



*This picture is in honor of Naroa Echevarría García,  
author and winner of the Technical and Engineering  
School's competition in Digital Pictures into the  
Digital Design knowledge area.*

## **IMAI RESEARCH GROUP COUNCIL**

Director - Dr. Rubén González Crespo, Universidad Internacional de La Rioja (UNIR), Spain

Office of Publications - Lic. Ainhoa Puente, Universidad Internacional de La Rioja (UNIR), Spain

Latin-America Regional Manager - Dr. Carlos Enrique Montenegro Marín, Francisco José de Caldas District University, Colombia

## **EDITORIAL TEAM**

### **Editor-in-Chief**

Dr. Rubén González Crespo, Universidad Internacional de La Rioja (UNIR), Spain

### **Managing Editor**

Dr. Elena Verdú, Universidad Internacional de La Rioja (UNIR), Spain

### **Associate Editors**

Dr. Óscar Sanjuán Martínez, CenturyLink, USA

Dr. Jordán Pascual Espada, ElasticBox, USA

Dr. Juan Pavón Mestras, Complutense University of Madrid, Spain

Dr. Alvaro Rocha, University of Coimbra, Portugal

Dr. Jörg Thomaschewski, Hochschule Emden/Leer, Emden, Germany

Dr. Vicente García Díaz, Oviedo University, Spain

Dr. Carlos Enrique Montenegro Marín, Francisco José de Caldas District University, Colombia

Dr. Manju Khari, Ambedkar Institute of Advanced Communication Technologies and Research, India

Dr. Francisco Mochón Morcillo, National Distance Education University, Spain

### **Editorial Board Members**

Dr. Rory McGreal, Athabasca University, Canada

Dr. Anis Yazidi, Oslo Metropolitan University, Norway

Dr. Nilanjan Dey, Techo India College of Technology, India

Dr. Abelardo Pardo, University of Sidney, Australia

Dr. Hernán Sagastegui Chigne, UPAO, Peru

Dr. Lei Shu, Osaka University, Japan

Dr. Ali Selamat, Malaysia Japan International Institute of Technology, Malaysia

Dr. León Welicki, Microsoft, USA

Dr. Enrique Herrera, University of Granada, Spain

Dr. Hamido Fujita, Iwate Prefectural University, Japan

Dr. Francisco Chiclana, De Montfort University, United Kingdom

Dr. Luis Joyanes Aguilar, Pontifical University of Salamanca, Spain

Dr. Ioannis Konstantinos Argyros, Cameron University, USA

Dr. Ligang Zhou, Macau University of Science and Technology, Macau, China

Dr. Juan Manuel Cueva Lovelle, University of Oviedo, Spain

Dr. Pekka Siirtola, University of Oulu, Finland

Dr. Peter A. Henning, Karlsruhe University of Applied Sciences, Germany

Dr. Vijay Bhaskar Semwal, Indian Institute of Information Technology, Dharwad, India

Dr. Sascha Ossowski, Universidad Rey Juan Carlos, Spain

Dr. Miroslav Hudec, University of Economics of Bratislava, Slovakia

Dr. Walter Colombo, Hochschule Emden/Leer, Emden, Germany

Dr. Javier Bajo Pérez, Polytechnic University of Madrid, Spain

Dr. Jinlei Jiang, Dept. of Computer Science & Technology, Tsinghua University, China

