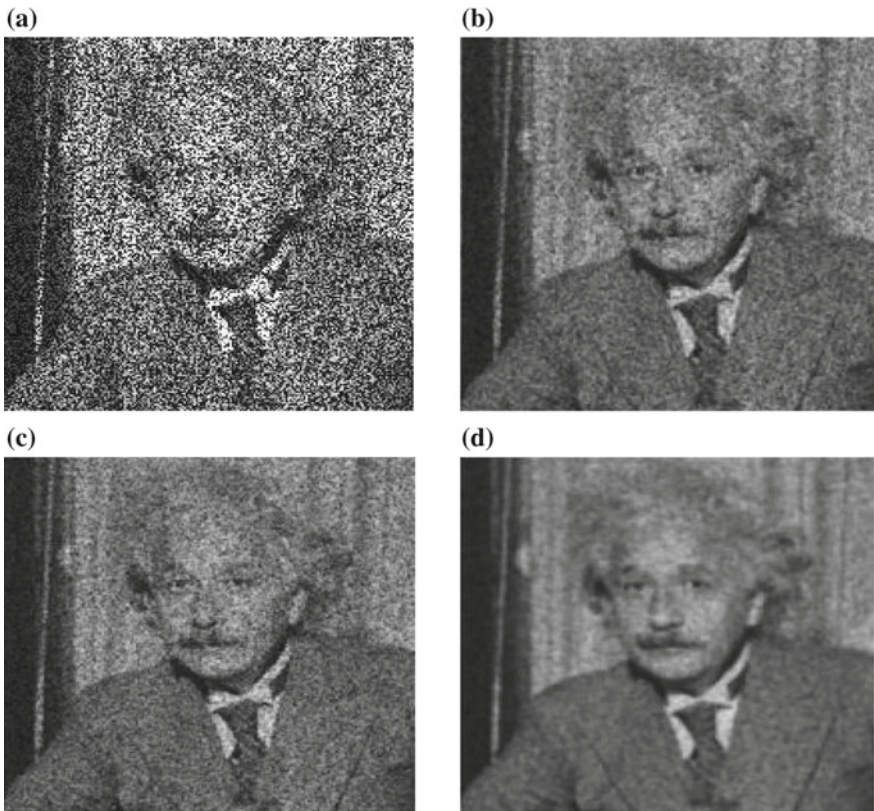


Fig. 4 a Noisy image, b Kuan filter, c Frost filter, d proposed

5 Conclusions

In this paper, a flower pollination algorithm-based two-dimensional FIR filter design is presented for image denoising. On performing the comparison, it was verified that the proposed algorithm exhibits better performance in terms of PSNR, SSIM, and IQI values under various noise considerations. From the figures, it could be estimated that the proposed approach produces better denoised image than the Median, Wiener, bilateral, Lee, and Kuan filters. In the future, we will put emphasis on the design of three-dimensional FIR filters for the video signal processing.

References

1. D.L. Donoho, De-noising by soft-thresholding. *IEEE Trans. Inf. Theory* **41**(3), 613–627 (1995)
2. D.L. Donoho, I.M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage. *J. Am. Stat. Assoc.* **90**(432), 1200–1224 (2012)
3. S. Pragada, J. Sivaswamy, Image denoising using matched biorthogonal wavelets, in *Proceedings of Sixth Indian Conference on Computer Vision, Graphics & Image Processing* (2008), pp. 25–32
4. V.N.P. Raj, T. Venkateswarlu, Denoising of medical images using dual tree complex wavelet transform. *Proc. Technol.* **4**, 238–244 (2012)
5. S.A. Shanthi, C.H. Sulochana, T. Latha, Image denoising in hybrid wavelet and quincunx diamond filter bank domain based on Gaussian scale mixture model. *Comput. Electr. Eng.* **46**, 384–393 (2015)
6. Lee JS (1980) Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* **2**(2), 165–168
7. D.T. Kuan, A.A. Sawchuk, T.C. Strand, P. Chavel, Adaptive noise smoothing filter for images with signal-dependent noise. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* **7**(2), 165–177 (1985)
8. V.S. Frost, J.A. Stiles, K.S. Shanmugan, J.C. Holtzman, A model for radar images and its application to adaptive digital filtering of multiplicative noise. *IEEE Trans. Pattern Anal. Mach. Intell. PAMI* **4**(2), 157–166 (1982)
9. J. Yang, J. Fan, D. Ai, X. Wang, Y. Zheng, S. Tang, Y. Wang, Local statistics and non-local mean filter for speckle noise reduction in medical ultrasound image. *Neurocomputing* **195**, 88–95 (2016)
10. S. Chitroub, Principal component analysis by neural network. Application: remote sensing images compression and enhancement, in *10th IEEE International Conference on Electronics, Circuits and Systems*, vol. 1 (2003), pp. 344–347
11. N. Mohamadi, A.R. Soheili, F. Toutounian, A new hybrid denoising model based on PDEs. *Multimedia Tools Appl.* **77**(10), 12057–12072 (2018)
12. D. Bharathi, S.M. Govindan, A new hybrid approach for denoising medical images. *Advances in computing and information technology. Adv. Intell. Syst. Comput.* **177**, 905–914 (2013)
13. P.D. Patil, A.D. Kumbhar, Bilateral filter for image denoising, in *Proceedings of International Conference on Green Computing and Internet of Things* (2015), pp. 299–302
14. K.N. Chaudhury, K. Rithwik, Image denoising using optimally weighted bilateral filters: a SURE and fast approach, in *Proceedings of IEEE International Conference on Image Processing* (2015), pp. 108–112
15. S. Kockanat, N. Karaboga, T. Koza, Image denoising with 2-D FIR filter by using artificial bee colony algorithm, in *Proceedings of International Symposium on Innovations in Intelligent Systems and Applications* (2012), pp. 1–4

16. F. Latifoğlu, A novel approach to speckle noise filtering based on Artificial Bee Colony algorithm: an ultrasound image application. *Comput. Methods Programs Biomed.* **111**(3), 561–569 (2013)
17. J. Hua, W. Kuang, Z. Gao, L. Meng, Z. Xu, Image denoising using 2-D FIR filters designed with DEPSO. *Multimedia Tools Appl.* **69**(1), 157–169 (2014)
18. J.L. de Paiva, C.F.M. Toledo, H. Pedrini, An approach based on hybrid genetic algorithm applied to image denoising problem. *Appl. Soft Comput.* **4**, 778–791 (2016)
19. Z.K. Huang, Z.H. Li, H. Huang, Z.B. Li, L.Y. Hou, Comparison of different image denoising algorithms for Chinese calligraphy images. *Neurocomputing* **188**, 102–112 (2016)
20. S.V. Kumar, C. Nagaraju, FFBF: cluster-based Fuzzy Firefly Bayes Filter for noise identification and removal from grayscale images. *Cluster Comput.* 1–23 (2018)
21. X.S. Yang, Flower pollination algorithm for global optimization, in *International Conference on Unconventional Computing and Natural Computation Lecture Notes in Computer Science*, vol. 7445 (2012), pp. 240–249

Image Enhancement Using Differential Evolution Based Whale Optimization Algorithm



Supriya Dhabal and Dip Kumar Saha

Abstract This paper proposes an enhancement method of digital images applying differential evolution based whale optimization algorithm (DEWOA). The enhancement is performed by improving the intensity of pixels in a given image. This is realized using a cost function that contains both the local and global information. A comparison of the proposed DEWOA is performed with few recently developed metaheuristic algorithms like PSO, ABC, CSA, and FPA. The simulation results are presented with respect to the background variance, detail variance, PSNR, entropy, and number of detected edges.

Keywords Image enhancement · BV · DV · PSNR · Entropy · WOA

1 Introduction

Image enhancement is a method which enhances the contrast level of a given image by sharpening the intensity of edge pixels as compared to its background [1, 2]. It also indicates the removal of noise and magnification of the pixel for the proper interpretation of the image. Image enhancement has important roles in medical image analysis, remote sensing, human vision, and so on. Basically, enhancement technique increases the quality of the image which helps to extract visual information from the enhanced image. Enhancement can be performed both in frequency domain and spatial domain. The spatial domain image processing is performed on the pixel values itself whereas in the frequency domain approach the transformations like Fourier or Wavelet of the image is taken.

S. Dhabal (✉) · D. K. Saha
Department of Electronics and Communication Engineering, Netaji Subhash Engineering
College, Kolkata 700152, West Bengal, India
e-mail: supriya_dhabal@yahoo.co.in

D. K. Saha
e-mail: saha.dip@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling
and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_54

The recent developments in image enhancement used few optimization-based approaches like genetic algorithm (GA) [3, 4], particle swarm optimization (PSO) [5–8], artificial bee colony (ABC) [9], flower pollination algorithm (FPA) [10], cuckoo search algorithm (CSA) [11], biogeography-based optimization (BBO) [12], and differential evolution (DE) [13]. It was found that most of the available image enhancement methods introduce abnormality in the natural looking of images. Thus, enhancement of images in different domains like medical, remote sensing, etc. is an open research area [3–8]. To solve this problem, here an improved image contrast enhancement algorithm is proposed, named as differential evolution based whale optimization algorithm (DEWOA). In most of the methods, an objective function based on image intensity, entropy, and edge information is formulated which is then maximized using the global optimization algorithm.

Recently, an efficient algorithm known as whale optimization algorithm (WOA) was developed by Mirjalili following the behavior of whales [14]. WOA demonstrated better performance in solving of benchmark problems than PSO [5–8], gravitational search algorithm (GSA) [15], GA, evolution strategy (ES) [16], and DE. Afterward, WOA has been used in solving several engineering problems like optimization of renewable energy resources in the distribution systems [17], feature selection [18], multilevel thresholding segmentation [19], and optimization of skeletal structures [20] and so on. For better searching behavior of WOA, an improved WOA based on DE is presented and it is verified on the enhancement of images.

The remaining part of the paper is arranged as follows: the problem formulation of image enhancement is discussed in Sect. 2. Section 3 presents an overview of whale optimization algorithm. The proposed DEWOA is discussed in Sect. 4. Finally, the simulation results of DEWOA and conclusions are presented in Sects. 5 and 6, respectively.

2 Problem Formulation

In spatial domain with the help of a transformation function, intensity of each pixel values is changed to a new value to improve the contrast. The transformation function can be defined as:

$$G(i, j) = T[F(i, j)] \quad (1)$$

where $G(i, j)$ and $F(i, j)$ denote the pixel intensity of the enhanced and the input image, respectively. The function T in Eq. (1) can be formulated as

$$T[F(i, j)] = K(i, j)[F(i, j) - c \times m(i, j)] + m(i, j)^a \quad (2)$$

where c, a are the unknown constants, and $m(i, j)$ represents local mean value of the original image calculated for a $n \times n$ window:

$$m(i, j) = \frac{1}{n \times n} \sum_{x=0}^{n-1} \sum_{y=0}^{n-1} F(x, y) \quad (3)$$

The enhancement function $K(i, j)$ can be defined as

$$K(i, j) = \frac{k.D}{\sigma(i, j) + b} \quad (4)$$

where b and k are unknown constants. The local standard deviation of the input image, i.e., $\sigma(i, j)$ and the global mean (D) is calculated as per Eqs. (5) and (6), respectively.

$$\sigma(i, j) = \sqrt{\frac{1}{N \times N} \sum_{x=0}^N \sum_{y=0}^N [F(x, y) - m(i, j)]^2} \quad (5)$$

$$D = \frac{1}{M \times N} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} F(x, y) \quad (6)$$

Now from Eqs. (1), (2), and (4), a new transformation function is obtained as follows:

$$G(i, j) = \left[\frac{k.D}{\sigma(i, j) + b} \{F(i, j) - c \times m(i, j)\} + m(i, j)^a \right] \quad (7)$$

The local mean value is assumed as the center of enhancement and it performs the enhancement along with the unknown quantities a, b, c , and k . After the enhancement, the quality of enhanced image is evaluated. For this, a cost function is formulated which involves the sum of edge intensities, entropy value, and number of detected edge pixels. Accordingly, the cost function $\phi(Z)$ can be defined as [5]:

$$\phi(Z) = \frac{\text{ne}(I(Z)) \times \log(\log(E(I(Z)))) \times H(I(Z))}{M \times N} \quad (8)$$

where $I(Z)$ denotes the image obtained after transformation, $\text{ne}(I(Z))$ is the total number of pixels appear at the edges of $I(Z)$, and $E(I(Z))$ is the edge intensity of enhanced image sensed by Sobel operator. The enhanced image entropy, denoted as $H(I(Z))$, is calculated using Eq. (9).

$$H(I(Z)) = - \sum_{i=0}^{255} e_i \quad (9)$$

where $e_i = h_i \log_2(h_i)$, $h_i (\neq 0)$ represents the probability of the i th intensity level in the enhanced image.

3 Whale Optimization Algorithm

WOA, a recently developed bio-inspired algorithm in the field of optimization, imitates the hunting pattern of whales. The prey-encircling, hunting, and prey-searching are the three primary mechanisms in WOA. The prey-encircling behavior in WOA is represented as

$$\vec{X}(t + 1) = \vec{X}^*(t) - \vec{A} \otimes \left| \vec{C} \cdot \vec{X}^*(t) - \vec{X}(t) \right| \tag{10}$$

where \vec{X}^* is the solution vector of best candidate and ‘ \otimes ’ signifies element-wise multiplication. The value of \vec{X}^* need to be adjusted in each evolution based on the quality of the output. The vectors \vec{C} and \vec{A} are evaluated as follows:

$$\vec{C} = 2 \cdot \vec{r} \tag{11}$$

$$\vec{A} = 2 \cdot \vec{a} \cdot \vec{r} - \vec{a} \tag{12}$$

where the value of \vec{a} is gradually reduced from 2 to 0 during the entire evolutions and $\vec{r} \in [0, 1]$. The hunting procedure of humpback whales is known as bubble net hunting which can be described by:

$$\vec{X}(t + 1) = \begin{cases} \vec{X}^*(t) - \vec{A} \otimes \left| \vec{C} \cdot \vec{X}^*(t) - \vec{X}(t) \right| & \text{if } p < 0.5 \\ e^{bl} \cdot \cos(2\pi l) \otimes \left| \vec{X}^*(t) - \vec{X}(t) \right| + \vec{X}^* & \text{if } p \geq 0.5 \end{cases} \tag{13}$$

The prey-searching behavior of whales represents the exploration phase of the algorithm and the corresponding search equation is given by:

$$\vec{X}(t + 1) = \vec{X}_{\text{rand}} - \vec{A} \cdot \left| \vec{C} \cdot \vec{X}_{\text{rand}} - \vec{X} \right| \tag{14}$$

where \vec{X}_{rand} is a randomly selected solution taken from the pool of existing whale population. The exploration and exploitation in WOA are achieved depending on the values of A ; $|A| > 1$ leads to exploration whereas $|A| < 1$ helps in exploitation behavior of whales.

4 Proposed Algorithm

The cost function $\phi(Z)$, mentioned in Eq. (8), can be optimized using the proposed DEWOA algorithm which involves the following steps:

Step 1: Begin with a random selection of NP solutions: $P = \{X_1, \dots, X_{NP}\}$ with $X_i = \{x_i^1, \dots, x_i^D\}$, $i = 1, \dots, NP$. Set the current iteration $\text{iter} = 0$, failure = 0,

and assign maximum number of iterations as $iter_max$. Define a set of crossover rate $CR \in (0.1, 0.2, \dots, 0.9)$ and scale factor $SF \in (0.3, 0.4, \dots, 0.9)$.

Step 2: Assign each member with a randomly selected crossover rate (CR) and scale factor (SF) from the set of CR and SF values.

Step 3: WHILE $iter \leq iter_max$ DO the following:

Step 3.1: For each solution in the population pool, i.e., $i = 1, \dots, NP$

- (a) Generate a mutated vector $V_i = \{v_i^1, \dots, v_i^D\}$ corresponding to each target X_i based on the random selection of following three equations:

$$V_i = X^* + sf \otimes (X_{r1} - X_{r2}) \quad (15a)$$

$$V_i = X_{r1} + sf \otimes (X_{r2} - X_{r3}) \quad (15b)$$

$$V_i = X^* + e^{bl} \cdot \cos(2\pi l) \otimes (X^* - X_i) + sf \otimes (X_{r1} - X_{r2}) \quad (15c)$$

where sf denotes a random vector with mean selected from the typical entries of SF, ' \otimes ' denotes the element-wise multiplication.

- (b) Generate a trial vector $U_i = \{u_i^1, \dots, u_i^D\}$ for each target vector X_i using the following equation:

$$U_i = \begin{cases} V_i & \text{if } \text{rand} < CR \\ X_i & \text{otherwise} \end{cases} \quad (16)$$

- (c) Check whether the entries of solution vector are within upper and lower bounds or not. If not, then assign the entry with its boundary limit.

Step 3.2: For each solution, update the target vector (X_i) if it improves the fitness and also store the corresponding settings of CR and SF values. Otherwise, set $failure = failure + 1$.

Step 3.3: Update the best member of the population (X^*) if the new mutated vector (U_i) is better than the old best solution of the population.

Step 3.4: Calculate the mean failure rate: if $\text{rand} < \text{mean}(failure)$ then accept the values of crossover and scale factor from the earlier stored settings of CR, SF; otherwise a completely new setting of CR and SF is produced.

Step 4: Exit with the best solution (X^*) obtained as the global optimum.

In this proposed algorithm, initially a random population is chosen and maximum number of iteration is fixed to a predefined value. The initial iteration value is set to zero along with a failure counter which is also initialized to zero. A set of value is defined as Crossover Rate (CR) and Scale Factor (SF) for mutation of population members to get a new generation of search agent. Each possible solution in the population is assigned with a particular value of CR and SF from the predefined set. A mutated search agent is generated using Eqs. (15a)–(15c). Equations (15a) and (15b) was evolved directly from the differential evolution (DE) algorithm where a population member is being compared with another one and multiplied with the

Table 1 Detail description of the original images

Image	Size	BV	DV	Entropy	No. of edges
Lena	512 × 512	0.7116	119.7131	7.4455	8308
Vegetable	256 × 256	0.7712	298.2435	7.5326	2447
Ship	512 × 512	0.7273	194.5524	7.1913	10,555

weighted factor (SF) corresponding to a target vector. In Eq. (15c), the bubble net hunting method of whales is introduced. These three equations together generate a new mutated search agent. The crossover is performed using Eq. (16). After the generation of a new solution, it is checked whether the entries of the vector lies within the specified boundary or not. For each solution, the target vector is updated and if the fitness of the target vector is maximized then the value of CR and SF is being stored and if the fitness didn't improve then the failure counter is incremented. Finally, the best member of the population is updated if the mutated vector has better potential than the old best candidate.

5 Simulation Results and Discussion

The algorithm proposed in Sect. 4 is checked on few standard images. The performance analysis of three gray-level images like Lena, Vegetable, and Ship are discussed here. The simulation results of DEWOA are compared with the four algorithms namely PSO, ABC, CSA, and FPA. The parameters such as image size, background variance (BV), detail variance (DV), entropy, and number of detected edges of original images are listed in Table 1. The cost function mentioned in Eq. (8) is evaluated using the above algorithms and the simulation outputs are presented in Tables 2, 3 and 4. The calculation of BV and DV is performed by considering the local variance over a $n \times n$ window for the enhanced image. Here the window size is assumed as 3. The average variance due to the pixels in the foreground of the image is denoted as DV and for the background pixels it is estimated as BV. An image is considered as properly enhanced if the DV value of an image increases without significant change in BV of the original image [3, 5].

Example I The enhanced image of Lena is analyzed and presented in this example. The performance results are tabulated in Table 2 and the images obtained after enhancement are displayed in Fig. 1a–f. The performance analysis concluded that the proposed DEWOA exhibits the highest values of PSNR, entropy, and number of detected edges. As an example, the value of entropy in PSO, ABC, CSA, and FPA is 7.434507, 7.63254, 7.425906, and 7.635562, respectively. It may be noted that the proposed DEWOA results in highest value of the entropy, i.e., 7.65664.

Example II In this example, the result of the enhanced Vegetable image is presented. The performance results are listed in Table 3 and the corresponding output images are

Table 2 Comparing performance results for Lena image

Lena					
Algorithm	BV	DV	PSNR	Entropy	No. of edges
PSO [5]	0.07895	1587.98	61.9143	7.434507	11,173
ABC [9]	0.13267	1583.01	62.0129	7.63254	11,492
CSA [11]	0.32239	1003.71	66.5622	7.425906	8939
FPA [10]	0.20176	1606.68	62.4227	7.635562	11,511
Proposed	0.31965	710.958	68.2845	7.65664	13,047

Table 3 Comparing performance results for Vegetable image

Vegetable					
Algorithm	BV	DV	PSNR	Entropy	No. of edges
PSO [5]	0.19188	1219.442	64.30093	7.503146	3305
ABC [9]	0.147743	1140.444	67.9031	7.716686	3269
CSA [11]	0.143499	1186.329	67.3958	7.701565	3196
FPA [10]	0.14234	1187.329	67.43424	7.703569	3226
Proposed	0.214058	1192.778	69.33608	7.789437	3460

Table 4 Comparing performance results for Ship image

Ship					
Algorithm	BV	DV	PSNR	Entropy	No. of edges
PSO [5]	0.124525	561	59.80224	7.184249	11,284
ABC [9]	0.132253	537.0010	60.44477	7.221875	11,784
CSA [11]	0.171819	564	63.83874	7.181698	10,069
FPA [10]	0.066419	626	60.69591	7.392591	11,724
Proposed	0.174153	503.6403	70.70454	7.43894	12,329

illustrated in Fig. 2a–f. The proposed DEWOA exhibits the highest value of PSNR, entropy, and number of detected edges. For example, the value of PSNR in PSO, ABC, CSA, and FPA is 64.30093, 67.9031, 67.3958, and 67.43424 dB, respectively whereas the proposed method yields the highest value of PSNR, i.e., 69.33608 dB.

Example III The third example demonstrates the performance output of the Ship image. The results obtained from these five algorithms are listed in Table 4 and the enhanced images are depicted in Fig. 3a–f. The performance comparison reveals that the proposed DEWOA produces the simulation results with highest values of PSNR, entropy, and number of detected edges. For example, the ‘number of detected edges’ for PSO, ABC, CSA, and FPA is 11,284, 11,784, 10,069, and 11,724, respectively whereas in case of proposed algorithm it is the maximum value, 12,329.



Fig. 1 Simulation results of Lena image: **a** input, **b** PSO, **c** ABC, **d** CSA, **e** FPA, **f** proposed

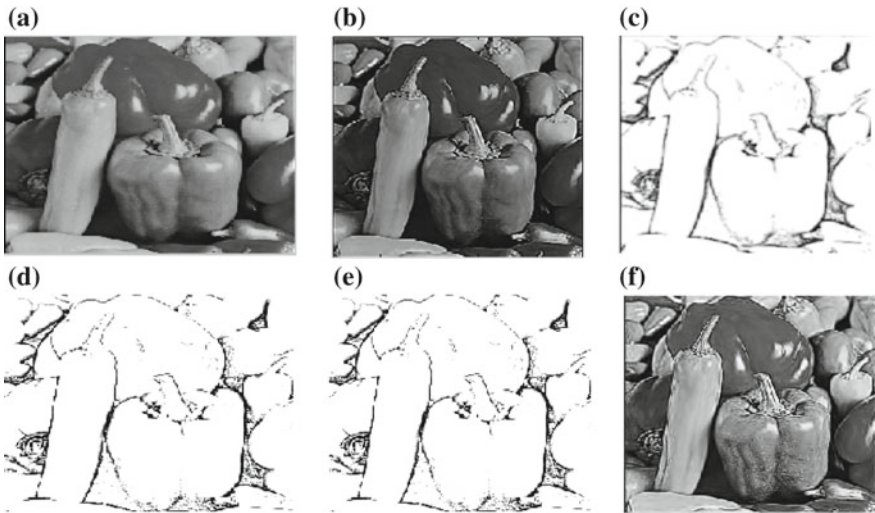


Fig. 2 Simulation results of Vegetable image: **a** input, **b** PSO, **c** ABC, **d** CSA, **e** FPA, **f** proposed

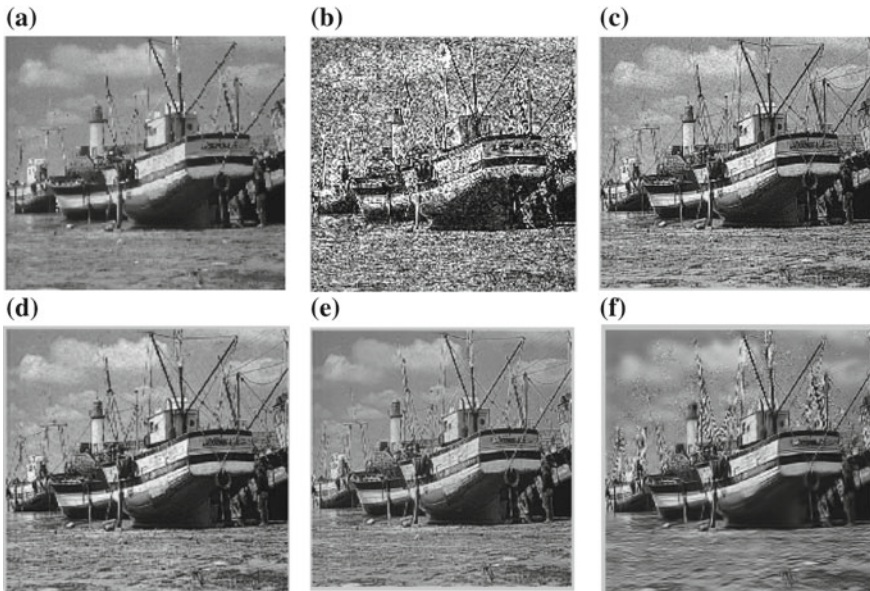


Fig. 3 Simulation results of Ship image: **a** input, **b** PSO, **c** ABC, **d** CSA, **e** FPA, **f** proposed

6 Conclusions

This paper addresses an alternative approach derived from the WOA to enhance digital images. The performance of DEWOA is evaluated on several gray images and the results are compared with few optimization-based algorithms like PSO, ABC, CSA, and FPA. The proposed algorithm produces assessment parameters of the enhanced images with highest value of PSNR, entropy, and number of detected edges than existing methods, which in turn validates robustness of the proposed DEWOA. The simulation results exhibit best optimal solution in almost all cases. Therefore, the proposed DEWOA could be the better heuristic algorithm to enhance gray images and it may be applied efficiently in solving other complicated image processing problems like denoising, segmentation, fusion, etc.

References

1. R.C. Gonzalez, R.E. Woods, *Digital Image Processing*, 3rd edn. (Prentice Hall, Upper Saddle River, NJ, 2008)
2. R.C. Gonzales, B.A. Fittes, Gray-level transformations for interactive image enhancement. *Mech. Mach. Theory* **12**(1), 111–122 (1977)
3. S. Hashemi, S. Kiani, N. Noroozi, M.E. Moghaddam, An image contrast enhancement method based on genetic algorithm. *Pattern Recogn. Lett.* **31**(13), 1816–1824 (2010)

4. F. Saitoh, Image contrast enhancement using genetic algorithm, in *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, vol. 4 (1999), 899–904
5. A. Gorai, A. Ghosh, A gray-level image enhancement by particle swarm optimization, in *World Congress on Nature & Biologically Inspired Computing (NaBIC)*, Coimbatore (2009), pp. 72–77
6. M. Braik, A. Sheta, A. Ayesh, Image enhancement using particle swarm optimization, in *Proceedings of the World Congress on Engineering WCE 2007* (2007)
7. N.M. Kwok, D. Wang, Q.P. Ha, G. Fang, S.Y. Chen, *Locally-Equalized Image Contrast Enhancement Using PSO-Tuned Sectorized Equalization. Computational Intelligence in Image Processing* (Springer, Berlin, Heidelberg, 2013), pp. 21–36
8. S.M.W. Masra, P.K. Pang, M.S. Muhammad, K. Kipli, Application of particle swarm optimization in histogram equalization for image enhancement, in *Proceedings of IEEE Colloquium on Humanities, Science and Engineering (CHUSER)*, Kota Kinabalu (2012), pp. 294–299
9. S.K. Mustafa, O. Findik, A directed artificial bee colony algorithm. *Appl. Soft Comput.* **26**, 454–462 (2015)
10. E. Nabil, A modified flower pollination algorithm for global optimization. *Expert Syst. Appl.* **57**, 192–203 (2016)
11. M. Mareli, B. Twala, An adaptive Cuckoo search algorithm for optimisation. *Appl. Comput. Inf.* **14**(2), 107–115 (2018)
12. J. Jasper, S.B. Shaheema, S.B. Shiny, Natural image enhancement using a biogeography based optimization enhanced with blended migration operator. *Math. Prob. Eng.* **Article ID 232796**, 11 p. <https://doi.org/10.1155/2014/232796>
13. P.P. Sarangi, B.S.P. Mishra, B. Majhi, S. Dehuri, Gray-level image enhancement using differential evolution optimization algorithm, in *Proceedings of the International Conference on Signal Processing and Integrated Networks (SPIN)*, Noida (2014), pp. 95–100
14. S. Mirjalili, A. Lewis, The whale optimization algorithm. *Adv. Eng. Softw.* **95**, 51–67 (2016)
15. S.M. Mahmoudi, M. Aghaie, M. Bahonar, N. Poursalehi, A novel optimization method, Gravitational Search Algorithm (GSA), for PWR core optimization. *Ann. Nucl. Energy* **95**, 23–34 (2016)
16. K. Weicker, N. Weicker, On evolution strategy optimization in dynamic environments, in *Proceedings of the 1999 Congress on Evolutionary Computation-CEC99 (Cat. No. 99TH8406)*, Washington, DC, vol. 3 (1999), p. 2046
17. P.D.P. Reddy, V.C.V. Reddy, T.G. Manohar, Whale optimization algorithm for optimal sizing of renewable resources for loss reduction in distribution systems. *Renew: Wind Water Solar* 1–13
18. M.M. Mafarja, S. Mirjalili, Hybrid whale optimization algorithm with simulated annealing for feature selection. *Neurocomputing* **260**, 302–312 (2017)
19. M.A.E. Aziz, A.A. Ewees, A.E. Hassanien, M. Mudhsh, S. Xiong, Multi-objective whale optimization algorithm for multilevel thresholding segmentation. *Adv. Soft Comput. Mach. Learn. Image Process. Stud. Comput. Intell.* **730**, 23–39 (2018)
20. A. Kaveh, M.I. Ghazaan, Enhanced whale optimization algorithm for sizing optimization of skeletal structures. *Mech. Based Des. Struct. Mach.* **45**(3), 345–362 (2016)

Advanced Portable Exoskeleton with Self-healing Technology Assisted by AI



Piyush Keshari and Santanu Koley

Abstract The intention of this paper is to exhibit the wonderful combination of the future artificial intelligent system capable of making important decision with prompt action installed in the mechanical exoskeleton. It will follow the orders from the operator of the suit and will be comprising of nanochips to control the various function of the system. Such a system will be controlled by the human beings encoded with a single person at a time. This is recommended for the safety of the suit itself. This exoskeleton will have self-healing technology. At the time when exoskeleton is getting damaged by any man-made source, then it will start the self-healing process. The exoskeleton will not only provide the protection of the person operating it but also has a wide variety of use. The most significant part of this exoskeleton will be its self-sustaining energy production to power up the suit for its functioning. The energy producer will be environment-friendly and will not produce any toxicants. Magnetic fields will be used to yield electricity. This suit will be capable of working everywhere in air, water and land. Since this will be the first prototype of the suit, some more characteristics with developments will be presented inside. This will also be the encroachment in the field of robotics where there will be the best coordination between a machine and a human. This modernisation will defend the precious human lives fighting for redeemable civilisation.

Keywords Exoskeleton · Nanochips · Self-sustaining energy · Self-healing technology

P. Keshari · S. Koley (✉)
Department of Computer Science & Engineering, Budge Budge Institute of Technology,
Kolkata, WB, India
e-mail: santanukoley@yahoo.com

P. Keshari
e-mail: kesharipiyush05124@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling
and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_55

1 Introduction

Here, we will elaborate the full function of the suit, technology that will be used, benefits and various prospects. The suit will comprise three layers, viz. the outer part, the middle part and the inner part.

The outer part will consist of metal and self-healing polymers by virtue of which if damaged artificially, the suit will regain its structure which will help the suit for the long run as well as less maintenance will be required. This will also help the suit if it is located in the region, where no mechanical support can be reached.

The middle layer will comprise of tiny nanochips which will be used for the functioning of the suit. This layer is vital as it has all the essential microcontrollers and microprocessors. These nanochips will be activated only when its genuine operator will try to turn on the suit; that is, the suit will only be operated by the person who is encoded with the suit. A nanochip will be injected to the operator's body in the elastic cartilage part which will not only be used to turn on the suit but also verify the operator by its retina scan and DNA test.

The inner layer will comprise of insulation sheet. This layer will protect the operator from external electric shocks and also have grounding mechanism; i.e., if the suit is in a region of thunderstorm and if it conducts thunderbolts, then the generated electricity will be passed to the ground and also small amount of that electricity will be absorbed and stored. This will prevent the damaging of the suit.

2 Software Requirements

Till now it was all about hardware. Focusing on software part, the suit will be accompanied by a human-like.

Artificial intelligence will be capable of taking an important crucial decision in adverse condition. Here, we are using several programming languages such as Python and C++ to make human-like AI. Python is an interpreted high-level programming language for general-purpose programming. It is an open-source programming language. C++ is a general-purpose programming language. It has imperative, object-oriented and generic programming features, which also offers the facilities for low-level memory manipulation. We will also be using cloud storage to store the information collected by the suit. We will also be using GPU's for pattern recognition and image processing. Speech recognition will be used for voice activation and recognition.

2.1 *Cloud System*

The cloud services can be utilised in terms of storage and access the data from anywhere [1]. Cloud services are Internet service-oriented computing concept, where hardware, software and data are shared [2]. Here, we will be using Fujitsu server for storing of the data.

Fujitsu server system prototype we are using in the cloud because of high geared speed and flexibility measured so far in simultaneous application. This has been tested and proved that the hardware prototype performs four times more than non-disk pool system. System performance is measured in every picosecond [3].

The next-generation server is using Resource Pool Architecture pooling or arranging of μ -processors; hard disk drives (HDDs) are done for high performance, high utilisation and serviceability.

Fujitsu cloud services provide a fully flexible model for IT infrastructure, platforms and applications, allowing companies to match technology systems and costs directly to changing business needs. This will reduce the usage of many devices to store the information. Also, we can retrieve it at any point of time from any part of the world. The multi-workflow [4] facility of cloud computing will make it faster than any other. The system is inspired by saving of soldiers' life as much as we can [5].

2.2 *Artificial Intelligence*

For this suit, artificial intelligence plays extremely important role in the functioning of the suit. The AI will possess various characteristics such as self-learning, various machine ethical standard [2, 6–8], capable of taking on spot decision, time-to-time calibration of the suit, automatic triggering of the healers. This AI with the help of various trained neural networks will recognise the operator who is operating the suit. The AI will also have fine image processing capability, i.e. clarity of the visual to recognise correctly [9]. Also with the help of various sensors, the AI can make certain assumptions and present it to the user which will be beneficial for any successful execution of the plan. The large amount of data will be handled through machine learning with the help of Python software. The AI will be capable of extracting and transmitting data from the cloud server by its own and also through deep learning can analyse them. It will also be capable of overriding another system by an on-board Internet connection. Regarding the security of the AI if anyone tries for security breach, the AI will alert the operator and will block itself and prevent it for further breach by isolating itself from all the networks.

For the field of medical science, the AI will have minor difference in its neural network. It will be trained according to the patients need. Since this will be self-learning AI, it can adjust its configuration according to the user's needs. Also by gathering data from day-to-day own experience, it will analyse those data and will recommend better solution to their problems. It will also consist of reminders like

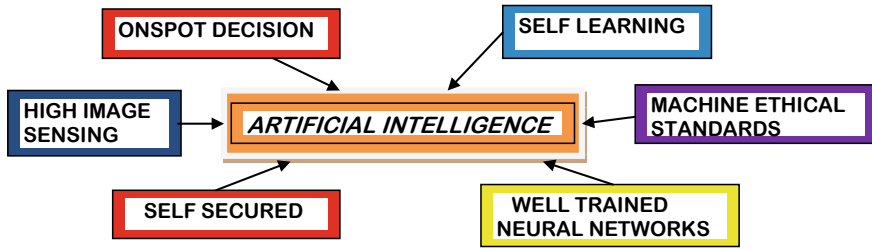


Fig. 1 Characteristic of the desired artificial intelligence

physiotherapy exercise, the movements, so that if the patient forgets that move, it will remind him/her. Also by some special set-ups, it can also remind the amount of water present in the body and update the patient from time-to-time to take water.

For defence purpose, the AI will play a key role in the plans successful execution. From accessing the satellites image and determining the position of the enemy to the successful making of the on-field plan, it can also determine the range of the enemy, and with the help of certain geographical calculations, it can tell the operator how to proceed to the targeted place. The war drill experience will also help the AI as it will gather the required information to be executed at the time of real war.

The components of the AI will be branched; that is, each module of it (machine learning, data science and analytics, deep learning, various neural and back-propagation neural networks) will be made separately and then it will be clustered together to form the desired AI with all the specified characteristics (Fig. 1).

3 Materials to Be Used and Their Specification

In the outer layer, we are using titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260) alloy. It is a near beta-alloy. We have chosen this material as it fits into our criteria, neodymium magnets for the generation of the electricity and SiO₂ for the insulation of the nanochips and to make it shock resistance. Figures 2 and 3 are given to describe its chemical composition and its properties.

3.1 Titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260)

3.1.1 Chemical Composition

Titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260) [10] comprises of various elements. Oxygen is present about 0.15 wt%, nitrogen 0.04 wt%, carbon

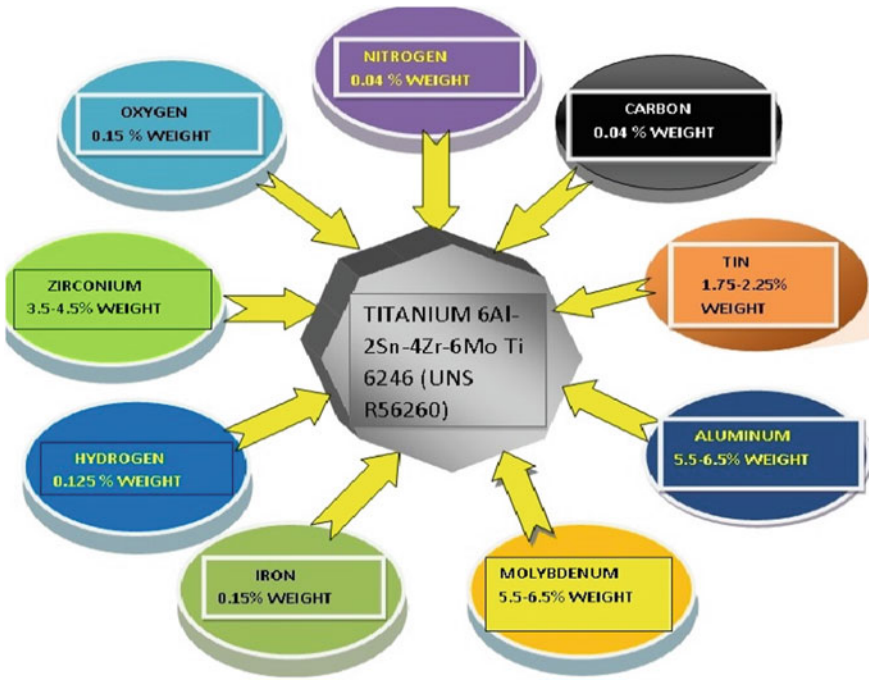


Fig. 2 Chemical composition of titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260)

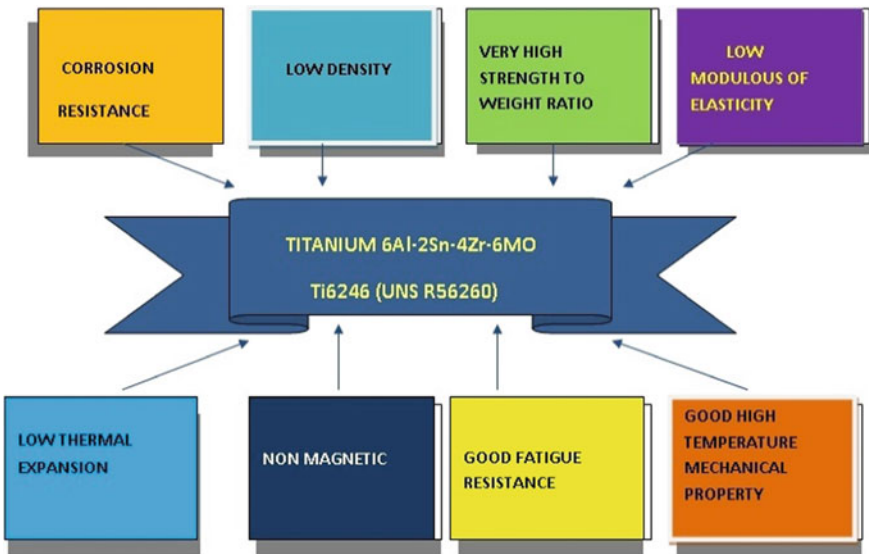


Fig. 3 Properties of titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260)

0.04 wt%, tin (sn) (1.75–2.5) wt%, aluminium (Al) (5.5–6.5) wt%, molybdenum (mo) (5.5–6.5) wt%, iron 0.15 wt%, hydrogen 0.125 wt% and zirconium (3.5–0.5) wt%.

Other residues are also been added which can be neglected. As it is present in very small quantity and will not affect the original composition of this metal.

3.1.2 Properties of Titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260)

We are choosing this element due to following reasons.

Titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260) [10–13] is corrosion resistance; that is, it will not get damaged easily by natural cause. It has very low density. It has very high strength to weight ratio; that is, packing efficiency of atoms is more. It has less weight and high tensile strength. It also has low modulus of elasticity; that is, it can recover any deformation under load. It also has low thermal expansion; that is, it will not get deformed by high temp. It is non-magnetic in nature; that is, magnetic flux will not put any effect on the metal.

Figure 3 shows the property of titanium 6Al-2Sn-4Zr-6Mo Ti 6246 (UNS R56260) [10]. It has good fatigue resistance; that is, it will not be damaged by large loads. It also has good high-temp mechanical property.

3.2 Neodymium Magnet

3.2.1 Chemical Composition

Neodymium magnets can be used to powering up the suit [14]. It comprises of various elements.

Figure 4 shows the chemical composition of neodymium magnets. It comprises of iron, boron and neodymium.

3.2.2 Properties of Neodymium Magnet

We have chosen this element like neodymium magnet because it is less expensive, it has very high magnetic strength, its magnetisation and de-magnetisation can be controlled. Neo-magnets are graded by two major criteria: magnetic value and resistance to temperature. Grades and properties of neodymium magnets are important to their performance (Fig. 5).

3.3 Silicon Dioxide (SiO_2)

3.3.1 Chemical Composition

Silicone dioxide (SiO_2) is a combination of silicon and oxygen atoms. It is a stable compound.

It appears naturally in several crystalline forms, one of which is quartz.

3.3.2 Properties of SiO_2

SiO_2 is a good insulation compound. It is grown on silicon wafers. It can be easily deposited on various metals. It is resistant to many chemicals. It can be used as a blocking material for ion implantation. It also has high dielectric strength and also has high-temperature stability up to $1600\text{ }^\circ\text{C}$ (Fig. 6).