Dr. B. Cristina Pelayo G. Bustelo, University of Oviedo, Spain

Dr. Masao Mori, Tokyo Institute of Technology, Japan

Dr. Daniel Burgos, Universidad Internacional de La Rioja (UNIR), Spain

Dr. JianQiang Li, NEC Labs, China

Dr. David Quintana, Carlos III University, Spain

Dr. Ke Ning, CIMRU, NUIG, Ireland

Dr. Alberto Magreñán, Real Spanish Mathematical Society, Spain  
Dr. Monique Janneck, Lübeck University of Applied Sciences, Germany  
Dr. Carina González, La Laguna University, Spain  
Dr. Mohammad S Khan, East Tennessee State University, USA  
Dr. David L. La Red Martínez, National University of North East, Argentina  
Dr. Juan Francisco de Paz Santana, University of Salamanca, Spain  
Dr. Héctor Fernández, INRIA, Rennes, France  
Dr. Yago Saez, Carlos III University of Madrid, Spain  
Dr. Guillermo E. Calderón Ruiz, Universidad Católica de Santa María, Peru  
Dr. Giuseppe Fenza, University of Salerno, Italy  
Dr. José Miguel Castillo, SOFTCAST Consulting, Spain  
Dr. Moamin A Mahmoud, Universiti Tenaga Nasional, Malaysia  
Dr. Madalena Riberio, Polytechnic Institute of Castelo Branco, Portugal  
Dr. Juan Antonio Morente, University of Granada, Spain  
Dr. Holman Diego Bolivar Barón, Catholic University of Colombia, Colombia  
Dr. Manik Sharma, DAV University Jalandhar, India  
Dr. Sara Rodríguez González, University of Salamanca, Spain  
Dr. Elpiniki I. Papageorgiou, Technological Educational Institute of Central Greece, Greece  
Dr. Edward Rolando Nuñez Valdez, Open Software Foundation, Spain  
Dr. Juha Röning, University of Oulu, Finland  
Dr. Luis de la Fuente Valentín, Universidad Internacional de La Rioja - UNIR, Spain  
Dr. Paulo Novais, University of Minho, Portugal  
Dr. Giovanni Tarazona, Francisco José de Caldas District University, Colombia  
Dr. Sergio Ríos Aguilar, Corporate University of Orange, Spain  
Dr. Fernando López, Universidad Internacional de La Rioja (UNIR), Spain  
Dr. Mohamed Bahaj, Settati, Faculty of Sciences & Technologies, Morocco  
Dr. Javier Martínez Torres, Centro Universitario de la Defensa de Marín, Escuela Naval Militar, Spain  
Dr. Abel Gomes, University of Beira Interior, Portugal  
Dr. Edgar Henry Caballero Rúa, Inforfactory SRL, Bolivia  
Dr. Víctor Padilla, Universidad Internacional de La Rioja (UNIR), Spain  
Ing. María Monserrate Intriago Pazmiño, Escuela Politécnica Nacional, Ecuador

## Editor's Note

**T**HE International Journal of Interactive Multimedia and Artificial Intelligence - IJIMAI (ISSN 1989 - 1660) provides an interdisciplinary forum in which scientists and professionals can share their research results and report new advances on Artificial Intelligence (AI) tools or tools that use AI with interactive multimedia techniques. This regular issue presents research works based on different AI methods such as deep networks, genetic algorithms or classification trees algorithms. These methods are applied into many and various fields as video surveillance, forgery detection, facial recognition, activity recognition, hand written character recognition, clinical decision, marketing, renewable energy or social networking.

The issue starts with a review article, written by A. Ghazvini, S. N. H. S. Abdullah and M. Ayob [1], which gives a view on current individual counting approaches based on clustering, detection, regression or density based methods. They describe their advantages and limitations, concluding that the use of convolutional neural networks with a density map approach will contribute in scheming more precise counting techniques, focusing not only on counting but also on localization of individuals in crowded scenes.

The second article describes a work of F. López Hernández, L. de-la-Fuente Valentin and I. Sarria Martínez de Mendivil [2] about a quick and simple method to detect brush editing in images, which can be used in image-tampering detection tools. Their two main contributions are: the design of a new approach to detect brush editing and the algorithm of the filter that detects this editing; and the introduction of intentions as subjective metric, in contrast to other classical objective forgery metrics.

N. Bouchra, A. Aouatif, N. Mohammed and H. Nabil [3] investigate deep learning models for face classification tasks. They propose using deep belief networks and stacked auto-encoder besides back propagation neural network to capture various latent facial features. The proposed approach shows better performance on two facial databases compared to other published methods.

A deep learning model is proposed by S. Jha, A. Dey, R. Kumar and V. Kumar-Solanki [4] for visual question answering, a process in which a machine answers to a natural language question related to an image. Specifically the model involves faster Region based Convolutional Neural Network (R-CNN) for extracting image features with an extra fully connected layer whose weights are dynamically obtained by Long Short Time Memory (LSTM) cell according to the question. Questions can be open ended or multiple choice questions and authors show that the visual question answering problem can be solved by a single R-CNN model.

Next work, described by F. Z. Benhacine, B. Atmani and F. Z. Abdelouhab [5], has the objective of improving the visualization of large sets of association rules to ease doctors' activity when using clinical decision support systems. The authors propose to use the CASI (Cellular Automata for Symbolic Induction) cellular machine together with the colored 2D matrices to improve the visualization of association rules. Effective interactivity between the human expert and the visualization matrix is a focus of the work to facilitate clinical decision.

In the field of renewable energy, A. H. A. Elkasem, S. Kamel, A. Rashad and F. Jurado [6] aim to allow doubly fed induction generator wind farms (DFIG), which are connected to the power system, to effectively participate in feeding electrical loads. To achieve this they propose a multiobjective optimization algorithm that is applied to a model of doubly fed induction generator wind farm (DFIG) with the

aim of determining the optimal values of the gains of the DFIG control system. The oscillation in power system is one of the problems of the interconnection of wind farms to the grid, and this proposal achieves lower oscillation in electrical power, so that DFIG wind turbines are more reliable.

In the following article, L. E. George and H. A. Hadi [7] propose the use of Electroencephalogram (EEG) signals for user identification and verification processes, which have shown effectiveness against forgery and theft. However, to make EEG applicable, the acquisition process should be easy and a minimum number of mental tasks must be asked to be performed by the user. Specifically, the authors propose methods using only two EEG channels when the user is performing one mental task, in order to reduce system complexity while maintaining high system accuracy.

Next article by S. Taleb Zouggar and A. Adla [8] presents a work on ensemble methods. Although these methods improve the performance of classifiers, they deteriorate the readability of the models. Therefore, some researchers propose to synthesize the structure of a tree from a set of classifiers, but prediction gets worse in this case. S. Taleb Zouggar and A. Adla propose an evaluation function combining performance and diversity for selection in a homogeneous ensemble used in a process of hill climbing. The method was evaluated on several benchmarks and compared to pruning homogeneous ensembles in literature, achieving better results.

In the field of activity recognition, A. Jalal and S. Kamal [9] propose a system that classifies the nature of 3D body postures obtained by Kinect. They present novel features suitable for depth data, which are robust to noise, invariant to translation and scaling, and capable of monitoring fast human body parts movements. Besides, an advanced hidden Markov model is used to recognize different activities. The system outperforms other methods when applied to three depth-based behavior datasets, in both posture classification and behavior recognition.

M. M. G. Ribeiro and A. J. P. Gomes [10] face the challenge of improving visual perception of deutan and protan dichromats, by introducing the first algorithm mainly focused on the enhancement of object contours in images. Typically the problem is mitigated by remapping colors to other colors but this does not help individuals to learn the naturalness of colors from past experience. Their algorithm increases the image contrast while keeping the naturalness of image colors, so the image perception improves but perceptual learning about the world is not disturbed.

Going back to recognition tasks but, this time, of handwritten characters, M. Daldali and A. Souhar [11] propose an Arabic text line segmentation approach using seam carving. The technique offers satisfactory results at extracting handwritten text lines without the need for the binary representation of the document image, even with documents presenting low text-to-background contrast such as degraded historical manuscripts. Although this work focuses on Arabic language, the method is language independent.

Next paper presents another work in the field of medicine, specifically a tool to help radiologists in breast cancer detection tasks [12]. L. Belkhodja and D. Hamdadou describe an automatic computer aided detection system, which combines medical image processing, bioinspired pattern recognition areas and others methods in computer vision. The system succeeded in automatic detection of abnormal regions.

H. M. Keerthi Kumar, B. S. Harish and H. K. Darshan [13] present



a work on sentiment analysis. They intend to identify sentiments analyzing short texts, specifically movie reviews. They propose the use of hybrid features obtained by concatenating machine learning features with lexicon features. Experiments show that the results obtained are highly promising both in terms of space complexity and classification accuracy.

Next article by M. Raees and S. Ullah [14] presents an approach on Three Dimensional (3D) interaction inside a Virtual Environment (VE). They describe an interaction technique where manipulation is performed by the perceptive gestures of the two dominant fingers, thumb and index. Experimental results show that the proposed approach has reliable recognition and accuracy rates. Moreover, the system neither needs training of images nor use any feature extraction, hence providing fast preprocessing.

W. Waheeb and R. Ghazali [15] describe a new application of ridge polynomial based neural network models in multivariate time series forecasting. Their main objective is to investigate and compare the forecasting efficiency of neural network models in forecasting the well-known multivariate time series called Box-Jenkins gas furnace time series.