3.4 Self-healing Polymers/Composites

Self-healing of the suit is our biggest achievement. We are using polydimethylsiloxane (PDMS) [15]. It is an extrinsic healer. Its healing approach is in the form of microcapsules and has the healing efficiency of more than 100%. As soon as it finds the crack, it will automatically reach the crack site and repair it [16]. After the combination of titanium and PDMS, we may say that it acts as a metal matrix composites (MMC) [17]. This technique will reduce the time of maintenance. The PDMS matrix composite contains encapsulated PDMS resin associated with a separate cross-linker

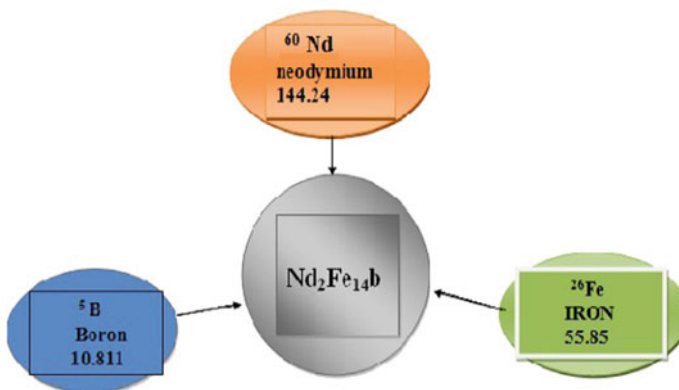


Fig. 4 Chemical composition of neodymium magnets

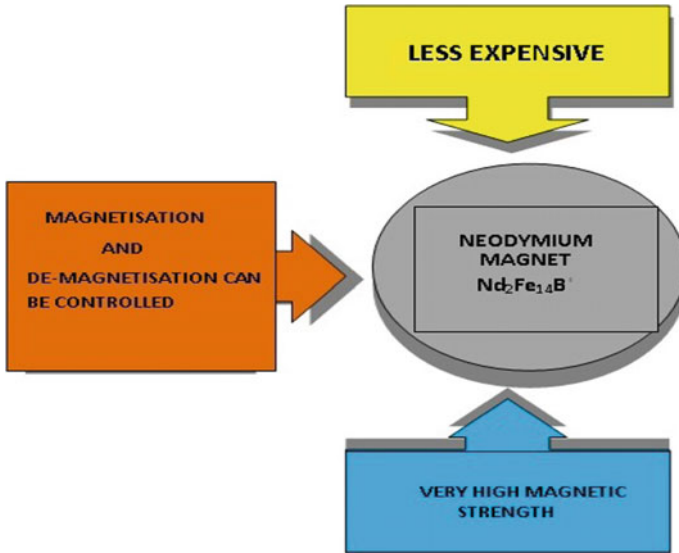


Fig. 5 Properties of neodymium magnets

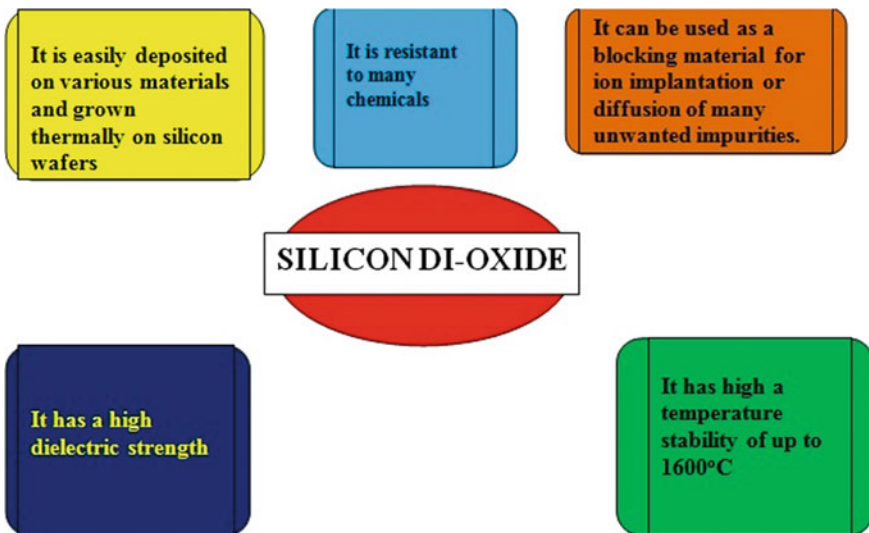


Fig. 6 Properties of silicon dioxide

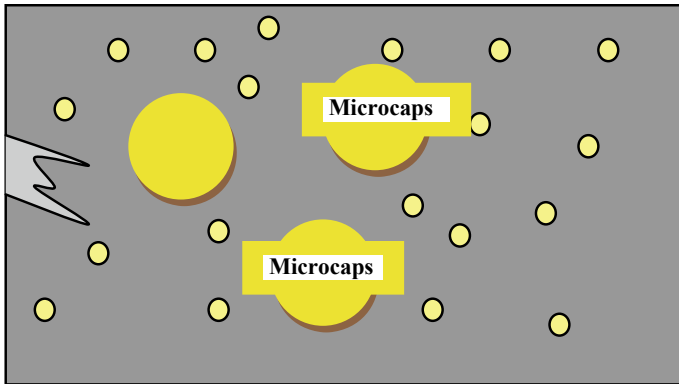


Fig. 7 Sample crack in the metal

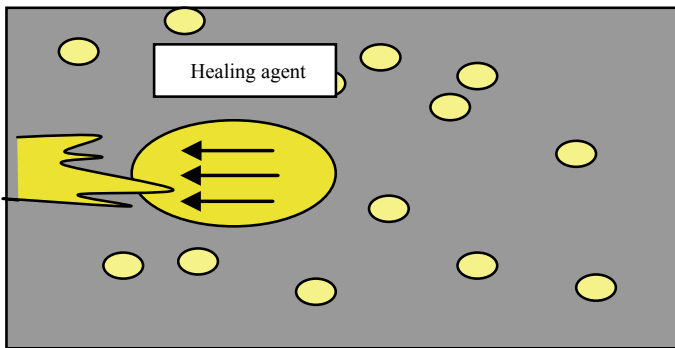


Fig. 8 Microcapsules coming into play and filling the crack site

[18]. The steps given below describe the functioning of the microcapsules self-healing agent. The self-healing microcapsules process is given below:

- Step-1 A crack has been occurred by external means in the metal having self-healing microcapsules. This crack can be intensive or extensive depending upon the external means (Fig. 7).
- Step-2 As the crack occurs in the metal, the self-healing microcapsules automatically shifts towards the crack site, and with the help of catalysts, the healing agent will be released into the crack. The time of healing will depend on the area affected by the crack (Fig. 8).
- Step-3 The healing agent heals the crack fully and maintains the stability of the metal. In this way, the metal regains its proper definite crystal structure (Fig. 9).

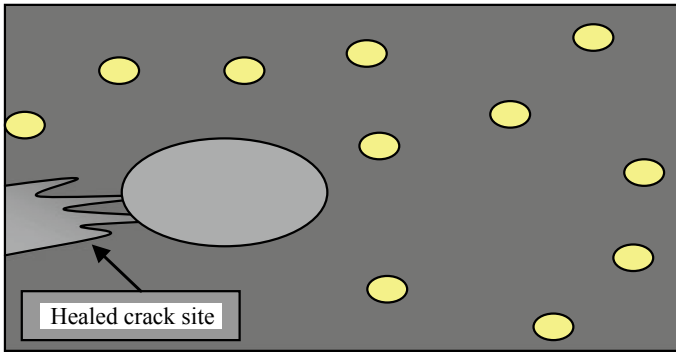


Fig. 9 Healed crack site of the metal

4 Procedure

There are three types of procedure we will discuss here as external and internal layer procedure, the middle layer procedure is in between. We have introduced another new thing which is energy production procedure.

a. External layer procedure

Getting the titanium metal in the form of alloy, it is made into thin sheets of 1.5 inch thickness [19]. The standard suit will be 6 ft. in height. After the metal is drawn into thin sheets, the self-healing polymers [14, 17, 18] will be implanted fully on the suit. In fact, we can inject the self-healing polymer atoms into the titanium sheet. After making the sheets and adding healers to it, the sheet has to be transformed into the respective dimensions of the operator, i.e. size of body parts from head to toe. Also in this layer, grounding mechanism will also be used. After it is transformed into replica according to the dimensions of the operator, we will move to the most important structural part, the middle layer. This external layer will provide covering and safety to the middle layer.

b. Middle layer procedure

This layer is the vital part of the suit. This layer will comprise of all the nanochips, gears and bearings for smooth movement of the suit. The titanium alloy will be further used to provide mechanical support to all the electrical and electronics components; that is, it will act as a chassis where these components will be embedded.

Now since our body comprises of joints, thus in this suit we will combine different parts with gears and bearings. These bearings and gears will be the size of operators joint; that is, the diameter of the gears and bearings will depend upon the size of the operators joint. The chassis of one part will be interlinked with the chassis of the other parts. Now these networks of metal will be used to embed the electrical and electronics components.

Moving on to nanochips, the printed circuit board (PCB) will be designed with the help of computerised numerical control (CNC) machines. After designing this PCB, all the required sensors and integrated circuits (IC's) will be fabricated. The PCB will comprise of voltage regulators, high kilo-ohm resistors, high energy storage capacitor, several simple IC's, several biological sensors such as retina scanner, DNA scanner, voice recognition amplifier. The PCBs containing all these sensors and IC's will be present in every part of the suit.

We will be using connectors to the power source to power up the PCBs. These power sources will be present in all the parts along with the PCBs. After the attachments of PCBs and power sources, we will put enhancers to each and every power sources. Now the generated electricity will be carried by high melting point wires. Every power source will be connected to each other to maintain stable electricity in the suit. Here, we will be using molybdenum wires for transmission of the electricity. Also we will fabricate all the nanochips with silicon dioxide because of its good insulation and anti-reflecting shocks. After the connections between PCBs and power sources are made, the connections between servo motors, mini-levers and PCBs will be connected.

c. Internal layer procedure

Once the PCBs and electrical components are embedded on the titanium chassis, the inner part (i.e. towards the body of the operator) will soon be designed. A layer of 1-inch foam will be used as insulator depending on the size of the operator. This layer will be specially designed to absorb heat as well as an emitter will be attached to this layer to drain the excess heat generated. The absorber will be attached to both electrical components as well as to the body of the operator. This emitter will be installed to the external side just opposite to the skin. A vacuum tube made of steel will be attached to this layer to drain out excessive heat through emitter.

d. Energy Production Procedure

For energy production, we are using a simple generator. We know mechanical energy can be converted into electrical energy [14, 20]. According to Faraday's law of induction, electricity can be induced through magnetic field/flux. Here, we are taking a miniature exhaust fan which is commonly used in laptops, personal computers (PCs). The cylindrical neodymium magnets are fixed to each fan blade. Another neodymium magnet is kept closer for the repulsion of the magnets fixed on the blades of fan. As the propeller will rotate, its speed will determine the flow of electricity.

If we can reduce voltage drop or losses, we can get more electromotive force (EMF) to power up the suit. Since we are using enhancers, the final voltage will be sufficient to power up the suit. The enhancers will be connected to each and every exhaust fans. Now we will combine these three layers, and each part of the suit will be joined through levers, gears and servo motors, resulting in the formation of humanoid structure of the suit.

5 Advantages of This Suit

The greatest advantage of this suit is that it will save human lives. Since it uses self-healing technology, it requires less maintenance. No toxicants are produced in the energy-making process. Portability and durability are discussed earlier which adds on to the list of advantages. There are many more advantages which will be added to its list. The suit is easily portable. It can be moved to even remote places where it has its important role. The most important function of this suit is that it can be assembled into fragmented parts and can be taken to the various places after dissemblance.

Another benefit of the suit is that it uses self-healing technology, which makes our suit more durable. At an average, the suit can last long up to 10 years without maintenance. This time frame can be improved in the near future which will help us in saving and of our natural resources.

Now the most important part of the suit is its energy production. In this suit, we are currently using magnetic fields to generate electrical power. This is a cleaner fuel to power up the suit. Moreover, it will not contain any toxicants making it an eco-friendly suit. This energy unit can produce unlimited power generation which will be beneficial for the suit.

There are also various types of exoskeleton present, but no country till date has tried this type of suit for defence or in any other field. It can boost up any country's technical as well as any other required field.

6 Disadvantage of the Suit

It requires one-time high amount to make this suit. Its cost can be reduced if the alternative of titanium is found which will be available widely and cost-effective. It causes very small amount of biological hazardness which can be cured very easily. Minor circuit problem may arise but again it can be resolved easily.

7 Datasets

7.1 *Live Records*

See Table 1.

Table 1 Live records data set to be stored in cloud

Date	Time (GMT)	Coordinates	Start time (GMT)	End time (GMT)	Duration (min)	Image id	Image	Live video footage
10-3-2014	1000	22° 25'' NE	1020	1040	20	IG-001	LV1.IMG	LV1.MPEG
25-03-2014	1600	45° 52'' 16' SNW	1605	1635	30	IG-002	LV2.IMG	LV2.MPEG
16-10-2014	0500	66° 29'' NES	0505	0515	10	IG-003	LV3.IMG	LV3.MPEG
22-12-14	2230	50° 45'' 12' WS	2245	2300	15	IG-004	LV4.IMG	LV4.MPEG
05-01-2015	1730	28° 4'' SE	1740	1750	10	IG-005	LV5.IMG	LV5.MPEG
26-02-2015	1900	21° WES	1905	1925	20	IG-006	LV6.IMG	LV6.MPEG
14-05-2015	0700	36° 25'' ENS	0710	0725	15	IG-007	LV7.IMG	LV7.MPEG

Table 2 Background dataset as and when required

Image id	Image	resolution	Past background	Current background	Last found location
IG-001	LV1.IMG	720p × 1080p	NIL	NIL	36° 22'' 45' NSE
IG-002	LV2.IMG	720p × 1080p	NIL	NIL	52° SE
IG-003	LV3.IMG	720p × 1080p	NIL	NIL	41° 21'' WS
IG-004	LV4.IMG	720p × 1080p	NIL	NIL	48° 16'' NS
IG-005	LV5.IMG	720p × 1080p	NIL	NIL	60° 14'' W
IG-006	LV6.IMG	720p × 1080p	NIL	NIL	28° SW
IG-007	LV7.IMG	720p × 1080p	NIL	NIL	33° ES

7.2 Background Record

These records are structured data set and are sorted by date. It can be retrieved by the operator as well as by the control room during any condition or situation as and when required (Table 2).

Here, we are making database through MY SQL. The software used to make this database is XAMPP. It is open-source software. By running simple queries, the data can be inserted as well as can be retrieved easily. The duplicity is reduced in this database by making some of its field as unique [21]. Generally, these records are stored in local host server, and after necessary manipulation, it will be stored on the cloud server.

8 Application in Various Fields

This suit has wide variety of uses. The utilisation starts from serving to the defence and not limited to medical fields. It can also be explored in various adventures sports. As this suit can be assembled into various parts, each part can be used in different fields. And together, it will be an overall protection of the person operating it.

Since it has more future advancement, this suit will come up to the expectations of the desired result.

For example, the suit can be used for specially challenged people. The upper section can be utilized by Mountaineers, as it has various features installed in it. The complete suit can be used for the protection of important personalities, and also for the defence. For rescue purpose also, it can be used. In this way, this suit has a wide variety of uses according to the demand of the particular field.

9 Conclusion

At the end, we can conclude that the suit will serve a great purpose to humanity. It will also be helpful in maintaining peace, eco-friendly, and can serve in any nation development. Since everything has some pros and cons, this suit also has it, but it can be resolved quickly and efficiently with the best advancement and improvisation it can have. It has its application in various fields as stated above. It will fulfil the desired criteria in any field of work.

10 Future Advancements

Since this will be the TYPE-I suit, it will have various types of future advancement with the demand of time and also with the demand of the field in which it will be utilised. If we get any type of help or assistance from any agency, it will help us to make this ideal exoskeleton of what we want to make.

References

1. S. Koley, S. Ghosh, Cloud computing with CDroid OS based on Fujitsu server for mobile technology. *Skit Res. J.* **4**(2), 1–6 (2014). ISSN 2278-2508
2. S. Koley, S. Ghosh, CDroid in Fujitsu server for mobile cloud, in *Data Analytics and Business Intelligence* (Emerging Paradigms, 2014), p 80
3. S. Koley, N. Singh, Cdroid: used in Fujitsu server for mobile cloud. *Ge-Int. J. Eng. Res.* **2**(7), (2014). ISSN 2321-1717
4. M. Adhikari, S. Koley, Cloud computing: a multi-workflow scheduling algorithm with dynamic reusability. *Arab. J. Sci. Eng.* (Springer), Journal: 13369, Article: 2739 (2017)
5. S. Koley, R. Agarwal, R. Jain (2015) Cloud computing: a new era of saving precious lives in the developing world. *Skit Res. J.* **5**(1), 1–7 (2015). ISSN 2278-2508
6. S.L. Anderson, in *Machine Metaethics*, ed. by M. Anderson (University of Connecticut, Cambridge University Press), pp. 21–27
7. S.D. Baum, Social choice ethics in artificial intelligence. *AI & Soc* 1–12 (2017). Springer, London, ISSN 0951-5666
8. D. Howard, I. Muntean, Artificial moral cognition: moral functionalism and autonomous moral agency. *Philos. Comput.* 121–159 (2017)
9. Rollin M. Omari, Masoud Mohammadian, Rule based fuzzy cognitive maps and natural language processing in machine ethics. *J. Inf. Commun. Ethics Soc.* **14**(3), 231–253 (2016)
10. A.A. Dawood, *A Study on the Sustainable Machining of Titanium Alloy* (Western Kentucky University, Masters Theses & Specialist Projects, Spring 2016), pp. 1–61
11. Detailed about this statistic/trend source, <http://www.timet.com/assets/local/documents/datasheets/alphaandbetaalloys/6246.pdf>. Last accessed Mar 2018
12. Y.M. Ahmed1, K.S.M. Sahari, M. Ishak, B.A. Khidhir, Titanium and its alloy. *Int. J. Sci. Res. (IJSR)* **3**(10), 1–11. , Department of Mechanical Engineering, University of Tenaga Nasional, Jalan IKRAM-UNITEN, 43000 Kajang, Selangor, Malaysia
13. A.A. Dawood, *A Study on the Sustainable Machining of Titanium Alloy* (Masters Theses & Specialist Projects Paper 1566, 2016), pp. 12–61

14. Prof. P.G. Shewane, A. Singh, A. Narkhede M. Gite, *An Overview of Neodymium Magnets over Normal Magnets for the Generation of Energy*, vol. 2, Issue 12 (International Journal on Recent and Innovation Trends in Computing and Communication, 2014), pp. 1–4
15. Detailed about this statistic/trend source, <https://research.wsulibs.wsu.edu/xmlui/bitstream/handle/2376/5545/52-self%20healing.pdf>. Last accessed Mar 2018
16. S.R. White, N.R. Sottos, P.H. Geubelle, J.S. Moore, M.R. Kessler, S.R. Sriram, E.N. Brown, S. Viswanathan, Autonomic healing of polymer composites. *Nature* **409**, 794–797 (2001). <https://doi.org/10.1038/35057232>
17. R. Das, C. Melchior, K.M. Karumbaiah, *Self-healing Composites for Aerospace Applications* (University of Auckland, Auckland, New Zealand; National Polytechnic Institute of Chemical Engineering and Technology (INP-ENSIACET), Toulouse, France, 2016), pp. 5–32
18. T.C. Mauldin, M.R. Kessler, *Self-Healing Polymers and Composites*, vol 55 (Institute of Materials, Minerals and Mining and ASM International, 2010), pp 1–30
19. T. Matsumuraa, S. Tamurab, Cutting simulation of titanium alloy drilling with energy analysis and FEM, in *15th CIRP Conference on Modelling of Machining Operations* (Tokyo Denki University, 5 Senjyu Asahi-cho, Adachi, Tokyo 120-8551, Japan Industrial Technology Center of Tochigi Prefecture, 1-5-20, Yuinomori, Utsunomiya, Tochigi, Japan, 2015), pp. 1–6
20. A. Manoharan, P. Suman, M. Sankara Pandian, J. Ganesan, An innovative way for power generation by using Magnet Mill. *Int. J. Theor. Appl. Res. Mech. Eng. (IJTARME)* **3**(3), 1–4 (2014). Department of Mechanical Engineering, Sree Sowdambika College of Engineering, Aruppukottai, India
21. D. Petkovic, *Temporal data in relational database systems: a comparison* (University of Applied Sciences Hochschulstr. 1, Rosenheim, 83024, Germany, Springer, Switzerland, 2016), pp. 1–12. https://doi.org/10.1007/978-3-319-31232-3_2

Crosstalk Minimization as a High-Performance Factor in Three-Layer Channel Routing



Sumanta Chakraborty

Abstract Evolution in fabrication technology has allowed the transistors, gates, and connecting wires to be placed in more close proximity in a VLSI chip. The transistors and connecting wires are being placed at more close proximity, leading to the reduction of the interconnection delay as well as crosstalk noise between the overlapping wires. Minimization of crosstalk is an important constraint in computing routing solution of a channel instance. In this paper, first we present a survey of two heuristic algorithms for producing two-layer minimum crosstalk VH channel routing solution. Then, we concentrate on crosstalk minimization for producing three-layer HVH channel routing solution of a given channel instance. Two different algorithms to produce minimum crosstalk three-layer HVH channel routing solutions have been proposed in this paper. The results computed from each of the algorithms are very promising and far better than the results computed for several two-layer routing solutions. For very small number of nets, our algorithms can produce sometimes 100% crosstalk reduction.

Keywords Crosstalk · High performance · Three layers · Channel routing · HVH · No-dogleg

1 Introduction

With the rapid scaling of the very large-scale integrated (VLSI) technology, millions of transistors can be organized on a single chip and a large number of transistors leads to a large number of connecting wires. The goal of the VLSI design is not only to increase the number of devices on a single chip, but also to make the chip size as small as possible (compaction). For the past two or three decades, fabrication technology has also been evolving very rapidly. So, day by day minimum feature sizes for VLSI chips are gradually decreasing and the transistors and connecting wires are

S. Chakraborty (✉)
Institute of Engineering and Management, Kolkata, India
e-mail: sumanta.chakraborty@iemcal.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_56

being placed at more close proximity, leading to the reduction of the interconnection delay and the routing area. Due to smaller interconnection delay and circuit switching delay, integrated circuits (ICs) are operating at higher frequencies. The closely placed devices and interconnection wires and higher frequencies of operation of VLSI chips cause coupling capacitance between the interconnection wires. These capacitive couplings induce a noise to a great extent between the wires. This noise is referred to as crosstalk. Generally, crosstalk between two wires is proportional to the coupling capacitance between them. The coupling capacitance between two parallel wires is proportional to the length of coupling which is the total length of overlapping of wires on adjacent tracks and the coupling capacitance between two orthogonal wires is considered to be negligible. When two wires are adjacent to each other, the coupling capacitance C_c between the two wires is proportional to the length L_o of the overlapping segment of wires and C_c is inversely proportional to the distance of separation D between the two wires:

$$C_c = k. (L_o / D^{1.34}) [1] \quad (1)$$

where k is a constant. So, crosstalk is inversely proportional to the distance of separation between the wires. Crosstalk is also proportional to the frequency of operation of the circuits.

Crosstalk induces signal delay in the circuits and causes a voltage spike which may result in wrong logic values at the input of logic gates. Thereby, circuit performance is degraded. So, in high-performance routing, crosstalk should be a constraint to develop good channel routing solutions. A channel routing solution with density (d_{\max}) number of tracks may not be a good routing solution in terms of crosstalk.

This paper has been organized as follows. In Sect. 2, we present a review of two heuristic algorithms proposed by Pal et al. for producing two-layer minimum crosstalk VH channel routing solution. In Sect. 3, we formulate the three-layer HVH minimum crosstalk channel routing problem. In Sect. 4, we propose three-layer HVH crosstalk minimization algorithms along with their time complexities and performances. Section 5 concludes the paper.

2 Two-Layer HVH Crosstalk Minimization Algorithms

2.1 Idea

Crosstalk minimization in no-dogleg two-layer VH channel routing is an NP-hard problem [2, 3], also for simple channel instances. As crosstalk minimization is a hard problem, different kind of approaches [4, 5] have been adopted, for example, approximation algorithm (mixed ILP formulation as presented by Gao and Liu [6]), exhaustive algorithm (as proposed by Jhang et al. [7]), heuristics algorithms (two algorithms as proposed by Pal et al. [8, 9]).

For each of the following algorithms, one thing is common. For each algorithm, the routing procedure is divided into two phases with two different objective functions in each phase. In the first phase, the objective function depends on the absolute positions of the nets and the objective function is the number of tracks. In this phase, the routing algorithm minimizes the number of tracks and produces minimum track, i.e., minimum area routing solution. In the second phase, the objective function depends on the relative positions of the nets and the objective function is the amount of crosstalk (sum crosstalk or bottleneck crosstalk). In this phase, the routing algorithm minimizes the amount of crosstalk on the minimum area routing solution. In this section, we will review Pal et al.'s heuristic algorithms for producing two-layer minimum crosstalk VH channel routing solution.

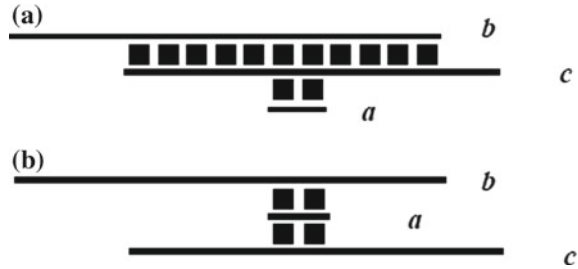
2.2 *Pal et al.'s Heuristic Algorithms*

The motivation of Pal et al.'s algorithms [8, 9] is the permutation of tracks. As the permutation of tracks in order to minimize the total amount of crosstalk is NP-hard (as proven in [6]), they chose heuristic approach. Actually, they proposed two heuristic algorithms: One is based on track permutations (track interchange algorithm), and the later is based on permutation of nets (net interchange algorithm). They proposed track interchange algorithm for both simple and general channel instances. Here, we review track interchange algorithm for general channel instances and then net interchange algorithm.

2.2.1 Track Interchange Algorithm

For general channel instances, they construct a modified vertical constraint graph (VCG) which is known as reduced vertical constraint graph (RVCG). In an RVCG, a vertex represents a track containing non-overlapping nets in the channel routing solution, S , and a directed edge between two tracks t_i and t_j indicates that track t_i is vertically constraint to track t_j . There should always be a source vertex because S is a feasible two-layer routing solution without any cyclic vertical constraints, and thus, the corresponding RVCG is also acyclic. Then in the first iteration, all the source vertices (that are freely interchangeable) in RVCG are taken. Then, the tracks are sorted in non-increasing order according to their effective intervals (EI). This is so because of the assumption that the amount of crosstalk between the nets in the first track and the fixed terminal in the top row of the channel is negligible. The EI of a track is computed by adding the net spans of all the nets assigned to the track. If two or more tracks have the same EIs, the tie is broken by arranging those tracks in ascending order according to their total intervals (TIs). The TI of a track is the interval between the leftmost terminal and the rightmost terminal of the net(s) assigned to a track. Then, the tracks are rearranged by sandwiching a net of smaller span with two nets of larger spans because the amount of crosstalk is mostly reduced in this way.

Fig. 1 Demonstration of crosstalk minimization after track permutation



So, for a set of t number of tracks $\{n_1, n_2, \dots, n_t\}$, the rearranged set is $\{n_1, n_t, n_2, n_{t-1}, n_3, n_{t-2}, \dots\}$. The rearranged tracks are assigned to the topmost $|T_1|$ number of tracks, where $|T_1|$ is the total number of source vertices in the RVCG. Then, a modified RVCG is obtained by deleting these vertices from RVCG. This iteration continues until the RVCG becomes empty. In Fig. 1 (as presented in [8]), it is shown how crosstalk can be minimized arranging (permuting) tracks in sandwiched fashion. In Fig. 1a, a feasible routing solution is shown with three intervals of three different overlapping nets a, b , and c that are in three different tracks. The horizontal overlap of nets b and c is 11 units of and that of nets c and a is 2 units. So, the total coupling length is 13 units. In Fig. 1b, another feasible routing solution with three tracks for the same channel instance, with a total coupling length of 4 units. Thus by reassigning the nets to tracks, a minimized crosstalk routing solution is obtained.

The overall time complexity of track interchange algorithm in the worst case is $O(n^2)$, where n is the total number of nets present in the channel.

2.2.2 Net Interchange Algorithm

In this heuristics, Pal et al. constructed a sequence of interchangeable nets and considered the interchangeable nets one after another. For each net, they searched for a track where the net can be fitted best leading to the overall crosstalk minimization with no vertical constraint violation. Since there is an exponential number of interchangeable sequences, they used a constant number of sequences which are computed by sorting of the interchangeable nets from top left to bottom right or from bottom left to top right in the routing solution S with respect to their starting column positions in the channel or by sorting interchangeable blank spaces from top left to bottom right or from bottom left to top right in S with respect to the starting column positions of the nets in the channel. This heuristic may also be repeatedly followed for the constant number of sequences if crosstalk is reduced in each case.

The overall time complexity of net interchange algorithm in the worst case is $O(n^2)$ where n is the total number of nets present in the channel.

Table 1 Performance of Pal et al.'s algorithms for two-layer general channel instances [9]

Number of nets	Average initial crosstalk	Crosstalk reduction after track interchange (%)	Crosstalk reduction after net interchange (%)
20	135	7.41	12.59
40	564	6.03	12.41
60	1363	5.28	11.81
80	2475	5.17	11.80
100	3997	4.73	11.58
150	9042	4.60	11.21
200	16,260	4.69	11.47
250	25,293	4.42	11.07
300	36,612	4.39	11.02
350	49,708	4.34	11.02
400	65,032	4.41	11.29
500	102,922	4.47	11.07
1000	412,991	4.77	11.37
2000	1,665,471	5.14	11.81

2.2.3 Performances

The results computed using Pal et al.'s algorithm are shown in Table 1.

3 Formulation of the Three-Layer HVH Minimum Crosstalk Channel Routing Problem

In three-layer channel routing model, there are two horizontal layers H1 and H2, respectively. That is, one more horizontal layer is present in the three-layer routing model. So, there is a high chance that crosstalk minimization in no-dogleg three-layer HVH channel routing [10] will also be a hard problem. In this section, we will present two different heuristics to compute three-layer HVH channel routing solutions with reduced crosstalk from given channel instances and corresponding two-layer VH minimum area routing solutions. Given a two-layer VH channel routing solution of ' t ' number of tracks or by computing a feasible two-layer VH channel routing solution of ' t ' number of tracks of a given channel instance, the first objective is to compute the total amount of crosstalk of the routing solution. Computing a two-layer VH routing solution of a given channel instance can be performed by using Track_Assignment_Heuristics (TAH) algorithm, Left_Edge_Algorithm (LEA), Constraint_Left_Edge_Algorithm (CLEA), Maximum_Clique_Cover_1 (MCC1) algorithm, or Maximum_Clique_Cover_2 (MCC2)

algorithm. Then, the second and prime objective is to compute a three-layer HVH channel routing solution for the given channel instance so that the total amount of crosstalk is further minimized or reduced. In this paper, we consider that crosstalk due to the non-adjacent and different layers are almost negligible and crosstalk due to the adjacent and the same horizontal layers are considered. Like the two-layer VH minimum crosstalk channel routing problem, we have not considered here the crosstalk due to the vertical wire segments of two nets assigned to adjacent tracks.

4 Three-Layer HVH Crosstalk Minimization Algorithms

4.1 Idea

Given a channel instance, C , and a two-layer minimum area routing solution, S , we can compute a three-layer routing solution, S^* by assigning only one set of non-overlapping nets to any of the two horizontal layers of a track and then the next set of non-overlapping nets to an alternate layer on the next track. While devising the algorithms, we have an important observation that by assigning the sets of non-overlapping nets to the alternate layers on the adjacent tracks, cent percent reduction in crosstalk can be achieved. Since crosstalk due to the non-adjacent and different layers is almost negligible and crosstalk due to the adjacent and the same horizontal layers is considered, therefore, in S^* crosstalk will be reduced to 100%. But area minimization problem is an important cost optimization problem in the physical design phase of VLSI design cycle and the lower bound on the number of tracks in the three-layer HVH routing solution is the maximum of $d_{\max}/2$ and v_{\max} where d_{\max} is the channel density and v_{\max} is the channel height, i.e., the longest path in the vertical constraint graph (VCG). Therefore, we assign at most two sets of non-overlapping nets to the two horizontal layers on the same track. We construct the reduced vertical constraint graph, RVCG. Thus, we assign at most two vertices of the RVCG to the H1 layer and the H2 layer on the same track.

4.2 Layer Assignment Algorithm Version One

Given a channel instance, C , and a two-layer minimum area routing solution, S , we have computed a three-layer minimum crosstalk routing solution, S^* from the reduced vertical constraint graph (RVCG) taking source vertices in descending order of their effective intervals (EI), if tie, breaking it in descending order of their total intervals (TIs).

4.2.1 Algorithm

Algorithm I. Layer Assignment Algorithm Version One

```

1: procedure LayerAssignmentAlgorithm_ver1 (C, S)
2:   Compute EI and TI of each track in the two-layer routing solution.
3:   Compute the total amount of crosstalk for the two-layer routing solution.
4:   Construct the RVCG where a vertex represents a track containing non-overlapping nets in S and a
   directed edge between two tracks  $t_i$  and  $t_j$  indicates that track  $t_i$  is vertically constraint to track  $t_j$ .

5:   while RVCG  $\neq$  NULL do // RVCG = NULL indicates that all the vertices in the RVCG are assigned to
   either H1 layer or H2 layer

6:     Take the source vertices from the RVCG.
7:     Sort them in descending order of their EI s (if tie, break it in descending order of their corresponding TIs).
8:     if iteration == 1 then // First iteration
9:       if There is one source vertex in RVCG then
10:        Assign it to any of H1 layer or H2 layer on the first track.
11:       else:
12:        Take the first two source vertices.
13:        Assign them to H1 layer and H2 layer, respectively, on the first track.
14:       end if
15:     end if
16:     while All the source vertices are not assigned do
17:       if There is only one new source vertex and only one source vertex on the previous track then
18:        Assign the new source vertex to a layer alternate to the layer to which the previous vertex is assigned.
19:       else
20:        Take at most two new source vertices from the RVCG.
21:        Compute the amount of crosstalk due to each of these two on the already assigned vertex (or on each of
   the already assigned vertices) on the previous track.
22:        Store all the computed crosstalks. //At most four crosstalks
23:        if Two new vertices are taken from the RVCG and there are two vertices assigned on the previous track
   and values of more than one computed crosstalks are the same as the maximum crosstalk then
24:          Take the next maximum crosstalk.
25:          Assign the new vertex to a layer alternate to the layer of any of the previously assigned vertices on
   which the computed crosstalk due to the new vertex is the next maximum.
26:          Assign the other new vertex to an alternate layer on the same track.
27:        else
28:          Take the new source vertex due to which crosstalk on a previously assigned vertex is maximum.
29:          Assign the new source vertex on the next track to a layer alternate to the layer to which the previous
   vertex is assigned.
30:          Assign the other new source vertex to an alternate layer on the same track.
31:        end if
32:       end if
33:       Proceed to the next source vertex or vertices.
34:     end while
35:     Delete all the source vertices from the RVCG.
36:   end while
37:   Compute the total amount of crosstalk for the final three-layer routing solution.
38: end procedure

```


4.2.2 Computational Complexity

At the line 2, computing EIs and TIs of all the tracks requires $O(n^3)$ where ‘ n ’ is the number of nets in the channel instance. At the line 3, the initial crosstalk can be computed in $O(n)$ time. At the line 4, the adjacency matrix of the RVCG of ‘ t ’ vertices can be constructed in $O(t^2)$ time. At the line 5 since the RVCG is not null, therefore, iteration will start. At the line 6, finding the source vertices requires $O(t^2) = O(n^2)$. At the line 7, sorting can be done in $O(n \log n)$ time. The total computational complexity of the lines 8–15 is $O(1)$. The total computational complexity of the lines 16–34 is $O(n)$ because for each source vertex or a pair of two source vertices, crosstalk is to be computed and computational complexity of computing crosstalk is $O(1)$. The line 35 requires $O(n^2)$. The iteration ends. The iteration continues until the RVCG is not null. If there is ‘ m ’ number of iterations, the total computational complexity of the lines 5–36 is $O(mn^2)$. At the line 37, the final crosstalk can be computed in $O(n)$ time. So, the algorithm requires $O(n^3)$ time.

4.2.3 Example

Given channel instance is as follows

```

Top :    0 6 4 4 0 0 8 10 5 0 8 0 4 9 7 5 10 0 8 6 2 0 0 2
Bottom : 9 0 1 5 0 10 1 3 1 6 7 0 7 5 0 5 5 6 8 4 3 0 7 0
    
```

Given two-layer routing solution of the channel instance is Fig. 2.

The reduced vertical constraint graph (RVCG) is in Fig. 3.

So, the three-layer routing solution computed using the algorithm is in Fig. 4. The amount of crosstalk for the given routing solution in Fig. 2 is 91 units. The final crosstalk in the computed three-layer routing solution in Fig. 4 is 54 units.

Fig. 2 Channel routing solution (two-layer) for the given channel instance

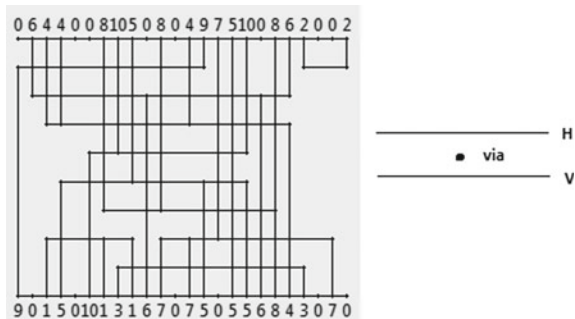


Fig. 3 RVCG of the two-layer routing solution

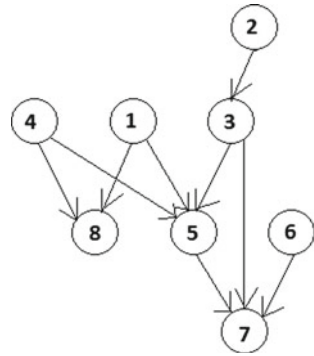


Fig. 4 Three-layer routing solution obtained using layer assignment algorithm version one

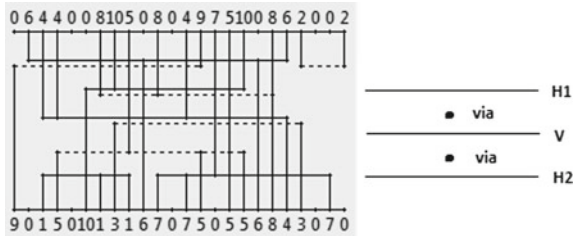


Table 2 Performance of the layer assignment algorithm version one for benchmark channel instances

Channel instance used	Crosstalk in the initial two layers	Crosstalk in the final three layers	Reduction in crosstalk (%)
RKPC1	18	5	72.22
RKPC2	11	5	54.55
RKPC3	15	7	53.33
RKPC4	14	5	64.29
RKPC5	16	7	56.25

4.2.4 Performances

Results computed using layer assignment algorithm version one are shown in Tables 2 and 3. Table 2 contains the results obtained for different benchmark channel instances. Table 3 contains the results obtained for random general channel instances. Hence, for the number of nets in each row 200 random instances have been generated and the data in each row are computed by taking average on all the 200 executed data.

Table 3 Performance of the layer assignment algorithm version one for random general channel instances

Number of nets	Average initial crosstalk in two layers	Best crosstalk reduction in three layers (%)	Average crosstalk reduction in three layers (%)	Time elapsed (s)
5	31	100	70.28	1.22
10	122	100	54.34	1.42
20	482	81.17	45.33	2.40
30	1070	74.70	39.51	3.87
40	1901	74.02	36.81	6.26
50	2940	59.70	33.65	9.27
60	4253	62.69	33.45	13.46
70	5779	60.26	31.54	19.19
80	7671	63.17	30.94	27.94
90	9856	51.72	29.48	36.78
100	12,081	56.09	29.29	47.63
150	27,060	44.08	27.10	143.49
200	48,306	43.07	24.99	325.03

4.3 Layer Assignment Algorithm Version Two

Given a channel instance, C , and a two-layer minimum area routing solution, S , we have computed a three-layer minimum crosstalk routing solution, S^* from the reduced vertical constraint graph (RVCG) taking source vertices in descending order of their effective intervals (EI), if tie, breaking it in descending order of their total intervals (TI), and then sandwiching the sorted source vertices in a fashion such that for a set of t number of tracks $\{n_1, n_2, \dots, n_t\}$, the reassigned set is $\{n_1, n_t, n_2, n_{t-1}, n_3, n_{t-2}, \dots\}$. Another reassigned sequence can be $\{n_1, n_t, n_3, n_{t-2}, \dots, n_{t-1}, n_2\}$.

4.3.1 Algorithm

Algorithm II. Layer Assignment Algorithm Version Two

```

1: procedure LayerAssignmentAlgorithm_ver2 (C, S)
2:   Compute EI and TI of each track in the two-layer routing solution.
3:   Compute the total amount of crosstalk for the two-layer routing solution.
4:   Construct the RVCG where a vertex represents a track containing non-overlapping nets in S and a
   directed edge between two tracks  $t_i$  and  $t_j$  indicates that track  $t_i$  is vertically constraint to track  $t_j$ .

5:   while RVCG  $\neq$  NULL do // RVCG = NULL indicates that all the vertices in the RVCG are assigned to
   either H1 layer or H2 layer

6:     Take the source vertices from the RVCG.
7:     Sort them in descending order of their EIs (if tie, break it in descending order of their corresponding TIs).
8:     Sandwich the sorted source vertices such that for a set of  $t$  number of tracks  $\{n_1, n_2, \dots, n_t\}$ , the
   rearranged set is  $\{n_1, n_t, n_2, n_{t-1}, n_3, n_{t-2}, \dots\}$ 
9:     if iteration == 1 then // First iteration
10:      if There is one source vertex in RVCG then
11:        Assign it to any of H1 layer or H2 layer on the first track.
12:      else:
13:        Take the first two source vertices.
14:        Assign them to H1 layer and H2 layer, respectively, on the first track.
15:      end if
16:    end if
17:    while All the source vertices are not assigned do
18:      if There is only one new source vertex and only one source vertex on the previous track then
19:        Assign the new source vertex to a layer alternate to the layer to which the previous vertex is assigned.
20:      else
21:        Take at most two new source vertices from the RVCG.
22:        Compute the amount of crosstalk due to each of these two on the already assigned vertex (or on each
   of
23:        the already assigned vertices) on the previous track.
24:        Store all the computed crosstalks. //At most four crosstalks
25:        if Two new vertices are taken from the RVCG and there are two vertices assigned on the previous
   track
26:        and values of more than one computed crosstalks are the same as the maximum crosstalk then
27:          Take the next maximum crosstalk.
28:          Assign the new vertex to a layer alternate to the layer of any of the previously assigned vertices on
   which the computed crosstalk due to the new vertex is the next maximum.
29:          Assign the other new vertex to an alternate layer on the same track.
30:        else
31:          Take the new source vertex due to which crosstalk on a previously assigned vertex is maximum.
32:          Assign the new source vertex on the next track to a layer alternate to the layer to which the previous
   vertex is assigned.
33:          Assign the other new source vertex to an alternate layer on the same track.
34:        end if
35:      end if
36:      Proceed to the next source vertex or vertices.
37:    end while
38:    Delete all the source vertices from the RVCG.
39:  end while
40:  Compute the total amount of crosstalk for the final three-layer routing solution.
41: end procedure

```

4.3.2 Computational Complexity

At the line 2, computing EIs and TIs of all the tracks requires $O(n^3)$ where n is the number of nets in the channel instance. At the line 3, the initial crosstalk can be computed in $O(n)$ time. At the line 4, the adjacency matrix of the RVCG of t vertices can be constructed in $O(t^2)$ time. At the line 5 since the RVCG is not null, therefore, iteration will start. At the line 6, finding the source vertices requires $O(t^2) = O(n^2)$. At the line 7, sorting can be done in $O(n \log n)$ time. At the line 8, sandwiching can be done in $O(n)$ time. The total computational complexity of the lines 9–16 is $O(1)$. The

Fig. 5 Three-layer routing solution obtained using layer assignment algorithm version two

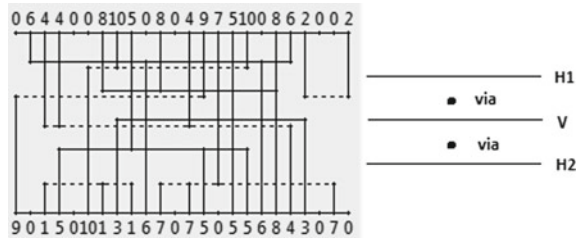


Table 4 Performance of the layer assignment algorithm version two for benchmark channel instances

Channel instance used	Crosstalk in the initial two layers	Crosstalk in the final three layers	Reduction in crosstalk (%)
RKPC1	18	5	72.22
RKPC2	11	5	54.55
RKPC3	15	7	53.33
RKPC4	14	5	64.29
RKPC5	16	7	56.25

total computational complexity of the lines 17–35 is $O(n)$ because for each source vertex or a pair of two source vertices crosstalk is to be computed and computational complexity of computing crosstalk is $O(1)$. The line 36 requires $O(n^2)$. The iteration ends. The iteration continues until the RVCG is not null. If there is m number of iterations, the total computational complexity of the lines 5–37 is $O(mn^2)$. At the line 38, the final crosstalk can be computed in $O(n)$ time. So, the algorithm requires $O(n^3)$ time.

4.3.3 Example

Given the previous general channel instance and the previous two-layer routing solution in Fig. 2 the three-layer routing solution obtained using the algorithm is in Fig. 5.

The amount of crosstalk for the given routing solution in Fig. 2 is 91 units. The final crosstalk in the computed three-layer routing solution in Fig. 5 is 56 units.

4.3.4 Performances

Results computed using layer assignment algorithm version one are shown in Tables 4 and 5. Table 4 contains the results obtained for different benchmark channel instances. Table 5 contains the results obtained for random general channel instances. Hence, for the number of nets in each row 200 random instances have been generated and the data in each row are computed by taking average on all the 200 executed data.

Table 5 Performance of the layer assignment algorithm version two for random general channel instances

Number of nets	Average initial crosstalk in two layers	Best crosstalk reduction in three layers (%)	Average crosstalk reduction in three layers (%)	Time elapsed (s)
5	30	100	70.27	1.14
10	120	100	57.07	1.23
20	475	90.12	44.25	2.12
30	1089	79.04	39.76	3.88
40	1926	66.34	35.74	5.77
50	3010	63.21	34.10	8.91
60	4260	53.56	33.82	13.52
70	5793	56.46	33.25	19.30
80	7774	53.37	31.38	26.99
90	9709	67.93	30.84	36.38
100	11,952	53.40	30.44	48.16
150	26,790	48.47	28.26	145.41
200	47,874	48.83	26.97	475.11

5 Conclusion

In this paper, we have presented and implemented two different algorithms for computing minimum crosstalk three-layer routing solutions. These algorithms have been implemented using Python 2.7 on Windows 7 64-bit platform with 2 GB RAM. We have two important observations. For each of the two versions of the layer assignment algorithms, the best reduction in crosstalk for the computed three-layer routing solutions for very small number of nets is up to 100%. For each of the four versions of the layer assignment algorithms, the average reduction in crosstalk in the computed three-layer routing solutions decreases with the increase in the number of nets. But still performance of the two algorithms is far better than the performances of other two-layer routing algorithms. When the number of nets is 200, average crosstalk reduction is twice than the average crosstalk reduction with Pal et al.'s algorithm [9].

In this paper, we have proposed crosstalk minimization algorithms for the three-layer no-dogleg routing solutions. The objective of the algorithms is to reduce the sum crosstalk. There can be several probable future works:

1. Devising crosstalk minimizing algorithms for the multi-layer channel routing solutions can be a possible work. Minimum crosstalk multi-layer channel routing problem can be a scope of research in the future.
2. Minimizing bottleneck crosstalk, the other variant of crosstalk, for the three-layer routing solution can be another possible work.

3. Three-layer minimum crosstalk routing solution can be designed with considering area minimization constraint. Since routing solution with minimum area may not lead to minimum crosstalk routing solution, therefore, whether it is possible to compute a minimum area and minimum crosstalk three-layer routing solution is a very interesting scope of research in the future.
4. To compute, a minimum wire length and minimum crosstalk three-layer routing solution is a very interesting scope of research in the future.

Acknowledgements I would like to thank Institute of Engineering and Management, Salt Lake, Kolkata, West Bengal, for giving me a very good opportunity and encouragement to present this paper.