This regular issue starts and ends with review works, specifically the last article is about marketing intelligence and big data [16]. Jan Lies describes how marketing intelligence has started to impact marketing practice with an important scope of social engineering techniques, concluding that marketing intelligence has led to a paradigm shift in marketing, from digital marketing to social engineering. Digital marketing includes data driven marketing, search engine marketing, recommender marketing, etc., while social engineering relates to content marketing, influencer marketing, social media marketing or creative marketing. Marketing intelligence is, without doubt, a current relevant issue. In fact, this is an advance of a next special issue of this journal that will focus on use cases of artificial intelligence, digital marketing and neuroscience.

Dr. Elena Verdú

## REFERENCES

- [1] A. Ghazvini, S. N. H. S. Abdullah, M. Ayob, "A Recent Trend in Individual Counting Approach Using Deep Network," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 7-14, 2019, doi: 10.9781/ijimai.2019.04.003.
- [2] F. López Hernández, L. de-la-Fuente Valentín, I. Sarria Martínez de Mendivil, "Detecting Image Brush Editing Using the Discarded Coefficients and Intentions," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 15-21, 2019, doi: 10.9781/ijimai.2018.08.003.
- [3] N. Bouchra, A. Aouatif, N. Mohammed, H. Nabil, "Deep Belief Network and Auto-Encoder for Face Classification," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 22-29, 2019, doi: 10.9781/ijimai.2018.06.004.
- [4] S. Jha, A. Dey, R. Kumar, V. Kumar-Solanki, "A Novel Approach on Visual Question Answering by Parameter Prediction using Faster Region Based Convolutional Neural Network," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 30-37, 2019, doi: 10.9781/ijimai.2018.08.004.
- [5] F. Z. Benhacine, B. Atmani, F. Z. Abdelouhab, "Contribution to the Association Rules Visualization for Decision Support: A Combined Use Between Boolean Modeling and the Colored 2D Matrix," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 38-47, 2019, doi: 10.9781/ijimai.2018.09.002.
- [6] A. H. A. Elkasem, S. Kamel, A. Rashad, F. Jurado, "Optimal Performance of Doubly Fed Induction Generator Wind Farm Using Multi-Objective Genetic Algorithm," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 48-53, 2019, doi: 10.9781/ijimai.2019.03.007.
- [7] L. E. George, H. A. Hadi, "User Identification and Verification from a Pair of Simultaneous EEG Channels Using Transform Based Features," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 54-62, 2019, doi: 10.9781/ijimai.2018.12.008.
- [8] S. Taleb Zouggar, A. Adla, "A Diversity-Accuracy Measure for Homogenous Ensemble Selection," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 63-70, 2019, doi: 10.9781/ijimai.2018.06.005.
- [9] A. Jalal, S. Kamal, "Improved Behavior Monitoring and Classification Using Cues Parameters Extraction from Camera Array Images," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 71-78, 2019, doi: 10.9781/ijimai.2018.07.003.
- [10] M. M. G. Ribeiro, Abel J.P. Gomes, "Contour Enhancement Algorithm for Improving Visual Perception of Deutan and Protan Dichromats," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 79-88, 2019, doi: 10.9781/ijimai.2018.05.003.
- [11] M. Daldali, A. Souhar, "Handwritten Arabic Documents Segmentation into Text Lines using Seam Carving," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 89-96, 2019, doi: 10.9781/ijimai.2018.06.002.
- [12] L. Belkhdja, D. Hamdadou, "IMCAD: Computer Aided System for Breast Masses Detection based on Immune Recognition," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 97-108, 2019, doi: 10.9781/ijimai.2018.12.006.
- [13] H. M. Keerthi Kumar, B. S. Harish, H. K. Darshan, "Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 109-114, 2019, doi: 10.9781/ijimai.2018.12.005.
- [14] M. Raees, S. Ullah, "GIFT: Gesture-Based Interaction by Fingers Tracking, an Interaction Technique for Virtual Environment," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 115-125, 2019, doi: 10.9781/ijimai.2019.01.002.
- [15] W. Waheeb, R. Ghazali, "Forecasting the Behavior of Gas Furnace Multivariate Time Series Using Ridge Polynomial Based Neural Network Models," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 126-133, 2019, doi: 10.9781/ijimai.2019.04.004.
- [16] J. Lies, "Marketing Intelligence and Big Data: Digital Marketing Techniques on their way to becoming Social Engineering Techniques in Marketing," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 5, no. 5, pp. 134-144, 2019, doi: 10.9781/ijimai.2018.05.002.

## TABLE OF CONTENTS

EDITOR'S NOTE.....	4
A RECENT TREND IN INDIVIDUAL COUNTING APPROACH USING DEEP NETWORK.....	7
DETECTING IMAGE BRUSH EDITING USING THE DISCARDED COEFFICIENTS AND INTENTIONS.....	15
DEEP BELIEF NETWORK AND AUTO-ENCODER FOR FACE CLASSIFICATION .....	22
A NOVEL APPROACH ON VISUAL QUESTION ANSWERING BY PARAMETER PREDICTION USING FASTER REGION BASED CONVOLUTIONAL NEURAL NETWORK.....	30
CONTRIBUTION TO THE ASSOCIATION RULES VISUALIZATION FOR DECISION SUPPORT: A COMBINED USE BETWEEN BOOLEAN MODELING AND THE COLORED 2D MATRIX.....	38
OPTIMAL PERFORMANCE OF DOUBLY FED INDUCTION GENERATOR WIND FARM USING MULTI-OBJECTIVE GENETIC ALGORITHM.....	48
USER IDENTIFICATION AND VERIFICATION FROM A PAIR OF SIMULTANEOUS EEG CHANNELS USING TRANSFORM BASED FEATURES.....	54
A DIVERSITY-ACCURACY MEASURE FOR HOMOGENOUS ENSEMBLE SELECTION .....	63
IMPROVED BEHAVIOR MONITORING AND CLASSIFICATION USING CUES PARAMETERS EXTRACTION FROM CAMERA ARRAY IMAGES .....	71
CONTOUR ENHANCEMENT ALGORITHM FOR IMPROVING VISUAL PERCEPTION OF DEUTAN AND PROTAN DICHROMATS.....	79
HANDWRITTEN ARABIC DOCUMENTS SEGMENTATION INTO TEXT LINES USING SEAM CARVING .....	89
IMCAD: COMPUTER AIDED SYSTEM FOR BREAST MASSES DETECTION BASED ON IMMUNE RECOGNITION .....	97
SENTIMENT ANALYSIS ON IMDB MOVIE REVIEWS USING HYBRID FEATURE EXTRACTION METHOD .....	109
GIFT: GESTURE-BASED INTERACTION BY FINGERS TRACKING, AN INTERACTION TECHNIQUE FOR VIRTUAL ENVIRONMENT.....	115
FORECASTING THE BEHAVIOR OF GAS FURNACE MULTIVARIATE TIME SERIES USING RIDGE POLYNOMIAL BASED NEURAL NETWORK MODELS.....	126
MARKETING INTELLIGENCE AND BIG DATA: DIGITAL MARKETING TECHNIQUES ON THEIR WAY TO BECOMING SOCIAL ENGINEERING TECHNIQUES IN MARKETING.....	134

## OPEN ACCESS JOURNAL

ISSN: 1989-1660

## COPYRIGHT NOTICE

Copyright © 2019 UNIR. This work is licensed under a Creative Commons Attribution 3.0 unported License. Permissions to make digital or hard copies of part or all of this work, share, link, distribute, remix, tweak, and build upon ImaI research works, as long as users or entities credit ImaI authors for the original creation. Request permission for any other issue from support@ijimai.org. All code published by ImaI Journal, ImaI-OpenLab and ImaI-Moodle platform is licensed according to the General Public License (GPL).

<http://creativecommons.org/licenses/by/3.0/>

# A Recent Trend in Individual Counting Approach Using Deep Network

Anahita Ghazvini\*, Siti Norul Huda Sheikh Abdullah, Masri Ayob

Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM) 43600, Bangi Selangor (Malaysia)

Received 18 February 2019 | Accepted 12 April 2019 | Published 17 April 2019



## ABSTRACT

In video surveillance scheme, counting individuals is regarded as a crucial task. Of all the individual counting techniques in existence, the regression technique can offer enhanced performance under overcrowded area. However, this technique is unable to specify the details of counting individual such that it fails in locating the individual. On contrary, the density map approach is very effective to overcome the counting problems in various situations such as heavy overlapping and low resolution. Nevertheless, this approach may break down in cases when only the heads of individuals appear in video scenes, and it is also restricted to the feature's types. The popular technique to obtain the pertinent information automatically is Convolutional Neural Network (CNN). However, the CNN based counting scheme is unable to sufficiently tackle three difficulties, namely, distributions of non-uniform density, changes of scale and variation of drastic scale. In this study, we cater a review on current counting techniques which are in correlation with deep net in different applications of crowded scene. The goal of this work is to specify the effectiveness of CNN applied on popular individuals counting approaches for attaining higher precision results.