References

1. T. Sakurai, K. Tamaru, Simple formulas for two- and three- dimensional capacitance. *IEEE Trans. Electron Devices* **30**(2), 183–185 (1983)
2. R.K. Pal, Approximate and bottleneck high performance routing for green electronics and computing VLSI Circuits, in *XTalk for VLSI D 2011*
3. T.T. Ho, S.S. Iyengar, S.-Q. Zheng, A general greedy channel routing algorithm. *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.* **10**, 204–211 (1991)
4. T. Yoshimura, E.S. Kuh, Efficient algorithms for channel routing. *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.* **1**(1), 25–35 (1982)
5. H. Manem, G.S. Rose, A crosstalk minimization technique for sub-lithographic programmable logic arrays, in *Proceedings of 9th IEEE Conference on Nanotechnology* (2009), pp. 218–221
6. T. Gao, C.L. Liu, Minimum crosstalk channel routing. *IEEE Trans. Comput.-Aided Des. Integr. Circ. Syst.* **15**(5), 465–474 (1996)
7. K.-S. Jhang, S. Ha, C.S. Jhon, Simulated annealing approach to crosstalk minimization in gridded channel routing. 16 Aug 1995
8. A. Pal, T.N. Mandal, A. Khan, R.K. Pal, A.K. Datta, A. Chaudhuri, Two algorithms for minimizing crosstalk in two-layer channel routing. *IJETTCS* **3**(6), 194–204 (2014)
9. A. Pal, D. Kundu, A.K. Datta, T.N. Mandal, R.K. Pal, Algorithms for reducing crosstalk in two-layer channel routing. *JPS* **6**, 167–177 (2006)
10. G.A. Schaper, Multi layer channel routing, Ph.D. thesis, Department of Computer Science, University of Central Florida, Orlando, 1989

A Human Intention Detector—An Application of Sentiment Analysis



Megha Dutta, Shayan Mondal, Sanjay Chakraborty
and Arpan Chakraborty

Abstract In recent past, our society has witnessed different types of criminal and unpleasant activities with the common people, and in most of the cases, women are being victimized. The women are getting harassed or threatened in different places. So there should be some means by which we can provide security or safety to women. Being responsible citizens of India, we feel there is an urgent need to take major steps to protect the ladies as well as the moral values of our society. These types of incidents motivated us to develop a real-time safety device for human beings, especially for women. It has been observed in the study that when a person is under threatening his/her body temperature changes. Taking this into account and with the help of sentiment analysis, we have developed a simple safety device which can be integrated with smartphones. Our device predicts the intention of a person by analysing the speech and checks the body temperature of the victim and combines these two facts together to detect the severity level of threatening. In this paper, we are going to describe our proposed approach to detect the threats. We shall also provide the experimental result analysis and accuracy.

Keywords Sentimental analysis · Temperature · Mood detection · Speech recognition · Intention · Human safety · AFINN list

M. Dutta (✉) · S. Mondal · A. Chakraborty
Institute of Engineering & Management, Kolkata, India
e-mail: meghadutta201@gmail.com

S. Mondal
e-mail: shayanmondal237@gmail.com

A. Chakraborty
e-mail: arpan.chakraborty0007@gmail.com

S. Chakraborty
Techno India College of Technology, Salt Lake, Kolkata, India
e-mail: schakraborty770@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_57

1 Introduction

Technological advancement has brought us to the world where everything seems to be possible. The ultimate use of any resource is for the development and well-being of the society. Therefore, the main objective of our project is to develop a system which can be used for various safety purposes in all fields of security. It is very obvious that the intention of a person becomes clear by the words he/she uses while speaking. With the help of an AFINN list, we can make almost accurate classification of positive and negative words. The positive words signify good attitude, while the negative words suggest that the speaker has a bad intention. This is the first alarming point. Negative intentions are a sign of possible future crimes. But again, this alone is not enough to assure that the situation is alarming. Here comes the second aspect of this project. It scientifically measures. Combining both these factors, our project can sense the possibilities of crime in a situation. Initially, the thought was women safety against such heinous crimes occurring in the society, but later we realized that it could be helpful for others as well.

1.1 Background Study

Body temperature is a complex clinical variable. This is among the few factors which can be measured accurately and quantitatively analysed. The normal body temperature is defined as 37 °C [1]. However, the temperature of a body keeps changing throughout the day in response to various factors like individual's metabolism rate, room temperature and place. It even varies for different parts of the body. There are five major clinical practices to measure the body temperature.

- Rectal
- Oral
- Axillary
- Forehead
- Ear.

Most convenient practice is measuring the oral site, while the standard one is rectal temperature. Sentimental analysis refers to the identification [2], extraction and study of affective states. It is also referred to as opinion mining. It has proved to be very useful in the commercial field, especially to study customer surveys and feedbacks. Sentimental analysis has three major approaches:

- Knowledge-based methods
- Statistical methods
- Hybrid methods.

In knowledge-based methods, texts are classified into obvious affect words and assign them a probable affinity to particular emotions. This list is referred to as AFINN list. Statistical methods take the help of machine learning, while hybrid methods are a combination of both. The most basic task of opinion analysis is to determine the polarity of a text message. It can be either positive, negative and neutral.

It is observed that the attitude or emotional response of the speaker can be evaluated or senses by quantifying the overall affinity of his message with respect to any given topic.

2 Theoretical Concepts

AFINN [3] is a list of English words rated for valence with an integer between minus five (negative) and plus five (positive). The words have been manually labelled by Finn Årup Nielsen in 2009–2011.

- Very negative (rating -5 or -4)
- Negative (rating -3 , -2 , or -1)
- Positive (rating 1 , 2 , or 3)
- Very positive (rating 4 or 5).

Google has a great speech recognition API. This API converts spoken text (microphone) into written text (Python strings), briefly speech to text. Speech can be entered in a microphone and Google API will translate this into written text. The API has excellent results for English language. The audio is recorded using the speech recognition module, and the module will include on top of the programme. Secondly, the recorded speech is sent to the Google speech recognition API which will then return the output.

r.recognize_google (audio) returns a string.

3 Proposed Work

The project aims at creating an application for human safety purpose. The proposed idea is to study the intention of a person based on different factors like speech and body temperature. It is a common observation that the body temperature of human being varies according to his mood. When a person is threatened, his body temperature lies within a certain range, while when he is happy, his temperature lies in a different range. This key factor has been taken into consideration in generating a result (positive or negative) and combined with the sentimental analysis of the speech of the third person who is trying to threaten him/her.

The proposed work can be divided into two modules:

1. **Temperature Module**

It is assumed that the data input for temperature has already been recorded through some hardware components. It is observed that if a person is threatened, there can be two cases, either the person becomes 'angry' or he becomes 'frightened'. Presently, the experiment has been conducted only with the first case and the data range for the mood 'angry' has been recorded.

2. **Speech Module**

Speech is taken as input with the help of a microphone. The Google speech recognition API has noise cancellation to some extent. This clear voice is recorded in an audio file while it is in turn converted to text with the help of Google speech to text conversion API. The text file is processed with the help of various Python libraries [2]. The stop words are removed and keywords are extracted. These keywords are run against the AFINN list to find the respective scores thus giving the sentiment of the input speech.

The result from both these modules is combined and if body temperature lies within the given range and the sentiment is negative, then some threat is suspected and emergency contact shall be informed with the help of phone's GPS.

Algorithm:

Input

temp: Body temperature of the victim

speech: Speech of the object

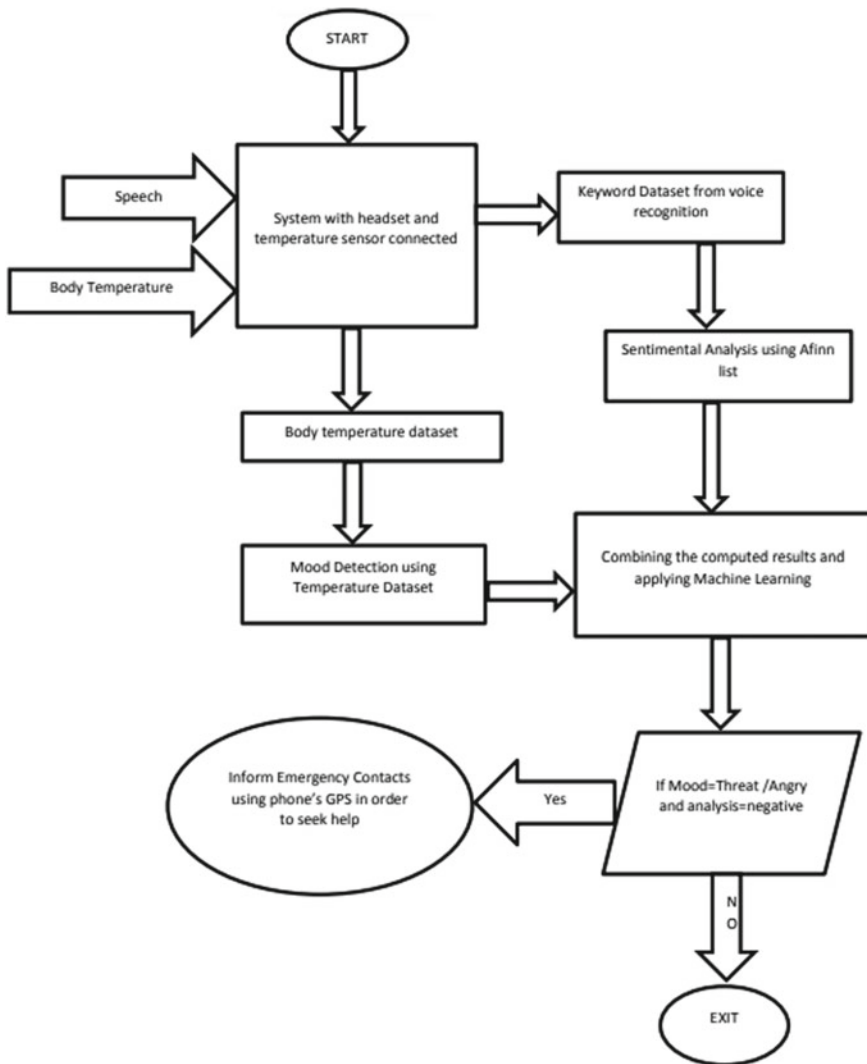
Output

- a) Sentimental analysis of the speech after converting it to text
- b) Mood corresponding to a given body temperature
- c) Final results after combining sentimental analysis and mood to detect seriousness of a situation.

Pseudocode:

1. P: total no of moods; counter: 0
2. Repeat steps 3 through 9 until counter $\leq P$
3. N: total no of input temperature for each mood
4. timestamp $\leftarrow 0$
5. loop: Repeat steps 4 and 5 until timestamp $\leq N$ for a particular mood
6. Record temp
7. timestamp \leftarrow timestamp + 1
8. end loop
9. Compute the average of all the recorded temperature and store in finalTemp[]
10. Record speech
 Speech to text conversion using suitable APIs and well-known fact that our nervous system reacts to different situations in different ways. Body temperature plays an important role in human body's response to various conditions. It varies according to place, climate, health, mood, etc. The variation of body temperature with mood is the key to sense threat. If a person is in threat, her body temperature will vary from what it usual
11. Removal of stop words and punctuations and generating clean text
12. Categorizing clean text and keywords as positive, negative and neutral
13. Storing the result in the form of an array textAnalysis[] which will have to columns one for keyword and the other for assigned values(0 \rightarrow neutral, 1 \rightarrow positive, -1 \rightarrow negative)
15. Combining finalTemp[] and textAnalysis[] array and applying machine learning to generate the desired result
16. If Fmood == threat and Fanalysis == negative then
 "Inform emergency contact and trigger an alarm"
17. else
 "No threat detected"
18. End

3.1 Block Diagram of the Proposed Idea



4 Experimental Observation

The data set was created by taking feedback from various audiences using Google docs form. All possible threat messages collected from the survey was tested against the proposed programme and the output was as follows (Figs. 1 and 2).

It is observed that as the number of sample data set is increased and as new words are added to the AFINN List, the accuracy % is increasing.

For example, as shown, if we input the speech as ‘I will kill you’, we get an overall score of -3 which indicates negative sentiment. This result is combined with temperature data set [4] as given in Table 1.

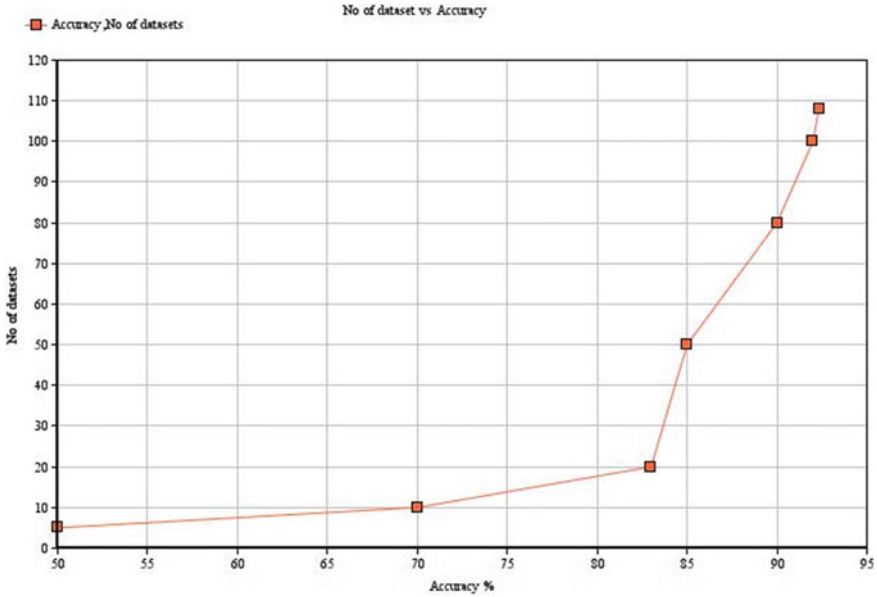


Fig. 1 No. of data set versus accuracy in %

```
cmd Command Prompt
Microsoft Windows [Version 10.0.16299.192]
(c) 2017 Microsoft Corporation. All rights reserved.
C:\Users\bhagyashree>cd anaconda3
C:\Users\bhagyashree\Anaconda3>cd envs
C:\Users\bhagyashree\Anaconda3\envs>cd speech
C:\Users\bhagyashree\Anaconda3\envs\speech>python speechbm.py
Say something!
C:\Users\bhagyashree\Anaconda3\envs\speech>python speechbm1.py
C:\Users\bhagyashree\Anaconda3\envs\speech>python test.py
['I', 'will', 'kill', 'you']
-3.0
C:\Users\bhagyashree\Anaconda3\envs\speech>
```

Fig. 2 Sample run

Table 1 Variation of body temperature with mood

Mood	Body temperature
	Range (in °C)
Happy	28–31.5
Sad	29–32
Angry	30–35

5 Conclusion and Future Scope

From the results of this project, it is clear that it is possible to predict the criminal intent of a person or the possibility of a person being under threat using existing technologies. And these can contribute tremendously towards security purposes and will help to reduce crimes. However, this is just the beginning. We can improve the output of such a system to a huge extent. The scope of such project is very extensible in future. Few improvements are stated below:

- (1) We worked with a limited data set. Collecting more number of data samples and adding more words to the AFINN list will increase the accuracy effectively.
- (2) Adding a few more factors will improve the system performance to a great extent. Some of the factors can be:
 - Heartbeat sensing
 - Humidity sensing
 - Automatic camera activation to take the video/image of the predicted culprit.
- (3) A GPS locator can be put to work to let the contacts of the predicted victim know her location and a danger alarm.
- (4) This whole idea can be converted to an android safety app and contribute to a better and safe society.

References

1. <https://www.ukessays.com/essays/physiology/literature-review-body-temperature-3097.php>
2. CNN, Cable News Network (No Date), <https://edition.cnn.com/2014/02/04/tech/innovation/this-new-tech-can-detect-your-mood/index.html>. Accessed 19 Feb 2017
3. Technology Review (2010), <https://www.technologyreview.com/s/421316/sensor-detects-emotions-through-the-skin/>. Accessed 15 Jan 2016
4. <https://www.sciencedirect.com/topics/earth-and-planetary-sciences/body-temperature>

Survey on Applications of Machine Learning in the Field of Computer Vision



Himanshu Shekhar, Sujoy Seal, Saket Kedia and Amartya Guha

Abstract Machine learning and computer vision have taken all over today's world which include medical diagnostics, statistical algorithm, logistic regression trees, etc. Such implementations are already been done in the arena such as smartphone applications, computer applications and online websites. We are getting a lot of features and advanced accessibility from machine learning and computer vision like Google Maps, Uber and Snapchat. We have put forth an introduction of some of the apps which use machine learning. We have also focussed on the algorithms used for machine learning implementation and their future uses. Some of the famous machine learning algorithms are random forest distribution algorithm, Naïve Bayes algorithm, decision tree. In today's world of advanced technologies, cybersecurity has become paramount. Online threat analysis uses machine learning from the very grass-roots level. We have also discussed advancement done in medical science field through machine learning in computer vision.

Keywords Machine learning · Computer vision · Algorithms · 3D-printing · Neural network

H. Shekhar · S. Seal (✉) · S. Kedia · A. Guha
Department of Computer Science and Engineering, Institute of Engineering & Management,
Kolkata, India
e-mail: sujoyseal11@gmail.com

H. Shekhar
e-mail: pandit98himanshu@gmail.com

S. Kedia
e-mail: saketk99hm@gmail.com

A. Guha
e-mail: amartya.aa99@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_58

1 Introduction

Machine learning deals in how we design our program such that it automatically improves the program performance over time. For example, programs concerning, speech recognition, optical text, recognition, autonomous unmanned cars, humanoid like Sophia, detecting credit card frauds could be improved upon through machine learning. Machine learning is dynamic and does not require any human intervention to process certain type of data after some training has been given to it. Whether it is in medical science, computer science, cybersecurity, crime investigation, data mining, artificial intelligence and retail shopping. Machine learning has made our lives more easier. For example, lots of mobile phone applications use machine learning to provide their services like Uber, Google, Facebook, Leafsnap, Twitter, Snapchat and Aipoly Vision. Undoubtedly, machine learning is affecting our day-to-day life. Is it not good enough, that we don't need any driver for driving a vehicle, just sit and it starts moving towards destination? At the time of boredom, we can make our faces funny by using different filters and virtual stickers. By just typing the destination, we can book a cab. By sitting at a place, we can hover through the streets of another place. Isn't our life easy?

Despite all these, we also have to take care of our own future and that of our future generations. When we perceive something, what does it really consists of? It consists of a bunch of information all around us. Alternatively, when we take a picture with a camera, it is just a bunch of coloured dots called pixels. Have you ever imagined, how we can edit these images? Do you make the computer understand these images? It might look easy but it is actually not...

2 Social Welfare

Many advancements have been done in last few decades and lots of improvements are going on nowadays. Today's world is based on machine learning, artificial intelligence and computer vision. We can do lots of work in very less time and more accurately as compared to last few decades. But, we also have to take care of our society and culture, privacy and security. Lots of computer applications like IBM's Blue-Whale, Google's AlphaGo has been developed using machine learning and artificial intelligence, which are more efficient than humans. Probably after few decades later, machines will be able to take their own decisions. Although machine learning and artificial intelligence are being used in the improvement and elevation of humankind, we can also work with machine learning and computer vision to make our life more joyful and productive. Many tech giants like Google, Facebook, Snapchat, Apple, IBM, Wipro, Uber and some small organizations are shifting their focus towards machine learning and computer vision for better productivity and services. For exam-

ple, Google has produced, Google Street View, a persuasive city mapped datasets application, from which we virtually explore cities and streets of some specified place. People may use it for threat purposes. Attackers may find their way or make a plan by simply using Google Street View to destroy some human lives. *The question arises "Is machine learning helpful for future of humanity? Does machine learning make our lives better or ruins our job?"*. This should not matter, how we use it, that matters. For example, lots of labourers lost their livelihood owing to 3D-printing. Previously, dozens of labours were needed to make a building, but now only a couple of skilled labour is enough to make a building. Hence machine learning has reduced chances of employment, has cut down people's salaries and also caused people to depend too much on machines. It's a great impact though negative. Our aim is to show whatever be the nature of the effect, it always covers a very wide scope. Hence, the motivation for social advancement involves a lot more than just mere computer vision.

Besides involving a lot more than the employment section, we also must consider that job market and monetary cycle is directly or indirectly affected by involving new tech giants. Machine learning has definitely brought an intense level of change. As mentioned earlier, it has brought a quality product in the market but the need of the situation may not have demanded so.

3 Urban Planning

Lot of changes are occurring in the advanced world of computer science and machine learning. With smartphones in our hand, it seems like the whole world is in our hand. We have a chance to do much more than what we want. Many creativities ideas are being put forth in the streets in urban areas like designing a natural space for pedestrian. Pavements embedded with LED lights could change the way of walking. Lights could show you the direction to where you want to go, like coffee shops, hospitals and restaurants. Machine learning could help us to organize cities.



A vector diagram showing the advancements in the world of machine learning and artificial intelligence

The arrangement of drainage systems has been greatly enhanced by machine learning. In developed countries like China, Japan and the USA as well as some developing countries use computer vision technologies to further enhance their shopping malls, cafeterias, multiplex buildings and other physically made mega construction projects. 3D-printing technologies are the most recent developments in civil and mechanical sciences. A great example that we can give is from China that, a Chinese company “Winsun” claimed to build 10 houses in just 24 h using a proprietary 3D printer and in next year they built a 5-storey building and a 1100 sqft. mansion! hence the utility is not only in terms of socio-economic development but also covers majority of gadgets which leads a country towards urbanization. Urban planning also is related to and in worst scenario leads to large-scale deforestation. Often, this results in wastage of land, money and precious resources. It is because our own vision is flawed when such large perspectives are at stake, computer vision hence implements technologies which in turn helps us to optimize any visual problem. Visual problems include road accidents, robbery which can all be resolved and prevented under constant closed-circuit television (CCTV) surveillance.

Machine learning features that are being used in South Bengal State Transport Corporation (SBSTC) bus and all other Android apps can be used to further develop urban planning. Hospitals use body scanner for different body parts viz. electrocardiogram (ECG) and computed tomography (CT) scanner which are highly accepted form of scanning techniques. The future of modernizing any city depends heavily on machine learning in field of computer vision.

4 Overview of Applications that Uses Machine Learning and Computer Vision

A lot of the applications available to us through Play Store, App Store and Amazon are based on machine learning. There are a significant number of organizations and start-ups which have turned towards optimum machine learning and have proven that investing in machine learning is the best in today's world.

4.1 Google Street View

It is a pervasive city imagery datasets application. It is an application from which we can virtually explore streets of cities. It uses a dense geo-sampling tool to show the streets of cities. Streets are captured through a fleet of vehicles equipped with a specialized camera.



A screenshot of Google Street View of Vineyard road, Roseville, California.

After collection of photos, they are digitally processed and combined together to look like a single image. From files reported for privacy, Google pixelated faces of pedestrian and licence plate which was captured. Web-based mapping technologies have been embraced by disciplines such as geography, archaeology and ecology, but also by several social scientific disciplines. Researchers working in the discipline of geography, archaeology and ecology quickly incorporated web-based mapping technologies into their research designs. There are various applications of Google Street View in research field, although the number still remains limited. It is also used for better estimation of fishing areas, estimation of forestry biomass in India,

estimation of area of different regions or lakes, etc. Google also helps in the criminological studies that have been implemented in the Google Maps and Streets View in their research design. Public and some law enforcement agencies and offenders are getting familiar with the power of online mapping technology through their day-to-day life and work. We can see Google Maps and its Street View can be used in various fields. It can be used in mapping or developing or maintaining cities' streets. We can use Google Street View to make an infrastructure of building or apartment, park, bridges and water reservoirs. Google Maps and Google Street View can be used in some research field of detecting the population or urbanization in some areas or throughout the globe.

4.2 Uber

Uber is one of the examples of using machine learning. It uses an algorithm which provides estimated time and real-time location on map, which is very useful and helpful for both drivers and riders. The company is also dealing with fraudulent behaviour like face detection and invalid stolen credit cards.

4.3 Google Keyboard

Almost all Android handset uses Google Keyboard. Gboard (Google Keyboard) uses the neural spatial model to determine the pixels touched on the screen and making relevant words and emojis in handwriting mode. It predicts the next word by matching the currently typed word with its dictionary set, which helps user to type fast and accurately.

4.4 Snapchat

Snapchat uses machine learning in face detection technology for applying filters on it. One may wonder about how Snapchat filter works? It first detects a face. Then locates facial features, and then creates a mesh of 3D mask (pyramidal shape) over the face. Snapchat not only applies the filters but also a list of things they are doing like, language detection for very short texts, names entity recognition and disambiguation using multi-modal named-entity recognition (NER) (sound, text, etc.), normalizing text misspellings (phonetic, orthographic, semantic representations), emotion analysis (from emoji to actual pictures), speech, music recognition (keyword spotting), personalized neural conversational models. We can use this technology for detection of culprit's face even if he or she made some facial changes.

4.5 *Virtual Voice Assistants*

The world is moving in the path of automation. People want their lives easier and comfortable like this hand free service provided by voice assistants. There are lots of virtual assistants available like Google Assistant, Apple's Siri, Cortana by Microsoft, Alexa by Amazon, Samsung's S voice. As there are more advances in machine learning, voice assistants are become more and more pseudo-emotionally attached to human beings. Voice assistants remind us on time so that we do not skip some important stuffs. Voice assistants along with computer vision can do many things that we even can't expect. It can assist in doing almost 70% of our daily work, from morning tea to evening supper.

4.6 *Evernote*

Evernote uses machine learning which automatically identifies the document file from the device storage and applies filters on it, such that it appears clear and tidy.

5 **Computer Vision in Medical Diagnostics**

Medical diagnosing techniques have fascinated us for a long time. It has been common for us to use them in our daily life and implement these technologies. Machine learning and especially computer vision contribute a lot in medical sciences, which make different difficult tasks easy for doctors and more tolerable for patients. They are widely useful in early detection of diseases, and hence are a valuable tool to save human life. Cardiographic techniques are a must for monitoring in the old age and for the infant safety.

Some of the computer vision includes:

Retinoscopy—Although primitive in approach, they are a must once in our lives. Retinoscopy is used to measure activities of rod and cone receptors in our eyes. Retina has three distinct areas for colours—erythrolabe, chlorolabe and cyanolabe which are analogical to pixel design and identification algorithms on machine learning.

Tumour detection—Around the world, cancer has been affecting billions of lives both in terms of life and money. Machine learning diagnosing systems apply their identification systems to further develop accurate detection in terms of size, location, quality of such tissues which are suspected to become malignant uncontrolled group of fast-dividing cells.

Amniocentesis—Though banned in India for social cause, amniocentesis can greatly affect man to woman ratio in any country if followed on such comparably large scale for 2–3 years. Drastic changes were brought during World War I and II

by Nazis and Fascists in Germany and Italy mis-balancing their population heavily for years.

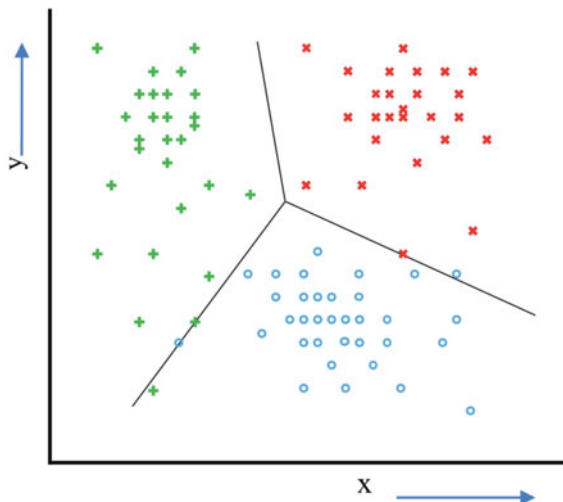
CT scan—A very common term for cancer patients which uses electromagnetic radiations under manually operated controlled computer vision gratings which are so accurate that it can measure a pigment called c-125 in our blood.

6 Algorithms

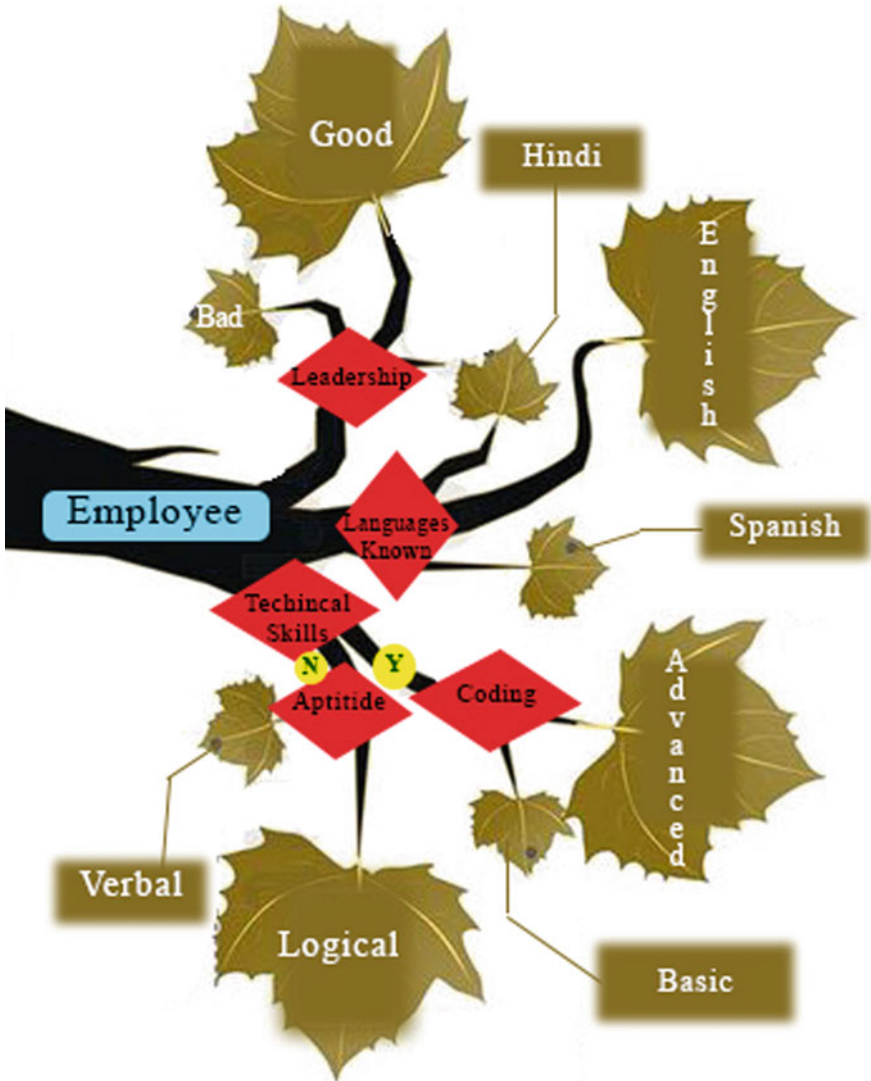
The subfield of artificial intelligence, machine learning has gained much popularity in last few couple of years. Many tech giants use machine learning algorithms, like Netflix's algorithms to make movie prediction from your previous watched movies. In this section, we would like to present some of the famous algorithms which are used frequently.

They are:

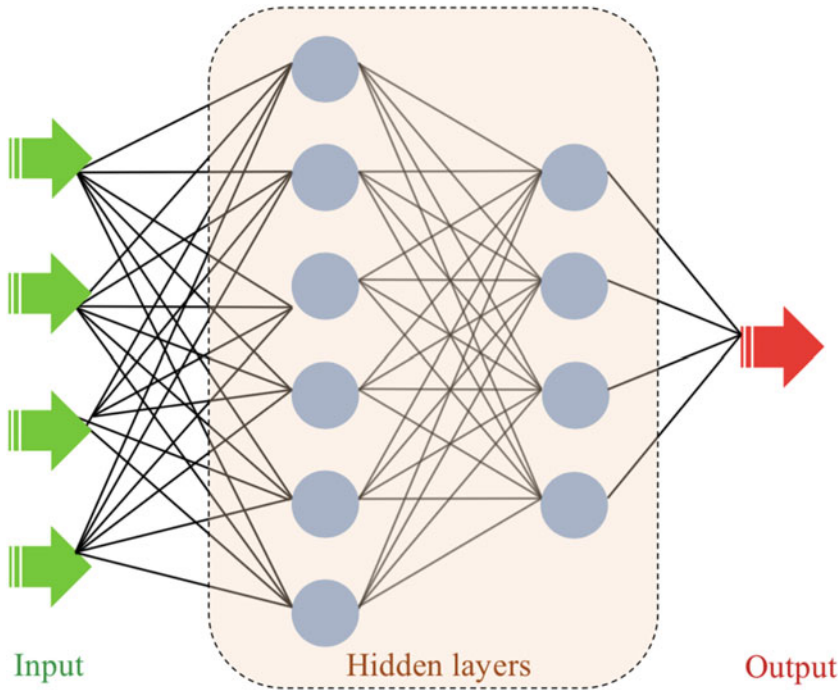
1. **Naïve Bayes' algorithm**—This is the algorithm mostly used in machines and hardware. It simply applies Bayes' theorem along with strong independence assumptions. For example, it could be used to mark an email as spam and used for face detection software. $p(C_k|x) = p(C_k) * p(x|C_k)/p(x)$; for each K possible outcomes or classes C_k .
2. **k-means clustering algorithm**—This is a type of unsupervised learning which has various uses including business and management. This algorithm also lets us know profit at each stage of the product. It is also referred as Lloyd's algorithm. This algorithm is also used in grouping of features into different labels.



- 3. **Decision trees**—These are trees in which decisions are made by the computer at each stage based upon recurrence relations.



A schematic diagram of decision tree for deciding quality of an employee



4. **Neural network**—Our neurons in the body play a major role in determining the steps to process a single task. Similarly, network of artificial neurons could be used to make a decision and executed consecutively. This again depends on activity of the neurons. An artificial neuron is an actual piece of hardware which helps the system to make a decision based on the receptors. This algorithm helps us to build any machine, emulating human reflex arcs.

Algorithms used in machines have several important implementations. The algorithms are also useful in healthcare industries, for example, random forest distribution algorithm. This algorithm is mostly derived from statistical studies, is useful in calculating people densities, mass or chunk density. The most important are the artificial neural networks algorithms. These algorithms combine artificial intelligence through the implementation of neural networking. With computer vision, these algorithms are able to judge system on basis of their reactance to external stimuli.

7 Cybersecurity

Cybersecurity refers to protecting your personal information from malicious activities. Machine learning has a lot of algorithms and systems which protect users from threats. Such as the Paypal app which was developed in December 1998, uses machine learning algorithms to protect its users from different threats and online

spoofing. It uses three types of machine learning algorithms that are linear, neural network and deep learning algorithm.

The threats are:

Waterhole—It is like a pit surrounded by greenery. Hackers access other people's information by using sites which are more accessible to the public more than anything else. For example, networks in a coffee shop are accessed by so many users such that these users load their personal computers with whatsoever data is provided to them. Like this, there are so many sites to put on viruses and worms. Machine learning algorithms, detect path of these malwares, blocking them with a firewall thereafter.

Web shell—These are piece of code which are loaded into a working device which provokes the user to misjudge and then taking advantage, entry is gained into the full database.

Ransomware—Similar to web shell, but here the user is vulnerably threatened externally by a group of software brokers who have corrupted the users' personal files. Such scenarios can be totally avoided by using machine level language which has early detection.

Remote exploitation—Here, user has to give a lot of permissions to externally attached embedded system which virtually gains a lot many access to the security software to nullify them permanently. Machine learning softwares present within, deny permissions irrespective of user's choice.

8 Conclusion

So far, we have surveyed the different domains in machine learning. There are other domains like cryptology in which machine learning can be applied to give it a wider perspective, but such developments need innovation and it is for us to reason what the future applications may be. Any scientist and researcher may demand to describe trajectories of motion of extra-parts of solar system like space stations of universal repute, comets and satellites. One may also apply computational nano-technology combined with computer vision to describe laws of genetics more accurately and recursively. References can be made to Charles Darwin who developed his laws of evolution on earth based upon vision, structural and morphological similarities in flora and fauna. We can extend his idea using machine learning.

Bibliography

1. L. Li, J. Tompkin, P. Michalatos, H. Pfister, Hierarchical visual feature analysis for city street view datasets
2. L. Fridman, D.E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, J. Kindelsberger, L. Ding, S. Seaman, H. Abraham, Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation (2017)

3. A. Pettinato, B. Seppelt, L. Angell, B. Mehler, B. Reimer, MIT Autonomous vehicle technology study: large-scale deep learning based analysis of driver behaviour and interaction with automation
4. C. Robert, *Machine Learning, a Probabilistic Perspective* (American Statistical Association, 2014)
5. L. Liu, H. Wang, C. Wu, A machine learning method for the large-scale evaluation of urban visual environment
6. G. Shakhnarovich, T. Darrell, P. IotrIndyk, *Nearest-Neighbor Methods in Learning and Vision: Theory and Practice (Neural Information Processing)* (2006)
7. S. Gong, S.J McKenna, Al. Psarrou, *Dynamic Vision: From Images to Face Recognition* (Imperial College Press, London, 2000)
8. J. Wright, Y. Ma, J. Mairal, G. Sapiro, T.S. Huang, S. Yan, Sparse representation for computer vision and pattern recognition, in *Proceedings of the IEEE*, June 2010
9. T. Caelli, W.F. Bischof, *Machine Learning and Image Interpretation* (Springer, 2013)
10. Y. Netzer, T. Wang, A. Coates, A. Bissacco, B. Wu, A.Y. Ng, Reading digits in natural images with unsupervised feature learning (2011)
11. S.S Haykin, *Neural Networks and Learning Machines* (New York, Prentice Hall, 2009)
12. S. Nowozin, C.H Lampert, Structured learning and prediction in computer vision (2011)
13. A.M. Gillies, Machine learning procedures for generating image domain feature detectors
14. A. Criminist, J. Shotton, E. Konukoglu, Decision forests: a unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning (2012)

Automatic Speech Recognition Based on Clustering Technique



Saswati Debnath and Pinki Roy

Abstract Automatic Speech Recognition (ASR) is defined as a computer-driven transcription of the spoken word into readable text. The main aim of ASR technology is to correctly identify the words spoken by a person. In this paper, we are using the clustering method in ASR to identify the digit spoken by the speaker. The importance of ASR is to allow a computer to recognize the words that are spoken by any human being independent of vocabulary size, noise, speaker characteristics or accent. This paper introduces ASR using clustering techniques to recognize the English digits. The clustering algorithms used here are K-means and Gaussian Expectation Maximization (GEM). We use Mel-frequency cepstral coefficients (MFCCs) for extracting features from speech. Performance is calculated for an individual digit using the hard threshold technique. This paper also compares the result using two different clustering techniques.

Keywords Digit recognition · MFCC · K-means clustering · Gaussian Expectation Maximization clustering · Hard threshold

1 Introduction

ASR is a technology that allows a machine to recognize the person's spoken words and convert it into written text. The speech signals are slowly timed varying signals (quasi-stationary), but in a short period of time (5–100 ms), its characteristics are fairly stationary. Speech is the most natural form of human communication. The first step of speech recognition is speech signal preprocessing and feature extraction where the relevant information from the speech sample is extracted to characterize the time-varying properties of an acquired speech sample. In feature extraction phase, feature vectors are extracted using feature extraction methods like cepstral coefficients of speech signal. After feature extraction, classification phase is carried out with the

S. Debnath (✉) · P. Roy

Computer Science and Engineering Department, NIT Silchar, Silchar, Assam, India
e-mail: dnsaswati@gmail.com; Debnath.saswati123@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_59

help of some classifiers. There are several different classes of speech recognition system like recognition of isolated words, digit recognition, continuous speech and spontaneous speech.

Digit recognition is the part of speech recognition with increasing importance among speech processing application. Digit recognition is getting more and more attention since last decade due to its wide range of application. This is because it has an importance in several fields and used in checks in banks or for recognizing numbers on car plates, railway PNR or many other applications.

In this paper, we introduce digit recognition system based on different clustering technique. MFCCs [1, 2] are extracted from speech signal and K-means [3, 4] and GEM [5, 6] are used to cluster the speech features. The threshold is generated for the individual digit during the training phase of the system. At the testing phase, calculate the recognition rate for every digit. The paper is written as follows: Sects. 2 and 3 give the literature review and the proposed methodology, experimental result and analysis are described in Sects. 4 and 5. Section 6 analyzes the comparison of result and finally concludes the paper in Sect. 7.

2 Literature Review

Farhat and O'Shaughnessy [7] proposed an idea that provides a trade-off between the complexity of the acoustic models and their trainability. The authors defined a shared-distribution approach in their HMM-based continuous speech recognizer.

Revathi and Venkataramani [8] explored the performance of perceptual features for performing isolated digits and continuous speech recognition. The perceptual features were extracted, and training models were developed by K-means clustering.

A novel approach for large speech databases quantization was proposed by Lazli and Boukadoum [9]. An unsupervised iterative process was used to regulate a similarity measure to set the number of clusters and their boundaries.

Thatphithakkul et al. [10] proposed an approach called noise-cluster HMM interpolation for robust speech recognition. In that method, a new HMM is interpolated from existing noisy speech HMMs that has been best matched to the input speech.

Context-dependent hybrid connectionist speech recognition system using a set of generalized hierarchical mixtures of experts (HME) was introduced by Fritsch et al. [11]. Context-dependent modeling was shown to significant improvements over simple context-independent modeling requiring only small additional overhead in terms of training and decoding time.

Saini and Kaur [12] described a study of basic approaches to speech recognition and how their result shows better accuracy. They also represent research that has been carried out for dealing with the problem of ASR. Three major approaches like acoustic phonetics, pattern recognition and artificial intelligence approaches are used for ASR. The acoustic phonetic approach uses knowledge of phonetics and linguistics to guide the search process.

Patel and Srinivasa Rao [13] proposed an efficient speech recognition system with the integration of MFCC feature with frequency sub-band decomposition using sub-band coding. It was HMM-based recognition approach.

It is important to detect speech endpoints accurately in the process of speech recognition. Li et al. [14] introduced a comparative analysis of various feature extraction techniques of endpoint detection in speech recognition of isolated words in noisy environments. An optimum set of characteristics was identified by combining parameters from both time domain and frequency domain.

Hanchate et al. [15] proposed a digit recognition system using ANN. This paper described the application of neural network (NN) in the pattern recognition approach. First, the features of the training data set were extracted automatically using the end-to-end detection function. The features were then used to train the NN.

Speaker independent, isolated word recognition carried out using MFCC and HMM was developed by Ananthakrishna Thalengala and Kumara Shama for Kananda language [16]. They have used phone level and syllable level HMM for recognition.

The development of an efficient speech recognition using different techniques such as MFCC, Vector Quantization (VQ) and HMM was proposed by Swamy and Ramakrishnan [17]. MFCC was used to extract the characteristics from the input speech signal with respect to a particular word uttered by a particular speaker. Then HMM was used on quantized vectors to identify the word by evaluating the maximum log-likelihood values for the spoken word.

3 Proposed Methodology

Proposed methodology includes the steps given below. The architecture of the proposed model is shown in Fig. 1.

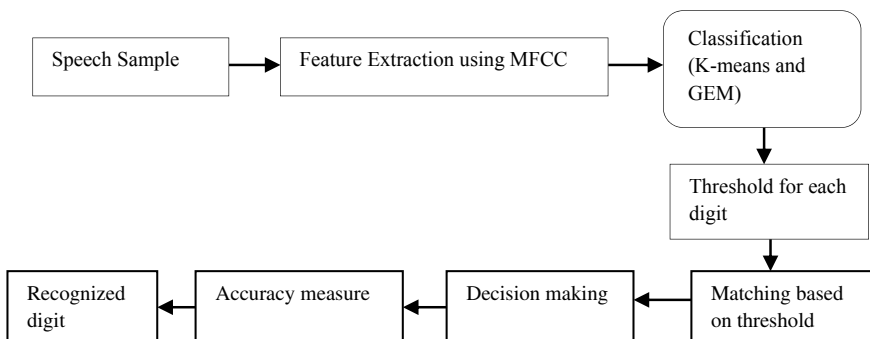


Fig. 1 Block diagram of proposed model

The steps involved are:

- Database collection.
- Feature extraction: MFCC.
- Classification: K-means and GEM Algorithms.
- Decision making: Define hard boundaries using a threshold.

3.1 Database Description

The speech database is recorded in a lab environment. English digit zero to nine (0–9) has been recorded from 10 adult male speakers of age group 24–30 years where each speaker has repeated each word 20 times. The sampling rate is 16 kHz, and the sampling bit resolution is 16 bits/sample. For every digit, 200 data samples were recorded. The database is uploaded and it is available on IITG webpage of Dr. P. K. Das [18].

3.2 Feature Extraction

This important step in ASR is to extract features from the audio signal. MFCCs are extracted from speech signal to calculate speech features.

The calculation of MFCC includes the following steps [2]:

- I. Framing: frame the speech signal at the initial stage.
- II. Windowing: after framing the most commonly used Hamming window is used for windowing to remove the discontinuity of the signal.
- III. Fast Fourier transforms (FFT): calculate FFT to convert the signal from time domain to frequency domain.
- IV. Mel-frequency warping: calculate the Mel filter bank energies.
- V. Discrete cosine transform (DCT): DCT is performed to convert the Mel spectrum to the time domain and capture the coefficients is called an acoustic vector (MFCCs).

3.3 Classification Using Clustering

Clustering is a method of machine learning where each data point or cluster is grouped into a subset or a cluster that contains a similar kind of data points. The two clustering algorithms have been used for this experiment which are K-means and GEM. Each cluster includes one centroid and a number of code vector.

3.3.1 K-Means

K-means [3, 4] is one of the simplest unsupervised learning algorithms. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed in prior.

Steps involved in the K-means algorithm:

- (1) Randomly select ' k ' cluster centers.
- (2) Calculate the distance between each data point and cluster centers.
- (3) Assign the data point to the cluster center whose distance from the cluster center is minimum of all the cluster centers.
- (4) Recalculate the new cluster center using Euclidean distance.

3.3.2 Gaussian Expectation Maximization (GEM)

If there is uncertainty about the data points where they belong to, we use GEM [5] instead of hard assigning of data points. It uses probability of a sample to determine its feasibility of belonging to a cluster. An expectation–maximization (EM) [6] algorithm is an iterative method to find maximum likelihood or maximum a posteriori (MAP) estimates of parameters in statistical models, where the model depends on unobserved latent variables.

Algorithm:

1. Initialize the mean μ_k , covariance and mixing coefficients and evaluate the initial value of the log likelihood.
2. Expectation (E) step: Evaluate the responsibilities using the current parameter values:

$$\gamma_j(x) = \frac{\pi_k \varphi(x|\mu_k, \sigma_k)}{\sum_{j=1} \pi_j \varphi(x|\mu_j, \sigma_j)} \quad (1)$$

3. Maximization (M) step:

$$\mu_j = \frac{\sum_{n=1}^N \gamma_j(x_n) x_n}{\sum_{n=1}^N \gamma_j(x_n)} \quad (2)$$

$$\sigma_j = \frac{\sum_{n=1}^N \gamma_j(x_n) (x_n - \mu_j)(x_n - \mu_j)^T}{\sum_{n=1}^N \gamma_j(x_n)} \quad (3)$$

$$\pi_j = N^{-1} \sum_{n=1}^N \gamma_j(x_n) \quad (4)$$

4. Estimate log likelihood.

If not converged, update $k = k + 1$ and return to step 2.

$$\ln p(X|\mu, \sigma, \pi) = \sum_{n=1}^N \ln \left\{ \sum_{k=1}^k \prod_k \varphi(x_n|\mu_k, \sigma_k) \right\} \quad (5)$$

Each iteration of the EM algorithm calculates two processes: the *E*-step and the *M*-step. In the *E*-step for giving parameter values, we can compute the expected values of the latent variable. The *M*-step updates the parameters of our model based on the calculated latent variable.

3.4 Decision Making

The centroids obtained from clustering algorithm are used to determine the threshold. Codebook is found out for each of the utterances. For each sample, we find the Euclidean distance from the codebook of the same digit. The mean (μ_1) and standard deviation (σ_1) of the distances are calculated. We also consider the mean (μ_2) and standard deviation (σ_2) of the distances calculated for other random selected digit's codebook and the digit in consideration. The threshold is calculated using the following formula.

$$\text{threshold} = \frac{\mu_1\sigma_1 + \mu_2\sigma_2}{\sigma_1 + \sigma_2} \quad (6)$$

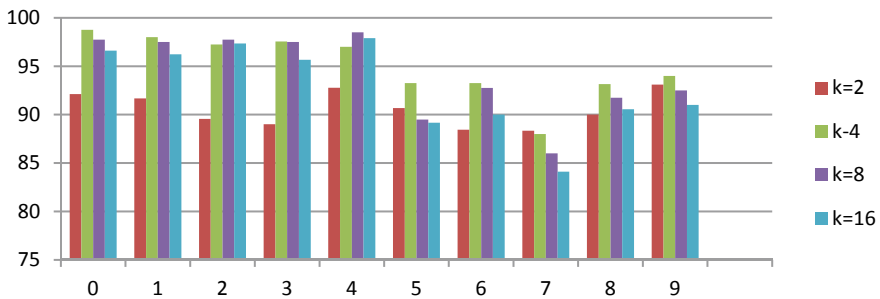
The threshold is calculated for each specific digit.

4 Experimental Results and Description Using K-Means Clustering

We have extracted the features for each audio signal using MFCC and the order of MFCC considered here is $n = 13$. After feature extraction classification has been done using K-means and GEM clustering algorithm. We have taken the cluster size 2, 4, 8 and 16 for both the algorithm. Audio samples from the speakers are taken into consideration. Total 200 samples are taken for the experiments where 140 samples are used for training and 60 samples are used for testing. The hard threshold is calculated for each digit in the training phase. At the time of testing, we use the Euclidean distance measure.

Table 1 Performance of digit recognition using K-means clustering for 2, 4, 8 and 16 codebook size

Digits	$k = 2$	$k = 4$	$k = 8$	$k = 16$
0	92.13	98.75	97.75	96.61
1	91.67	98.00	97.50	96.23
2	89.56	97.25	97.75	97.34
3	89.01	97.55	97.50	95.67
4	92.78	97.00	98.50	97.89
5	90.67	93.25	89.50	89.16
6	88.45	93.25	92.75	90.00
7	88.34	88.00	86.00	84.11
8	90.00	93.15	91.75	90.56
9	93.11	94.00	92.50	91.00

**Fig. 2** Performance graph of k-means algorithm using different codebook size

4.1 Result Analysis

Using K-means clustering, accuracy achieved greater than 85% for each of the digits from 0 to 9. We have selected a different number of centroids, i.e., $k = 2$, $k = 4$, $k = 8$ and $k = 16$. When the number of centroid increases from $k = 2$ to $k = 4$, accuracy increases. But as codebook size increases from 4 to 8 and 8 to 16, accuracy decreases. More number of centroid increase complexity and that decreases the accuracy. It is observed that better accuracy is achieved with 4 codebook size. Thus, increasing the number of codebook size after 4 decreases the system accuracy. Table 1 gives the performance of individual digit recognition using K-means and MFCC (Fig. 2).

5 Experimental Results and Analysis Using GEM

Table 2 shows the individual digit recognition using GEM and MFCC for 2, 4, 8 and 16 codebook size.

5.1 Result Analysis

Using GEM, we acquire the accuracy greater than 87% for each of the digits from 0–9 with $k = 4$. We get more accuracy with 4 centroids clustering. When the codebook size increases from $k = 2$ to $k = 4$, accuracy increases. But after a certain number, i.e., codebook size $k = 4$ accuracy decreases with increased codebook size because the higher codebook size increases the complexity of representation and also the sparse distribution of data. The number of centroids also increases the complexity of the system. Thus for both the clustering techniques, K-means and GEM, system accuracy depends on the size of codebook (Fig. 3).

Table 2 Performance of digit recognition using GEM clustering

Digits	$k = 2$	$k = 4$	$k = 8$	$k = 16$
0	97.42	99.00	97.25	96.44
1	96.11	98.5	97.00	97.21
2	96.67	99.25	97.25	96.62
3	98.10	98.75	98.25	98.13
4	94.36	98.75	97.25	97.32
5	92.65	93.75	87.25	87.22
6	92.23	94.00	92.75	91.35
7	89.47	87.00	84.75	82.22
8	90.43	93.45	91.75	91.13
9	92.22	94.00	90.25	90.00

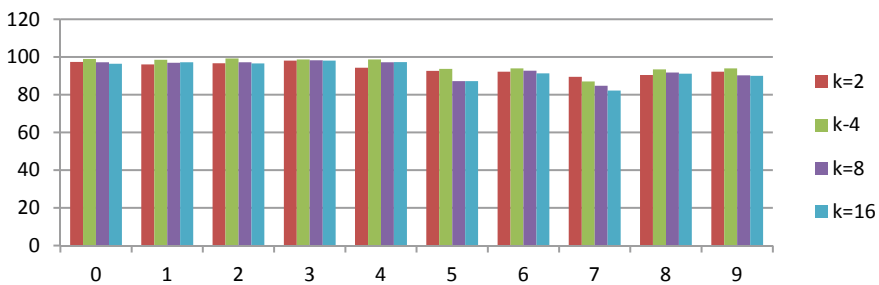


Fig. 3 Performance graph of GEM algorithm using different codebook size

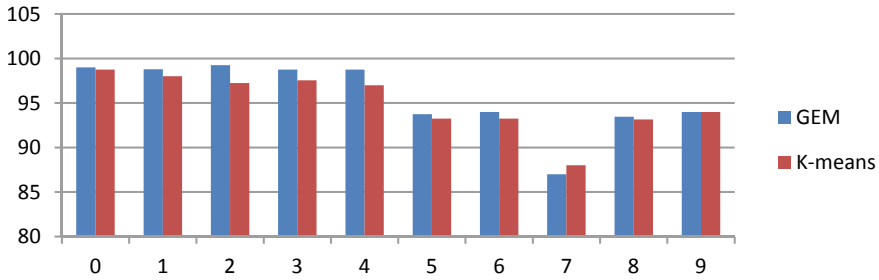


Fig. 4 Comparison graph of GEM and k-means clustering for digit recognition

6 Comparison Between K-Means and Gaussian Expectation Maximization

We have achieved more accuracy when the codebook size is 4 for both the algorithms. So, we have considered $k = 4$ for comparing both the algorithms. After comparison, we conclude that the GEM gives more accuracy than K-means. This is because K-means is hard clustering algorithm, it means a single point in the matrix should belong to only one cluster. The problem arises when a point lies in midway, i.e., equidistant from more than one centroid. Assigning such a point to a specific cluster may not give accurate results. Whereas in GEM calculates the weighted distance, i.e., it takes both mean and variance into consideration as compared to only mean in K-means algorithm. Comparison graph is shown in Fig. 4.

7 Conclusion

In this work, we have extracted the speech features using MFCC and then we have applied clustering algorithms K-means and GEM. We have taken the codebook size 2, 4, 8 and 16 for both the clustering algorithms. Both the algorithms perform better results when codebook size $k = 4$. Higher accuracy is achieved using the GEM algorithm as compared to K-means algorithm with codebook size $k = 4$. The reason behind this is the GEM calculates the weighted distance, but K-means is hard clustering algorithm. The hard threshold technique has been applied to measure the performance because it gives the robustness of the system.

Acknowledgements The author(s) acknowledge the traveling support of Technical Education Quality Improvement Programme (TEQIP-III) of the Government of India for attending the conference.

References

1. S.B. Davis, P. Mermelstein, Comparison of parametric representation for monosyllabic word recognition in continuously spoken sentences. *IEEE Trans. Acoust. Speech Sig. Process.* **28**(4), 357–365 (1980)
2. B. Soni, S. Debnath, P.K. Das, Text-dependent speaker verification using classical LBG, adaptive LBG and FCM vector quantization. *Int. J. Speech Technol.* **19**(3), 525–536 (2016)
3. T. Kanungo, D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman, A.Y. Wu, An efficient k-means clustering algorithm: analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 881–892 (2002)
4. J. Yadav, M. Sharma, A review of K-mean algorithm. *Int. J. Eng. Trends Technol. (IJETT)* **4**(7), 2972–2976 (2013)
5. M. Nadif, G. Govaert, Block clustering via the block GEM and two-way EM algorithms, in *The 3rd ACS/IEEE International Conference on Computer Systems and Applications* (2005)
6. V.-E. Neagoe, V. Chirila-Berbentea, Improved Gaussian mixture model with expectation-maximization for clustering of remote sensing imagery, in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)* (2016)
7. A. Farhat, D. O’Shaughnessy, The use of a distribution-clustering technique in HMM-based continuous-speech recognition, in *Canadian Conference on Electrical and Computer Engineering* (IEEE Xplore, 1995). <https://doi.org/10.1109/ccece.1995.526598>
8. A. Revathi, Y. Venkataramani, Perceptual features based isolated digit and continuous speech recognition using iterative clustering approach networks and communications, in *First International Conference on Networks & Communications* (IEEE Xplore). <https://doi.org/10.1109/netcom.2009.32>
9. L. Lazli, M. Boukadoum, HMM/MLP speech recognition system using a novel data clustering approach, in *Electrical and Computer Engineering (CCECE), 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE)* (IEEE Xplore). <https://doi.org/10.1109/ccece.2017.7946644>
10. N. Thatphithakkul, B. Kruatrachue, C. Wutiw WATCHAI, Robust speech recognition using noise-cluster HMM interpolation, in *ICSP 2008. 9th International Conference on Signal Processing* (IEEE Xplore). <https://doi.org/10.1109/icosp.2008.4697203>
11. J. Fritsch, M. Finke, A. Waibel, Context-dependent hybrid HME/HMM speech recognition using polyphone clustering decision trees *Acoust. Speech Sig. Process.* (1997) ICASSP-97. IEEE xplore
12. P. Saini, P. Kaur, Automatic speech recognition: a review, in *2013 International Journal of Engineering Trends and Technology*
13. I. Patel Dr. Y. Srinivasa Rao, Speech recognition using hidden Markov model with MFC-C—Subband Technique, in *2010 International Conference on Recent Trends in Information, Telecommunication and Computing*
14. X.-g. Li, M.-f. Yao, W.-t. Huang, Speech recognition based on K-means clustering and neural network ensembles, in *2011 Seventh International Conference on Natural Computation*
15. D.B. Hanchate, M. Nalawade, M. Pawar, V. Pophale, P.K. Maurya, Vocal digit recognition using artificial neural network, in *2010 2nd International Conference on Computer Engineering and Technology*
16. A. Thalengala, K. Shama, Study of sub-word acoustical models for Kannada isolated word recognition system. *Int. J. Speech Technol.* **19**, 817–826 (2016). <https://doi.org/10.1007/s10772-016-9374-0>
17. S. Swamy, K.V Ramakrishnan, An efficient speech recognition system. *Comput. Sci. Eng. Int. J.* **3**, 21 (2013)
18. <http://www.iitg.ernet.in/pkdas/digits.rar>

A Study of Interrelation Between Ratings and User Reviews in Light of Classification



Pritam Mondal, Amlan Ghosh, Abhirup Sinha and Saptarsi Goswami

Abstract Natural language processing (NLP) is a broad area of study where human language and their emotion can be characterized and analyzed by computer. Types of NLP problems can be, but not limited to, categorization, classification, analytical, etc. A major application of NLP is at sentiment analysis of rich texts. In this study, we have worked on sentiment analysis and categorization aspects of NLP. Here, we have performed analysis, lemmatization, sampling, and classification on Yelp text reviews, which is an open-source dataset, and then compared the results of the different types of classifiers. Categorization of different texts has been considered as a response of classification problems. After validation, we have achieved a significant level of robustness in these methods. An attempt was made to study NLP classification problems under various situations.

Keywords Natural language processing · Machine learning · Classification · Lemmatization · Sampling

1 Introduction

With the advent of modern computing, scientists have long been busy in empowering computers by imitating human decision-making and brain functions. It has given birth to artificial intelligence, a field of study based on the claim that natural intelligence ‘can be so precisely described that a machine can be made to simulate it’ [1].

Machine learning can be thought of as a subset of artificial intelligence and differs from the traditional rule-based approach of AI. The rule-based approach demands

P. Mondal

Department of Industrial and Systems Engineering, IIT Kharagpur, Kharagpur, India

A. Ghosh

Department of Industrial Engineering and Operations Research, IIT Bombay, Mumbai, India

A. Sinha (✉) · S. Goswami

A. K. Chowdhury School of Information Technology, University of Calcutta, Calcutta, India

e-mail: Abhirup.rupai@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling*

and Graphics, Advances in Intelligent Systems and Computing 937,

https://doi.org/10.1007/978-981-13-7403-6_60

specific pre-fixed rules on some decision problem, and the system can act accordingly. Human intervention is necessary for this. Machine learning enables computer systems to learn from past known observations and results. Such systems can take decisions about unseen problems of similar type when encountered. It is a significant improvement over the rule-based approach of classical AI problems, as human intervention is not that much necessary. The system learns from historical data and tries to figure out a correlation between past data and results and builds rules on that, all by itself. It can be thought synonymous to intelligence, so many prefer machine learning for building AI systems. We, as humans, can take a decision based on past experiences, and machine too is reflecting that ability of ours and adapting to changing environments, that qualifies them as ‘Intelligent Machines’ [2].

Here, our primary purpose is to find the relationship between the user’s text review and their given star ratings in any business profile. Taking the Yelp training set as an example, we first created a corpus from the whole document by using lemmatization and then further removing the common words, namely ‘stopwords’ from the text document along with some useless expressional words which occur very less. Then, using the corpus, we used four different classifiers and compared the results among themselves, thus making a stable base for future semantic and sentiment analysis.

2 Related Works

In a paper by Tetsuya Nasukawa and Jeonghee Yi, they described extracting the sentiments associating with positive and negative feedbacks from any document without classifying the text as positive or negative [3].

The same previous authors along with Razvan Bunescu and Wayne Niblack presented sentiment analyzer (SA) that extracts sentiment from online text documents where the analysis consists of topic-specific feature term extraction, sentiment extraction, and (subject, sentiment) association by relationship analysis. SA utilizes the sentiment lexicon and the sentiment pattern database [4].

The paper by Bo Pang et al. also is based on text classification using machine learning techniques. Their objective was to classify the documents by overall sentiment, not by topic [5]. Also, the paper by Subhasree Bose et al. discussed techniques to gain insights from textual data [6].

On the Yelp dataset, many types of research were done before. Some proposed a hierarchical attention network for document classification [7]. Researches were also done to detect the aspects and sentiment in this customer review [8].

3 Procedure

3.1 Data Source

Before going to the detailed procedure, we are giving an overview on why and what data we have chosen to analyze. As we are going to do some natural language processing, we have to find a data where a decent amount of text reviews are available for processing and validation purpose. We have chosen the dataset from ‘Yelp Dataset Challenge’ where rich text reviews are available along with the corresponding star rating of those reviews.

3.2 Data Description

Many other features like ‘date,’ ‘business_id,’ ‘review_id,’ and ‘votes’ are also available in the Yelp dataset, but we are only concerned about the ‘text’ reviews and ‘stars’ ratings here. The whole dataset contains as many of 229,907 reviews. The text review has texts in mostly English language, and the star ratings have rating ranges from 1 to 5. The star ratings of 1–5 can be treated as ‘very bad,’ ‘bad,’ ‘average,’ ‘good,’ and ‘very good,’ respectively. We considered all 4- and 5-star reviews as ‘good’ and all 2- and 1-star reviews as ‘bad.’ 3-star reviews were considered as ‘average.’ It was done to preserve the similarity in 4- and 5-star reviews or in 2 and 1-star reviews.

3.3 Corpus Creation

We created our corpus of reviews from the dataset after removing common punctuations and reducing every inflected word to its root form. We have used lemmatization based on WordNet [9, 10] for this purpose. Use of lemmatization over stemming was preferred, as algorithmic stemmers often reduce the inflected words to an incomplete form like ‘was’ can be reduced to ‘wa.’ Also, for words having the same root, algorithmic stemmers might not detect it. Like, ‘is’ and ‘was’—both are inflected form of the verb ‘be.’ While WordNet can detect it, algorithmic stemmers might fail. Thus, the use of lemmatizer has helped us in sparsity control too.

3.4 Sparsity Control

Sparsity control is a significant issue while dealing with large language corpora. We have already mentioned how the use of lemmatization over stemming has helped us control the number of words in our corpus, without altering sentence’s meaning.

To further reduce the corpus size, we filtered out common English words, known as ‘stopwords,’ before processing. Also, we removed less occurring words, which occurred less than 10% in the whole corpora. We only considered words having more than three letters, without having any numeric characters. Only unigrams were taken into consideration.

3.5 Sampling

Upon careful inspection of the dataset and based on our preliminary findings, we came to understand that the dataset in hand is more biased toward higher ratings, i.e., toward ‘good’ ratings. The situation is depicted in Fig. 1. So, we had to perform over-sampling first. The well-known synthetic minority over-sampling technique (SMOTE) was used for over-sampling. Imbalanced-Learn [11] is a python library that provides implementations of several sampling methods. We used it for our sampling purpose. After the performance of the sampling, the resulting classes became balanced, as shown in Fig. 2.

Fig. 1 Rating distribution before sampling

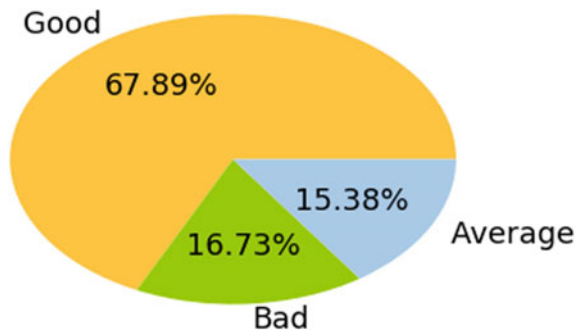
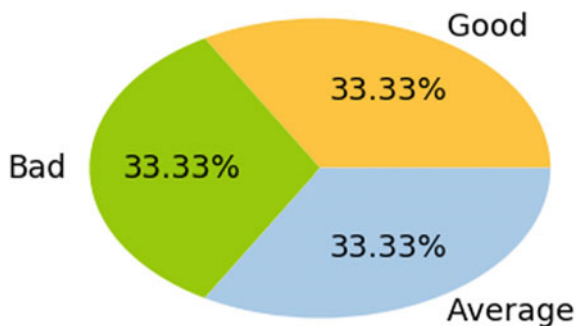


Fig. 2 Rating distribution after sampling



3.6 Classification

After all these preprocessing, we came to our chief task at hand. The collection of words was transformed into a sparse matrix. We got different classifiers from Scikit-learn [12] to deal with the main corpus. Here, we have used four different classifiers and compared the results in each type. The classifiers are random forest classifiers, decision tree classifiers, K nearest neighbors classifiers, and extra trees classifiers. We have kept 25% of total data for validation and testing purpose. We tried to classify different reviews into the mentioned three categories, i.e., ‘good,’ ‘average,’ and ‘bad.’ For example, in our data, a 5-star review: ‘*Best food, super friendly staff, and great prices. Loved it!*’ is focused and classified on words like ‘best,’ ‘super,’ ‘friendly,’ ‘great,’ and ‘love,’ whereas a 1-star review: ‘*Bad experience...Terrible service, terrible food.*’ is mainly focused on words like ‘bad and ‘terrible.’ A 4-star review is: ‘*loved the gyro plate. Rice is so good and I also dig their candy selection,*’ which is focused on ‘love’ and ‘good’; a 2 star review: ‘*decent Filibertos knockoff but the place feels like a major health hazard considering how new it is*’ is focused on ‘decent,’ ‘major,’ ‘health,’ ‘hazard,’ and ‘new.’ That is why we put the 5- or 4-star reviews in the ‘Good’ category and the 1- or 2-star review in ‘Bad’ category. An ‘Average’ or 3-star review looks like ‘*Hit or miss food. Corned beef was WAY too salty, the side veggies were good,*’ focused on ‘Hit,’ ‘Miss,’ ‘salty,’ and ‘good.’ The presence of both positive and negative aspects in a sentence helps the machine to classify it as an average review. The performance of the chosen classifiers at different sample sizes is shown in Fig. 1. The same thing has also been summarized in tabular format in Table 1.

Table 1 Table of accuracy scores at different sample sizes

Samples	Resampling	DT	KNN ($k = 10$)	RF	ET
30,000	SMOTE	0.591	0.644	0.715	0.768
50,000	SMOTE	0.669	0.717	0.763	0.808
70,000	SMOTE	0.567	0.636	0.687	0.749
90,000	SMOTE	0.552	0.632	0.677	0.744
110,000	SMOTE	0.557	0.633	0.672	0.739
130,000	SMOTE	0.562	0.640	0.678	0.745
150,000	SMOTE	0.554	0.628	0.672	0.733
170,000	SMOTE	0.561	0.635	0.673	0.736
190,000	SMOTE	0.544	0.626	0.664	0.729
210,000	SMOTE	0.547	0.627	0.661	0.729
Whole	SMOTE	0.556	0.633	0.670	0.735

3.7 Validation

After the validation work, we got the accuracy of the prediction and different other metrics like precision, recall, F -score from Scikit-learn's classification report. Imbalanced-Learn's classification report gave us some more measures like specificity, geometric mean, and index-balanced accuracy of the geometric mean for each class.

Precision score is the ability of the classifier not to label as positive a sample that is negative. It deals with the factor of false positivity in our classification. [12] Mathematically, precision is given as:

$$\text{Precision} = \text{No. of true positive} / (\text{No. of true positive} + \text{No. of false positive}) \quad (1)$$

The recall is the ability of the classifier to find all the positive samples. It deals with the factor of false positivity in our classification. [12] Mathematically, recall is given by:

$$\text{Recall} = \text{No. of true positive} / (\text{No. of true positive} + \text{No. of false negative}) \quad (2)$$

The F -score or F -measure can be regarded as a weighted harmonic mean of the precision and recall. [12] It reaches its best score at 1 and worst score at 0.

The geometric mean is given by the root of the product of class-wise sensitivity or recall. It tries to maximize the accuracy on each of the classes while keeping these accuracies balanced [11].

4 Results

Before delving into the analysis of results we got from our classifiers, let us first take a look at the distribution of ratings across the dataset. We have presented below distributions of both conditions, before sampling and after sampling, in the following figures. All 4- and 5-star reviews are cumulatively considered as 'Good' reviews. Collection of 2- and 1-star reviews is 'Bad' falls in 'Bad' category. We considered 3-star reviews as 'Average' reviews.

As we can see from the figures, i.e., Figs. 1 and 2, before the application of SMOTE algorithm, nearly 70% of the total samples were positive, i.e., good reviews. If we had applied machine learning techniques directly on this imbalanced dataset, our classifier would have got less trained on average or neutral reviews and bad or negative reviews. To overcome this challenge, we applied the over-sampling technique. Over-sampling methods like SMOTE synthetically create new samples of the minority classes to balance the major class. On application of SMOTE, overall sample size increases [13].

Below, we present the results that we got from our classifiers at different sample sizes.

We have tested four different classifiers using over-sampling over 10 different sample sizes and once over the whole dataset, i.e., 30 thousand, 50 thousand, 70 thousand, 90 thousand, 1.1 lakh, 1.3 lakh, 1.5 lakh, 1.7 lakh, 1.9 lakh, 2.1 lakh and the whole dataset and the results obtained are shown in Table 1. It is evident from the results that extra trees classifier gives the best results among the classifiers mentioned above. In Table 1, DT, KNN ($K = 10$), RF, and ET represent decision tree classifier, K nearest neighbor classifier ($k = 10$), random forest classifier, extra trees classifier, respectively. For KNN, we used five neighbor points, ten neighbor points, and fifteen neighbor points during fit and testing. Finally, we chose ten neighbor points for KNN as it gave us best accuracy score and precision value among them.

In Fig. 3, the comparison of accuracies given by different classifiers at different sample sizes is shown. DT, KNN ($K = 10$), RF, and ET represent decision tree classifier, K nearest neighbor classifier ($k = 10$), random forest classifier, extra trees classifier, respectively.

It is pretty much evident from the classification report in Table 2, we have achieved a decent precision across every class. Also, we have a quite good recall and F-score across the classes too.

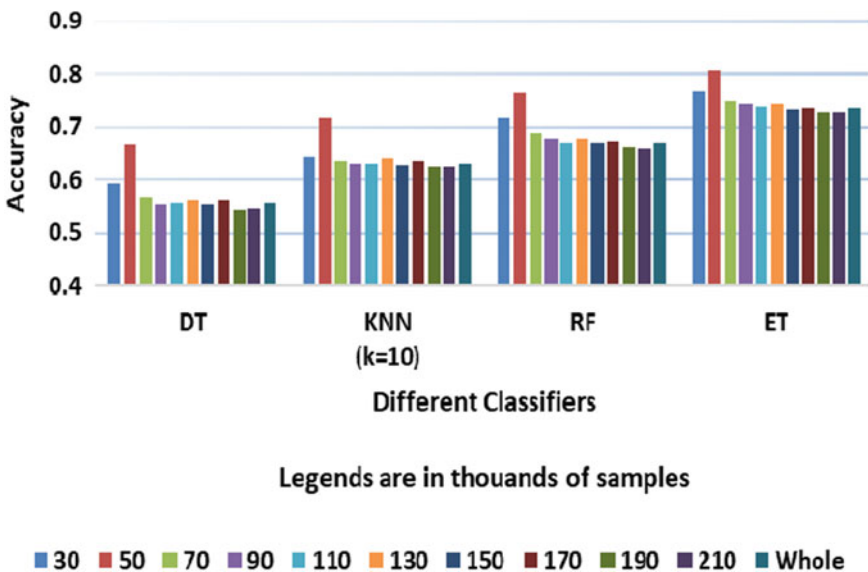


Fig. 3 Accuracy versus different classifiers plot

Table 2 Classification report of extra trees algorithm on the whole dataset

	Precision	Recall	Specificity	F1-score	Geometric mean	Index-balanced accuracy	Support
Good	0.748	0.656	0.889	0.699	0.791	0.621	39,142
Average	0.757	0.756	0.880	0.756	0.816	0.657	38,818
Bad	0.705	0.794	0.834	0.747	0.792	0.616	39,043
Avg./total	0.737	0.735	0.868	0.734	0.800	0.631	117,003

5 Future Works

In our case, without resampling the data, we got poor results. As a reviewer's intention behind a given rating may sometimes remain unclear or sarcastic, the machine might not get trained well. There might be no apparent difference between a 4-star review and a 3-star or 5-star review. Also, the initial inspection of the data revealed the dataset to be positively skewed. Resampling was done to address this issue of skewness. We plan to try out other various sampling methods in future, under-sampling or over-sampling or both, based on the nature of data and fine-tuned those using hyperparameter tuning. Better results might be yielded this way.

Parameter tuning tasks can be applied to classification methods too. Here, we have used several popular classifiers and drew a comparison between them. However, if we use other available classifiers, it might be very fruitful. An ensemble of different classifiers might prove useful too. Such works have been scheduled to get done in future.

We also intend to incorporate models for corpus-specific stopwords extraction. Such types of words are insignificant regarding the description of the corpus. Simple term-based frequency calculation and filtration can be followed. Also, filtering out words based on their inverse document frequencies can be a good approach to corpus-specific stopword extraction task. Such filtering will further reduce sparsity without hampering the useful information present. It may prove beneficial in case of large data handling.

6 Conclusions

It was interesting to find out the relation between user reviews and user-given star ratings to a business. Rather than merely classifying it between two major classes, i.e., positive and negative, classification into three levels has given a more elegant look into the reviewer's perspective. It lays a stronger foundation to further semantic and sentiment analysis. We also performed over-sampling to avoid under-fitting problems in minority classes. Classification into three sets, positive, negative, and neutral; good, bad, and average, also mimics real-world scenarios. Often, we clas-

sify some sentences as neutral sentences in practice. These sentences are not entirely negative, neither completely affirmative. Rather than classifying in positive and negative, our three-level classification helped the learning model to capture the essence of neutrality. We think this approach is very suitable for real-world applications.

References

1. J. W. McCarthy, M. Minsky, N. Rochester, C.E. Shannon, A proposal for the dartmouth summer research project on artificial intelligence. *AI Magazine* **27**(2006), 12–14 (August 31, 1955)
2. L.C. Jain, A. Quteishat, C.P. Lim, Intelligent machines: an introduction, in ed. by J.S. Chahl , L.C. Jain, A. Mizutani, M. Sato-Ilic. *Innovations in Intelligent Machines—1. Studies in Computational Intelligence*, vol 70 (Springer, Berlin, Heidelberg)
3. T. Nasukawa, J. Yi, Sentiment analysis: capturing favorability using natural language processing, in *Proceedings of the 2nd International Conference on Knowledge Capture* (ACM, 2003)
4. J. Yi et al., Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques (IEEE, 2003)
5. B. Pang, L. Lee, S. Vaithyanathan, Thumbs up?: sentiment classification using machine learning techniques, in *Proceedings of the ACL-02 Conference on Empirical Methods in Natural Language Processing*, vol. 10 (Association for Computational Linguistics, 2002)
6. S. Bose, U. Saha, D. Kar, S. Goswami, A.K. Nayak, S. Chakrabarti S, RSentiment: a tool to extract meaningful insights from textual reviews, in *Proceedings of the 5th International Conference on Frontiers in Intelligent Computing: Theory and Applications. Advances in Intelligent Systems and Computing*, vol 516, ed. by S. Satapathy, V. Bhateja S. Udgata, P. Pattnaik (Springer, Singapore, 2017)
7. Z. Yang et al., Hierarchical attention networks for document classification, in *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (2016)
8. S. Kiritchenko et al., NRC-Canada-2014: Detecting aspects and sentiment in customer reviews, in *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)* (2014)
9. G.A. Miller, WordNet: a lexical database for English. *Commun. ACM* **38**(11), 39–41 (1995)
10. C. Fellbaum (ed.), *WordNet: An Electronic Lexical Database* (MIT Press, Cambridge, MA, 1998)
11. G. Lemaître, F. Nogueira, C.K. Aridas, Imbalanced-learn: a python toolbox to tackle the curse of imbalanced datasets in machine learning. *JMLR* **18**(17), 1–5 (2017)
12. F. Pedregosa et al., Scikit-learn: machine learning in python. *JMLR* **12**, 2825–2830 (2011)
13. N.V. Chawla, K.W. Bowyer, L.O. Hall, W.P. Kegelmeyer, SMOTE: synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **16**, 321–357 (2002)

A Study on Spatiotemporal Topical Analysis of Twitter Data



Lalmohan Dutta, Giridhar Maji and Soumya Sen

Abstract In this new era of Web 2.0, people around the world are expressing their feelings, sentiments, thoughts, daily activities, and local and global events happening around them in different social networking sites like Twitter, Facebook, etc. This generates vast amount of data in social media by registered users which are geographical and temporal information-oriented. This rich data could be potentially useful information and is being extensively used nowadays for different applications like user's sentiment analysis, product or service reviews, real-time information extraction like traffic, disaster reporting, personalized message or user recommendation, and other areas. Extracting topic distribution from social media in spatial and temporal dimensions is an important research area. Hence, our focus of this study is on discussing various topical modeling techniques and their uses in different recent research works. This chapter gives a brief overview of the recent updates of spatiotemporal topical analysis using Twitter data. This study categorizes a large number of recent studies and articles in relevant area to get a summarized view of the state of the art in this field. This survey will help researchers, who are new to the domain, and provide a quick baseline for further research.

Keywords Topical modeling · Twitter data analysis · Text mining · Topical analysis

L. Dutta · S. Sen (✉)

A.K. Choudhury School of I.T University of Calcutta, Kolkata 700106, India
e-mail: iamsoumyasen@gmail.com

L. Dutta

e-mail: Lalmohan.Dutta@gmail.com

G. Maji

Department of Electrical Engineering, Asansol Polytechnic, Asansol 713302, India
e-mail: Giridhar.Maji@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_61

1 Introduction

In recent days, social networks, especially, Twitter and Facebook, have become valuable resource for sharing people's views and thoughts in diversified fields. People around the world are expressing their views on what is happening around them including any local or global events that spread awareness. It includes various fields like weather, crime, disasters like earthquake, politics, entertainment, sports, etc. Information obtained from these social networks such as Twitter and Facebook has been shown to be extremely valuable to marketing research companies, public opinion organizations, trend analysis, and other text mining entities. Twitter is a free social networking micro-blogging site that allows registered users to broadcast short posts called "tweets" of maximum length up to 280 characters, which is recently updated from 140 characters. News from these tweets is spread almost in real time all over the world. Spatiotemporal distribution of such data will be very useful to get important insights which can be used for trend analysis and future prediction through different text mining techniques.

Initial works with Twitter data were mainly focused on study of number of aspects and characteristics of Twitter started in 2007. For instance, Java et al. in [4] analyzed geographical and topological properties of Twitter data and found out the correlation between them in terms of messages (or tweets) and users. They also identified user intention and community structure. Since then, a numerous number of research works have been done with different approaches and different techniques. A good survey on usefulness of Twitter data stream for real-time event detection has been done by Hassan et al. in [3]. Major research works have been categorized based on approach (like supervised, unsupervised, historical, real-time) and techniques or tools used (like LDA, author-topical model, etc.). They are discussed in Sect. 4 in detail.

The rest of this chapter is organized as follows: Sect. 2 contains a discussion on different text mining techniques mainly focusing on Twitter data mining. Section 3 discusses a detailed overview of Twitter data and APIs available. Section 4 talks about different topical analysis and modeling approaches. Section 5 explains different challenges related to Twitter data mining. In Sect. 6, a comparative study and discussion of different research studies are depicted. Finally, Sect. 7 concludes the chapter with a brief summary and challenges ahead.

2 Text Mining Techniques

Text mining is a non-trivial process. Extracting important insights from large set of text data is fair complex and difficult due to couple of reasons especially for Twitter data. Like (1) variety of data in terms of language, short word, emoticon. (2) Complexity of natural language (3) there are many ways to represent similar concepts, leads to ambiguity and (4) High dimensionality and finally (5) text data source is unstructured and heterogeneous. (6) Tweets are not usually grammatically structured

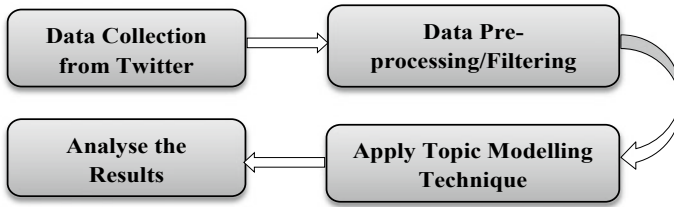


Fig. 1 Twitter text mining process

compared to regular document texts. In spite of above difficulties, text mining is widely used to get interesting facts from large dataset. Text mining for Twitter data process consists of mainly four steps: (1) gather data, (2) apply preprocessing, (3) apply topical modeling techniques, and (4) analyse the result (Fig. 1).

2.1 Collecting Data from Twitter

Data collection is the very first step for text mining. Data selection is to be done as per the application requirement. To collect data, i.e., tweets and other related information from Twitter micro-blogging site, the first task would be to create an app with Twitter API. Twitter provides three different APIs [16]: (1) Search API, to get old tweets (2) Rest API, to get user profile details, and finally, (3) Streaming API, which provides real-time tweets. In order to access those APIs, there are some keys required like Consumer Key, Consumer Secret Key, Access Token, and Access Token Secret. The APIs can be called from any programming languages like Java, Python, etc.

2.2 Data Preprocessing

Once the data is collected, that data could be noisy or inconsistent. To get more accurate result, the quality of gathered data is to be checked and refined if required. This process is called preprocessing. Preprocessing methods can contain four steps: data cleaning, data integration, data transformation, and data reduction [18]. Data cleaning is mainly used for filtering out the redundant data and inconsistent one. There are many steps involved in preprocessing like clearing stop words, removing punctuation and URLs, stemming of words, lemmatization or converting all text to lowercase, etc. [10].

2.3 *Text Mining*

Text mining is the process to extract important insights or patterns for large set of text data. There are different kinds of text mining methods which are used in the analysis of data [18]. Those are as follows: (1) information extraction, (2) categorization, (3) clustering, and (4) visualization. Information extraction is a process to retrieve information from large text-based documents or sources. Categorization is a process to categorize the retrieved information into some predefined classes (in case of supervised) or cluster to different classes (in case of unsupervised).

2.4 *Result and Analysis*

The analysis of result is the final step. The result of text mining is depicted using some graphs and charts, to have an interesting view of the extracted information. This could be based on temporal distribution and spatial distribution and is very useful for different applications. How and why results are useful to the research community and society as a whole is discussed in this section.

3 *Twitter Data Overview*

Twitter is a one of the most widely used free social networking micro-blogging sites, launched in 2006, and allows its registered users to share short posts called “tweets.” Next, we discuss how this data can be collected and used for analysis.

3.1 *Twitter API*

Twitter API is the application programming interface which can be used for posting any tweet, searching tweets with query string, getting real-time tweets, home timeline, and many other details. Twitter API is made up of different entities which are tweet object, user object, entities object, extended entities object, geo objects [16]. Figure 2 shows the object structures with major attributes and relationship among them. Table 1 contains the description of each objects.

Twitter API provides three different services: (1) Search API—used to get old tweets of last 7 days (2) Rest API—to get user profile details with friends and followers, and finally, (3) Streaming API—this is used to get real-time tweets. It provides 1% random tweets from the total volume of tweets at a particular moment [13].

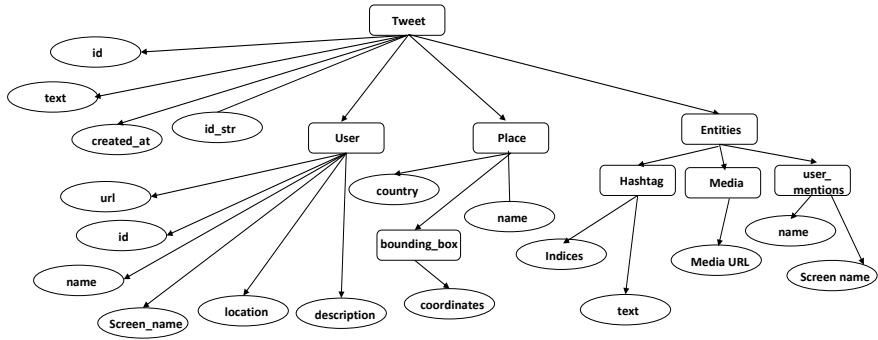


Fig. 2 Tweet object structure with major attributes

Table 1 Tweet objects and their descriptions

Object name	Description	Important data fields
Tweet	Main building block of Twitter. Contains text and other information	created_at, id, id_str, text, user, entities, source, place
User	Contains details of the author of the tweet	Id, id_str, name, screen_name, location, friend_count, language, followers_count, etc.
Entities	Consists of common tweet elements such as hashtags, URLs, mentions, polls	Hashtags, URLs, user_mentions, symbols
Extended entities	Includes photos, video, or animated GIF	Media, media_url, type, size, etc.
Place	Specifies the exact location of the tweets	Name, country, place type, bounding box, etc.

Table 2 Programming language and library for Twitter data

Programming language	Library name	Description
Java	Twitter4J, Temboo	Unofficial library from Apache
Python	Tweepy, Temboo, python-twitter	It is open-source, hosted on GitHub
C++	Twitcurl	C++ twitter API library
PHP	tmhOAuth	Opensource Twitter library for PHP
Ruby	Twitter, twitter-ads	A Twitter supported and maintained Ads API SDK for Ruby
VB Script	VBScriptOAuth	Library to perform OAuth in VBScript

3.2 *Accessing Twitter API*

To access Twitter API services, four keys would be required as follows: (1) Consumer Key, (2) Consumer Secret Key, (3) Access Token, and (4) Access Token Secret. These keys can be retrieved from user's account, after creating an app in Twitter Web portal. Every registered user can get those keys for each app they create in portal. Library to access such APIs is available in many programming languages as listed in Table 2.

3.3 *Application of Twitter Data*

As Twitter is capturing a huge set of events, product reviews, or feelings per minute, it is a good resource for detecting and analyzing different events happening around the world. Twitter data is undoubtedly useful, and there are variety of applications using the data which includes user's sentiment analysis, product or service reviews, real-time information extraction like traffic, disaster reporting, personalized message or user recommendation, and other areas.

There are quite a number of challenges in gathering the required Twitter data which includes ethical, methodological, and legal issues [1]. Authors in [14] defined the challenges as (a) **Volume**—the huge dataset, leading to vastness in volume; (b) **Velocity**—the speed of data creation in real time; (c) **Variety**—the available data can be of many different forms like unstructured or structured for a data source; and (d) **Veracity**—uncertainty with regard to data quality.

4 **Topical Modeling of Twitter Data**

Research on topical modeling is still an evolving area. Our reviewed researches can be broadly categorized on different categories based on different characteristics. We categorize into two parts like (1) approach-based and (2) technique-based. Approach-based can be classified into different types like real-time analysis, historical/ retrospective analysis, supervised and unsupervised. Technique-based could be classified based on different tools and techniques used like LDA-based modeling, author-topic-based modeling, etc. (1) Event or topic type which is detected by Twitter data stream like pre-specified events or unspecified events (2) The detection coverage like New Event Detection (real time) and Retrospective Event Detection (historical) and (3) event or topic detection mechanism like supervised, unsupervised or hybrid.

4.1 Approach-Based

The used approach and analysis tools and techniques vary with the application. Different approach-based topical modelings are described below.

4.1.1 Real-Time Topical Analysis

This kind of study uses Twitter Stream API for twitter live data collection, filtering with specific topics or specific region. There are many applications of real-time topical analysis like traffic condition monitoring, real-time news monitoring, etc. Authors in [17] proposed a framework for real-time Twitter data analysis related to FIFA World Cup 2016 for detecting relevant topical extraction. The analysis of the tweets is done using improved soft frequent pattern mining (SFPM) technique.

4.1.2 Historical Topical Analysis

Historical or retrospective topical analysis is another aspect which gives temporal topical distribution over a long period of time. Twitter provides a Search API which allows collection of historical data of last 7 days. Historical analysis could be used in many applications including identifying trends of any topic at any geographical region. Researchers in [18] extracted stream data for IPLT20 hashtags and analyzed temporal distribution of tweets with discussion on IPLT20. Researchers in [4] proposed a Geo-H-SOM (geographic, hierarchical, self-organizing map) model to do analysis on geographical, temporal, and semantic distribution of geo-referenced tweets. The said model considered tweets to be “similar” when the distance between two tweets is small in semantic, in geographic and timely manner. An efficient novel micro-blog event discovery algorithm is proposed and discussed in [15]. The algorithm used the concept of Symbolic Aggregate AppRoXimation algorithm (SAX), which is used for time-series analysis. They demonstrated that the proposed algorithm is able to detect and trace patterns from collective tweet dataset in a very precise way with lower complexity compared to other similar algorithms.

4.1.3 Topical Analysis on Pre-specified Topics

Most of the reviewed researches are based on pre-specified topics. They defined few topics and classified the collected tweets in different topics. Like [5] used 768 topics spread across 18 classes to analyze 23,000+ tweets. In [12], Sriram et al. classified collected dataset into a predefined set of classes like events, news, product reviews, opinions, and private messages based on author demography and domain-based features extracted from dataset such as short words, slang words, time phrases, hashtag, and user mentioned using @ sign within the tweet.

4.1.4 Topical Analysis on Unspecified Topics

Many researches have been done where there is no pre-specified topic to classify tweets. Rather the tweets are clustered into different groups using unsupervised method. So, the list of topics is not known in advance. Authors in [2] did analysis if Twitter can replace the traditional newswire sources (BBC, Google News, CNN, Guardian, Reuters, The Register, New York Times, etc.) for breaking news. They performed a rigorous analysis of different major and minor events collected for a period of 2 months in newswire and tweets using Stream API. To accomplish this, they analyzed two sets of events. They did a comparative study of news source to find out, which one is more in real time, which type of news are spread in Twitter and can Twitter replace newswire. They found that Twitter has comparatively better coverage of news for local and relatively less important events that are mainly sometimes ignored by the newswire.

4.2 *Technique-Based*

Till now researchers have used wide variety of modeling techniques to overcome certain limitations over another. Most widely used techniques are Naïve Bayes (NB), Naïve Bias multinomial (NBM), LDA, modified versions of LDA, named entity recognition, PageRank algorithm, SAX, neural networks, etc. Some of the most used techniques are explained below.

4.2.1 LDA Topical Modeling

The most frequently used probabilistic topical model is Latent Dirichlet Allocation (LDA). In LDA, each document (tweets in our context) is considered as a mixture of topics where each word in the tweet has certain probabilities. LDA is a bag-of-words model mainly used in unsupervised fashion, where the list of topics is pre-specified. This is widely used to extract topics from tweets because of the fact that mainly a single tweet represents a single topic. Steiger et al. in [13] used LDA model extensively to identify spatiotemporal and semantic distribution and co-relation of tweets.

4.2.2 Naïve Bayes Multinomial

Naïve Bayes multinomial (NBM) is a supervised learning method, which is a variation of Naive Bayes (NB) text classification technique. Equation 1 shows the equation for NBM, where $P(c | d)$ is the probability of a tweet d falls under class c , $P(c)$ is prior probability of a tweet occurring in class c , and $P(t | c)$ is the conditional prob-

ability of any term t occurring in a tweet of class c . Lee et al. in [5] used NBM to categorized trending topics into 18 different categories.

$$P(c | d) \propto P(c) \prod_{1 \leq k \leq n_d} P(t | c) \tag{1}$$

4.2.3 Graph-Based Clustering Algorithms

There are several graph-based clustering algorithms that have been proposed and used in Twitter data mining. Long et al. in [7] used hierarchical divisive clustering method on a co-occurrence graph to categorize topic words into event clusters.

4.2.4 PageRank and Probabilistic Scoring Model

PageRank is another important technique used in many applications. Zhao et al. [19] proposed a context-sensitive PageRank algorithm for ranking keywords within a document and finding the co-relation or interestingness between words. They not only focus on single keyword, but also identify potential keyphrases which defines a topic. The page rank of their proposed method is defined as:

$$R_t(w_i) = \lambda \times \sum_{j:w_j \rightarrow w_i} et(w_j, w_i)/ot(w_j)R_t(w_j) + (1 - \lambda)P_t(w_i) \tag{2}$$

Here, they computed from w_j to w_i propagation in the context of any topic t . $et(w_j, w_i)$ is basically the edge weight between word w_j to w_i , which is nothing but the number of co-occurrences of these words.