## KEYWORDS

Analysis of Individuals, Automatic Video Surveillance, Counting Individuals, CNN, Deep Learning.

DOI: 10.9781/ijimai.2019.04.003

## I. INTRODUCTION

**I**NDIVIDUAL counting scheme has an extensive application in various domains such as public stations, shopping mall, universities, etc. [1]. The analysis of crowd has been presented surprising assessment to get knowledge about the crowded scene, though comprising the behavior analysis of crowd [2][3], tracking individuals [4][5], and segmenting the crowd [6][7].

The region of interest (ROI) and line of interest (LOI) are considered as two broad categories which are employed for individuals counting in the video. The ROI approximates the total value of individuals in some areas at a specific time whereas LOI obtains the total value of each individual in a certain period of time when they cross a detecting line [8]. To deal with these two categories, researchers employed either human detection, feature regression, or clustering techniques respectively [8]. Numerous schemes are proposed to count the individual on the basis of ROI and LOI categories. The general techniques are clustering [9], detection [10], and regression [11][12].

The crowd analysis has contributed profusely towards understanding the scene with overpopulation via linking the behavior analysis [2][3], tracking [4][5], and segmenting the crowd [13][6]. Based on the literature [14][15], the different crowd schemes had similar principles which enable them to precisely describe by the attributes. Currently [16], multiple studies struggle on profiling attributes of the crowd, due to the limitation to the value of the attribute and the indifference to the

dissimilarity in the scene.

The research of [17], suggested a deep multitask method to establish a good comprehension towards crowd by attaining information, mix procedure and features movement. The consequence of this letter demonstrates significant advances on the evaluation of recognizing the attribute in cross-scene by the suggested deep scheme.

The letter of [18] compared the most well-known counting approaches such as detection, clustering, and the latest technique, regression, which is gaining traction in handling the overpopulated area. The results of [18] showed that the regression scheme presented a higher performance in contrast to other approaches. Among the well-known counting schemes, the performance of detection and clustering based approaches are significantly marred on overpopulated scenes. In contrast to these approaches of the regression techniques, [19] is applicable to the overpopulated area and overlapping among the objects in crowded scenes. On the other hand, of all the regression techniques, the partial least square regression (PLSR) can resolve the collinearity difficulty with fewer factors, also requires fewer computations and converges speedily compared to other methods [20]. The major restriction of PLSR is at higher risk of overlooking 'real' correlations and sensitivity to the relative scaling of the descriptor variables [21]. Still, these approaches fall short of localization task due to their ineffectiveness to specify the location data of each individual in the scene.

The density maps technique exhibits great success in tackling the issue of individual counting in comparison to regression-based methods, specifically where the scene consists of a few complications such as heavy inter-occlusion between individuals, unclear resolution in videos [22]-[25][7]. Additionally, these methods can specify spatial

\* Corresponding author.

E-mail address: p86698@siswa.ukm.edu.my

data about the crowd.

Density estimation-based approach is normally utilized in public places like malls, the squares, and stations. This approach is highly effective in establishing a better understanding of individuals' behavior, which can elevate the individuals' safety in a public area. As the crowds may vary in distributions and shape patterns, the problem of recognizing pattern arises [26]. The crowd density is attested to be at greater advantage in comparison to crowd counting methods as it provides location data of the crowd. Counting individuals can be predicted with greater ease by utilizing a density map for some specific region. The total number of individuals in the video frame/image was achieved simply by determining the integral of density function over that frame/image [27].

Deep learning has shown a great performance in various area like speech and pattern recognition while among these techniques of deep learning, the CNN fares the best in the task of the image classification. The study of [28] proposed a novel approach based on deep techniques to classify a face. This work boosts the network performance through three various techniques of Deep belief Net (DBN), Stacked Auto-Encoder (SAE) besides of Back Propagation Neural Networks (BPNN) while utilizing sigmoid objective function. The high-level features are extracted via a deep algorithm which were utilized as inputs of the classifier to distinguish among the faces and non-faces. The outcome of this letter [28] attested to the robustness and proficiency of proposed technique for facial classification task on two datasets of BOSS and MIT.

The study of [29] suggested the estimation of density methods to assess the density maps on dissimilar crowd analysis tasks. Since the proficiency of density methods is firmly based on the features forms [22], this paper estimated a classical CNN to produce the original resolution of density map in contrast to current CNN approaches, which caused the reduction of density maps resolution through down-sampling strides in convolution processes. The obtained result from this work proved that the performance of counting remained unaffected on the unclear light of density maps, it, however, did not work for the localization tasks, like detection and tracking.

By producing the information of high-density, the utmost of research concentrated on an approximation of a density map to obtain the total number of the individuals [30]. These approaches were totally depending on the forms of features while CNN was utilized to obtain the valuable info from input automatically and also very efficiently to deal with overlapping among objects, non-uniform lighting, and altering scales [30]. However, the current CNN approaches may err in resolving the two issues of non-stable density distributions and dissimilarity in scale. The letter of [30] suggested the multi-column multi-task convolutional neural network (MMCNN) to solve these problems. This paper suggested three innovative methods, namely to offer a new density map which could focus on location and full data, propose a multi-column CNN to get beneficial data from different scales and lastly, to estimate the density map, level of crowded environment, as well as background or foreground mask. The accuracy of the density map improved by additional tied objectives. The suggested scheme of this work displayed to be more proficient in contrast to conventional Gaussian density map and the recently utilized shaped of individual density map. The method of this work proved that current CNN-based individual counting methods were capable in dealing with the issue of non-uniform distributions and scale variations. The study of [13], proposed a framework to disentangle the individual counting difficulties in cross-scene by using CNN with no additional annotations for a new target scene. The CNN model of this letter was trained based on two objectives of the density map and individual counts where these objectives assisted each other to gain higher local optima. Also, this model could learn certain crowd features more efficiently compared to the handcraft features. To tackle the gap

among various scenes, the pre-trained CNN was adjusted to all scenes. The texture of the crowd would be taken through the CNN technique which yielded an accurate counting outcome without considering the foreground segmentation results, as the method of this study was solely based on appearance information. The experiment result of this study showed the efficiency and consistency of the proposed method. As a consequence, in comparison to different counting approaches, density map methods were more proper for an overpopulated area, overlap scenes. Moreover, it seemed very potential in specifying spatial data to locate each individual in the overcrowded scene. This technique is accurate in defining the crowds through density distribution, as it provides the data on the location.

Nevertheless, it may collapse where only heads of individual are obvious. As the performance of this method is strictly dependent on the features types, the CNN can provide useful information automatically in contrast to shallow techniques. Moreover, the CNN techniques are highly beneficial to deal with difficulties that involve overlapping among objects, non-uniform lighting, and altering scales. Although the MMCNN is able to resolve these difficulties, they are restricted to the scales which are applied in overtraining and their ability to understand well-generalized schemes. In this letter, we aim to provide a review on recent individuals counting techniques which use deep networks in various applications of the crowded scene, and also to specify the effectiveness of deep CNN for attaining higher precision while utilizing popular individuals counting approaches. This paper is segregated into five parts, specifically, introduction, analysis of individuals counting, discussion, outcomes and conclusion which are specified in parts I, II, III, IV, and V correspondingly.

## II. INDIVIDUAL COUNTING ANALYSIS

Individuals counting approaches are classified into three techniques, namely, "Detection techniques", "Clustering techniques", "Regression techniques", and "Density Map techniques".

### A. Detection Technique

The detection-based technique assigns the detector to recognize the individuals in the video scene for obtaining the total value of individuals with their location [31].

The well-known detection methods commonly include detection based on the body [32], shoulders [33] and heads [34]. In heavy occlusion condition, only the heads of individuals might be appearing so, head detection methods mostly show well performance in contrast to the other schemes of shoulder and body detection. The problem in detection technique is to identify an individual by itself, mostly in the existence of crowds and overlapping among objects [35]. Although this kind of method shows promising outcomes, the carefully hand-engineered designed indicator is not precise in the blurred scene where the quality of utmost videos is relatively low [1][35][36]. Furthermore, this scheme is unable to deal with the overpopulated region or unstable climate as well as to count the total value of individuals with the different flow direction, and it limits its operation in real scenes [15]. These schemes are efficient to produce appropriate detections in sparse scenes, providing the total value of individuals and location, in addition, to posture of all individuals in a scene. However, these approaches are travail from sight clutter. A multi-camera adjustment with overlying views is not available in many circumstances, and also scene not being able to promise satisfactory cross-dataset generalization [10], while training of a specific scene indicator for counting is difficult [18]. The tracking and detecting of individuals are getting more complex while the crowded environment is getting denser and larger [37]. Nevertheless, these approaches require huge computing resource and are often limited by an overlapping and complex background in