5 Major Challenges

There are quite number of challenges in Twitter data mining, where lots of researches are going on. All major challenges are discussed in brief here.

1. **High-Quality Data Collection:** Twitter data is huge is size, and also the speed of creating new data is very fast. Finding out the required data from the vast volume is one of the major challenges. The variety of data distributed over different dimensions creates the situation more complex. Like tweet can be of different languages, different formats, like “hashtags,” “user-mentioned,” link, image, video, emoticon, and short form in place of original vocabulary word. The presence of such data makes the source data noisy and incomplete. So, data cleaning is to be properly done to get the only required data. Otherwise, the presence of such irrelevant data might impact the result.

2. **Tweet Size Limitation:** Tweet can be maximum of 140 characters. This limitation forces users to post the tweet using short words, a large number of spelling and grammatical errors which make it difficult to find out the exact topic under discussion. Although recently the size is increased to 280 now, it is currently in beta testing phase and not available for all users.
3. **Finding Tweet Location:** To perform spatiotemporal analysis of Twitter data, the location of tweet is an important information. But, sometimes, user does not reveal the location of tweet. Initially, till 2012, Twitter API used to provide exact locations of the tweet in latitude and longitude for each tweet if the user made the location public. But later on, due to some security issues, this was no more allowed. This is another challenge to derive the location of the tweet, when the user does not reveal it. Alternatively, it can be derived from the location of user or location of majority of friends or location of retweets or finally, location mentioned in the tweet itself. However, this derivation might not be 100% correct.
4. **Ironic or Sarcastic Tweets:** Ironic and sarcastic tweet has completely different meaning and is very difficult in topical modeling or event extraction. It is quite difficult also to identify if a certain tweet is sarcastic or not. It can affect the analysis result in a different way.
5. **Software Architecture:** Usage of proper software architecture is another challenge. There are quite number of tools and software available to text mining. Each of them has different advantage and disadvantages. Choosing the right tools for a particular analysis is very much important, and it is truly based on the problem statement to be worked upon.

6 Discussion

The studies discussed in preceding section mainly work on the principle of supervised and/or unsupervised method. Each of them has an uniqueness in terms of approach or dataset used or the methodology applied or maybe the application area. But the central theme of the discussed studies is to have efficient topical modeling of Twitter data so that important insight could be obtained from noisy data. Increasing the efficiency and scalability is still an open area in topical modeling. Performance analysis and complexity evaluation of different approaches is another important concern. Although topical modeling described in the selected representative studies spans across variety of applications, it is very new with respect to other trending research problems like fake message/news identification or fake network detection, recommendation problem, etc. The efficient and powerful algorithms need to be devised and applied in such type of scenarios. Table 3 shows a summarized comparative study of some important representative articles in the field of topical modeling on Twitter data.

Table 3 Summary of recent representative studies on topical modeling

Paper Reference	Category	Input	Outcome	Methodologies used
Li et al. [6]	Event detection and analysis	Twitter data stream for crime and disaster events (1 M tweets)	GUI based tool for event detection and visualization	Crawling with iterative rules, classification model
Zhao et al. [19]	Topical extraction	Large dataset of Twitter streams	Summarized Twitter content	PageRank and probabilistic scoring model
Yadav et al. [18]	Term frequency analysis of real-time events	Twitter stream for sports like IPLT20	Time-series and term frequency analysis	Filtering, information extraction for text mining
Soto et al. [11]	Data quality challenges in Twitter content	Twitter streams related to HealthCare	Solution strategies of the challenges	Named entity recognizer
Lee et al. [5]	Trending topic classification	23000+ tweets from 18 classes	Categorizing trending topic into 18 categories	Naive Bayes multinomial classifier
Petrović et al. [9]	Twitter and Newswire comparison for real-time breaking news	Newswire article (4.7 K) and tweets (51 M)	Coverage of Twitter is more. But Newswire are more realtime most of the cases	Pair similarity, event detection algorithm
Gaglio et al. [2]	Framework to analyze teal-time Twitter data	Tweets related to FIFA World Cup 2014	Framework for topic detection	Soft frequent pattern mining
Stilo et al. [15]	Temporal mining of Twitter data	Twitter stream collected for 2 months (160 M English tweets) related to sports, politics, disasters, health and celebrities	A micro-blog event discovery algorithm	SAX for characterization for the temporal behavior of events Symbolic Aggregate Approximation algorithm

7 Conclusion and Future Work

Twitter data analytics is a relatively new research area, but many researchers are working in this field. This chapter contributes to the overall idea on topical modeling on Twitter data along with a summary of the main challenges, and difficulties currently are there in different steps of the Twitter data analytics. We discussed about different steps in text mining including gathering Twitter data, data cleaning, information extraction, and knowledge discovery. These findings are quite relevant to

practitioners; current businesses are increasingly looking for extracting the meaningful fact from Twitter data. This chapter would further help the researchers as a literature survey endeavoring to work in this area. Although there are various types of researches going on in this field, the following can be different future scope of work in this area. (1) As tweets are sometimes written in short words, due to limitation of the size of tweets, for topical extraction, it is important to get original words from such short words. Although certain algorithm exists to get this, still they are not 100% reliable and accurate. There are so many scopes to work on this field to improve the accuracy of finding original words from short words. (2) Integrating Twitter data with other supporting demographic data into a data warehouse [8] and then a comprehensive analysis can be considered. (3) Considering language distribution for topical extraction is another aspect. Topical extractions till now was language-dependent. Further NLP could be used to have generic topical extraction solution multiple languages. (4) Further researches to be done on better NLP filtering to extract only related information of interest.

References

1. W. Ahmed, Challenges of using twitter as a data source: an overview of current resources, <http://blogs.lse.ac.uk/impactofsocialsciences/2015/09/28/challenges-of-using-twitter-as-a-data-source-resources/>. Accessed 3 Apr 2018
2. S. Gaglio, G.L. Re, M. Morana, A framework for real-time twitter data analysis. *Comput. Commun.* **73**, 236–242 (2016)
3. M. Hasan, M.A. Orgun, R. Schwitter, A survey on real-time event detection from the twitter data stream. *J. Inf. Sci.* **44**(4), 443–463 (2017)
4. A. Java, X. Song, T. Finin, B. Tseng, Why we twitter: understanding microblogging usage and communities, in *Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 Workshop on Web Mining and Social Network Analysis* (ACM, 2007), pp. 56–65
5. K. Lee, D. Palsetia, R. Narayanan, M.M.A. Patwary, A. Agrawal, A. Choudhary, Twitter trending topic classification, in *2011 IEEE 11th International Conference on Data Mining Workshops (ICDMW)* (IEEE, 2011), pp. 251–258
6. R. Li, K.H. Lei, R. Khadiwala, K.C.C. Chang, Tedas: a twitter-based event detection and analysis system, in *2012 IEEE 28th International Conference on Data Engineering (ICDE)* (IEEE, 2012), pp. 1273–1276
7. R. Long, H. Wang, Y. Chen, O. Jin, Y. Yu, Towards effective event detection, tracking and summarization on microblog data, in *International Conference on Web-Age Information Management* (Springer, 2011), pp. 652–663
8. S. Mandal, G. Maji, Integrating telecom CDR and customer data from different operational databases and data warehouses into a central data warehouse for business analysis. *Int. J. Eng. Res. Technol.* **5**(2), 516–523 (2016)
9. S. Petrović, M. Osborne, V. Lavrenko, Streaming first story detection with application to twitter, in *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics* (Association for Computational Linguistics, 2010), pp. 181–189
10. P. Sharma, A. Agrawal, L. Alai, A. Garg, Challenges and techniques in preprocessing for twitter data. *Int. J. Eng. Sci. Comput.* **7**(4), 6611–6613 (2017)
11. A. Soto, C. Ryan, F. Peña Silva, T. Das, J. Wolkowicz, E. Milios, S. Brooks, Data quality challenges in twitter content analysis for informing policy making in health care (2018)

12. B. Sriram, D. Fuhry, E. Demir, H. Ferhatosmanoglu, M. Demirbas, Short text classification in twitter to improve information filtering, in *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* (ACM, 2010), pp. 841–842
13. E. Steiger, B. Resch, A. Zipf, Exploration of spatiotemporal and semantic clusters of twitter data using unsupervised neural networks. *Int. J. Geogr. Inf. Sci.* **30**(9), 1694–1716 (2016)
14. S. Stieglitz, M. Mirbabaie, B. Ross, C. Neuberger, Social media analytics-challenges in topic discovery, data collection, and data preparation. *Int. J. Inf. Manage.* **39**, 156–168 (2018)
15. G. Stilo, P. Velardi, Efficient temporal mining of micro-blog texts and its application to event discovery. *Data Mining Knowl. Discov.* **30**(2), 372–402 (2016)
16. Twitter, Twitter developer platform, <https://developer.twitter.com/en/docs/basics>. Accessed 29 Mar 2018
17. W. Xie, F. Zhu, J. Jiang, E.P. Lim, K. Wang, Topicsketch: real-time bursty topic detection from twitter. *IEEE Trans. Knowl. Data Eng.* **28**(8), 2216–2229 (2016)
18. G. Yadav, M. Joshi, R. Sasikala, Twitter data analysis: temporal and term frequency analysis with real-time event, in *IOP Conference Series: Materials Science and Engineering*, vol. 263 (IOP Publishing, 2017), p. 042081
19. W.X. Zhao, J. Jiang, J. He, Y. Song, P. Achananuparp, E.P. Lim, X. Li, Topical keyphrase extraction from twitter, in *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, vol. 1 (Association for Computational Linguistics, 2011), pp. 379–388

A Secure Steganography Scheme Using LFSR



Debalina Ghosh, Arup Kumar Chattopadhyay,
Koustav Chanda and Amitava Nag

Abstract Steganography is a technique to hide the secret data inside some other cover files (like text, image, audio, etc.) in such a way that it prevents the detection of the secret data within the cover files. In this paper, a new scheme of concealing secret data has been introduced where the locations of hiding the secret bits will be generated by linear feedback shift register (LFSR). Using LFSR, we will compute random numbers (within a given range) which will determine the specific pixels where secret bits of the secret data will be concealed. We use least significant bit (LSB) technique where the secret bit will be stored at the least significant bit position of the pixel. Choosing the cover pixels from cover image randomly will strengthen the security of LSB technique.

Keywords Steganography · LSB · LFSR · Grayscale image · Secret text

1 Introduction

The term steganography [1] came from Greek words “*stegos*” meaning “cover” and “*graphia*” meaning “writing” that is covered writing. It is a technique to hide information in a cover medium like text, image, audio or video file. There is always a comparison between cryptography and steganography. In cryptography, the secret data will be encoded in some unintelligible form (meaningless) with some specific key. An attacker without having the key does not gain any knowledge about the secret data. But the entire secrecy of the information depends on the key. Once the key is exposed, encrypted data are not secured any more. But in steganography, it hides

D. Ghosh (✉) · A. K. Chattopadhyay
Institute of Engineering & Management, Kolkata, India
e-mail: debalinag1986@gmail.com

K. Chanda
Academy of Technology, Kolkata, India

A. Nag
Central Institute of Technology, Kokrajhar, India

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_62

the existence of the secret data in the cover media. The secret is concealed in such a manner that the existence of the secret data remains unknown to the observer or attacker. In this paper, we will use LSB steganographic technique. Among steganography methods, LSB [2] is a very well-known technique. In LSB, the least significant bits of the cover image pixels are used to hide the secret data. The traditional LSB technique is simple, but not very secure. In some modified LSB schemes like [3], a few bits from the most significant side determine the position at least significant side where to hide the secret bit.

To increase the security of LSB steganography, we use the concept of LFSR (a random number generator) [4] in steganography. Linear feedback shift register (LFSR) is a shift register where the input bit is a linear function of the previous state. The initial value of the LFSR is called seed. The operation of the shift register is deterministic. If the current state is known, then the next sequence of values can be determined. LFSR having a well-chosen feedback function can generate a large sequence of random bits. The bit stream generated by LFSR is pseudo-random and also satisfies the cryptographic randomness criteria. Two main parts of LFSR are the shift register and the feedback function. The task of a shift register is to shift the contents of the register to their adjacent places in one direction, such that one position at the other end becomes empty. That position remains empty unless a new content will be entered into the register. The new content will be generated by a linear function. The inputs are the contents of the filled positions. There is an exception in LFSR—if all the contents of the shift register are zeros, then the next state cannot be generated.

In traditional LSB technique, consecutive bytes are used to store the secret. In our scheme, we will use LFSR-based random number generator [5], and these random numbers will determine the pixels within the cover image, where the secret bits will be hidden. As a result, the sequence of the pixels that hide the secret is a random sequence.

The rest of the paper is divided into following major parts: in Sect. 2, we have discussed a few previous works done in the domains of steganography and LFSR; Sect. 3 comprises the proposed algorithm, and in Sect. 4 we conclude the paper.

2 Related Study

Steganography has been used in variety of domains like transform domain, spread domain, image domain, etc. In image steganography, the secret message will be hidden in some cover image or images. The cover image selection is also very important for a few algorithms, because at the time of fetching secret data, the recovery of the secret should be lossless. There are different types of methods available in digital image steganography to hide secret data within the cover images. One of them is least significant bit (LSB) technique. Kumar et al. [6] discussed about file format in steganography commonly used for cover medium like text, image, audio, video, protocol, etc. They have also performed a comparative study of steganography al-

gorithms where image is the cover object. Most significant among these techniques are least significant bit (LSB) and discrete cosine transform (DCT). The detailed analysis of these two techniques is performed in [7].

2.1 Least Significant Bit (LSB) Method

In this method, the secret data are concealed at least significant bit position of the consecutive pixels of the cover image. Following is the algorithm [8] to embed secret text message using grayscale image as cover.

Method of Embedding of Secret

Step 1: Read the secret message and the cover image.

Step 2: Convert the secret message to binary sequence.

Step 3: Find out the LSB of each pixels in the cover image.

Step 4: Replace the least significant bit of each cover pixel in the cover image with a secret bit (from the secret message) one by one.

Step 5: Generate the stego-image.

Method of Extraction of Secret

Step 1: Read the stego-image.

Step 2: Calculate the least significant bits from each pixel of the stego-image.

Step 3: Retrieve the LSBs and convert each eight bits into a character. Arrange the characters in sequence to reconstruct the secret message.

2.2 A Brief Review of LFSR

An LFSR can be considered as a finite automaton. A finite automaton contains finite number of states. Instead of states, LFSR consists of a number of stages. We are considering a LFSR of length L over \mathbb{F}_q . LFSR having length L contains L stages, where stage $S = S_t$, $t \geq 0$ and generates semi-infinite sequence of elements of \mathbb{F}_q . It must satisfy the linear recurrence relation of degree L over \mathbb{F}_q :

$$s_{t+L} = \sum_{i=1}^L c_i s_{t+L-i} \quad \forall t \geq 0.$$

where c_1, c_2, \dots, c_L are the feedback coefficients of LFSR. The LFSR of length L consists of L delay cells called stages. The contents of L stages are $s_t, s_{t+1}, \dots, s_{t+L-1}$ which form one state of LFSR. Let the initial content be s_0, s_1, \dots, s_{L-1} .

An external clock is used to control the LFSR. In each clock pulse, the bits will be shifted toward right and the content of rightmost stage S_t is the output. Now the new content at the leftmost stage is the feedback bit s_{t+L} , and it is computed as:

$$s_{t+L} = \sum_{i=1}^L c_i s_{t+L-i}.$$

The main weakness of traditional LSB steganography is that the secret bits are stored in consecutive pixels of the stego-image. For strengthening the security of LSB method, randomness in selection of cover pixels is important. LFSR is a random number generator and popularly used in many modified steganography schemes [9–11] and their applications. Random numbers are of two types, one is “truly random” and the other one is “pseudo-random.” The random numbers generated using some mathematical algorithms are of the pseudo-random type. Digital random number generators are used in different cryptography applications. The authors in [12–14] have proposed different types of 8-bit and 16-bit LFSR-based random number generators for high-secured multipurpose operations.

In [12], a polynomial modulator has been proposed to avoid the predictability in random numbers. Because periodic and predictable random sequences will help the intruder to find the secret. The authors in [9] implemented steganography with random number generator. They have introduced an algorithm where we found four stages. First stage generates pseudo-random sequences (pseudo-random sequences are generated by linear feedback shift register and standard chaotic map), permutation and XORing using pseudo-random sequences conducted in second stage. Encryption using Rabin cryptosystem and steganography using the improved diagonal queues are the last two stages. In [10], the authors proposed the use of random bit sequences for the purpose of steganography. The random bit sequence has been generated by linear feedback shift registers (LFSRs), and a new factor named beta factor has been created. The task of beta factor is to select cover from the database of covers.

3 Proposed Method

Let the cover image be I_c (of size $w \times h$) and the secret text be T_s which is in binary form of any given length l , $T_s \in \{0, 1\}^l$. The steps in the construction of stego-image are as follows.

3.1 Embedding of Secret Bits

3.1.1 Initialization

Use padding technique (add enough rows and columns of zeros) such that cover-image matrix I_c can be divided into blocks of $n \times n$ matrices (n must be chosen such that $n = 2^k$ where $k \geq 2$ and $n \leq \min\{\frac{w}{2}, \frac{h}{2}\}$). The number of such matrices will be as follows:

Number of matrices row-wise, $row_count = \lceil \frac{w}{n} \rceil$

Number of matrices column-wise, $column_count = \lceil \frac{h}{n} \rceil$

Total number of matrices, $N = row_count \times column_count$. Let the matrices are represented as $M_{(i,j)}$ where $1 \leq i \leq row_count$ and $1 \leq j \leq column_count$.

3.1.2 Selection of Blocks for Bit Insertion

We consider $M_{(i,j)}$ as cover-matrix, and the secret bits will be inserted at the least significant position of each pixel belongs that block.

The block selection method is as follows:

```

for  $i = 1$  to  $row\_count - 1$  do
  | for  $j = 1$  to  $column\_count - 1$  do
  | | select the matrix  $M_{(i,j)}$  as cover matrix
  | end
end

```

We intentionally do not use any of the matrices from row_count^{th} row and $column_count^{th}$ column as due to padding those may be filled with zeros.

3.1.3 Selection of Different Seeds for LFSR at Different Blocks

To ensure randomness to choose the cover pixels in different blocks, we compute the seed for each matrix $M_{(i,j)}$ as follows.

Initialize the first seed of r bits (r depends on the polynomial used for LFSR) from MSB part of the first pixel ($P_{[(1,1),(1,1)]}$) of first matrix ($M_{(1,1)}$). The seed for rest of blocks or matrices will be generated as follows.


```

seedprev = 0
for for  $i = 1$  to row_count - 1 do
  for for  $j = 1$  to column_count - 1 do
    select the matrix  $M_{(i,j)}$  as cover matrix
    select the  $r$  bits from the MSB of pixel  $P_{[(i,j),(1,1)]}$  as  $seed_{new}$ 
     $seed = seed_{prev} \oplus seed_{new}$ 
     $seed_{prev} = seed$ 
  end
end

```

3.1.4 Insertion of Length of Secret Bit Stream for the Secret Text

The length of the bit stream for the secret text which is l will be inserted into the first block, i.e., $M_{(1,1)}$. The l is converted into binary form, and the bits will be inserted into the least significant position of the cover-pixels selected by the order generated by the LFSR algorithm. The LFSR generates $2 \times k$ bits at a time, where first k bits decide the x-coordinate (p) and second k bits decide the y-coordinate (q); the cover pixel hence chosen is $P_{(p,q)}$ of matrix $M_{(i,j)}$.

3.1.5 Selection of Cover Pixel from $M_{(i,j)}$ for Bit Insertion

Consider a flag matrix $FM_{(i,j)}$ corresponding to each $M_{(i,j)}$, which stores the value one or zero. We initialize all flag matrices with value zero. From LFSR, we generate $2 \times k$ random bits. The first k bits will determine the row number (p), and next k bits will determine the column number (q) within the matrix and presented as pixel $P_{[(p,q),(i,j)]}$. The secret bit from T_s will be inserted at LSB of pixel $P_{[(p,q),(i,j)]}$ (pixel $P_{(p,q)}$ of matrix $M_{(i,j)}$). Each time a secret bit is inserted at the least significant position of the pixel at matrix position (p, q), the value at corresponding position at flag matrix will be changed to one. The flag matrix keeps the trace of which pixels are already used to store secret bits.

If the same pixel selected more than once as a cover pixel (a collision occurs), we utilize *linear probing technique*. In *linear probing technique*, if the cover pixel is already utilized, we scan the matrix for next unused pixel to insert the secret bit. The process will be continued until either we utilized all the pixels of the same matrix or we are exhausted with secret bits. In case all the pixels of the same matrix are already used, the process will be repeated for the next matrix.

3.1.6 Construction of Stego-Image

Once we are exhausted with all the secret bits (all the secret bits are already inserted into cover image I_c), we generate the stego-image I_{stego} .

3.2 Extraction of Secret Bits

The stego-image I_{stego} is considered as input.

3.2.1 Initialization

Generate the block matrices $M_{(i,j)}$ of size $n \times n$ (where $1 \leq i \leq row_count$ and $1 \leq j \leq column_count$). The row_count and $column_count$ computation is as discussed in the Sect. 3.1.1.

3.2.2 Selection of Block for Bit Extraction

Select the block matrices $M_{(i,j)}$ sequentially as discussed in Sect. 3.1.2.

3.2.3 Selection of Different Seed for LFSR at Different Blocks

The selection of the seed for LFSR at different blocks as discussed in Sect. 3.1.3.

3.2.4 Extraction of the Length of Secret Bit Stream for the Secret Text

The length of the secret bit stream is concealed in the first block $M_{(1,1)}$. Using LFSR, we generate the random bit sequence. We combine each $2 \times k$ bits to find the coordinate $P_{(p,q)}$ (where p and q are calculated from first and second k bits) where the secret bit is hidden. Extract the secret bit from the least significant position of the selected pixel. Hence, combine all the secret bits to construct the length (l) of the secret text (in binary form).

3.2.5 Identification of Cover-Pixel from $M_{(i,j)}$ and Bit Extraction

For each block matrix $M_{(i,j)}$, the cover-pixels are identified by the LFSR algorithm. The LFSR is used to generate $2 \times k$ bits, where first k bits construct the p , the x -coordinate and the second k bits construct the q , the y -coordinate. The (p, q) gives us the pixel position. The secret bit is at the least significant bit of pixel $P_{(p,q)}$, which is extracted. If all the l secret bits are extracted, then those can be combined to generate the secret text T_s .

4 Conclusion

In the proposed scheme, we have considered a cover image to conceal a secret text message. The cover image is first segmented into equal size blocks. Then, those blocks are selected one by one to hide secret bits of the text message. Within the block, each pixel stores one bit of secret message. Each secret bit is inserted at the least significant bit position of one of the pixels. The sequence of the cover-pixels within each block is generated by LFSR-based random number generator. It also ensures that the seeds selected for the LFSR are different for different blocks. The use of pseudo-random sequence of the cover pixels to store the secret bits makes the scheme more secure.

References

1. F.A.P. Petitcolas, R.J. Anderson, M.G. Kuhn, Information hiding—a survey. *Proc. IEEE*. **87**(7), 1062–1078 (1999)
2. N. Kaur, S. Behal, A survey on various types of steganography and analysis of hiding techniques. *Int. J. Eng. Trends Technol.* **11**(8), 388–392 (2014)
3. P. Pathak, A.K. Chattopadhyay, A. Nag, A new audio steganography scheme based on location selection with enhanced security, in *First International Conference on Automation, Control, Energy and Systems (ACES)* (IEEE, Hooghly, India, 2014), pp. 1–4
4. W. Mao, Y. Li, C.H. Heng, Y. Lian, Zero-bias true random number generator using LFSR-based scrambler, in *2017 IEEE International Symposium on Circuits and Systems (ISCAS)* IEEE, Baltimore, MD, USA (2017), pp. 1–4
5. A. Kaur, H.K. Verma, R.K. Singh, 3D playfair cipher using LFSR based unique random number generator, in *2013 Sixth International Conference on Contemporary Computing (IC3)*. IEEE, Noida, India (2013), pp. 18–23
6. R. Kumar, K. Choudhary, N. Dubey, An introduction of image steganographic techniques and comparison. *Int. J. Electron. Comput. Sci. Eng.* **1**(3), 1000–1005 (2012)
7. E. Walia, P. Jain, N. Navdeep, An analysis of LSB & DCT based steganography. *Global J. Comput. Sci. Technol.* **10**(1), 4–8 (2010)
8. R. Garg, T. Gulati, Comparison of Lsb & Msb based steganography in gray-scale images. *Int. J. Eng. Res. Technol.* **1**(8) (2012)
9. M. Jain, A. Kumar, R.C. Choudhary, Improved diagonal queue medical image steganography using Chaos theory, LFSR, and Rabin cryptosystem. *Brain Inform.* **4**(2), 95–106 (2017)
10. I.A. Sattar, M.T. Gaata, Image steganography technique based on adaptive random key generator with suitable cover selection, in *2017 Annual Conference on New Trends in Information & Communications Technology Applications (NTICT)*. IEEE, Baghdad, Iraq (2017), pp. 208–212
11. T. Jamil, A. Ahmad, An investigation into the application of linear feedback shift registers for steganography, in *IEEE SoutheastCon 2002*. IEEE, Columbia, SC, USA (2002), pp. 239–244
12. P. L'Ecuyer, F. Panneton, A new class of linear feedback shift register generators, in *2000 Winter Simulation Conference Proceedings*. IEEE, Orlando, FL, USA (2000), pp. 690–696
13. M. Sahithi, B.M. Krishna, M. Jyothi, K. Purnima, A.J. Rani, N.N. Sudha, Implementation of random number generator using LFSR for high secured multi purpose applications. *Int. J. Comput. Sci. Inf. Technol.* **3**(1), 3287–3290 (2012)
14. M. Han, Y. Kim, Unpredictable 16 bits LFSR-based true random number generator, in *2017 International SoC Design Conference (ISOCC)*. IEEE, Seoul, South Korea (2017), pp. 284–285

Crowd Behavior Analysis and Alert System Using Image Processing



Sayan Dutta, Sayan Burman, Agnip Mazumdar and Nilanjana Dutta Roy

Abstract This paper deals with a prototype model, which aims to detect pedestrians and analyze the crowded scene. In a real-time environment, there are many practical issues that need to be taken into consideration while analyzing a crowded scenario such as occlusion and failure in tracking people. Certain method and algorithm that can deal with such problem and provide an efficient solution is discussed. Computer vision and image processing algorithm are utilized for human tracking and analysis for crowd behavior based on contour detection and background modeling. The aim of this paper is to implement a successful algorithm that can predict any abnormal activity in a crowded scenario and helps to prevent it.

Keywords Crowd management · Computer vision · Image processing · Background subtraction · Contour detection · Optical flow

1 Introduction

With increasing growth in population, the problem of crowd surveillance has become a major issue. Various strategies are being issued to detect and maintain crowd induced anomaly. Crowd detection from a static platform holds the promise of increased safety for pedestrians. This is a difficult problem as the concerned object is in different shape, size, and color against cluttered background and varying illumina-

S. Dutta (✉) · S. Burman · A. Mazumdar · N. D. Roy
Institute of Engineering & Management, Salt Lake Electronics Complex,
Gurukul, Y-12, Sector V, Kolkata 700091, India
e-mail: sayan7048@gmail.com

S. Burman
e-mail: sayanburman2@gmail.com

A. Mazumdar
e-mail: agnip.mj2011@gmail.com

N. D. Roy
e-mail: nilanjana.duttaroy@iemcal.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_63

tion conditions. Computer vision is an integral part of today's surveillance industry. Certain computer vision algorithms are present that can help us to cope with such issue and create an effective crowd disaster management system. The system can be implemented in order to predict future crowd motivated disaster and help handle and provide solution to the situation efficiently. Human monitoring is a very time-consuming, inefficient, and expensive approach to crowd surveillance. A rather automated and efficient procedure is proposed in this paper that can make crowd surveillance a much easier task. Object detection is a vital part of this process. In order to detect the object of interest, which in this case is a pedestrian, certain background modeling algorithm can be put into effect. Various pre-processing steps are also suggested that can be performed to remove various distortions or noises that might be present while extracting video frames from the surveillance video. In a crowded environment, it is very difficult for humans to analyze each and every individual. In such situation, video surveillance system is the need of the hour. This kind of system can be implemented in areas where the crowd density is particularly high such as stadium, markets, or malls. Constant video surveillance system can be used to keep track of individuals, and whenever any abnormal situation is detected by our proposed algorithm, an alert signal is sent as a text message (SMS) to the corresponding authorities after which the desired action can be taken to contain the crowd induced disaster. The Crowd detection algorithm immediately detects crowd formations, and an alarm is triggered when a specified number of people are observed to be cluttered as a group, predicting the buildup of an unwanted situation.

2 Related Works

The motivation and the study of certain established systems in this domain are described in this particular section. Certain methods [1, 2] utilize crowd flow and semantic scene to trace and detect unusual activity. But these methods can only be applied for non-complex scenario.

2.1 *Crowd Density Analysis*

It is a prerequisite to all the measures of crowd analysis, and lot of emphasis has been given over crowd density estimation because of its importance and effectiveness in managing and controlling any anomaly in crowded situation. In video surveillance system, crowd analysis can be performed by monitoring all closed circuit televisions (CCTV) and implementing image processing techniques related to crowd detection, tracking and inspecting the anomalies. However, a lot of these systems fail when presented with hundreds of people for processing [3]. Whenever the amount of people present for detection is in question, these systems provide unreliable results.

2.2 Crowd Flow Segmentation and Stability Analysis

With the dawn of digital age, the need for constant vigilance and surveillance of public places is increasing considerably. The need for closed-circuit security systems aka CCTV is increasing to keep constant monitoring over areas such as airports, malls, or stadiums. It is necessary to have an infrastructure that integrates a chain of cameras covering all the major hotspots in a city. A number of intelligent surveillance systems have thus been developed that can undertake the processing of the video footage generated by these surveillance cameras and take necessary actions. However, despite the development of such systems, the desired potential of scrutiny required for real-time monitoring and detection has not yet been achieved.

To tackle such issue, a framework was developed which recreates the crowded scenario and bypasses the low-level object localization and tracking altogether. This was achieved by treating the crowded scene as a fluid flow where Lagrangian particle dynamics is employed to detect dominant crowd flow segments [4].

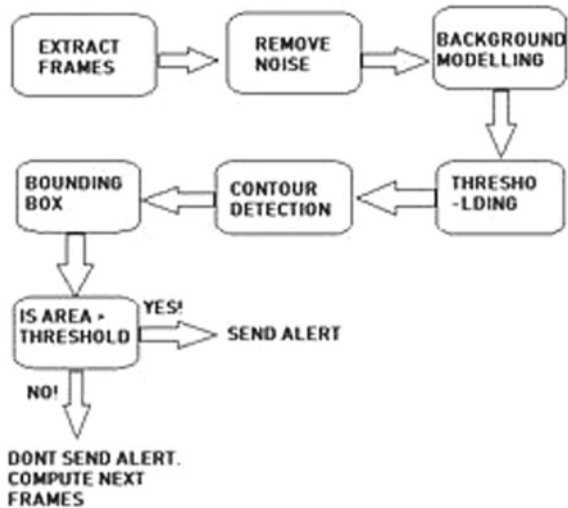
3 Methodology

The proposed method consists of four stages: (i) Pre-processing, (ii) video processing, (iii) pedestrian identification, and (iv) alert system.

A proposed framework of the architecture has been provided.

Figure 1 describes a flow diagram of the proposed architecture starting from extracting the initial frames from the video to performing all the processing algorithms to detecting a crowded region followed by an alert signal.

Fig. 1 Framework of the proposed architecture



3.1 Pre-processing

Image noise is any unwanted variation of information in images and is usually an aspect of electronic noise. It can be produced by the internal circuitry of a digital camera. Image noise is an unwanted by-product of image capture that hampers the processing of images/frames and results in unreliable data [5]. For our case study, the video obtained from the surveillance camera will be certainly subjected to a lot of noise and hence certain steps need to be ensured to guarantee a noise-free video and reliable processing. Various algorithms are presented that will help in image/frame de-noising and are provided by OpenCV. Any random variable having its mean as zero can be considered as noise. Consider a noisy pixel

$$p = p(0) + n \quad (1)$$

where $p(0)$ denotes the true value of pixel and n is the noise in that pixel. Ideally, we should get

$$p = p0 \quad (2)$$

since the mean of noise is zero. This is the basic concept of non-local means of de-noising [6]. The function that is provided by OpenCV for de-noising of colored images/frames is: `cv2.fastNlMeansDenoisingColored()` [6]. There are various parameters associated with this function which is described briefly as: the parameter for deciding the filter strength. A higher value for this parameter results in higher removal of noise but also results in loss of image details. The other two parameters include the template and search window size which are generally preferred as odd value.

Once the pre-processing stage is complete, the extracted frames that have undergone pre-processing can now be further evaluated.

3.2 Video Processing

Consecutive pre-processed frames which are extracted from the surveillance video are then subjected to a number of algorithms. Initially, the background from the extracted frames is subtracted by computing the change in the consecutive frames. Since the background for our case study is static thus any change in the foreground will be considered as our potential object and will be treated as a pedestrian. Background modeling plays a very important part in a video analysis system. Background subtraction is used to fragment the objects in a video frame.

Background subtraction is a very reliable approach whenever any static camera is used. Background subtraction is a technique which is used in the field of image processing and computer vision where an image's foreground is extracted for further

Fig. 2 Image of busy road with multiple pedestrians without any filter



Fig. 3 Image of busy road with multiple pedestrians with background subtraction algorithm



processing (object recognition etc.) from its background. Generally, it is an image's regions of interest (humans, cars, text etc.) in its foreground upon which any further processing is to take place [7].

Figure 2 shows an image of a busy road with multiple pedestrians walking. Figure 3 shows the processed image after background subtraction algorithm was performed on it. The above-displayed figure has been accessed from the OpenCV documentation [8]. Once the background has been extracted from the consecutive frames, the resultant frame is then ready for further processing and analysis.

3.3 Pedestrian Identification

The extracted foreground (pedestrians) in the processed frames has a very prominent contour. Contours can be well demonstrated as a curve joining a group of points, having same pixel value or intensity. The contours are a useful tool for shape analysis and object detection. For better accuracy, it is preferred to use binary images. So before finding contours, thresholding is applied [9]. Thresholding is a process which separates out regions of an image corresponding to objects which needs to be analyzed. This separation is based on the variation of intensity between the object pixels and the background pixels [10]. Figure 4 shows the application of thresholding. The image on the left side is the actual test image and the image on the right side

Fig. 4 An image of an apple subjected to thresholding



shows the threshold algorithm applied to it. The following displayed figure has been accessed from the OpenCV documentation [10]. Thresholding can be performed by the following function present in OpenCV: `cv.threshold (src, thresh, maxval, type[, dst])`. The parameters associated with this function are described:

`src` input array.
`dst` output array of the same size and type and the same number of channels as `src`.
`thresh` threshold value.
`maxval` maximum value to use with the `THRESH_BINARY` and `THRESH_BINARY_INV` thresholding types.
`type` thresholding type.

Once the thresholding operation is performed, the contour detection algorithm is applied which identifies the contour in the processed frame. Since this algorithm is being applied on the extracted foreground (pedestrians), the enclosed object bounded by the contour when mapped onto the actual video frame will show us individual pedestrian. Each such contour can then be enclosed in a bounding rectangle to represent an individual object which in this case will be a pedestrian.

The contour function that can be used is also provided by OpenCV. The function is: `cv.findContours()`, `cv.drawContours()`. The functions are used to find and then draw the contours on images/frames, respectively, [9].

Thus an algorithm that can detect individual pedestrian to enclose them in a bounding box is implemented. The next stage is to identify the situation when there will be a crowd. This can be achieved by the concept of overlapping of the bounding boxes. Whenever a certain number of bounding boxes will overlap, it will signify that two or more pedestrians are standing together. We can implement the concept of intersection over union [11] to find the overlap of bounding boxes and then further bound them with a single bounding box. This can help us in grouping a number of pedestrians or crowd together. Once the crowd has been successfully detected, we can then simply calculate the total area of the bounding box and compare it with a threshold value that can be estimated and updated explicitly by a user depending on the environment in which the system is being operated. The threshold value will vary according to the place in which the system is being set up. It will be a different value

for mall or market place or stadium. Or, we can also provide a normalized value. If the area of the maximum bounding rectangle is above the threshold value, we can trigger an alarm. There can be multiple non-overlapping bounding rectangles, which signify that there will be more than one crowd patch. If there are multiple bounding rectangles which do not overlap, then the area of each such bounding rectangles needs to be individually compared with the threshold value. If any of the bounding boxes have an area greater than the threshold value, an alarm is triggered.

The proposed concept of optical flow can also be implemented with this algorithm once the bounding box recognizing individuals has been developed, to find the movement of pixels within these boxes.

3.4 Optical Flow

The optical flow describes the direction and time rate of the pixels in a time sequence of two consequent images. This method generally describes the time rate and direction in a time between two consequent images. A 2D velocity vector is now imputed to each and every pixel present in a particular place. To make computation simpler and fast working, we now add three-dimensional objects to it, i.e., 3D + time. Then only, we can describe any image as the sum up of 2D brightness function of location and time.

$$I(x, y, t) = I(x + \delta x, y + \delta y, t + \delta t) \tag{3}$$

Now using Taylor series for the right-hand side of (1), we get:

$$\begin{aligned}
 I(x + \delta x, y + \delta y, t + \delta t) &= I(x, y, t) + \frac{\partial I}{\partial x} \delta x \\
 &+ \frac{\partial I}{\partial y} \delta y + \frac{\partial I}{\partial t} \delta t + H.O.T.
 \end{aligned} \tag{4}$$

Neglecting higher order terms,

$$I_x \cdot v_x + I_y \cdot v_y = -I_t \tag{5}$$

Optical flow methods are extremely demanding.

Horn–Schunck and Lucas–Kanade methods are extremely popular which introduce us to an error term in each pixel. Generally, it is determined by taking the sum of weighted smallest squares of gradient present in the closest neighborhood for that specific pixel.

$$\rho_{LK} = \sum_{x,y \in \Omega} W^2(x, y) [\nabla I(x, y, t) \vec{v} + I_t(x, y, t)]^2 \tag{6}$$

where W is the neighborhood of pixel concerned and $W(x, y)$ are the individual weights of different pixels. For finding the minimum error, we need to derive error term by the single components of velocity and then by putting the resultant to zero.

$$\vec{v} = [A^T W^2 A]^{-1} A^T W^2 \vec{b} \quad (7)$$

So now, we get the solution for the system as resultant velocity vector for each pixel. Horn–Schunck method in addition to gradient constraint as in Lucas–Kanade, Horn–Schunck method adds another error term called as global term for smoothing and is thus used for limiting greater changes of optical flow components in W .

$$\rho_{\text{HS}} = \int_D (\nabla I \vec{v} + I_t) + \lambda^2 \left[\left(\frac{\partial v_x}{\partial x} \right)^2 + \left(\frac{\partial v_x}{\partial y} \right)^2 + \left(\frac{\partial v_y}{\partial x} \right)^2 + \left(\frac{\partial v_y}{\partial y} \right)^2 \right] dx dy \quad (8)$$

Here, D is the domain for the total image which is the consequence for the added term.

3.5 Alert System

Once we have positively detected patches of crowd, an alert system is implemented that will direct a text message or SMS as soon as the detection is triggered. The numbers of the local authorities will be stored in a database that will be notified as soon as any unusual activity is observed and immediate response can be taken against the situation.

4 Conclusion and Future Work

This paper concludes how it is possible to use simple, well-established image processing algorithm and perform certain key processing techniques on video frames to create a crowd behavior management system that aims to alert authorities whenever there is any abnormal behavior in crowd pattern. The processor specification upon which the algorithm is to be tested is: Intel(R) Core(TM) i5-6200U CPU @ 2.30 GHz, 2401 MHz, 2 core(s), and 4 logical processor(s) with 8 GB of internal memory. All of the above-mentioned algorithm can be implemented using OpenCV and Python. Open Source Computer Vision (OpenCV) is a library of programming functions aimed at real-time computer vision. The library is cross-platform and free to use under the open-source BSD license [12]. Thus, using open-source software

and existing image processing techniques, an efficient and automated crowd detection system can be developed which can detect crowd gathering and also notify the authority to prevent any unwanted crowd-related hazard.

References

1. J. Li, S. Gong, T. Xiang, Scene segmentation for behavior correlation, in *Proceedings of European Conference on Computer Vision*, pp. 383–395 (2008)
2. R. Mehran, B.E. Moore, M. Shah, A streakline representation of flow in crowded scenes, in *Proceedings of European Conference on Computer Vision*, pp. 439–452 (2010)
3. P.V.V. Kishore, R. Rahul, K. Sravya, A.S.C.S. Sastry, Crowd density analysis and tracking, in *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)* (2015), pp. 1209–1213
4. S. Ali, M. Shah, A Lagrangian particle dynamics approach for crowd flow segmentation and stability analysis, in *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)* (2007)
5. R. Verma, J. Ali. A comparative study of various types of image noise and efficient noise removal techniques (2013)
6. OpenCV, Image Denoising (Online). Available—https://docs.opencv.org/3.4/d5/d69/tutorial_py_non_local_means.html. Accessed July 2018
7. S.S. Mohamed, et al., Background modelling and background subtraction performance for object detection, in *2010 6th International Colloquium on Signal Processing & its Applications* (2010), pp. 1–6
8. OpenCV, Background Subtraction (Online). Available—https://docs.opencv.org/3.3.0/db/d5c/tutorial_py_bg_subtraction.html. Accessed: July 2018
9. OpenCV, Contour Detection (Online). Available-https://docs.opencv.org/3.4.0/d4/d73/tutorial_py_contours_begin.html. Accessed: July 2018
10. OpenCV, Thresholding (Online). Available <https://docs.opencv.org/2.4/doc/tutorials/imgproc/threshold/threshold.html>. Accessed July 2018
11. Pyimagesearch, Intersection of Union (Online). Available <https://www.pyimagesearch.com/2016/11/07/intersection-over-union-iou-for-object-detection/>. Accessed July 2018
12. OpenCV, About OpenCV (Online) Available—<https://opencv.org/>. Accessed July 2018

Review Article on Magnetic Resonance Imaging



**Shatadru Majumdar, Rashmita Roy, Madhurima Sen
and Mahima Chakraborty**

Abstract MRI is serving human beings over the past 30–40 years as a chief investigator for many clinical problems. Since the development of MRI in the early part of 1970, MRI has progressed a lot and occupies an important position in the modern medical field. MRI is a robust imaging technique that furnishes internal images of living organisms, thus helping in easy identification of various diseases in our body and inflammation of different organs [1]. This article aims to describe history of MRI in brief, basic sequences of procedure, algorithm used in MRI, applications of MRI, future prospects of MRI and certain demerits of MRI.

Keywords Submagnetization · Larmor frequency · Relaxation time · Nuclear magnetization resonance

1 Introduction

MRI is a non-intrusive imaging technology helping us in taking internal images of our body [2]. MRI is an important tool in disease detection, diagnosis and in treatment monitoring by producing three-dimensional detailed anatomical view without using detrimental radiation. MRI depends on magnetic property of atomic nuclei and it detects the change in direction of rotational axis of proton found in the water molecules that build up the living organisms. Very detailed images can be made of

S. Majumdar (✉) · R. Roy · M. Sen · M. Chakraborty
Department of Computer Science and Engineering,
Institute of Engineering and Management, Kolkata, India
e-mail: shatadru1999@gmail.com

R. Roy
e-mail: rashmitaroy2000@gmail.com

M. Sen
e-mail: senmadhurima99@gmail.com

M. Chakraborty
e-mail: mahimaC75@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling
and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_64

soft tissues like soft tissues, muscles and brain. MRI is consequently very flexible technique that provides measure of both structure and function.

2 History of MRI

A brief history of MRI is given below in a tabular form is given below [3].

1857–1952	Larmor relationship-Sir Joseph Larmor
1930	Isidor Isaac Rabi succeeded in detecting single state rotation of atoms & molecules.
1946	MR phenomenon-Bloch & Purcell
1952	Nobel Prize-Bloch & Purcell
1950, 1960, 1970	NMR developed as analytical tool
1972	Computerized tomography
1973	Back projection MRI-Lauterbur
1977	Echo-planar imaging
1980	FT MRI demonstrated-Edelstein
1987	MR Angiography-Dumoulin
1991	Nobel prize-Ernst
1992	Functional MRI
2003	Nobel prize-Lauterbur & Manfield

MRI is basically a medical imaging technique based on NMR [Nuclear Magnetic Resonance] phenomenon. Almost three decades after the discovery of Nuclear Magnetic Resonance simultaneously and independently by Bloch and by Purcell in 1946, the 1st imaging experiment was held in 1970 and by Lauterbur and by Damadian.

With the advent of computer tomography in 1973 by Hounsfield and echo-planar imaging (a rapid imaging technique) in 1977 by Mansfield after this many scientists over the next 20 years developed MRI into the technology that we now have known today.

3 Algorithm

The basic idea of this algorithm is to obtain the final image by creating a matrix containing the evolution of global magnetization vector with time, for all values of gradient setting. The flowchart is given below. Each voxel of the simulated object has an associated list containing the coordinates of center, proton density, T_1 and T_2 values, stationary magnetic field and chemical shift (in ppm units). The user is said to specify the pulse sequence. The sequence is described by a table with six

columns: one for elapsing time, three for the values of gradients (G_x , G_y , G_z), one for Radio Frequency pulse, one for the number of acquired sample. The signal acquired by MRI scanner depends on the global magnetization, but the algorithm cannot deal with the complicated nature of global magnetization as a whole. So, it takes the magnetization of each voxel, for a particular time instant, from a vast collection of magnetization vectors. The magnetization vector of each voxel is called “submagnetization” (sub_mag). The global magnetization vector (write_list) is the summation of all the sub_mag values. A particular time instant is denoted by time_inst. The present values of gradient are denoted by grad_comb. The direction of submagnetization is modified by the application of Radio Frequency pulse. The rotational phenomena associated with submagnetizations are explained with the help of Bloch equations. Associated with submagnetization rotation are the relaxation phenomena, described by well-known exponential functions of time, whose time constants are T_1 , T_2 . The relaxations are applied by multiplying the submagnetizations, calculated according to Bloch equations, by relaxation function (Applicate Bloch and relaxation). When the acquisition process starts, the submagnetization components (the two components which are at 90° to the applied main magnetic field and defined as “real” and “imaginary” components) are added to the present contents of a linked list (add sub_mag to write_list) [4]. This process is repeated for all values of time_inst corresponding to a given value of sub_mag. This entire operation is repeated for all values of sub_mag. Thus, the resulting list gives the values of global magnetization vector corresponding to different values of time_inst. The resulting signal represents the simulated MRI signal collected by the given acquisition sequence, from the given sample, for a particular value of gradient setting or grad_comb. Now, the content of write_list is collected in a matrix (img_mat), write_list is re-initialized, the value of grad_comb is updated and the whole acquisition process is repeated for this new value of grad_comb. The acquisition finally ends when all values of grad_comb have been explored. The final image is obtained by 2D FFT of img_mat, after filtering img_mat (Fig. 1).

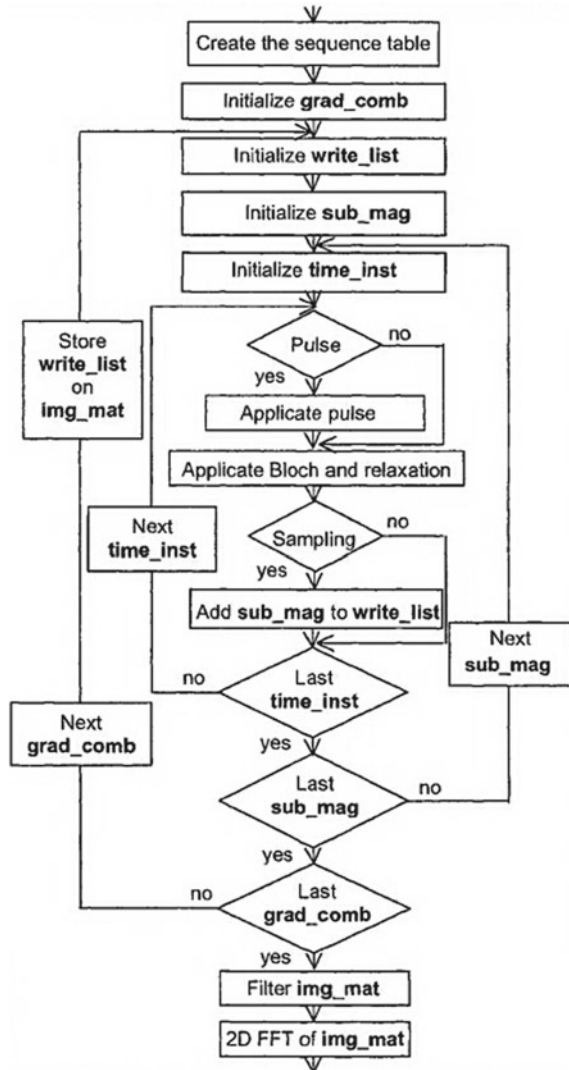
Another useful algorithm is the Correction Algorithm (Fig. 2).

4 Procedure

The process of MRI can be broadly broken down into four main points which are as follows [6]:

1. Preparation
2. Excitation
3. Spatial Encoding
4. Signal Acquisition.

Fig. 1 Flow diagram of the MRI simulation algorithm. Variables are indicated in bold: **grad_comb** is the index of the current gradient setting, **write_list** contains the sum of submagnetizations for the current **grad_comb**, **sub_mag** and **time_inst** are the indexes of the current submagnetization and of the current time instant, respectively, and **img_mat** is the matrix containing raw data [4]



4.1 Preparation

MRI scanner produces a static magnetic field, and the patient’s body is kept inside this field. 60% of human body consists of water (H₂O) molecules, thus contains a lot of hydrogen atoms. Hydrogen atoms are also present in some other compounds of human body.

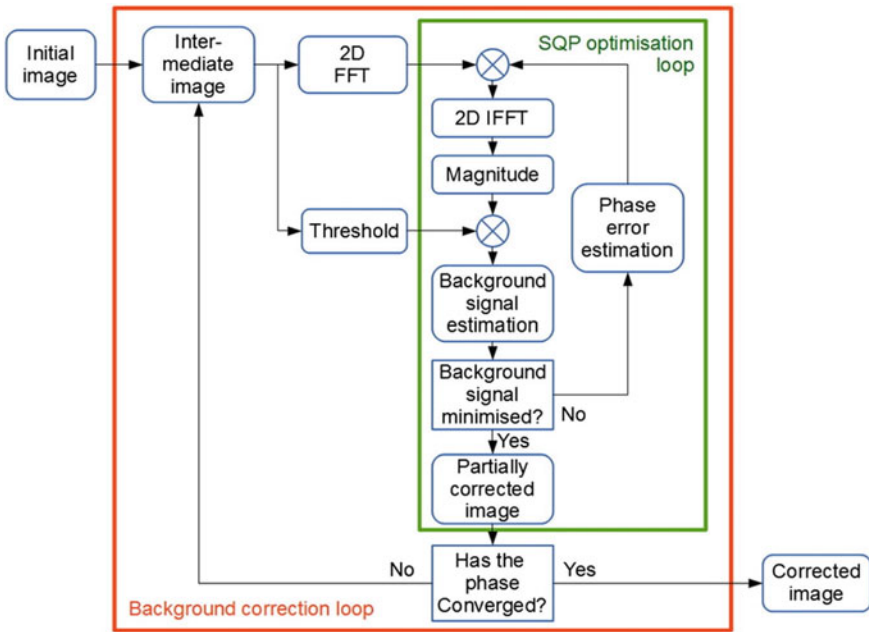


Fig. 2 Correction algorithm [5]

Magnetic Field B_0

The main static magnetic field in MRI is represented by B_0 and is measured in tesla. The majority of the MRI systems in clinical use are 1.5T, with increasing number of 3T systems being installed.

Larmor frequency

The application of B_0 produces a spin or precession of nuclei around it. The Larmor frequency is a measure of this rate of precession [7]. It depends on B_0 according to the equation

$$\omega = \gamma * B_0 \tag{1}$$

where, ω = Larmor frequency in MHz, γ is the gyromagnetic ratio in MHz/T. B_0 is the main magnetic field in T.

“Spin up” and “spin down” states

Due to application of B_0 , the hydrogen atoms align themselves parallel to as well as antiparallel to the B_0 .

The parallel and antiparallel states of the electrons correspond to the low and high energy state, respectively. These two states have an energy difference (ΔE) proportional to B_0 , γ and Plank constant (h).

$$\Delta E = \gamma * B_0 * h / (2 * \pi) \tag{2}$$

Net magnetization vector

In presence of B_0 , the parallel and antiparallel nuclei with equal and opposite magnetic moments cancel each other. But, the number of nuclei parallel to B_0 is always more (by a small amount) than those antiparallel to it and this difference is termed as NMV (Net Magnetization Vector) denoted by M [8].

4.2 Excitation

Scanner emits an RF (Radio Frequency) pulse at the time of image acquisition process. If the Larmor frequency matches with the RF pulse, then only the RF pulse is at resonance.

Resonance and Radio Frequency

When a nuclei gets exposed to RF radiation at Larmor frequency, then it jumps from a lower energy state to a higher energy state, this causes NMV to spiral away from B_0 . The NMV after a certain time rotates 90° from longitudinal position and lies in the transverse X - Y plane. It is the position that can be detected on MRI [9].

The application time of Radio Frequency pulse should be such that it makes the direction of NMV 90° to B_0 . When a receiving coil is put in the surrounding area of the tissue, the transverse magnetization will produce an electric current by induction, known as Nuclear Magnetic Resonance (NMR) signal. The loss of coherence of spin system attenuates NMR signal with a time constant, transverse relaxation time T_2 . Consecutively NMV slowly relaxes toward its equilibrium location parallel to B_0 , with time constant T_1 relaxation time. The contrast in MR results from the fact that T_1 , T_2 values are different for different tissues [10].

4.3 Spatial Encoding

Spatial encoding is done by using a varying magnetic field gradient (varies linearly depending on the spatial location) because the resonant frequency of the hydrogen atoms depends on the strength of the applied field. Lorentz force acts on the coil because the gradient coils are at 90° to B_0 . The gradient coils are quickly turned on and off to create the varying magnetic field which causes vibration and a loud noise during acquisition process.

4.4 Signal Acquisition

When magnetic field gradients are used, the obtained NMR signal contains various frequencies. This is called MRI signal. It is converted from analog to digital prior to processing. The mathematical operation linked with it is called Fourier transform.

5 Applications of MRI

1. MRI is a really important tool in radiology used to form pictures of anatomy and physiological processes of our body [11].
2. MRI detects infection in different organs.
3. MRI detects various kinds of degenerative diseases.
4. MRI plays a major role in the detection of strokes and tumors and chemical imbalances in the body.
5. MRI helps in identifying different kinds of musculoskeletal disorders.
6. MRI also helps in the identification of abnormalities in the central nervous system, specially in posterior fossa and brain stem (MRI identifies better than computed tomography) [12].

6 Current Role and Future Prospects of MRI

1. MRI plays an important role in analyzing several types of cancer including breast cancer. MRI has already been successful in case of detecting ovarian malignancies and tumor. MRI can play a vital role in detecting breast cancer at an early stage, including ductal carcinoma in situ. Mammography is not ideal for detecting breast cancer, for this issue in the future MRI can be used as the chief tool to detect breast cancer [13].
2. MRI should enhance its sensitivity and specificity with higher field strength should be enabled to detect vast array of metabolites.

In the future, progress in MRI will help in detecting cancer in early stages and will improve our ability in diagnosis.

7 Demerits of MRI

1. Due to the involvement of a large amount of electricity, MRI scans produce very loud noises while processing [14].
2. MRI scanners are really very expensive.
3. MRI scanners can be affected by movements so they are not suitable for problems such as mouth tumors as coughing or gulping can make the images hazy [15].
4. For an MRI scan, a patient has to remain in an enclosed environment, which can be a serious problem for claustrophobic patients [16].
5. MRI sometimes fails to differentiate between malignant and benign tumors, thus giving incorrect results [17].

8 Conclusion

Magnetic Resonance Imaging (MRI) uses a highly imaging technology to produce highly detailed images of the organs and tissues of living organisms. It helps in detection of problems related to blood circulation and is safe as there is no involvement of any kind of radiation. It is a relatively advanced technology. MRI has got wide applications in different clinical fields but there are limited applications in the pharmaceutical sciences. Despite all the advantages of MRI, there are also certain disadvantages such as: production of very loud noises while processing and very high expenses of MRI scans. But, the number of advantages is much more than the number of disadvantages. Day by day, there is rapid development in the field of imaging. Hence, MRI is becoming more reliable with time.

References

1. Wikipedia Article on Magnetic Resonance Imaging, https://en.m.wikipedia.org/wiki/Magnetic_resonance_imaging
2. Dr. T. Peter, (Research Paper) Name: MRI—An Insight. *Int. J. Sci. Res.* **7**(4) (2018)
3. Girish Katti, Syeda Arshiyaa Ara, Ayesha Shireen, Magnetic resonance imaging (MRI)—a review. *Int. J. Dent. Clin.* **3**(1), 65–70 (2011)
4. G. Placidi, M. Alecci, A. Sotgui. A general algorithm for magnetic resonance imaging simulation: a versatile tool to collect information about imaging artefacts and new acquisition techniques. IOS Press Ebooks, in *Health Data in the Information*, Society Studies in Health Technology and Informatics, Vol. 90, pp.13–17
5. L.M. Broche, P.J. Ross, G.R. Davies, D.J. Lurie, Simple algorithm for the correction of MRI image artefacts due to random phase fluctuations. *Mag. Reson. Imaging.* **44**, 55–59 (2017)
6. Dr. D.J. Bell and Dr. J.R. Ballinger et al., MRI physics. Radiopaedia, org, <https://radiopaedia.org/articles/mri-physics>
7. Dr. D.J. Bell and Dr. J. Jones et al., Larmor frequency. Radiopaedia, org, <https://radiopaedia.org/articles/larmor-frequency>
8. Dr. D.J. Bell and Dr. J. Jones et al., Net magnetization vector. Radiopaedia, org, <https://radiopaedia.org/articles/net-magnetisation-vector>
9. Dr. D.J. Bell and Dr. J.R. Ballinger et al., Resonance and radiofrequency. Radiopaedia, org, <https://radiopaedia.org/articles/resonance-and-radiofrequency>
10. Dr. J. Ray Ballinger et al., Relaxation. Radiopaedia, org, <https://radiopaedia.org/articles/relaxation>
11. D. Cammoun, K.A. Davis, Clinical applications of magnetic resonance imaging-current status. *West. J. Med.* **143**(6), 793–803 (1985)
12. University of Windsor Website, http://web2.uwindsor.ca/courses/physics/high_schools/2006/Medical_Imaging/mriapplication.html
13. J. Klostergaard, K. Parga, R.G. Raptis, Current and future applications of magnetic resonance imaging to breast and ovarian cancer patient management. *P. R. Health Sci. J.* **29**(3), 223–31 (2010)
14. Magnetic Resonance Imaging (MRI) Scans Website, <https://mriscans.weebly.com/advantages-and-disadvantages.html>
15. MRI of Trinidad and Tobago Ltd Website, <http://mritnt.com/education-centre/faqs/what-are-the-advantages-and-disadvantages/>

16. Sharecare Website, <https://www.sharecare.com/health/mri-magnetic-resonance-imaging/are-there-drawbacks-mri-scans>
17. Cancer Quest Homepage, <https://www.cancerquest.org/patients/detection-and-diagnosis/magnetic-resonance-imaging-mri>

An Intelligent Traffic Light Control System Based on Density of Traffic



Kriti Dangi, Manish Singh Kushwaha and Rajitha Bakthula

Abstract Traffic congestion is one of the severe problems in all urban cities. In recent years, there was a vast increment in a number of vehicles, but simultaneously the roads are not developed or widened on the same phase due to which it leads to traffic congestion. This further raised unnecessary waiting at traffic signals which caused frustration in pedestrians as well as drivers. Thus, this paper proposes an efficient and intelligent traffic control system based on the density of the traffic along with the safe pedestrian crossing (which was not considered a major issue in the literature methods). The performance of the proposed method has shown better results in comparison with the literature methods.

Keywords Traffic congestion · Traffic signal control · Image processing · Image difference · Auto Canny · Background updation · Starvation

1 Introduction

Traffic management is an important issue which impacts our daily life routine. The common reason for traffic congestion is poor traffic prioritization. As the number of vehicles is increasing in a very fast phase, the cities' infrastructure should also get upgraded simultaneously. But, this does not happen in real time and hence leads to traffic jams, especially during rush hours. So during rush hours, people had to wait for long hours by wasting their valuable time. Improper regularization of traffic may also lead to more carbon emissions, leading to an unhealthy environment for the people.

K. Dangi · M. S. Kushwaha · R. Bakthula (✉)
Motilal Nehru National Institute of Technology, Allahabad, Uttar Pradesh, India
e-mail: rajitha@mnnit.ac.in

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_65

1.1 Current Mechanisms Deployed in Traffic Control System

The density of the traffic can be managed or controlled by using the following three mechanisms.

Manual Controlling System: Manual controlling system, the name itself indicates that it requires a human power to control the traffic signals. Various countries employ the manpower in different roles depending on the traffic density in various cities. This man has the signboard, sign light and whistle to perform the task of controlling the traffic. The main problems with manual controlling systems are: First, they do not have the longer view of the traffic density; second, they may not be available 24×7 h; thirdly, a huge manpower is required during rush hours.

Fixed time Controlling Systems: Fixed time traffic light system uses electric sensors to control the traffic. The traffic control signals are automatically controlled (for on/off) based on the inbuilt timer. They usually range from 35 to 120 s. These static systems are not capable to handle the issues of starvation of the traffic due to the high density of vehicles.

Dynamic Time Control Systems: The traffic signals are controlled automatically and dynamically for on/off based on the density of the traffic. These systems are still in the development stage. Few researchers had proposed some schemes for the same in recent years. They used microcontrollers, sensors, neural networks, machine learning algorithms, etc., for achieving the success in this task. A dynamic video surveillance system is also used to detect traffic density (using image processing techniques).

2 Literature Review

A lot of techniques are proposed in the literature to detect traffic density and control traffic signals statically/dynamically. Traffic density has been measured using various techniques in the literature like RFID and GSM which are used for monitoring the traffic by Vidhya [1]. Here, a monitoring software system had been developed using wireless coordinator, router, RFID tags and GSM modems. M.P Sinhmar [2] also developed an intelligent system using IR sensors and microcontrollers. As these two methods are developed using sensors, so any object which is not of the vehicle may also be treated as a vehicle. Video surveillance systems are also been developed by many researchers [3–5] for the same. Evolutionary algorithms and fuzzy logic [6–12] are also been used by many researchers. Neural networks [13–15] also played an important role in the development of automatic traffic control systems. Reinforcement learning [16] is also proposed for the traffic signal control system. Edge detection method is used to control traffic in [17].

3 Proposed Methodology

Proposed method works in two phases: one for training and one for testing. The overview of these two phases is shown in Fig. 1, and the same is described detailedly in the following sections.

3.1 Training Phase

At every traffic signal, an initial image is captured and named as background_image when there is no traffic at all.

Image Preprocessing: The obtained background image is pre-processed for further processing. Here, the color image is converted to gray scale in order to save computational time. Then, a Gaussian blur method is applied to reduce the noise from it (as shown in Figs. 2 and 3).

3.1.1 Edge Detection Using Auto Canny Method

Auto Canny edge detection [18] is applied to the grayscale image as shown in Fig. 4. This algorithm simply takes the median of the image and then constructs upper and lower thresholds based on a percentage of this median. A sigma is defined, as a scalar value that specifies the deviation allowed from this median. Sigma can be used to vary the percentage thresholds that are determined based on statistics. A lower value

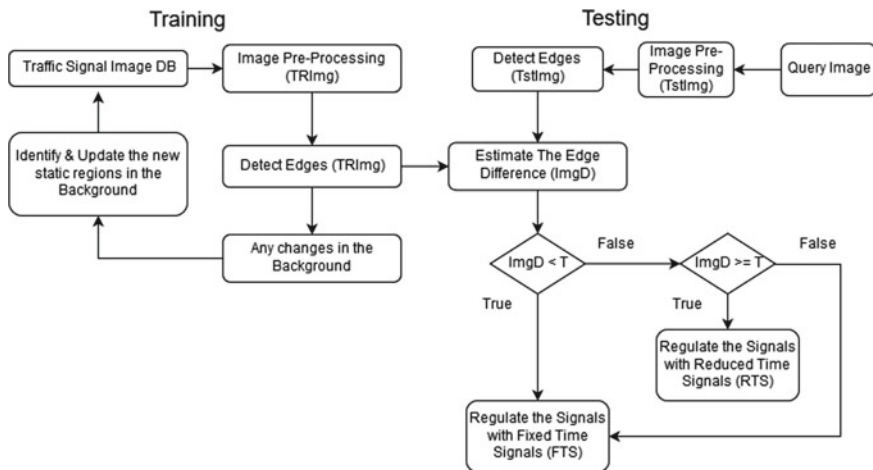


Fig. 1 Proposed method flow graph

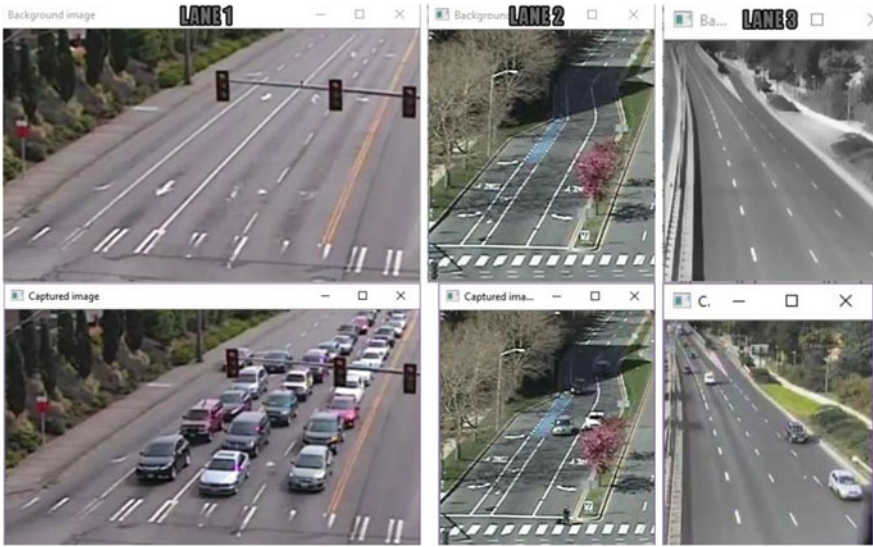


Fig. 2 Sample images of traffic



Fig. 3 Preprocessed images of Fig. 2 (grayscale and blurred images)

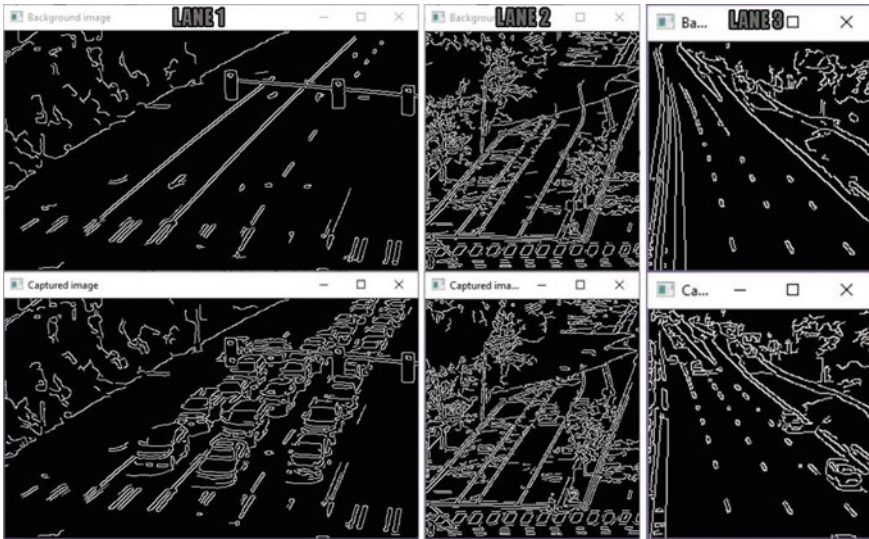


Fig. 4 Images after applying auto Canny

of sigma indicates a tighter threshold, whereas a larger value of sigma indicates a wider threshold. In this paper, sigma is chosen as 0.33 (as it tends to give good results on most of the dataset used).

Image Dilation: After applying Canny edge detection method, there could be broken edges between the edge segments. These broken edges may not depict the outline of the vehicles clearly. Thus, the edge image is dilated to fill in broken gaps between edge segments and preserve the details of the vehicles as shown in Fig. 5.

Background Update: This is the most important step in computer vision applications, including real-time object tracking and event analysis from a video. This application software is developed using a common share frame which acts as a background frame. Later, this frame is compared with the upcoming/dynamic frames to detect the foreground objects. Here in this paper, also a background_image is initialized as the first frame. Thus, this background_image frame needs to be updated dynamically based on any static change in the environment.

The proposed update method works on the assumption that background is stationary; i.e., only very small background motions may occur. A short processing delay is allowed subsequently to detect the object. Here, a temporary background image model is managed to store the pixels and their locations that are to be updated. At every short delay, background_image is subtracted from the current input image. If the pixel difference is the same (i.e., the difference is 0), then no change is updated. If pixel difference is not same (i.e., the difference is greater than a certain threshold), the pixels in temporary background_image are updated as the arithmetic average of the previous pixel value in temp background_image and the current pixel value.

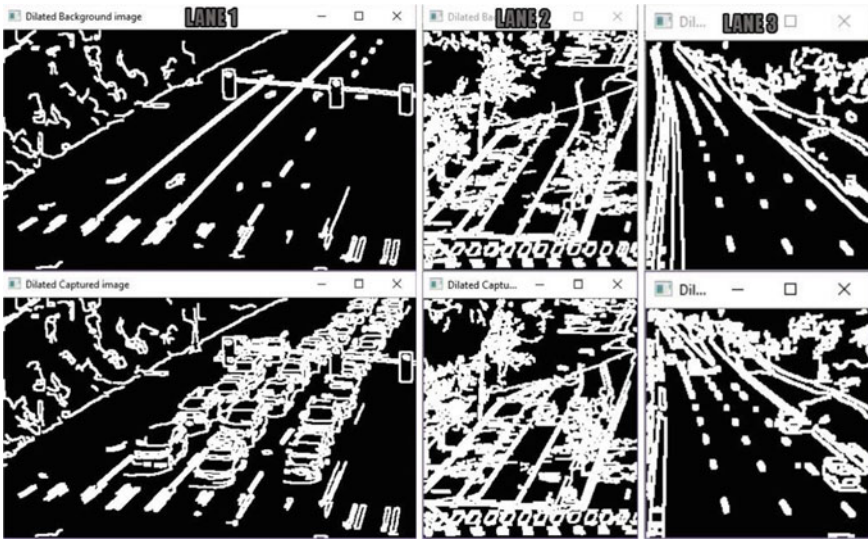


Fig. 5 Images after dilation

Simultaneously, a frequency of change for this pixel is maintained in an array by incrementing with 1. If these values in frequency array go over a certain threshold, then the pixel is copied from temp background_image to current frame and frequency is reset to zero. This sub-process is shown in Fig. 6. Again a threshold is applied in the number of pixels updated. If this number is greater than this threshold, then the original background image is saved in the disk for further use.

3.2 Testing Phase

Once the system is trained with background images, dynamically at every fixed interval the current status image of the traffic signal is captured as a test image. This test image will undergo the same preprocessing stages to reduce/remove the distortions from it. Then, the same edge detection algorithm is applied to this image along with dilation to obtain a clear and clarified image of the traffic. Later, the density of the traffic is estimated using these edge features: histogram matching [19] is applied first and followed by a template matching [20]. Finally, these two estimates are weighed based on 20% histogram matching and 80% template matching (since histogram of two different images might be same, template difference should be high as shown in Fig. 7).

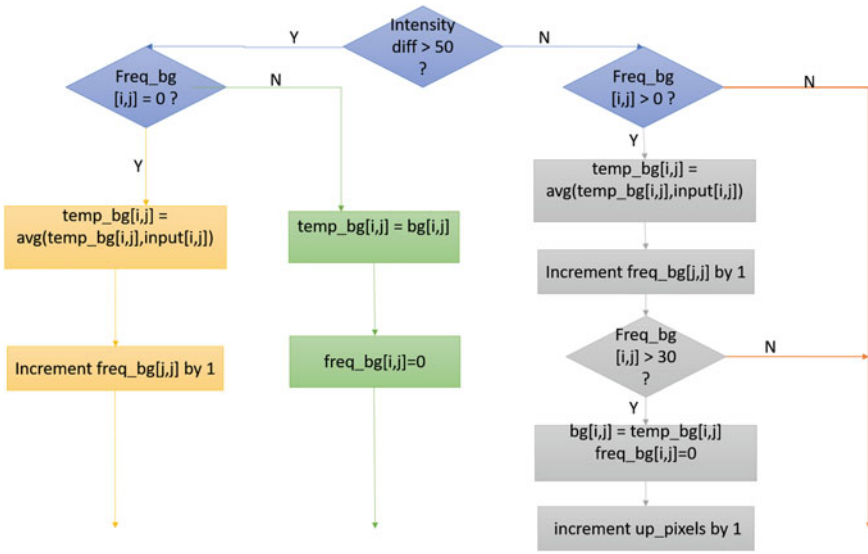
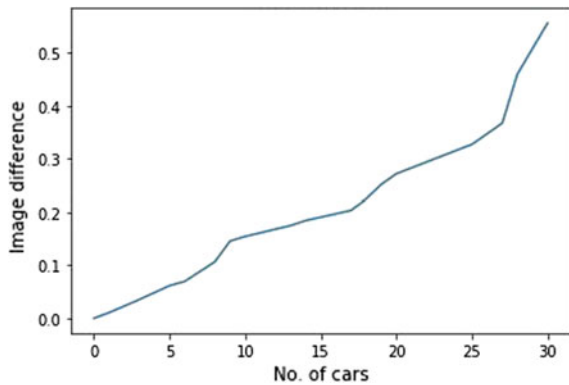


Fig. 6 Training image background update process

Fig. 7 Graphical analysis



Allocation of Lights: Based on the analysis of image difference (compared among background_image and current status of all lanes), a green light is allocated to the lane accordingly. That is, the lane with the highest image difference, $diff_{max}$ (scaled up by 100), is allotted green light for the duration $diff_{max} * 3$. This process is repeated for every fixed interval.

3.3 *Starvation Problem and Solution*

Starvation is the name given to the indefinite postponement of a process because it requires some resource before it can run, but the resource, though available for allocation, is never allocated to this process. However it happens, starvation is self-evidently a bad thing; more formally, it is bad because we do not want a non-functional system. If starvation is possible in a system, then a process which enters the system can be held up for an arbitrary length of time. The proposed system also encounters the same issue. In order to resolve it here, a frequency count is maintained for each of the lanes for assigning the green light. The lane with a max count for any lane for more than a specified time bound is treated as starvation lane. Here in this paper, it has been chosen as 180 s. Thus, the starvation problem has also been tried to resolve efficiently.

4 Result Analysis

4.1 *Dataset Used*

The dataset used in this paper is MIT traffic dataset. It has images from UK government taken from various JamCams installed at signals. This traffic dataset is for research on activity analysis in crowded scenes. For this paper, the images of JamCam at Westminster Bridge, London, are taken for analysis.

4.2 *Traffic Light Allocation to Different Lanes*

Figure 8 shows the images at three lanes along with its respective background_images. As in Fig. 8c, it is clear that the image difference is very high in comparison to the other two lanes. Hence, the green light is given for this lane. Similarly, in Fig. 9a the traffic density is high when compared to the remaining two lanes; thus, lane 1 is given the green light dynamically. At next interval, the traffic density is high in lane 3 as shown in Fig. 10 so instead of lighting the green light in lane two it is been lighted in lane 3. Likewise, the lane 2 is given the green light in the next interval. The proposed method reduces the traffic delays. From the Figs. 9–11 it is observed that, the lane-2 is not at all given the priority for green light thus, leads to starvation of vehicles. Proposed method also resolves this issue by assigning frequency counts for every green light (at different lanes dynamically). Figure 12b shows this update by assigning green light to lane 2 as this lane has large waiting time, image difference greater than zero (states that there is a moderate traffic) and a lower frequency count for the green light.

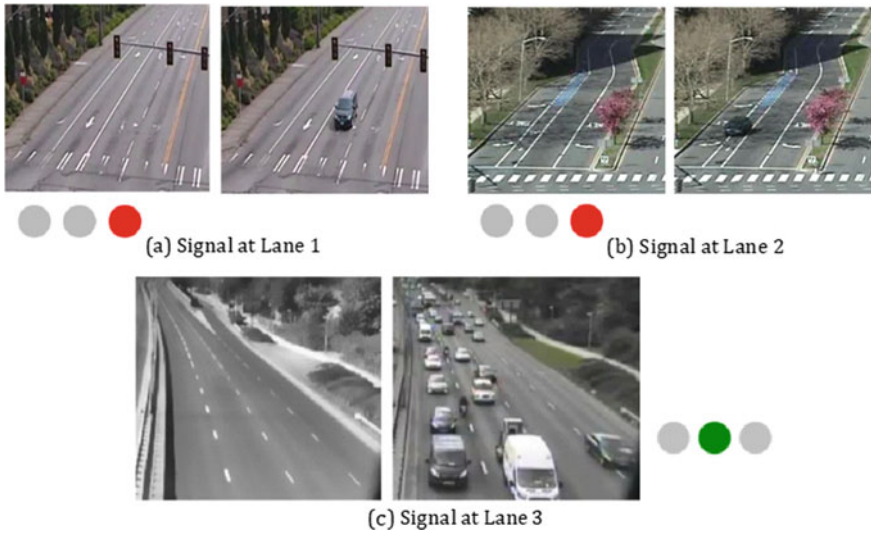


Fig. 8 Traffic at lane 1 (a), lane 2 (b) and lane 3 (c) along with background_ground image. After analyzing the edge estimation, lane 3 is given the green signal

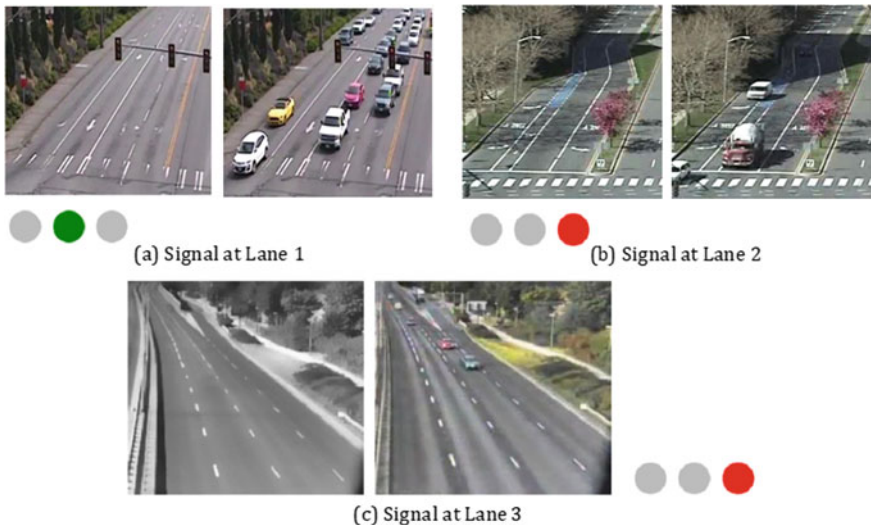


Fig. 9 Traffic at lane 1 (a), lane 2 (b) and lane 3 (c) along with background_ground image. After analyzing the edge estimation, lane 1 is given the green signal

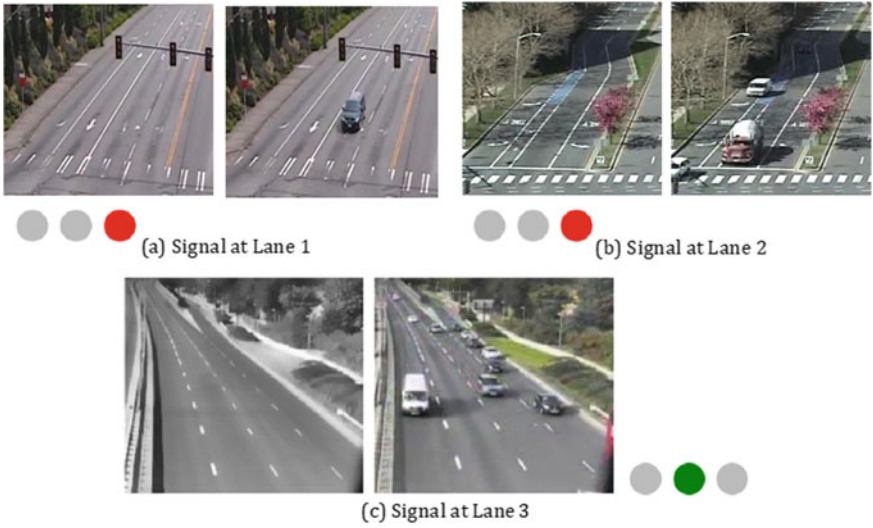


Fig. 10 Traffic at lane 1 (a), lane 2 (b) and lane 3 (c) along with background_ground image. After analyzing the edge estimation, lane 3 is given the green signal

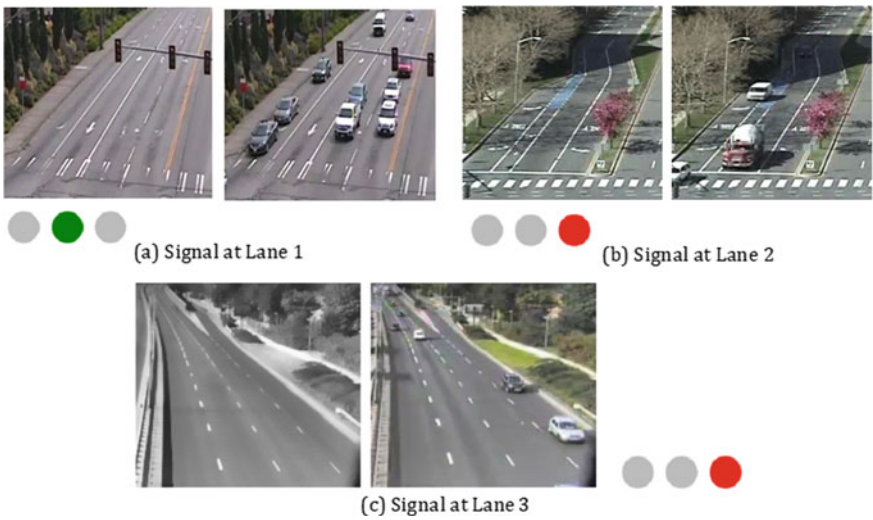


Fig. 11 Traffic at lane 1 (a), lane 2 (b) and lane 3 (c) along with background_ground image. After analyzing the edge estimation lane 1 is given the green signal

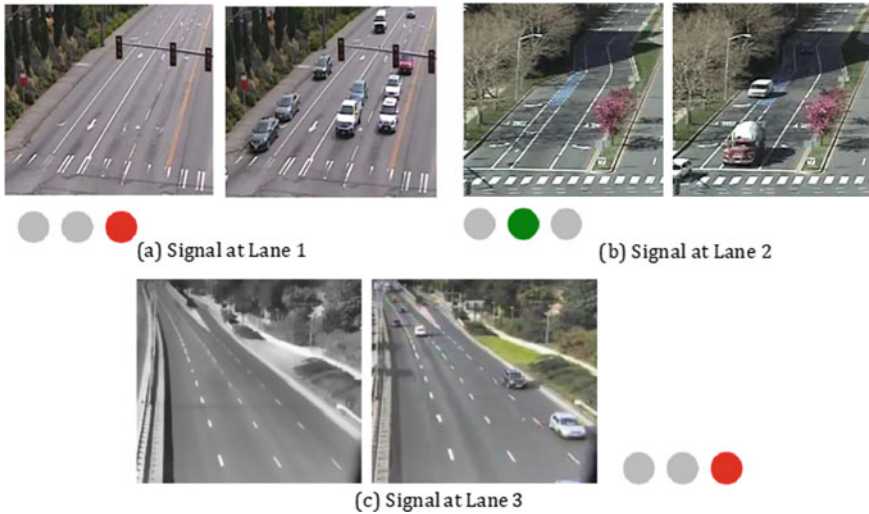


Fig. 12 Traffic at lane 1 (a), lane 2 (b) and lane 3 (c) along with background_ground image. After analyzing the edge estimation, lane 2 is given the green signal to resolve the starvation issue

5 Conclusion

Building or moderating the civil infrastructures like roads, underpasses and flyovers may lead to traffic congestion in most of the rural and urban cities. This leads to time delays for passengers. In order to resolve this issue, a fixed automatic traffic control system use of timer had been used for many years, but it may lead to unwanted green light assignment on the empty roads/lanes. The proposed method in this paper has aimed to solve this challenge using video surveillance camera and video processing techniques. The proposed method is inexpensive and easy to install.

References

1. K. Vidhya, A. Bazila Banu, Density based traffic signal system. *Int. J. Innovative Res. Sci. Eng. Technolo.* **3**(3), 2218–2222 (2014)
2. P. Sinhmar, Intelligent traffic light and density control using IR sensors and microcontroller. *Int. J. Adv. Technol. Eng. Res. (IJATER)* **2**(2), 30–35 (2012)
3. J.M. Blossville, C. Krafft, F. Lenoir, V. Motyka, S. Beucher, TITAN: A traffic measurement system using image processing techniques, in *Second International Conference on Road Traffic Monitoring (IET)*, pp. 84-88 (1989)
4. R. Cucchiara, M. Piccardi, P. Mello, Image analysis and rule-based reasoning for a traffic monitoring system, in *Proceedings. 1999 IEEE/IEEJ/JSAI International Conference on Intelligent Transportation Systems*, IEEE (1999)

5. Y. Iwasaki, An image processing system to measure vehicular queues and an adaptive traffic signal control by using the information of the queues, in *Proceedings of IEEE Conference on Intelligent Transportation System, 1997. ITSC'97*, IEEE (1997)
6. A. Tzes, W.R. McShane, S. Kim, Expert fuzzy logic traffic signal control for transportation networks, in *1995 Compendium of Technical Papers. Institute of Transportation Engineers 65th Annual Meeting. Institute of Transportation Engineers (ITE)* (1995)
7. Q. Lin, B.W. Kwan, L.J. Tung, Traffic signal control using fuzzy logic, in *1997 IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation*, IEEE, Vol. 2 (1997)
8. J. Niittymäki, Installation and experiences of field testing a fuzzy signal controller. *Eur. J. Oper. Res.* **131**(2), 273–281 (2001)
9. Y. Li, X. Fan, Design of signal controllers for urban intersections based on fuzzy logic and weightings, in *Proceedings of the International Conference on Intelligent Transportation Systems, 2003*, IEEE. Vol. 1 (2003)
10. W. Wen, A dynamic and automatic traffic light control expert system for solving the road congestion problem. *Expert Syst. Appl.* **34**(4), 2370–2381 (2008)
11. C. Dong, Z. Liu, X. Liu, Chaos-particle swarm optimization algorithm and its application to urban traffic control. *Int. J. Comput. Sci. Netw. Secur.* **6**(1B), 97–101 (2006)
12. S.C. Chang, M.W. Tsai, G.W. Huang, A GA based intelligent traffic signal scheduling model, in *2007 IEEE Symposium on Computational Intelligence in Scheduling, SCIS'07*, IEEE (2007)
13. R.K. Saraf, Adaptive traffic control using neural networks. *Transp. Res. A* **1**(30), 57–58 (1996)
14. C. Ozkurt, F. Camci, Automatic traffic density estimation and vehicle classification for traffic surveillance systems using neural networks. *Math. Comput. Appl.* **14**(3), 187–196 (2009)
15. C. Yeshwanth et al., Estimation of intersection traffic density on decentralized architectures with deep networks, in *2017 International Smart Cities Conference (ISC2)*, IEEE (2017)
16. Z. Li et al., Signal controller design for agent-based traffic control system, in *2007 IEEE International Conference on Networking, Sensing and Control*, IEEE (2007)
17. M. Arora, V.K. Banga, Real time traffic light control system using morphological edge detection and fuzzy logic, in *2nd International Conference on Electrical, Electronics and Civil Engineering (ICEECE'2012)*, Singapore (2012)
18. P. Srinivas, Y.L. Malathilatha, Dr. M.V.N.K Prasad, Image processing edge detection technique used For traffic Control Problem. *Int. J. Comput. Sci. Inf. Technol.* **4**, 17–20 (2013)
19. R. Maini, H. Aggarwal, A comprehensive review of image enhancement techniques. arXiv preprint [arXiv:1003.4053](https://arxiv.org/abs/1003.4053) (2010)
20. J.P. Lewis, Fast template matching. *Vis. Interface.* **95**(120123) (1995)

Scheduling in Cloud Computing Environment using Metaheuristic Techniques: A Survey



Harvinder Singh, Sanjay Tyagi and Pardeep Kumar

Abstract Cloud computing has become a burning topic for research these days. To reap the full benefits of cloud computing, a wider perspective of thinking and study is required from a researcher. A vital part of this cloud innovation is that it has given the solid development and movement to the computational environment like scheduling. Scheduling tends to the proficient administration and usage of the extensive amount of processing assets. Being an example of N-P category, scheduling maps the tasks to appropriate resources for delivering an optimized solution. Cloud technology paves a way to provides sub-optimal solutions within a short time period. Metaheuristic-based techniques are one of the ways through which solutions can be achieved efficiently even for hard optimization problems. The present work provides an analysis and discussion based on extensive study of various metaheuristic techniques. In addition, various parameters are identified and explored, that plays a vital role in the scheduling of tasks in a cloud environment.

Keywords Task · Resources · Scheduling · Metaheuristic scheduling · Nature-inspired algorithm

H. Singh (✉)

Department of Informatics, School of Computer Science,
University of Petroleum and Energy Studies, Dehradun, Uttarakhand, India
e-mail: hsingh@ddn.upes.ac.in

S. Tyagi · P. Kumar

Department of Computer Science and Applications,
Kurukshetra University, Kurukshetra, Haryana, India
e-mail: tyagikuk@gmail.com

P. Kumar

e-mail: mittalkuk@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_66

1 Introduction

Scheduling is one approach to achieve optimal resource allocation among all given tasks in limited time to get wanted QoS. Scheduler finds ways to assign appropriate task to limited resources for optimization of one or more objectives. Task is scheduled on some resource which is bounded by some constraints to find some optimal solution for objective function. The target of scheduling algorithm is to determine which task will be executed on which resource [1]. May it be a scheduling algorithm or the concept of threads in operating system, both remained an interesting and important topics of the research area. Not limiting, this paradigm supports access to pool of shared resources with the guarantee of QoS to the respective users. Achievement of desired performance is the key to user satisfaction and this can be done by considering the scheduling as the central theme in systems working on cloud computing [2].

Task assignment from the Internet user to the limited resource is one of the important applications of modern scheduling in distributed systems for computing. Several changes have been incorporated to these systems including the use of clusters which integrate multiple standalone computers to one single unit giving high performance. To beat the issue of cluster systems being just ready to utilize nearby assets, the following change, grid, has been produced to join all the accessible heterogeneous resources from topographically distributed organizations. Cloud computing systems use the qualities of cluster computing and grid computing [3]. Cloud computing addresses the NP-hard problems where the objective is to resolve the mapping of task on apparently large number of computing resources in cloud computing. No algorithm exists that provides an optimal solution for these types of problems within polynomial time [4, 5].

Common methods for scheduling include exhaustive and deterministic algorithms. Another type of algorithm which can be used for scheduling is metaheuristic algorithms. Practically, DAs show much better results than traditional exhaustive methods since they provide faster solutions for problems on scheduling [6]. The DAs are limited for few data distribution and do not fit to large-scale distributions of collected data. On the other hand, metaheuristic algorithms employ approximation methods and use iterative strategies to achieve optimized solutions in reasonable time. Metaheuristic algorithms have high efficiency and effectiveness in solving comparatively larger problems that incorporate complexity among them [7–9].

The QoS-oriented task scheduling issue is an optimization issue which guarantees the ideal mapping between each task and accessible resources, which can impact the QoS of the entire cloud computing systems [10–12]. Dependency forms the basis to differentiate the scheduling algorithms; i.e., task scheduling is subjected to the completion of all the parent tasks. Such type of dependency is cited under work flow scheduling. When the tasks do not follow any sequence of scheduling the process is named as independent scheduling.

As of late, to get an optimum solution, most researchers concentrated on creating nature propelled metaheuristic algorithms like ACO, PSO, GA, ABC to take care of multi-objective workflow scheduling problem thinking about different param-

ters [13–17]. To additionally raise the viability, researchers endeavor to incorporate numerous criteria into the optimization algorithms, known as multi-objective task scheduling. MOTS [18] is a most imperative point in the cloud computing by including numerous parameters like makespan, cost, QoS, load and network parameters. The objective in the cloud environment is to plan the present tasks inside the given imperatives like QoS, resource utilization cost, load balancing, and makespan [1, 19]. The current work analysis presents the review of few scheduling algorithms including traditional and metaheuristic algorithms.

2 Metrics for Scheduling Optimization

Quantification to find the quality and determine the efficiency calls for the use of some empirical analysis. In cloud computing also, we need to quantify the work analysis on some basis. These bases form the optimization metrics for the work. Cloud computing revolves around two types of entities which are service provider for the cloud and the other one is the cloud consumer. The resources are provided on the rental basis by the cloud service provider to the consumers. Later do their task submission for processing the resources provides by the former. Both components of the cloud environment have their own motivations while achieving the objectives. Consumers prefer performance while providers seek for efficiency in resource utilization. The fundamental scheduling parameters considered in the already said techniques are recorded underneath:

1. **Makespan:** It is the aggregate completion time of all tasks in a job queue. A decent algorithm of scheduling dependably endeavors to diminish the makespan.
2. **Resource utilization:** It means keeping assets as occupied as could be expected under the circumstances. Maximization of resource usage must be achieved. This rule is picking up criticalness as service providers need to acquire the greatest benefit by leasing a predetermined number of resources [20].
3. **Resource utilization cost:** It demonstrates the aggregate sum the client needs to pay to a service provider for asset usage.
4. **Quality of Service:** QoS incorporates numerous client input limitations like meeting execution cost, due date, performance, makespan, and so forth [21, 22].
5. **Load balancing:** It is the strategy for dissemination of the whole load in a cloud network crosswise over various nodes so that at once no node stay under loaded while a few nodes are over-burden [23].
6. **Energy Consumption:** Energy utilization is an issue that ought to be considered with more care nowadays. Numerous scheduling algorithms were created for reducing power utilization and enhancing performance and thus making the cloud services green [24].
7. **Performance:** Performance shows the overall effectiveness given by the scheduling algorithm in order to give great administrations to the clients according to their requirements.

8. **Deadline:** It is characterized by the timeframe from presenting a task to the time by which it must be finished. A decent scheduling algorithm dependably attempts to keep the tasks executed inside the due date imperative.

3 Traditional Scheduling

Scheduling can be characterized as designating a set of given tasks to a set of given machines subject to the imperatives of optimizing objective functions. At the point when there is a single machine, the scheduling problem is alluded to as a single-processor scheduling problem. For in excess of one machine, the scheduling problem is viewed as a multiprocessor scheduling problem. The performance for scheduling algorithms can be quantified on the basis of resource utilization cost, makespan, load balancing and QoS and its variants. Since scheduling has been extensively utilized as a part of numerous research problem domains, a few examinations endeavored to exhibit a scientific categorization of scheduling problem as far as the job depiction is concerned.

4 Metaheuristic Scheduling

The metaheuristic scheduling can be studied under a unified framework showing the initial, transition, evaluation and determination evaluators employed to the framework. The major issue that is to be addressed is applying the encoded solution for metaheuristic algorithms to optimize the scheduling problems. The transition operator changes the current state of the solution to the next by using perturbative and constructive transition methods for solving combinatorial problem [25, 26]. The complexity of the operator depends upon the design of the metaheuristic algorithm which may vary from easy to complex. The evaluation operator takes responsibility for evaluating the object function value in the problem statement. For example, this operator determines the value of makespan for the scheduling problem. Yet another operator called determination guides the search and intent to quantize the research for intensification and diversification that implicitly or explicitly influence the convergence speed.

4.1 *Ant Colony Optimization Scheduling Algorithm*

Ant colony optimization scheduling algorithm is based on real behavior of ants. It finds the shortest path between the source food and the ant colonies. Originally called as ant system, the approach was introduced by Dorigo in 1992. The approach makes the use of the fact that ants leave pheromone on the way while walking amid

their colony and the food source. The intensity of the pheromone is increased on the passage due to the number of ants passing through and it decreases due to evaporation with time [27].

ACO algorithms are efficient in finding a solution for the discrete optimization problems which requires path finding to reach the goals. An ant colony optimization scheduling algorithm is studied in terms of its representation, operators, control parameters, evolutionary mechanism, performance metric, areas of applications as shown below.

1. **Algorithm:** Ant Colony Optimization Scheduling Algorithm
2. **Representation:** Undirected graph.
3. **Operators:** Pheromone update and measure, trail evaporation.
4. **Control Parameters:** Number of ants, iterations, pheromone evaporation rate, amount of reinforcement.
5. **Evolutionary Mechanism:** $P_{ij} = \frac{(\tau_{ij})^\alpha (\eta_{ij})^\beta}{\sum_{k \in allowed} (\tau_{ik})^\alpha (\eta_{ik})^\beta}$
6. **Performance Metric:** Makespan, load balancing, resource utilization cost, execution cost, deadline constraint, energy conservation, response time, throughput.
7. **Areas of Application:** Traveling salesman problem, classification problem in data mining, job-shop scheduling problem.

4.2 Particle Swarm Optimization Scheduling Algorithm

Particle swarm optimization algorithm is an evolutionary computation technique which is motivated by the social behavior of the particles. It exploits the motion of particles which have velocity and are aligned in some specific direction with the position vector. Their movement is traced as multidimensional motion in the search space. It follows the iterative approach where after each transition, particle adjusts its velocity on the basis of best position with respect to current position of self and the best particle in the whole population. It is the merger of global search and local search where the technique tries to balance the exploration and exploitation by the former method [28, 29].

The method incurs low-computational cost and is rapidly growing in terms of popularity due to simplicity in understanding and implementation. An algorithm is studied in terms of its representation, operators, control parameters, evolutionary mechanism, performance metric, areas of applications as shown below.

1. **Algorithm:** Particle Swarm Optimization Algorithm
2. **Representation:** D dimensional vector for position, speed, best state.
3. **Operators:** initializer, updater and evaluator.
4. **Control Parameters:** number of particles, Dimension of particles, Range of particles, inertia weight, maximum number of iterations.
5. **Evolutionary Mechanism:** $V_{i+1} = \omega V_i + c_1 rand_1 * (pbest - x_i) + c_2 rand_2 * (gbest - x_i)$

6. **Performance Metric:** Makespan, load balancing, resource utilization cost, communication cost, execution cost, throughput, energy conservation.
7. **Areas of Application:** Multimodal biomedical image registration, various scheduling problems, vehicle routing problems, color image segmentation, sequential ordering problem.

4.3 Genetic Scheduling Algorithm

Genetic scheduling algorithm (GA) is definitely a standout among the most critical population-based algorithms as far as its execution as well as the effectiveness of applying it to numerous issue areas. In GA, every chromosome (individual in the population) speaks to a conceivable answer for an issue and is made out of a series of genes. The underlying population is taken arbitrarily to fill in as the beginning stage for an algorithm. A fitness function is defined to check the appropriateness of the chromosome for the environment. Based on a value of fitness, chromosomes are chosen and crossover and mutation tasks are performed on them to deliver offspring for the new population. The fitness function assesses the quality of every offspring. The procedure is reiterated until sufficient offspring are made.

GA utilizes fitness function to assess the answers for the best fit. Fitness function works likewise to separates solutions in light of rank or proportion. The selection operator decides the search directions for the following cycle. While the crossover expects to exchange the data between the solutions, the mutation operators empower the algorithm to get away from the local optima amid the transitions.

Evaluation and determination operators of metaheuristic can be viewed as the determination technique of the GA. Distinctive outline contemplations of the fitness function can be drawn from different suspicions. The cost of data transfer, correspondence, and calculation, and in addition the profit, have all been considered in the outline of the fitness function. An algorithm is studied in terms of its representation, operators, control parameters, evolutionary mechanism, performance metric, areas of applications as shown below.

1. **Algorithm:** Genetic Scheduling Algorithm
2. **Representation:** Binary, real numbers, permutation of elements, list of rules, program elements, data structure, tree, matrix.
3. **Operators:** Crossover, mutation, selection, Inversion.
4. **Control Parameters:** Population size, maximum generation number, cross over probability, mutation probability, length of chromosome, chromosome encoding.
5. **Evolutionary Mechanism:** Depends on the maximization or minimization of Fitness function
6. **Performance Metric:** Makespan, Load Balancing, Resource Utilization Cost, Execution Cost, QoS, Energy Conservation.
7. **Areas of Application:** drug design, Web site optimizations, data mining problems, control system design, pattern recognition.

4.4 Artificial Bee Colony Optimization Scheduling Algorithm

Artificial bee colony optimization scheduling algorithm is an optimization technique to locate the ideal solution motivated by the foraging behavior of honey bees. There are three kinds of a population of the honey bee: Scout honey bees have an obligation to arbitrarily search for the new sustenance sources; employed honey bees need to search for the food sources and bring back the sustenance sources data to illuminate the onlooker honey bees which are holding up at the hives. At that point, the onlooker honey bees compute the fitness function and pick the ideal food source for sending the employed honey bees to gather. In the event that any food source was picked and all the food was gathered, the employed honey bees which are the proprietors of the food source will turn into the scout honey bees and they will again arbitrarily search for the new food sources. In the genuine nature, honey bees will look at the nature of food sources on the dance floor. They will move, called “Waggle Dance” to illuminate others about the direction, the distance, and the measure of honey. The onlooker honey bees will contrast all food sources for locating the best one [30, 31].

As of late, numerous scientists had used ABC in taking care of task scheduling problem. An algorithm is studied in terms of its representation, operators, control parameters, evolutionary mechanism, performance metric, areas of applications as shown below.

1. **Algorithm:** Artificial Bee Colony Optimization Scheduling Algorithm
2. **Representation:** D-dimensional vector $x_i = x_1, x_2$ to x_d .
3. **Operators:** Reproduction, replacement of bee, selection.
4. **Control Parameters:** number of food sources (SN), the value of limit, the maximum cycle number (MCN).
5. **Evolutionary Mechanism:** $p_i = \frac{fitness_i}{\sum_{i=1}^{NP} fitness_i}$
6. **Performance Metric:** Makespan, Load Balancing, Resource Utilization Cost, Execution Cost, Energy Conservation.
7. **Areas of Application:** clustering, scheduling problems, assembly line balancing problem, image segmentation, training of neural networks.

4.5 Crow Search Optimization Scheduling Algorithm

Scheduling optimization problems can be resolved by using crow search optimization scheduling algorithm, which tries to reproduce the astute conduct of the crows to find the optimized solution. There are numerous similitudes between optimization process and the flocks of crows. The crows used to hide the food available with them at specific locations (concealing spots) and during their requirement, crows used to recuperate the same. In order to search for better food sources, crows have a tendency to follow another crow in a flock. Finding food source concealed by a crow is not a straightforward task since if a crow finds another is following it, the

crow attempts to trap that crow by taking off to another area of nature. As per an optimization point of view, the searchers correspond to crows, the search space is complemented by the environment, possible solution corresponds to each position of the environment, objective function is represented by the quality of the food source and the global solution to the scheduling optimization problem is represented by the best food source of the environment [32].

An algorithm is studied in terms of its representation, operators, control parameters, evolutionary mechanism, performance metric, areas of applications as shown below.

1. **Algorithm:** Crow Search Optimization Scheduling Algorithm
2. **Representation:** D dimensional search space.
3. **Operators:** Flock size (N), maximum number of iterations $iter_{max}$, flight length (fl) and awareness probability (AP).
4. **Control Parameters:** Flight length (fl) and awareness probability (AP).
5. **Evolutionary Mechanism:** $x^{i,iter+1} = x^{i,iter} + r_i * fl^{i,iter} * (m^{j,iter} - x^{i,iter})$.
6. **Performance Metric:** Makespan, Load Balancing, Resource Utilization Cost, Energy Conservation, QoS.
7. **Areas of Application:** Solving complex engineering design problem, nonlinear and multimodal real-world optimization problems.

4.6 Penguin Search Optimization Scheduling Algorithm

The problem of scheduling optimization can be resolved by using penguin search optimization algorithm, which tries to find out the optimized solution by reproducing the hunting practice of the penguins. The penguins have attractive hunting technique since they can cooperate with each other with their endeavors and synchronize their dives to improve the global energy during the time spent in entire hunting practice [33].

All penguins speak to a result are scattered in gatherings, and each gathering looks for meal in characterized holes with differences levels. Penguins arranged all together for their groups and begin the pursuit in a particular hole & level as per meal disponibility probability. In every iteration, in like manner, the position of the penguin with each advanced solution is altered and gives three results, the best local, the last, and the new solution. The counts in the refresh, an answer is rehashed for every penguin in each gathering, and after a few doves, penguins pass on to one another the best solution which speaks to by the amount of fishes eaten, and new dispersion probability of gaps and levels is figured. An algorithm is studied in terms of its representation, operators, control parameters, evolutionary mechanism, performance metric, areas of applications as shown below.

1. **Algorithm:** Penguin Search Optimization Algorithm
2. **Representation:** 3D graph of the nonlinear multimodal Rastigin function.
3. **Operators:** the number of penguins, the number of holes and the number of levels should be large enough.
4. **Control Parameters:** distribution probability of holes and levels.
5. **Evolutionary Mechanism:** $D_{new} = D_{LastLast} + rand() | X_{LocalBest} - X_{LocalLast}$
6. **Performance Metric:** Makespan, load Balancing, resource utilization cost, energy conservation, QoS.
7. **Areas of Application:** Solving NP-hard problems, applications in many fields such as aeronautics, medical imaging.

5 Discussion

Based on the survey of metaheuristic task scheduling algorithms, the following observations and challenges are identified:

1. The way to a task scheduling methodology is to discover a trade-off between client necessities and resource usage. Nonetheless, tasks which are put together by various clients may have diverse prerequisites on the computing time, memory space, data traffic, reaction time, and so forth.
2. One of the real difficulties in the present cloud arrangements is to give the required services as per the QoS level expected by the client. Cloud service providers need to affirm that adequate measure of assets is provisioned to guarantee that QoS prerequisites of cloud service consumers, like due date, execution time, energy consumption, and budget limitations are met.
3. In multi-objective task scheduling optimization, it is expected to use the scheduling design with various criteria, to meet the prerequisite of clients and additionally service providers. Be that as it may, the schedulers would offer ascent to the issue of high load, serious resource rivalry, and the wasteful cooperation.
4. The techniques utilized different objectives for task scheduling, but the utilization of multiple objectives like QoS, cost, makespan, load within one algorithm was not considered. This consideration aims to provide optimal services that too at a minimal expense to the users.
5. Also, the recent and effective optimization algorithms are required to solve the multiple objective-based task scheduling because the traditional algorithm is only utilized in most of the works.

Ideas extracted from the key observations must be harnessed in order to develop a new multi-objective metaheuristic optimization scheduling algorithm that will minimize the shortcomings of existing metaheuristic optimization scheduling algorithm.

6 Conclusion & Future Scope

The primary intention of this research is to have a deep insight into various multi-objective task scheduling algorithms in the cloud computing environment. For this purpose, a number of metaheuristics scheduling algorithms are studied in terms of their representation, operators, control parameters, evolutionary mechanism, performance metric, areas of applications for techniques. In the future, the comparison of various metaheuristic scheduling algorithms such as ACO, PSO, GA, ABC, CSA, and PeSOA can be done quantitatively based on various parameters identified from the detailed study of the literature like makespan, resource utilization cost, load, and QoS.

References

1. Y. Liu, C. Zhang, B. Li, J. Niu, DeMS: A hybrid scheme of task scheduling and load balancing in computing clusters. *J. Netw. Comput. Appl.* **83**, 213–220 (2017). <https://doi.org/10.1016/j.jnca.2015.04.017>
2. T. Goyal, A. Singh, A. Agrawa, Cloudsim: simulator for cloud computing infrastructure and modeling. *Procedia Eng.* **38**, 3566–3572 (2012). <https://doi.org/10.1016/j.proeng.2012.06.412>
3. H. He, G. Xu, S. Pang, Z. Zhao, AMTS: adaptive multi-objective task scheduling strategy in cloud computing. *China Commun.* **13**(4), 162–171 (2016). <https://doi.org/10.1109/CC.2016.7464133>
4. X. Mao, C. Li, W. Yan, S. Du, Optimal scheduling algorithm of mapreduce tasks based on QoS in the hybrid cloud (2016). <https://doi.org/10.1109/PDCAT.2016.37>
5. X. Sheng, Q. Li, Template-based genetic algorithm for QoS-aware task scheduling in cloud computing, in *2016 International Conference on Advances in Cloud Big Data*, pp. 0–5 (2016). <https://doi.org/10.1109/CBD.2016.37>
6. Vijindra, S. Shenai, Survey on scheduling issues in cloud computing. *Procedia Eng.* **38**, 2881–2888 (2012). <https://doi.org/10.1016/j.proeng.2012.06.337>
7. K. Pradeep, T.P. Jacob, Comparative analysis of scheduling and load balancing algorithms in cloud environment, in *2016 International Conference on Control Instrumentation, Communication and Computing Technology ICCICCT*, pp. 526–531 (2017). <https://doi.org/10.1109/ICCICCT.2016.7988007>
8. T. Arora, Y. Gigras, A survey of comparison between various Meta. *Int. J. Comput. Eng. Sci.* **3**(2), 62–66 (2013)
9. M. Kalra, S. Singh, A review of metaheuristic scheduling techniques in cloud computing. *Egypt. informatics J.* **16**(3), 275–295 (2015)
10. K. Kaur, N. Kaur, K. Kaur, Data engineering and intelligent computing **542**, (2018). <https://doi.org/10.1007/978-981-10-3223-3>
11. Y. Xin, Z.Q. Xie, J. Yang, A load balance oriented cost efficient scheduling method for parallel tasks. *J. Netw. Comput. Appl.* **81**, 37–46 (2017). <http://linkinghub.elsevier.com/retrieve/pii/S1084804516303496>
12. Z. Li, J. Ge, H. Yang, L. Huang, H. Hu, H. Hu, B. Luo, A security and cost aware scheduling algorithm for heterogeneous tasks of scientific workflow in clouds. *Futur. Gener. Comput. Syst.* **65**, 140–152 (2016). <https://doi.org/10.1016/j.future.2015.12.014>
13. S.R. Shishira, A. Kandasamy, K. Chandrasekaran, Survey on meta heuristic optimization techniques in cloud computing, in *2016 International Conference on Advanced Computation Communication and Informatics, ICACCI*, pp. 1434–1440 (2016). <https://doi.org/10.1109/ICACCI.2016.7732249>

14. A. Verma, S. Kaushal, A hybrid multi-objective particle swarm optimization for scientific workflow scheduling. *Parallel Comput.* **62**, 1–19 (2017). <https://doi.org/10.1016/j.parco.2017.01.002>
15. S. Binitha, S.S. Sathya, Others: A survey of bio inspired optimization algorithms. *Int. J. Soft Comput. Eng.* **2**(2), 137–151 (2012)
16. T. Mathew, K.C. Sekaran, J. Jose, Study and analysis of various task scheduling algorithms in the cloud computing environment, in *2014 International Conference on Advances in Computer Communication and Informatics (ICACCI)*, IEEE, pp. 658–664. (2014)
17. C.W. Tsai, J.J.P.C. Rodrigues, Metaheuristic scheduling for cloud: a survey. *IEEE Syst. J.* **8**(1), 279–291 (2014)
18. Lakra, A.V., Kumar Yadav, D.: Multi-objective tasks scheduling algorithm for cloud computing throughput optimization. *Procedia Comput. Sci.* **48**(C), 107–113 (2015). <https://doi.org/10.1016/j.procs.2015.04.158>
19. HGEDG Ali, I.A. Saroit, A.M. Kotb, Grouped tasks scheduling algorithm based on QoS in cloud computing network. *Egypt. Informatics J.* **18**(1), 11–19 (2017). <https://doi.org/10.1016/j.eij.2016.07.002>
20. Y. Liu, X. Xu, L. Zhang, L. Wang, L., R.Y. Zhong, Workload-based multi-task scheduling in cloud manufacturing. *Robot. Comput. Integr. Manuf.* **45**(September 2016), 3–20 (2017). <https://doi.org/10.1016/j.rcim.2016.09.008>
21. L. Shakkeera, L. Tamilselvan, QoS and load balancing aware task scheduling framework for mobile cloud computing environment **10**, 4 (2016)
22. X. Wu, M. Deng, R. Zhang, B. Zeng, S. Zhou, A task scheduling algorithm based on QoS-driven in cloud computing. *Procedia Comput. Sci.* **17**, 1162–1169 (2013). <http://linkinghub.elsevier.com/retrieve/pii/S1877050913002810>
23. P.P.G. Gopinath, S.K. Vasudevan, An in-depth analysis and study of load balancing techniques in the cloud computing environment. *Procedia Comput. Sci.* **50**, 427–432 (2015). <http://www.sciencedirect.com/science/article/pii/S1877050915005104>
24. S. Singh, I. Chana, QRSF: QoS-aware resource scheduling framework in cloud computing. *J. Supercomput.* **71**(1), 241–292 (2014). <https://doi.org/10.1007/s11227-014-1295-6>
25. I. Boussaïd, J. Lepagnot, P. Siarry, A survey on optimization metaheuristics. *Inf. Sci. (Ny)* **237**, 82–117 (2013). <https://doi.org/10.1016/j.ins.2013.02.041>
26. P. Agarwal, S. Mehta, Nature-inspired algorithms: state-of-art, problems and prospects. *Nature* **100**(14), (2014)
27. L. Zuo, L. Shu, S. Dong, C. Zhu, T. Hara, A multi-objective optimization scheduling method based on the ant colony algorithm in cloud computing. *IEEE Access* **3**, 2687–2699 (2015)
28. X. Zuo, G. Zhang, W. Tan, Self-adaptive learning pso-based deadline constrained task scheduling for hybrid iaas cloud. *IEEE Trans. Autom. Sci. Eng.* **11**(2), 564–573 (2014). <https://doi.org/10.1109/TASE.2013.2272758>
29. A.I. Awad, N.A. El-Hefnawy, H.M. Abdel_kader, Enhanced particle swarm optimization for task scheduling in cloud computing environments. *Procedia Comput. Sci.* **65**(Iccmit), 920–929 (2015). <http://linkinghub.elsevier.com/retrieve/pii/S187705091502894X>
30. B. Andrea, A. Bennedeto, Nuevas alternativas para pensar el desarrollo de los territorios rurales. Posibilidades y riesgos I. *Cuad. Desarro. Rural* **57**(57), 101–131 (2006). <https://doi.org/10.1109/SCIS>
31. L.D. Dhinesh Babu, P. Venkata Krishna, Honey bee behavior inspired load balancing of tasks in cloud computing environments. *Appl. Soft Comput. J.* **13**(5), 2292–2303 (2013). <https://doi.org/10.1016/j.asoc.2013.01.025>
32. A. Askarzadeh, A novel metaheuristic method for solving constrained engineering optimization problems: Crow search algorithm. *Comput. Struct.* **169**, 1–12 (2016). <https://doi.org/10.1016/j.compstruc.2016.03.001>
33. Y. Gheraibia, A. Moussaoui, Penguins search optimization algorithm (PeSOA). *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **7906 LNAI**, 222–231 (2013). https://doi.org/10.1007/978-3-642-38577-3_23

Crown Detection and Counting Using Satellite Images



Rebeka Bhattacharyya and Avijit Bhattacharyya

Abstract Shade trees are the trees which are grown in the tea gardens in order to shade the tea for better quality of tea. In this paper, we have discussed the method for detection of shade trees and its counting using satellite images. As the trees are very crowded in our study area so their trees overlap. We use a number of sample images for training and optimizing the CNN, and labeling for all the images in a dataset which we have collected through the sliding window technique. Then, we have done merging of the predicted shade tree coordinates corresponding to the same shade tree and have obtained the final detection results. By this detection technique, we have obtained an overall accuracy of more than 96% in our study area.

Keywords Deep learning · CNN model · Shade trees detection

1 Introduction

Remote sensing is important in shade trees productivity and shade mapping, etc. Nowadays, satellite images have become popular for many applications. Previously, the detection of shade trees in a particular area was very easy and good detection results have been obtained as the trees are not crowded in the concerned area. But in this research work, the detection is a bit difficult because the trees are very crowded in our study area.

CNN is a deep learning model which is used for object detection, image classification, etc. Nowadays, these methods can also be used in high-resolution images for image classification, object detection, etc. and have achieved better performance.

R. Bhattacharyya (✉)

Electronics and Communication Engineering, Institute of Engineering and Management, Kolkata, India

e-mail: rebeka.bhattacharyya@gmail.com

A. Bhattacharyya

Tata Consultancy Services, Kolkata, India

e-mail: avijit.bhattacharyya@gmail.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharyya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937, https://doi.org/10.1007/978-981-13-7403-6_67

765

In this paper, we have discussed the detection of shade trees based on deep learning model. We have formed a CNN model for the detection of shade trees using satellite images from Jalpaiguri region in West Bengal. The detection is a bit difficult in our study area because the trees are very crowded in our study area. In our deep learning method, we have collected a large number of images and have separated them into training and testing images. Then, the CNN model is formed based on these training images. After forming the model, we have done parameter optimization by tuning the main parameters in order to obtain the best CNN model. We have then used this best CNN model for labeling all the images in the dataset that we have collected through sliding window technique. After label prediction, we have merged the same shade trees coordinates into one shade tree coordinate and have finally obtained the detection results. By this method, we have achieved an accuracy of prediction of shade trees more than 96%. The future scope of work will include the increase of sample size work to improve the accuracy of current suggested method.

2 Data and Preprocessing

Raw data consisted of frames extracted from footage of tea gardens on the globe. The view was from the top of the earth's surface. The study area is situated in the Jalpaiguri district, West Bengal, shown in Fig. 1. We have collected the samples from two different regions of Jalpaiguri district denoted by red and blue rectangles in order to evaluate the performance of our suggested method. The detected images are then compared with the ground truth of the samples collected manually. The training data set consisted of frames from the portions of tea gardens which consisted of shade trees and the testing data set consisted of frames from the portions of tea gardens which did not contain any shade trees. The final data images consisted of 200 training images and 100 testing images (Fig. 2).

3 Implementation

In this method, two phases are there: (1) training phase and (2) detection phase. During the training, a two-layered model is trained to capture the components of an object. The first layer in each case is trained with a seed. The first layer acts as a binary classifier. This output acts as input to the second layer. After classification, the detection is done by drawing bounding boxes around the detected images.



Fig. 1 Overall study area



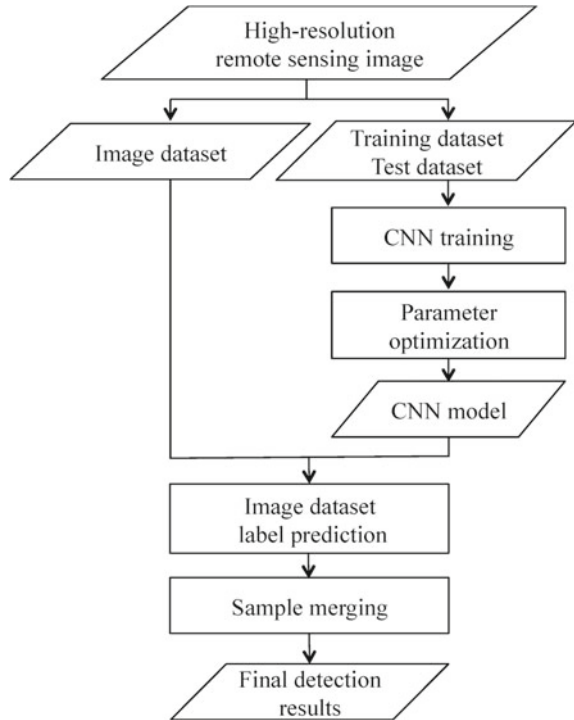
Fig. 2 Study area—two regions for samples

3.1 Overview

The sequence of steps that we need to perform to do object detection is depicted in Fig. 3. The CNN model was put into effect implemented by the Tensorflow model. At first, we had collected many images and had separated them into training and testing sets. The CNN model is then trained with the training images.

The main parameters of the CNN were adjusted and tuned continuously to obtain the best CNN model. Then, the images for detecting the shade trees were collected by sliding window of size 17×17 and with sliding step three pixels. Then, we have used

Fig. 3 Flowchart of our proposed method



the CNN model for labeling the images. After label prediction, we had done sample merging to merge same shade trees coordinate into one coordinate and obtained the final results.

3.2 Training CNN Model

We had used the LeNet CNN in our research work which consists of two convolutional layer, two pooling layers and a fully connected layer. The structure of convolutional neural network is shown in Fig. 4. The Rectified Linear Unit is the activation function of CNN. We have collected 5000 shade tree samples and 4000 non-shade trees samples from two different regions of Jalpaiguri district. Then, 7200 images are used for training, and 1800 images are used for testing. The main parameters of the CNN were adjusted and tuned continuously to obtain the best CNN model. Then, we had used the best CNN model for labeling the images. After label prediction, we had done sample merging to merge same shade trees coordinate into a single coordinate and obtained the final results.

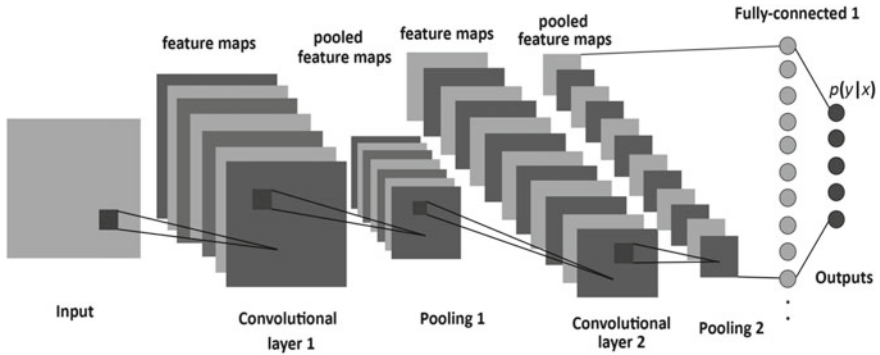


Fig. 4 CNN model



Fig. 5 Sliding window technique

3.3 Label Prediction

The images are collected by detecting the images by sliding window through the image. The size of the sliding window is 17×17 pixels and the sliding step is three. It should be noted that the sliding step is very important for shade trees detection. If the sliding step is too large, then many shade trees will be missed and it will give wrong detection results. If the sliding step is too small, then one shade tree will be detected repeatedly which will again produce wrong detection results. For this, through experimental tests we have selected the sliding step to be three so that we will obtain the correct detection results. After this, the CNN model is used for training (Fig. 5).



Fig. 6 Sample merging

3.4 Sample Merging

After label prediction of all the images, the spatial distance between the shade trees is collected. Here, a problem may arise that the number of predicted shade trees can be greater than the actual number of shade trees. This problem is due to the overlapping of the crowns as the trees are very crowded. To solve this problem, the same shade trees coordinates are merged into a single shade tree coordinate. In this research work, we have assumed that the spatial distance between two shade trees is greater than 8 pixels. Accordingly, we have set the merging process to be completed in six iterations. In each iteration, the coordinates with Euclidean distance less than a certain threshold is merged into single coordinates. The original coordinates are thus replaced by the average coordinates. The remaining coordinates represent the actual detected shade trees coordinates (Fig. 6).

4 Results

4.1 Accuracy

The accuracy of our CNN model was tested by 1800 test images. In order to obtain the highest accuracy, we have to optimize the parameters. For our research work, we have taken convolutional kernel size to be six, max-pooling kernel size to be three and mini-batch size to be 10 and 8500 iterations. We have adjusted the parameters in order to optimize the model. Thus, we can obtain an overall accuracy of 95% after

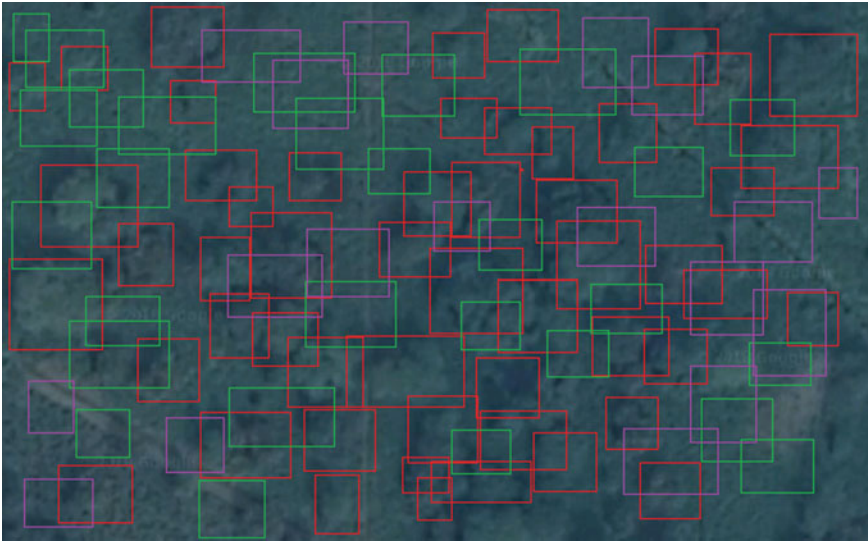


Fig. 7 Detection image of region 1. Each red square represents the shade trees that are detected correctly. The green squares denote the shade trees that cannot be detected correctly. The purple squares denote the background sample that is detected as a shade tree by mistake

8000 iterations if the number of kernels in two layers to be 40 and 60 and the number of hidden units to be 700.

4.2 Detection Results

In order to evaluate the accuracy of CNN model, we have to calculate precision, recall and accuracy. The respective formulas are given by the three equations below. Precision is the probability of detected shade trees to be valid. Recall is the probability of shade trees in ground truth to be detected correctly. Accuracy is then obtained by taking the average of precision and recall.

$$\text{Precision} = \frac{\text{The number of correctly detected shade trees}}{\text{The number of all detected objects}} \tag{1}$$

$$\text{Recall} = \frac{\text{The number of correctly detected shade trees}}{\text{The number of shade trees in ground truth}} \tag{2}$$

$$\text{Overall Accuracy} = \text{Precision} + \text{Recall}/2 \tag{3}$$

Table 1 shows the overall accuracy of the three regions as 96.22, 96.28 and 98.70%. These results are useful for both shade trees detection and counting (Fig. 7).

Table 1 Results of the accuracy of CNN

Index	Region 1	Region 2	Region 3
Detected shade trees count	1655	1610	1685
All detected objects count	1730	1700	1710
Actual shade trees count	1709	1645	1705
Precision (%)	95.66	94.70	98.53
Recall (%)	96.78	97.87	98.88
Accuracy (%)	96.22	96.28	98.70

5 Conclusion

In this paper, we have explained the detection and counting of shade trees by satellite images. From the table above, we have seen that the overall accuracy for the three regions is 96.22, 96.28 and 98.70%, respectively. We have also obtained the difference between the number of predicted shade trees and the actual shade trees to be less than 4%. Thus, we can say that the CNN model which we have formed through training and parameter optimization has huge accuracy. In our future research work, we want to further increase the accuracy by increasing the number of training and testing samples.

Bibliography

1. S. Suhaily, M. Jawaid, H.A. Khalil, A.R. Mohamed, F. Ibrahim, A review of oil shade bio-composites for furniture design and applications: potential and challenges. *BioResources* **7**, 4400–4423 (2012)
2. S. Malek, Y. Bazi, N. Alajlan, H. AlHichri, F. Melgani, Efficient framework for shade tree detection in UAV images. *IEEE. J. Sel. Top. Appl. Earth Obs.* **7**, 4692–4703 (2014)
3. A.P. Cracknell, K.D. Kanniah, K.P. Tan, L. Wang, Evaluation of MODIS gross primary productivity and land cover products for the humid tropics using oil shade trees in Peninsular Malaysia and Google Earth imagery. *Int. J. Remote Sens.* **34**, 7400–7423 (2013)
4. K.P. Tan, K.D. Kanniah, A.P. Cracknell, A review of remote sensing based productivity models and their suitability for studying oil shade productivity in tropical regions. *Prog. Phys. Geogr.* **36**, 655–679 (2012)
5. K.P. Tan, K.D. Kanniah, A.P. Cracknell, Use of UK-DMC2 and ALOS PALSAR for studying the age of oil shade trees in southern peninsular Malaysia. *Int. J. Remote Sens.* **34**, 7424–7446 (2013)
6. K.D. Kanniah, K.P. Tan, A.P. Cracknell, UK-DMC2 satellite data for deriving biophysical parameters of oil shade trees in Malaysia, in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, Munich, Germany (2012), pp. 6569–6572
7. Y. Cheng, L. Yu, A.P. Cracknell, P. Gong, Oil shade mapping using Landsat and PALSAR: a case study in Malaysia. *Int. J. Remote Sens.* **37**, 5431–5442 (2016)
8. A.P. Cracknell, K.D. Kanniah, K.P. Tan, L. Wang, Towards the development of a regional version of MOD17 for the determination of gross and net primary productivity of oil shade trees. *Int. J. Remote Sens.* **36**, 262–289 (2015)

9. D.A. Pouliot, D.J. King, F.W. Bell, D.G. Pitt, Automated tree crown detection and delineation in high-resolution digital camera imagery of coniferous forest regeneration. *Remote Sens. Environ.* **82**, 322–334 (2002)
10. H.Z. Shafri, N. Hamdan, M.I. Saripan, Semi-automatic detection and counting of oil shade trees from high spatial resolution airborne imagery. *Int. J. Remote Sens.* **32**, 2095–2115 (2011)
11. Y. Ke, L. Quackenbush, J. 2011 A review of methods for automatic individual tree-crown detection and delineation from passive remote sensing. *Int. J. Remote Sens.* **32**, 4725–4747 (2011)
12. P. Srestasathiern, P. Rakwatin, Oil shade tree detection with high resolution multi-spectral satellite imagery. *Remote Sens.* **6**, 9749–9774 (2014)
13. A. Manandhar, L. Hoegner, U. Stilla, Shade tree detection using circular autocorrelation of polar shape matrix. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* III-3, (2016) 465–472
14. A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, in *Proceedings of the Advances in Neural Information Processing Systems*, Lake Tahoe, NV, USA, 3–8 Dec 2012, pp. 1097–1105
15. D. Ciregan, U. Meier, J. Schmidhuber, Multi-column deep neural networks for image classification, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Rhode Island, RI, USA, 16–21 June 2012, pp. 3642–3649
16. Y. Sun, X. Wang, X. Tang, Deep learning face representation from predicting 10,000 classes, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA*, 24–27 June 2014, pp. 1891–1898
17. H. Li, Z. Lin, X. Shen, J. Brandt, G. Hua, A convolutional neural network cascade for face detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 7–12 June 2015, pp. 5325–5334

Hand Segmentation from Complex Background for Gesture Recognition



Soumi Paul, Arpan Bhattacharyya, Ayatullah Faruk Mollah,
Subhadip Basu and Mita Nasipuri

Abstract Hand gesture recognition has become very popular in the field of human–computer interface in recent years. Hand region identification is a fundamental task in most vision-based gesture recognition systems, since the subsequent detection and then segmentation depends on the quality of segmentation. If the background is complex and the illumination is varying, then the segmentation can be a difficult task. Different physical controllers like data gloves, color bands, mouse, and joysticks are used for human–computer interaction, which hinders natural interface as there is a strong barrier between the user and the computer. In such environments, most hand detection techniques fail to obtain the exact region of the hand shape, especially in cases of dynamic gestures. Meeting these requirements becomes very difficult, due to real-life scenarios. To overcome these problems, in this paper, we propose a new method for static hand detection and contour extraction from a complex background. We employ a new technique, histogram thresholding which gives better result over depth thresholding to improve hand region extraction.

Keywords Complex background · Hand gestures · Vision-based · Histogram thresholding · Sign language

S. Paul (✉) · A. Bhattacharyya · S. Basu · M. Nasipuri
Department of Computer Science & Engineering,
Jadavpur University, Kolkata 700032, India
e-mail: spaul.cse.rs@jadavpuruniversity.in

A. Bhattacharyya
e-mail: 2013arpan@gmail.com

S. Basu
e-mail: subhadip@cse.jdvu.ac.in

M. Nasipuri
e-mail: mnasipuri@cse.jdvu.ac.in

A. F. Mollah
Department of Computer Science and Engineering, Aliah University, Kolkata, India
e-mail: afmollah@aliah.ac.in

1 Introduction

Human–computer interaction arose as a small topic, but over the last couple of decades, it has grown up as an important field of its own. Modern computer systems need to understand human gestures, speech, facial, and bodily expressions and respond accordingly. The idea is that the user would remain as natural as possible and behave with the computer as if it is behaving with another human being. When it comes to the emerging fields such as sign language recognition, the interaction between the computer and the human being becomes more and more challenging.

1.1 Existing Works

There have been several works in vision-based hand gesture recognition algorithms. For a comprehensive review, refer to Mitra et al. [4]. Previously used techniques for capturing input gestures used different types of hand-held devices where one of the most promising one was data gloves [3]. To read hand movements, hand belt with gyroscope, accelerometer, and a Bluetooth was deployed [1]. The recent development of depth cameras, such as Microsoft Kinect [6], Creative Senz3D [10], or Mesa Swiss-Ranger [2], etc., opens up new avenues for hand gesture recognition. For perfect contour extraction, Ravikiran et al. have used boundary tracing and fingertip detection to recognize one-handed static finger spellings of American Sign Language (ASL) [8]. Vision-based features have been the preliminary means for hand detection. They are sensitive to variation in subject and surroundings, such as change of color of skin and the ambience light and shade. On the contrary, depth sensor opens the avenue of much simpler approach of depth thresholding [7] to isolate the hand. After localizing and segmenting the hand, several hand features can be extracted from either the depth maps, or the color maps, or combined.

2 Hand Detection in 3D

The steps involved in static hand gestures segmentation from complex background is described below.

2.1 Getting the Depth Image from Kinect

Microsoft Kinect was initially launched as a gaming device. With time, it has evolved into a high-resolution yet low-cost, depth as well as visual (RGB) sensing device. It has an infrared camera and PrimeSense sensor to capture the depth of the object upto

a certain distance. Microsoft Kinect V1 was released in February 2012 and Kinect V2 in the year 2013 with lots of feature improvements. It also has a RGB camera to capture the color images.

With Kinect device, Microsoft has also provided an SDK and some applications to help the users to understand how the color and depth images are getting captured by Kinect.

2.2 Data Collection

We use two publicly available datasets, collected using Microsoft Kinect sensor V1. The first one is the NTU dataset of [9]. Ten different subjects, each with 10 gestures for the numbers 0–9, contributed toward this. This yields 1000 cases each consisting of the pair: a color image and the corresponding depth image.

The second one is SPEMD dataset by [11]. In this dataset, they captured 10 gestures with 20 different poses from 5 subjects. Therefore, a total of 1000 cases for testing, labeled from 0 to 9 is available. Both these datasets have been collected using Kinect V1 sensor.

2.3 Hand Localization

In the previous depth camera-based approaches, the hand was required to be the front-most object from the depth camera. Moreover, some markers like a colored gloves or black belt on the gesturing hand's or wrist were also required to locate the hand position. In case of static images, simple background was also important to extract only hand and not any other body parts. Even with improper illumination, silhouette of the hand can impact on the region extraction from RGB images. In our system, these restrictions are relaxed by making use of the stable joints from depth sensors skeleton tracking and accepting only the depth images as input data. The joints getting from the sensor help in locating the hands, wrists, and also the elbows.

One implicit assumption in the whole process is that the hand is visible to the camera without any occlusion. This assumption helps us to quickly separate the hands from the background objects using depth information alone. Quick hand segmentation is performed based on color texture as well as depth map, using the hand joint point as the center.

2.4 Segmentation and Noise Removal

Same gestures when captured in camera can have varied representations. The hand-to-camera distance changes from person to person, so does the palm size. Also, there may

be different rotations of body postures. Thus, before cognition, some normalization is mandatory to have uniform and rotation-invariant, scale-invariant shapes suitable for recognition. So for extracting the hand region from its background correctly, we applied two distinct methods: depth thresholding and histogram thresholding.

In depth thresholding, which is a simple segmentation method, thresholding is applied to obtain the hand image as region of interest from the grayscale image. Each pixel intensity value (I) is replaced by black color if it is below some threshold (T). Otherwise, it is replaced by white color.

The operation of thresholding can be represented as shown in Eq. 1

$$I'(i, j) = \begin{cases} 0, & \text{if } I(i, j) > T \\ 255, & \text{otherwise} \end{cases} \quad (1)$$

If the input pixel intensity $I(i, j)$ is greater than the threshold value T , then the new intensity of the pixel $I'(i, j)$ is set to 0. Otherwise, it is set to the maximum possible value 255.

We use a threshold value (T) of 127 in our work for the above operation. Two types of thresholding are implemented, namely, inverted binary thresholding and Otsu's thresholding [5]. Inverted binary thresholding performs color inversion, i.e., it replaces an intensity I with $255-I$. As a result, we get a white image in a black background. Whereas for histogram thresholding, we need to calculate the mean and the standard deviation of the depth values and extracted grayscale cropped hand shapes.

Mathematically speaking, if D is the maximum depth, then for $i = 0$ to D , we count the number of pixels of depth i in the variable n_i .

Now, we take a sliding window of length w (an odd number) along the sequence n_0, \dots, n_D of size $D + 1$, and calculate the average within each of $D - w + 2$ such windows. Taking $k = (w - 1)/2$, we index these average values as v_k, \dots, v_{D-k} against depths k to $D - k$, respectively.

Next, we take another sliding window of length w along the sequence v_k, \dots, v_{D-k} and calculate the standard deviation (SD) within each of $D - 2w + 2$ such windows. We index these SD values as s_{2k}, \dots, s_{D-2k} against depths $2k$ to $D - 2k$ respectively.

Then we plot the v and SD values against the corresponding depth indices and look for $w + 1$ many consecutive zero SD values, while moving from left (depth 0) to right (depth D). We set our depth threshold d_{th} to be the first depth from where such a long zero SD sequence starts.

By empirical fine-tuning and optimization, the most suitable value of w was found to be 5 in our case. In Fig. 1, the graph of average v_i 's and standard deviation s_i 's are plotted against the depth values. It is observed that $d_{th} = 100$, since a long consecutive sequence of zero SD begins from this depth onwards.

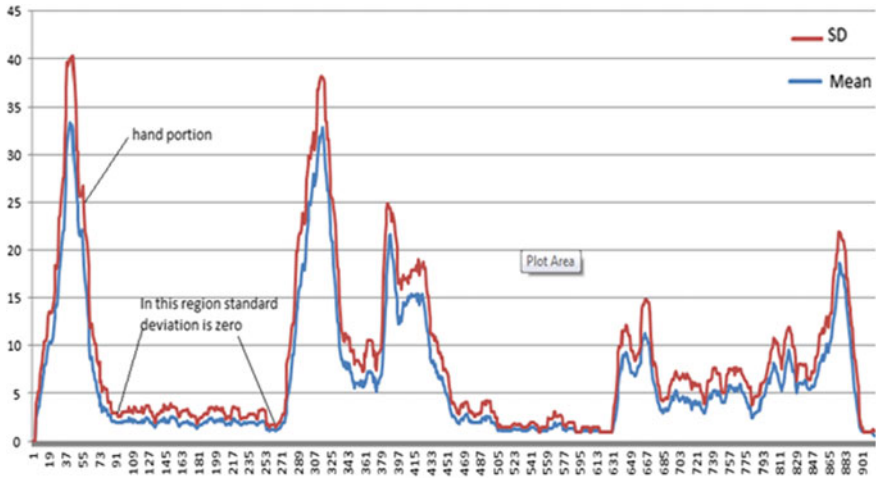


Fig. 1 Plotting of mean and standard deviation for preprocessing of hand gesture

3 Results

In our proposed method of histogram thresholding over depth thresholding, we have experimented over 10 gestures which are 0–9 from American Sign Language. Here, we have experimented our method on two publicly available datasets, NTU [9] and SPEMD [11].

Figures 2, 3, and 4 correspond to raw depth image, depth thresholded image, and histogram thresholded image of NTU dataset, respectively. Figures 5, 6, and 7 correspond to raw depth image, depth thresholded image and histogram thresholded image of SPEMD dataset, respectively. It is observed that for both the datasets, hand region extraction from histogram thresholded image is giving better results than that extracted from depth thresholded images.

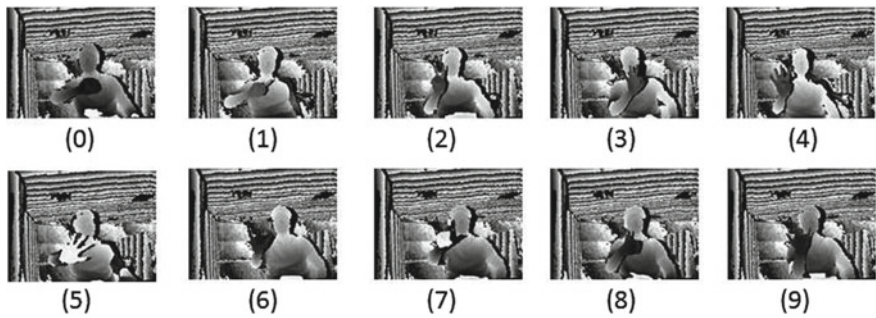


Fig. 2 Raw depth images from NTU Dataset

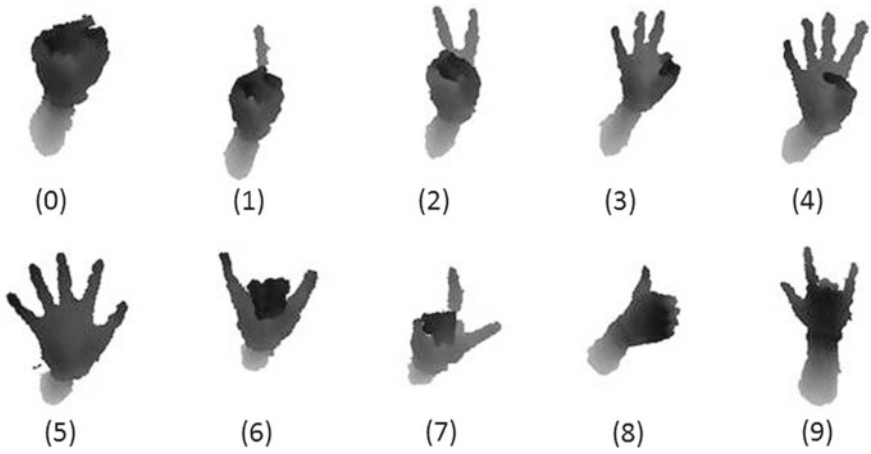


Fig. 3 Depth images after depth thresholding on NTU dataset

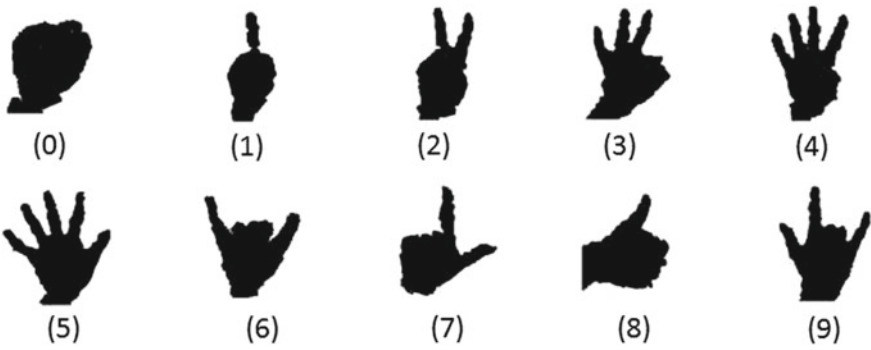


Fig. 4 Depth images after histogram depth thresholding on NTU dataset

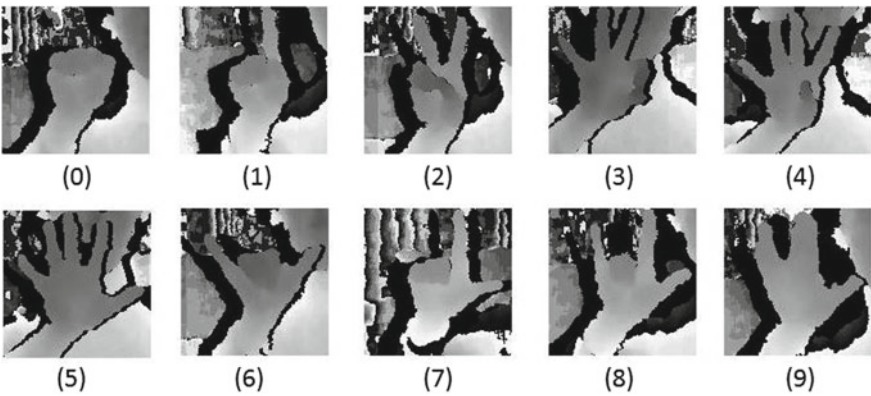


Fig. 5 Raw depth images from SPEMD dataset

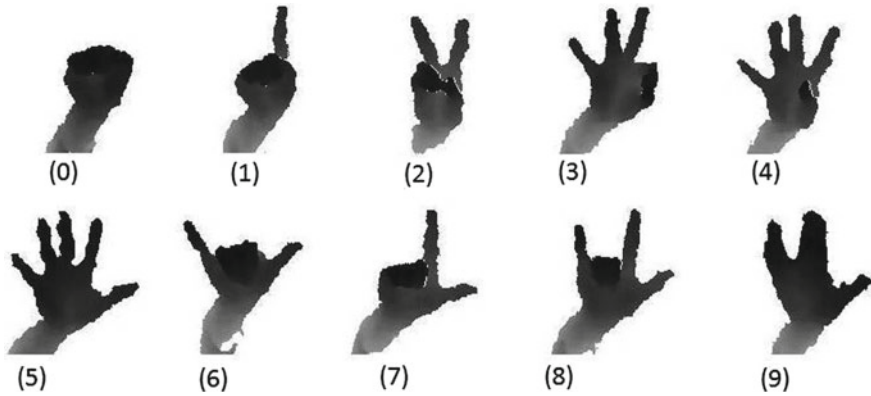


Fig. 6 Depth images after depth thresholding on SPEMD dataset

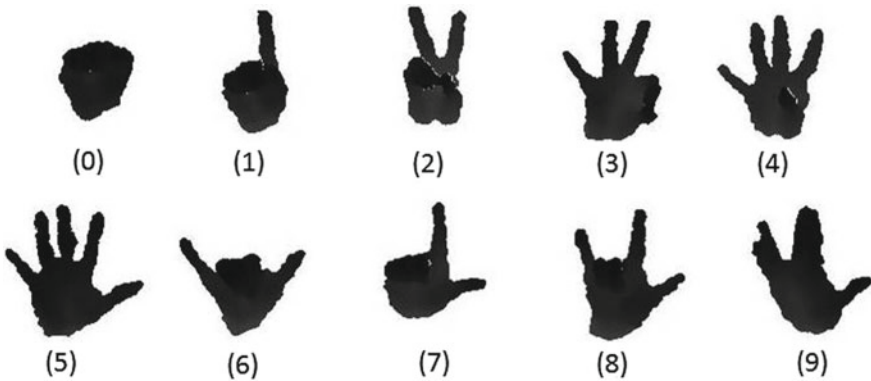


Fig. 7 Depth images after histogram depth thresholding on SPEMD dataset

4 Conclusion

As our current approach of histogram thresholding is showing promising results over normal depth thresholding, this technique can be adopted for hand segmentation. As future work, we may consider other variants of thresholding for possibility of finer extraction of hand region from complex background and also for some dynamic images in real time. Since not much data is publicly available for Microsoft Kinect sensor V2, we are in the process of creating our own benchmark datasets. We plan to experiment with thresholding, new feature extraction and then gesture recognition on our datasets.

Acknowledgements This research was being carried out at Multimedia Lab, Jadavpur University, INDIA, as a part of our current project “Development of Some Techniques for Automatic Indian Sign Language Recognition”. This work is partially supported by the Department of Science and Technology (DST) WOS-A Schema. Authors would also like to thank the Jadavpur University administration for providing the research facilities in the university.

References

1. C.-H. Hung, Y.-W. Bai, H.-Y. Wu, Home outlet and LED array lamp controlled by a smart-phone with a hand gesture recognition, in *2016 IEEE International Conference on Consumer Electronics (ICCE)* (IEEE, New York, 2016), pp. 5–6
2. T. Kapuscinski, M. Oszust, M. Wysocki, Recognition of signed dynamic expressions observed by TOF camera, in *2013 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)* (Sept 2013), pp. 291–296
3. G. Luzhnica, J. Simon, E. Lex, V. Pammer, A sliding window approach to natural hand gesture recognition using a custom data glove, in *2016 IEEE Symposium on 3D User Interfaces (3DUI)* (IEEE, New York, 2016), pp. 81–90
4. S. Mitra, T. Acharya, Gesture recognition: a survey. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **37**(3), 311–324 (2007)
5. N. Otsu, A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man Cybern.* **9**(1), 62–66 (1979)
6. S. Paul, S. Basu, M. Nasipuri, Microsoft kinect in gesture recognition: a short review. *Int. J. Control Theory Appl.* **8**(5), 2071–2076 (2015)
7. S. Paul, H. Nasser, M. Nasipuri, P. Ngo, S. Basu, I. Debled-Rennesson, A statistical-topological feature combination for recognition of isolated hand gestures from kinect based depth images, in *18th International Workshop on Combinatorial Image Analysis (IWCI)* (Springer LNCS, 2017), pp. 256–267
8. J. Ravikiran, K. Mahesh, S. Mahishi, R. Dheeraj, S. Sudheender, N.V. Pujari, Finger detection for sign language recognition, in *Proceedings of the International MultiConference of Engineers and Computer Scientists*, vol.1 (2009), pp. 18–20
9. Z. Ren, J. Yuan, J. Meng, Z. Zhang, Robust part-based hand gesture recognition using kinect sensor. *IEEE Trans. Multimedia* **15**(5), 1110–1120 (2013)
10. Y. She, Q. Wang, Y. Jia, T. Gu, Q. He, B. Yang, A real-time hand gesture recognition approach based on motion features of feature points, in *Proceedings of the 2014 IEEE 17th International Conference on Computational Science and Engineering* (IEEE Computer Society, 2014), pp. 1096–1102
11. C. Wang, Z. Liu, S.-C. Chan, Superpixel-based hand gesture recognition with kinect depth camera. *IEEE Trans. Multimedia* **17**(1), 29–39 (2015)

A Study on Content Selection and Cost-Effectiveness of Cognitive E-Learning in Distance Education of Rural Areas



Anindita Chatterjee, Kushal Ghosh and Biswajoy Chatterjee

Abstract Distance learning/education is to be carried out remotely. It focuses to deliver education to the learners who are not present physically in a traditional classroom or campus. The courses offered in distance learning helps rural people to understand the domain of rural development which in turn enhance their career scope and quality of life. It is more cost-effective than traditional classroom approach. Incorporation of cognitive study in distance education helps to judge the human nature by understanding their thought processes.

Keywords Distance education · Cognitive learning · Cost-effective · Learning content · Rural area

1 Introduction

The feasibility of implementation and enhancement of the rural distance education can be done with the help of courses. These courses help them to understand the domain of rural development and enrich themselves for a better life. The enhanced and enriched knowledge helps people to deal with the global challenges in a better way. These courses help them develop a different perspective and outlook toward their society, development issues, and attitudes toward the other people in need. A well-educated and established population equipped with knowledge not only helps drive the socioeconomic development but it also establishes personal growth. Rural

A. Chatterjee (✉)

Tata Consultancy Services Ltd, Kolkata, West Bengal, India
e-mail: anindita_star24@rediffmail.com

K. Ghosh

DCG Datacore Systems, Kolkata, West Bengal, India
e-mail: kushalcrickushal@gmail.com

B. Chatterjee

University of Engineering and Management, Kolkata, India
e-mail: biswajoy.chatterjee@iemcal.com

© Springer Nature Singapore Pte Ltd. 2020

J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937, https://doi.org/10.1007/978-981-13-7403-6_69

areas account for a larger part of the geographical area in many countries. Provisioning of basic infrastructure facilities for this large rural area has been a major challenge because infrastructure plays a vital role for not just the country's economic growth but also it brings progress in human development [1, 2]. Therefore, gaps in rural infrastructure need to be identified and addressed properly in order to achieve redistributive growth.

For the ease of learning, we need to judge the cognitive level of learners so that according to their cognitive level we can categorize them and based on this level, learning materials can be selected [3]. It is good to analyze the cognitive level of learners first and then to provide guidance and materials according to their merit. By surveying the area, the best media and medium to deliver the distance learning program can be identified. In distance learning, since there is no concept of classroom teaching hence the cost involved decreases. Also there is no travel time for the learner which reduces cost [4, 5].

2 Cognitive Learning

The cognitive learning is a process that mostly focuses on the processes which are involved in learning rather than on the observed behavior. It focuses mainly on the internal processes and connections that take place during the learning process. Learners are active participants in any learning process. They use various techniques to process and to construct their understanding of the content to which their minds are exposed. It is often not easy to understand and to identify the knowledge that already exists in the minds of learners. Most of the times, the knowledge and experience of the students are too complicated that is to be identified thoroughly [6, 7]. A teacher or instructor can ask question which in turn help students refine their thinking process and recognize where they may be wrong. Failure may be considered as a good thing because it helps learners to realize that there is a gap in their understanding and they need to learn more [8]. Role of the instructor is to monitor their progress and asking lots of questions in order to confirm better understanding of the learner. In order to judge learner's merit, cognitive study should be performed for a successful e-learning program [9, 10].

3 Discussion

The content should be based on the age group of learners and again it should be categorized according to the merit level of each age group. Merit of each group will be judged based on the cognitive level. Proper survey should be done in order to justify the cognitive level of the learners in their regional language. In this way, the best media and medium to deliver the distance learning program for rural people can be determined. The parameters which should be considered in order to get good result

are as follows: geographical location, density of population, no. of schools (private and government), percentage of literacy, and availability of Internet, availability of electricity and communication language. The content of course materials can be updated timely for better understanding. Efficiency of information delivery is also high for distance education. Subject Matter Experts (SMEs) can help manage the course content and context, and review them for efficiency and accuracy [11].

Distance learning is cost which in turn the production of content into a durable learning look which does not require much infrastructure. Study materials can be easily packaged in “just-in-time” format. This education is flexible and can be web-based and can be accessed anytime from anywhere as long as there is a computer connected to the Internet. It can also be accessed via smart phones. Interesting contents create an immersive and effective learning environment that motivates and engages the learners.

In distance learning of rural area, face-to-face interaction is not possible frequently hence traditional classroom type teaching methodology is not possible all the time. It is one of the required essentials in order to achieve success in teaching program. In case of distance learning, this classroom type teaching process is possible via video conferencing, VoIP, etc. But this kind of techniques mostly violets travel expenditure and time. In rural areas, it is difficult to arrange video conferencing, VoIP, etc. techniques. In order to perform virtual interaction instead of traditional interaction, it needs to judge proper cognitive level so that proper study materials can be assigned as per their merits [12].

4 Conclusion

For distance education, content selection is one of the vital aspects in order to make it successful. Traditional classroom type concept is not present for distance education that is why the courses play a vital role. If the content is not selected properly, then there will be a gap between learner and teaching procedure because there is hardly interaction as that of traditional approach. Most of the time, rural people cannot manage their time for completing education as due to financial problem or they have very little scope. Since distance education is cost-effective it helps rural people better opportunity today life and fulfills their needs.

References

1. R. Guoqiang, The research on distance education for rural off-farm employee, IEEE, (2009), pp. 648–652
2. Y. Wang, Y. Cheng, F. Wang et al., Design and implementation of WEB-based distance teaching support platform for rural teachers, IEEE, (2009), pp. 88–90
3. Y. Cheng, Z. Zheng, Y. Wang et al., Study on the instructional model in rural distance education for K-12, IEEE, (2009), pp. 100–103

4. Y. Cheng, Y. Wang et al., Strategic research on constructing and developing resources for modern distance education project, *IEEE*, (2009), pp. 275–278
5. M.R. Syed, H. Rahman, M.S. Alam, Study on the instructional model in rural distance education for k-12, *IEEE*, (2009), pp. 100–103
6. Z. QU, X. Wang, Application of PCA in assessing influential factors and characteristics of modern distance education in rural areas, *IEEE*, (2009), pp. 460–464
7. Quqizhong, Study on the resource construction of modern distance education in rural schools, *IEEE*, (2009), pp. 1796–1799
8. F.-S. chen, C.W. Liao, Tsaihsiuchen, Adult distance education student's perspective on critical success factor of e-learning, *IEEE*, (2009), pp. 140–143
9. E.R. Fowler, W.B. Hudson, Distance education from faculty perspective, *IEEE*, (1994), pp. 583–585
10. T.T. Lau, The impact of web-based training system for distance education: a corporate perspective, *IEEE*, (1999), pp. 257–260
11. A.F. Sharef, Kinshuk, Distance education model for secondary schools in Maldives, *IEEE*, (2003), pp. 479–483
12. B. Li, N. Zhang, The Analysis and reform of the distance education in China's rural area, *IEEE*, (2009), pp. 173–180
13. T.K. Shih, Distance education technologies current trends and software system, *IEEE*, (2002), pp. 38–43
14. A. Debapriya, P. Durgaprasad, Rural development management education in India: perspectives and potentials, in *African-Asian Journal of Rural Development*, Vol. XXXXV (2), (2012)
15. D. Das, S. Samanta, Rural education in India: as an engine of sustainable rural development. *Int. J. Res. Humanit. Arts Lit. ISSN* 2(10), 2347–4564 (2014)
16. T.A. Angelo, K.P. Cross, *Classroom assessment techniques: a handbook for college teachers*, 2nd edn. (Jossey-Bass Publishers, San Francisco, 1993)
17. J. Willis, A. Thompson, W. Sadera, Research on technology and teacher education: current status and future directions (1999), pp. 29–45
18. L. Gang, C. Haiyun, J. Anquan, A new teaching model of project orientation and distance collaboration (2008), pp. 459–461
19. K.M. Conn, Identifying effective education interventions in sub-Saharan Africa: a meta-analysis of impact evaluations (2017)
20. C.H. Peterson, N.A. Peterse, K.G. Powell, Cognitive interviewing for item development: validity evidence based on content and response processes (2017)

An Automatic Method for Bifurcation Angle Calculation in Retinal Fundus Image



Suchismita Goswami and Sushmita Goswami

Abstract Retinal vascular angles are found to be one of the crucial parameters for biometric template generation. Many of the previous literature have used bifurcation angles for arteriovenous classification also. Acute angles are preferred as bifurcation angles in most of the cases. To prove vascular angle at a bifurcation point as an acute angle is itself a challenging task. In this paper, an initiative has been taken to prove bifurcation angle is acute followed by distinguishing the actual and distorted bifurcation angle. The accuracy of the proposed approach has been evaluated from a publicly available database using the experimental observation.

Keywords Vascular angle · Bifurcation point · Bifurcation angle

1 Introduction

The tendency to adopt biometric authentication technology shows a better solution against vulnerable threats in traditional security system. Retina, being one of the secure authentication trait, is found to be reliable in this phenomena. Retinal vessels are anatomically kept safely under conjunctiva of human eye, and it is hard to tamper with it. Due to the anatomical structure of it and hassle-free capturing process, retinal authentication has gained its popularity among other biometric features. In many literature, vascular angles, which bear significant information, have been used as a successful parameter for making retinal template [1]. When a retinal vessel bifurcates, two other child vessels are generated forming an angular difference with the parent vessel. The point from where this division occurs is termed as bifurcation point. Angle of bifurcation is used in some literature as a biometric parameter. Finding bifurcation angle is a challenging task in such an environment where often transformation occurs

S. Goswami (✉) · S. Goswami
Hyland Software Solutions India LLP, Kolkata, India
e-mail: suchismita.g24@gmail.com

S. Goswami
e-mail: sushmita.g24@gmail.com

© Springer Nature Singapore Pte Ltd. 2020
J. K. Mandal and D. Bhattacharya (eds.), *Emerging Technology in Modelling and Graphics*, Advances in Intelligent Systems and Computing 937,
https://doi.org/10.1007/978-981-13-7403-6_70

on the image. Also, it has been seen that the bifurcation angles are acute in nature. The statement, nevertheless is true, needs some justification against it. In this paper, we have taken an initiative to prove that bifurcation angles are acute in nature. To accomplish the goal, retinal images are passed through several image processing steps for proper segmentation. The vessels are thick white set of pixels on black background, and finding centerline pixels of the vessels is an essential task in angle calculation. Organization of the paper is as follows. After this introduction written in Sect. 1, we present a brief review of the related works in Sect. 2. The proposed method is described in Sect. 3 including image preprocessing and segmentation. Section 4 furnishes the experimental results along with the performance analyses. Finally, concluding remarks are made in Sect. 5. The complete flow of the proposed approach has been described pictorially in Fig. 1.

2 Related Work

An automated approach has been proposed to find various bifurcation features of retina including bifurcation point, vessel width, bifurcation angle which is very competent in clinical conditions [2]. Proposed method in this paper applied Ribbon of Twins active contour model to calculate various bifurcation features and attained an accuracy with slightly different bias from manual measurement. Bifurcation angle measurement from retinal vascular structure based on orientation estimation is proposed in [3]. This approach involves identifying all the bifurcation and crossover points followed by measurement of bifurcation angle based on orientation estimation at each pixel of a gray retinal image. This approach eliminates the skeletonization of retinal image to avoid significant variation in the bifurcation angle.

A manual method of identifying retinal bifurcation feature including bifurcation angle has been proposed by drawing a line from the bifurcation point along the three segments centerline [4].

A novel approach to measure bifurcation angle using tracking-based method has been presented [5]. This approach involves measurement of the angles made between the child vessels and the direction of the parent vessels keeping the bifurcation point as source. Then the total bifurcation angle is calculated by the summation of these two angles.

3 Proposed Method

In this paper, a novel approach has been adopted to measure the bifurcation angle around each distinct bifurcation point. From the literature review, acute angle is preferred as bifurcation angle at each bifurcation point. But few exceptional condition arises due to severe diseases like diabetic retinopathy (DR) which causes the widening of the bifurcation angle. DR can be classified into two stages, non-proliferative

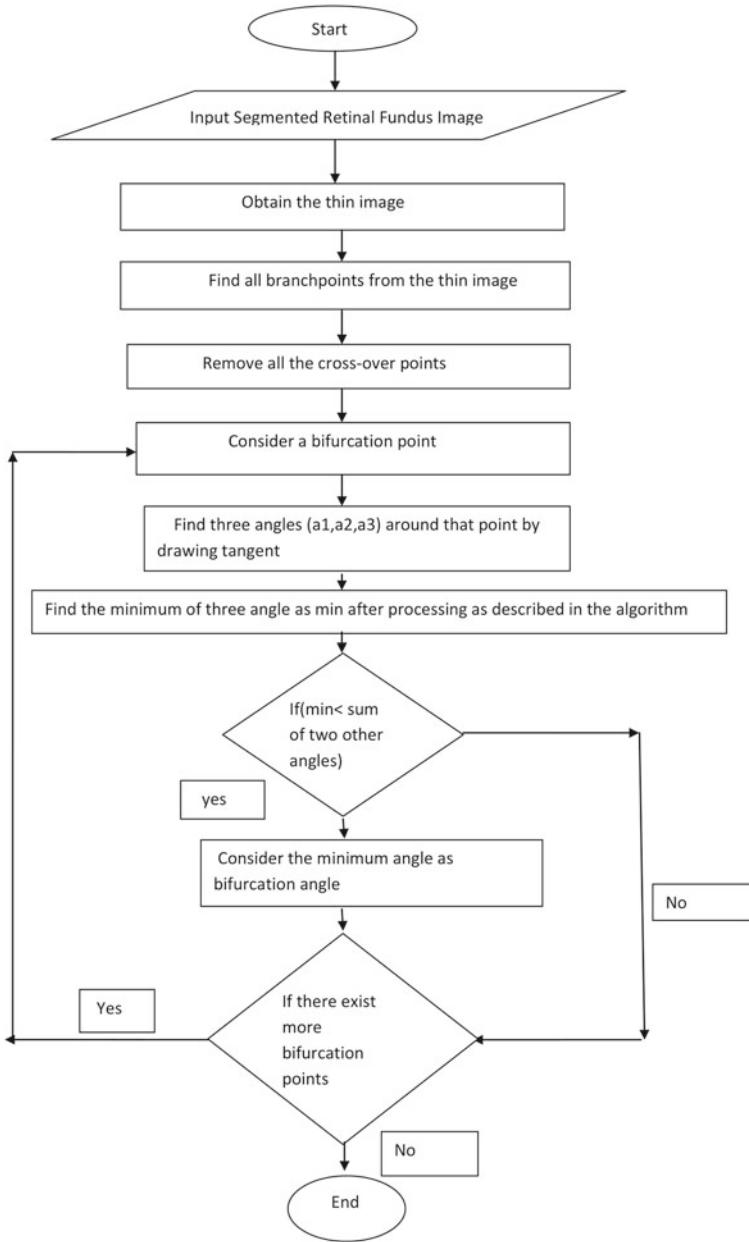


Fig. 1 Flow diagram of the proposed method

diabetic retinopathy (NPDR) and proliferative diabetic retinopathy (PDR). NPDR is defined to be the early stages of DR where blood vessels leak causing swelling of the retina. At the advance stage of DR, i.e., PDR, new fragile blood vessels develop and may cause vision loss. Hence, these widened (non-acute) bifurcation angle significantly contributes to the diagnosis and early detection of DR. So an approach has been taken to distinguish the acute and non-acute angles from retinal vascular structure. The proposed work consists of three major chunks of execution that are (i) image preprocessing and segmentation, (ii) distinct bifurcation point detection, and (iii) measurement of bifurcation angles. The individual steps are elaborated in the following sections.

3.1 Image Segmentation and Thinning

A captured retinal fundus image may be of various qualities, viz. low contrast, high contrast, or suffering from noise. To work with these kinds of input images effectively, we need an enhancement as preprocessing. Here, we have employed Otsu's thresholding followed by a sharpening method using CLAHE to obtain the enhanced image. Finally, 2D median filtering techniques are used to get a smooth textured image in a binary platform. The anatomical structure of blood vessels always does not resemble straight lines. Segmented binary image through this set of operations is shown in Fig. 2. Sometimes, the thickness of the blood vessels affects the accuracy of angle measurement process. So, the necessity arises to assess every vessel as a single-pixel blood vessel thread for the ease of the method. To do so, vessel width is measured here based on no. of white pixels present on black background. Any vessel with less than 3 pixel width is not considered as major vessels and need to be removed as they do not carry relevant information always. Midpoint of the total vessel width is the desired output here which resembles an almost straight line at some point in a small 3×3 scanning window.

Fig. 2 Showing the segmented result from RGB fundus image



3.2 Detection of Distinct Bifurcation Points

In an eight-neighborhood window, a pixel point with exactly three neighbors is the bifurcation point. Four neighbors show a possibility of being real crossover point and a situation which occurs due to an arteriovenous nicking which has been dealt here. The second category of the crossovers is sometimes presented as two close bifurcations often in a thinned image. Please refer Fig. 3. However, this situation leads to a false detection of real scenario as two close bifurcation points. Initially, all the branch points, including crossovers and bifurcations, are detected from the thinned image. Then the points with four or more neighbors are removed from the located branch points to get an image with only the distinct bifurcation points present. But the above method does not able to distinguish the crossover points appeared as closely connected bifurcation points due to arteriovenous nicking. Hence, to deal with this situation, distance between two close bifurcation points and the difference between angle at each of these two close bifurcation points are calculated. If the difference between these angles is less than a threshold value, provided the distance is also less than a certain threshold, and then the midpoint of those two close bifurcation points is marked as crossover point [6].

3.3 Measurement of Bifurcation Angle

Now, we are concerned with the angle measurement of the distinct bifurcation points on the thinned image as a result of the previous step. Around each bifurcation point within a 3×3 scanning window, all three vessels are traced and marked up to a certain distance as three distinct points. Keeping bifurcation point as the source coordinate and other marked points as the ending points, three tangents were drawn along the three branches using the formula $m = \tan \theta = (y1 - y2)/(x1 - x2)$. Here, $(x1, y1)$ is the bifurcation point and $(x2, y2)$ is one of the three points found by the previous method, and m is the gradient of the tangent. Then the angle between three vessels was calculated using the formulae $\theta = \tan^{-1}(m1 - m2)/(1 + m1 * m2)$, where θ

Fig. 3 **a** Crossover point with arteriovenous nicking and **b** crossover point falsely appeared as two closely connected bifurcation points after thinning operation



is the angle between two tangents, θ lies between -90 and $+90$, and m_1 and m_2 are the gradients of two tangents, respectively. By following the above-mentioned formula for angle calculation, three acute angles will be always obtained in such a way that bifurcation angle will always be greater than other two angles as shown in the Fig. 4. In support of the above statement, the absolute value of all three angles was calculated first. Then the maximum acute angle (θ_1) was considered as the angle of bifurcation point. Next for other two angles (θ_2) and (θ_3), corresponding obtuse angles are found which are $\theta_2 = 180 - \theta_2$ and $\theta_3 = 180 - \theta_3$. Hence, the minimum of these three angle is the bifurcation angle, and it satisfies the rule, i.e., ($\theta_1 < \theta_2 + \theta_3$). Please refer Fig. 5. There are some exceptional situations which arise in the image (shown in Fig. 5b, c), where summation of three consecutive angles does not produce 360° altogether. Identifying bifurcation angle from this category of points is a challenging task. Marking the measured acute angle always as bifurcation angle does not produce correct result here when the corresponding obtuse angle is the true bifurcation angle in reality. Distortion in bifurcation point causes the above scenario as shown in Fig. 5b, c. So $180 - (\text{bifurcation angle found previously})$ will be the actual bifurcation angle. Hence, the minimum of these three angle is the bifurcation angle, and it satisfies the rule, i.e., ($\theta_1 < \theta_2 + \theta_3$). These bifurcation points can be recognized as distorted bifurcation too. It has been found statistically that the value of the bifurcation angle, calculated by the above-mentioned formulae, will always yield acute angles between two child vessels. But it is noticed that the summation of all three angles ($\theta_1 + \theta_2 + \theta_3$) (as shown in this Fig. 5b) is incapable to produce 360° as a whole and (≤ 360). Hence, we conclude with the fact that it is a distorted bifurcation point. Since it is a distorted bifurcation point (i.e., bifurcation angle ≥ 90) as shown in Fig. 5b, in order to find out the exact bifurcation angle, corresponding obtuse angle (as shown in the previous figure) between two child nerves is to be taken, i.e., $180 - \text{acute angle}$.

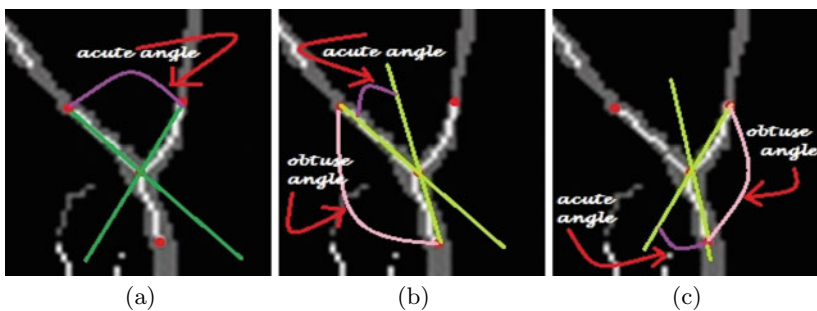


Fig. 4 **a** Angle between two child vessels, **b** angle between first child and parent vessels, **c** angle between second child and parent vessels

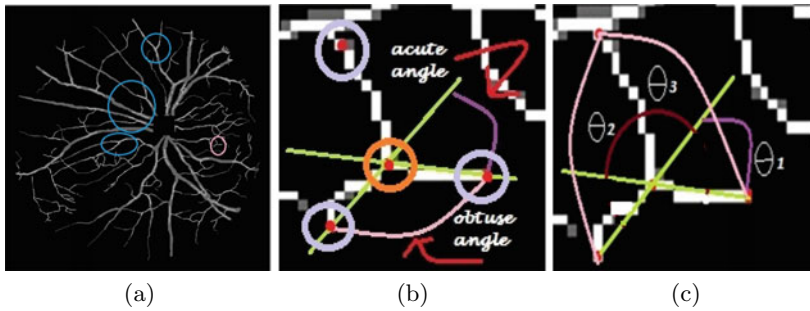


Fig. 5 **a** Bifurcation points, **b** distorted angle at bifurcation point ex_1 , and **c** distorted angle at bifurcation point ex_2

4 Results and Discussion

Following the above-mentioned method, an observation has been drawn that whenever there is an acute angle found around a bifurcation point, it is found to be the bifurcation angle. Figure 5b shows the glimpses of a real scenario on a blood vessel with bifurcation where minimum of all the angles around bifurcation point is found to be the desired one. Figure 5c also shows some more categories of the vasculature in support to the above-mentioned statement. Experimental results of the first image from DRIVE database [7] are summarized in a nut shell and presented in Table 1 for the first ten bifurcation points. In all the ten points, the minimum angle which is acute in nature is found to be the bifurcation angle by comparing with hand-driven vessel tracking mechanism (Fig. 6).

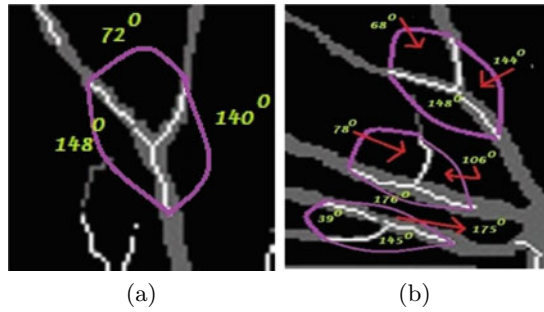
Experimental setup

To evaluate the performance of the proposed system, experiments have been conducted on the DRIVE [7] database. The images of the DRIVE database were obtained

Table 1 First ten angles (in degree) of the first image from DRIVE database

	Image 12	Image 26	Image 19	Image 3
Actual bifurcations present	73	64	51	84
Bifurcation point not detected (FN)	0	1	0	1
Correctly measured bifurcation angle (TP)	72	63	51	82
Incorrectly calculated bifurcation angle (FP)	1	0	51	1
No of acute bifurcation angle	68	58	48	76
No of distorted bifurcation angle	5	6	3	8
Sensitivity	1	0.984	1	0.988
Specificity	0	0	0	0
Accuracy (%)	98.63	98.43	100	97.62

Fig. 6 Bifurcation angle measurement



from a diabetic retinopathy screening program in The Netherlands. Each image was captured using 8 bits per color plane at 768 by 584 pixels. For the training images, a single manual segmentation of the vasculature is available. For the test cases, two manual segmented images are available; one is used as gold standard, the other one can be used to compare computer generated segmentation with those of an independent human observer.

Evaluation Parameter

To evaluate the proposed algorithm manually, following parameters are defined in order to compare the results from the algorithm with other algorithms.

1. TP (true positive) This refers to the pixels of the blood vessels that are correctly recognized by the algorithm. Let TP denotes the number of true positives.
2. FP (false positive) This refers to the pixels of the blood vessels that are incorrectly recognized as positive pixels by the algorithm. Let FP denotes the number of false positives.
3. TN (true negative) This refers to the pixels of the blood vessels that are correctly discarded by the algorithm. Let TN denotes the number of true negatives.
4. FN (false negative) This refers to the positive pixels of the blood vessels that were mislabeled as negative pixels by the proposed algorithm. Let FN denotes the number of false negatives.

Equations 1 and 2 help in calculating the final performance of the proposed method.

$$\begin{aligned}
 \text{Sensitivity} &= \frac{TP}{(TP + FN)} \\
 \text{Specificity} &= \frac{TN}{(TN + FP)} \\
 \text{Accuracy} &= \frac{(TP + TN)}{(P + N)}
 \end{aligned} \tag{1}$$

Table 2 Overall result on DRIVE database

Database	Sensitivity	Specificity	ROC	Accuracy (%)
DRIVE	0.996	0	0.991	98.7

Table 3 Description of evaluation parameters

Experimental result	Present	Absent
Positive	True positive (TP)	False positive (FP)
Negative	True negative (TN)	False negative (FN)

$$\text{Performance} = \frac{\text{(truly detected points(spurious points + missed out points))}}{\text{ground truth points}} \times 100 \tag{2}$$

5 Conclusion

Here, the fundus images are taken from the available database [7], and the experimental results are done on all the images of DRIVE database. The angle calculation between bifurcation points gives us approximately 98.75% result. Acute angles are preferred as bifurcation angles in most of the cases. To prove any vascular angle as a bifurcation angle is itself a challenging task. In this paper, an initiative has been taken to distinctly identify the acute and non-acute (distorted) bifurcation angles. Experimental results are shown in Table 1, where a snapshot of measured bifurcation angles from a retinal vascular structure taken from DRIVE database is given. Table 2 shows the overall performance of the method on DRIVE database (Table 3).

References

1. N. Dutta Roy, A. Biswas, Detection of bifurcation angles in a retinal fundus image, in *Eighth International Conference on Advances in Pattern Recognition (ICAPR)* (Indian Statistical Institution, Kolkata, IEEE Xplore Digital Library, 2015)
2. B. Al-Diri, A. Hunter, Automated measurements of retinal bifurcations, in *World Congress on Medical Physics and Biomedical Engineering, IFMBE Proceedings* (Springer, Berlin, 2009)
3. S. Morales, I. Legaz Aparicio, V. Naranjo, R. Verd-Monedero, Determination of bifurcation angles of the retinal vascular tree through multiple orientation estimation based on regularized morphological openings (2015)
4. B. Al-Diri, A. Hunter, D. Steel, M. Habib, Manual measurement of retinal bifurcation features, in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEEE Engineering in Medicine and Biology Society, Conference*, Aug, 2010
5. M. Zamir, J.A. Medeiros, T.K. Cunningham, Arterial bifurcations in the human retina (1979)

6. N. Dutta Roy, S. Goswami, S. Goswami, S. De, A. Biswas , Extraction of distinct bifurcation points from retinal fundus images, in *Proceedings of the First International Conference on Intelligent Computing and Communication*, Springer AISC Series, vol. 458 (Kalyani University, India, 2016)
7. The DRIVE database, Image sciences institute, university medical center utrecht, The Netherlands. <http://www.isi.uu.nl/Research/Databases/DRIVE/>. Last accessed 7 July 2007

Author Index

A

Agarwal, Arushi, 1
Aggarwal, Alakh, 327
Agnihotri, Ayushi, 137
Alia, Firdosh, 191

B

Bakthula, Rajitha, 741
Bal, Abhishek, 359
Banerjee, Minakshi, 295, 359
Banerjee, Tanmoy, 519
Banik, Sayantan, 217
Basak, Shyanka, 55
Basu, Subhadip, 775
Bhadury, Priyanka, 61
Bhalla, Monika, 231
Bhattacharyya, Arpan, 775
Bhattacharyya, Avijit, 765
Bhattacharyya, Debika, 557
Bhattacharyya, Rebeka, 203, 765
Bhattacharyya, Sayari, 181
Bhimjyani, Priya, 265
Bhunia, Soumi, 203
Biswas, Arindam, 485
Burman, Sayan, 721

C

Chakraborty, Arpan, 659
Chakraborty, Biswarup, 239
Chakraborty, Mahima, 731
Chakraborty, Prerana, 113
Chakraborty, Sanjay, 659
Chakraborty, Srija, 607
Chakraborty, Sumanta, 645

Chakraborty, Tamal, 547
Chanda, Koustav, 713
Chatterjee, Anindita, 783
Chatterjee, Biswajoy, 783
Chatterjee, Sankhadeep, 21
Chattopadhyay, Arup Kumar, 713
Chattopadhyay, Ayon, 191
Chattopadhyay, Pratik, 85, 327
Chattopadhyay, Soham, 249
Chaturvedi, Vaibhav, 61
Chaudhary, Pratibha, 1
Choudhury, Amitava, 447
Chowdhary, Sunil Kumar, 1

D

Dangi, Kriti, 741
Das Choudhury, Sruti, 567
Das, Abhirup, 191
Das, Amit Kumar, 29, 69, 265, 395, 567
Das, Dipankar, 137, 341, 419, 433
Das, Kaustav, 595
Das, Khakon, 295
Das, S., 309
Das, Sampriiti, 519
Das, Sanghamitra, 409
Das, Soma, 217
Dash, T.P., 309
Dash, Tara Prasanna, 409
Daw, Nilanjan, 485
De, Sayantan, 519
Deb, Koushik, 595
Debnath, Saswati, 679
Debnath, Shubham, 217
Dey, Monalisa, 129

Dey, Nibhash, 265

Dey, S., 309

Dey, Souvik, 147

Dey, Yajushi, 55

Dhabal, Supriya, 607, 619

Dutta, Debarati, 171

Dutta, Lal Mohan, 699

Dutta, Megha, 659

Dutta, Nixon, 319

Dutta, Sayan, 721

G

Ganguli, Runa, 583

Ganguly, Debankan, 265

Ghosal, Attri, 69

Ghosh, Amlan, 689

Ghosh, Ayan, 171

Ghosh, Debalina, 713

Ghosh, Debraj, 39

Ghosh, Kushal, 783

Ghosh, Sheersendu, 61

Ghosh, Soham, 21

Ghosh, Sourav, 567

Ghosh, Souvik, 129

Ghosh, Subhajit, 159

Gomes, Rohan Mark, 129

Goswami, Partha Sarathi, 547

Goswami, Saptarsi, 29, 69, 395, 567, 583, 689

Goswami, Suchismita, 9, 787

Goswami, Sushmita, 9, 787

Guha, Amartya, 667

Guha, Anubhav, 21

Gupta, Rajarshi, 533

Gupta, Sanjay Kumar, 85

H

Hajra, Mahimarnab, 99

J

Jain, Gitesh, 29

Jana, Uddipto, 249

Jena, J., 309

Jha, Aditi, 459

Jotheeswaran, Jeevanandam, 383

K

Kaushik, Keshav, 447

Kedia, Saket, 667

Keshari, Piyush, 629

Khare, Ankit, 447

Koley, Santanu, 629

Kumar, Ashiwani, 45

Kumar, Gulshan, 29

Kumar, Pardeep, 753

Kumari, Juhi, 191

Kundu, Subhradeep, 279

Kushwaha, Manish Singh, 741

L

Lahiri, Rajarshi, 147

M

Mahapatra, Priya Ranjan Sinha, 203

Mahata, Sainik Kumar, 137

Maiti, C.K., 309

Maitra, Mausumi, 295, 359

Maitra, Promita, 419, 433

Maji, Giridhar, 699

Majumdar, Rana, 1

Majumder, Sayan, 557

Majumdar, Shatadru, 731

Makhija, Sushnat, 137

Mal, Sandip, 45

Moitra, Tiyasa, 171

Mollah, Ayatullah Faruk, 775

Mondal, Anupam, 55, 61, 319

Mondal, Pritam, 689

Mondal, Shayam, 659

Mukherjee, Shreejita, 9

Mustafi, Joy, 395

N

Nag, Amitava, 713

Nag, Soumyadip, 147

Naman, Sumiran, 181

Nandy, Arunima, 69

Narnolia, Vishal, 249

Nasipuri, Mita, 775

Nayak, Laboni, 203

Neogi, Pinaki Prasad Guha, 395

P

Paira, Ramkrishna, 373

Pal, Mrinmoyi, 159

Panday, Mrityunjoy, 29, 69

Pandey, Shreelekha, 287

Paul, Abriti, 567

Paul, Ankur, 265

Paul, Debapriya, 485

Paul, Dip, 279

Paul, Pritam, 319

Paul, Soumi, 775

Paul, Souptik, 595

Paul, Sutanu, 341

Prakash, Sumit, 541

Pramanik, Arunangshu, 61

R

Rai, Ayush Kumar, 541
Ramchandani, Rohit, 447
Roy, Ayushi, 21
Roy, Nilanjana Dutta, 9, 485, 721
Roy, Pinki, 679
Roy, Ranajit, 265
Roy, Rashmita, 731
Roy, Sekhar Kumar, 583
Roy, Shramana, 249
Roy, Shubhasri, 9
Roy, Somshubhra, 533
Roy, Soumit, 147

S

Sabharwal, Simran, 471
Sachdeva, Himani, 287
Sadhukhan, Subham, 113
Saha, Dip Kumar, 619
Saha, Sourav, 203, 217
Saha, Suman, 265
Saha, Tufan, 181
Sarkar, Rajnika, 159
Sarkar, Susmit, 519
Seal, Sujoy, 667
Sen, Madhurima, 731
Sen, Pratap Chandra, 99
Sen, Soumya, 567, 699

Sengupta, Sushanta, 501
Shamoon, Mohammad, 447
Sharma, Ajay, 231
Sharma, Punit, 295, 359
Sharma, Shilpi, 459, 471
Shaw, Bipul, 239
Shaw, Sweta, 533
Shekhar, Himanshu, 667
Shrivastava, Shailesh, 327
Sikdar, Prosenjit, 607
Sil, Sayan, 129
Singh, Harvinder, 753
Sinha, Abhirup, 689
Sinha, Sharob, 55
Somya, Abbas, 541
Srivastava, Abhishek, 1
Sultaniya, Gaurav Mahesh, 85
Suman, Ankur Kumar, 239
Swarnakar, Spandan, 519

T

Tyagi, Sanjay, 753

U

Upadhyay, Ashutosh, 383
Upadhyay, Nityasree, 113