realistic situations, subsequent a low accuracy [38]. This owes to the fact that these methods would have to scan each frame via a trained detector, which normally takes more time to obtain the total value of individuals [39]. The individuals counting through detection-based approaches need more time whilst applied to the scene with occlusion and lighting changes [40]. These approaches are able to provide higher accuracy in the crowd scene with low density in comparison to the scene with high crowd density [41]. Additionally, it is essential to have scene with high resolution to get high precisions [42]. The pros and cons of prevalent detection techniques are shown in Table I.

### B. Clustering Technique

The total number of individuals in the clustering-based techniques are obtained through tracking techniques algorithms [54]. In these approaches the useful information/features are followed out frame by frame, then by using spatial and temporal constancy heuristics, they are able to cluster the direction and also use extra elements to accomplish the spatial path for each person [9][54][55]. The total number of individuals is indicated based on the number of clusters [56]. As these methods highly depend on tracking approaches, they are time-consuming and need high computation resource. The performance of these techniques degrades excessively while they are applied under certain conditions of variations in illumination, low resolution, and motion imaging platforms, which reduces the stability of the tracking algorithm [30][54][57]. These techniques are based on extracting and counting the individuals' blobs of a temporal slice of the video, the certain blobs which consist of many individuals are not able to provide precise individual counting due to severe occlusion [58][59]. Furthermore, the accuracy of these techniques is influenced by the inaccuracy of coherently unbalanced info which is not matched with the exact object [60]. Mainly in a real scene, the handcraft information is hardly capable to adapt with different climate conditions, and lighting [58][59]. However, a model propounds movement coherency, thus the probability of improper approximation increases when objects staying constant in a scene, signifying objects which allocating similar info at a certain period [61]. These methods are capable to handle the

consecutive image frames while other counting approaches are not facing this restriction [18]. Also, these approaches are able to provide an accurate result when dependable trajectories can be extracted [39].

### C. Regression Technique

The regression-based techniques, by gaining the low-level features map, are able to count the individuals. Typically, the regression algorithms encompass the appropriate features such as texture features [11][62][63] and key points [11][62]-[64] which are extracted from the foreground through background subtraction approaches. These techniques obtain the correlation amongst features and count individuals through the training of extracted features without considering the individuals' identification [54]. These approaches provide superior performance under over-crowded scene circumstances [65] by escaping the issue of hard detection. However, these approaches lack of the ability to provide the information about the individual count, and also, disability to specify the position of each individual, which restricts these approaches for localization tasks [7][66]. It is crucial for video surveillance systems to have information about the position of each person in the scene in order to acquire the spatial distribution of individuals [11][24][67]. The computational cost of these approaches is very low as these methods do not need to detect and trace the individual [37]. Table II summarizes the advantages and disadvantages of six popular regression-based techniques.

### D. Density Based Technique

In the Density-based techniques, the total value of the individuals is equivalent to the integral of the density map around sub-region. These approaches are employed for both counting and localization by retaining the spatial information which renders these methods very effective for describing the density distribution of crowds. Notwithstanding, these approaches might collapse where only heads of individuals can be observed. These methods are able to tackle the individual's counting issue, under high occlusion, and low resolution in the video scene [22]-[25][7]. The performance of this technique is extremely based on the types of features [22]-[25][7], where the convolutional neural network

TABLE I. THE PROS AND CONS OF INDIVIDUALS' DETECTION-BASED TECHNIQUES

No	Detection Approach	Function	Pros	Cons	Paper
1	Monolithic	<ul style="list-style-type: none"> <li>Use the appearance of full body to train the classifier. The quality and speed of detection is on the bases of classifier's choice.</li> </ul>	<ul style="list-style-type: none"> <li>Provides reasonable detection in sparse scenes.</li> </ul>	<ul style="list-style-type: none"> <li>Not applicable under overlapping situation among the individuals in scenes.</li> <li>Not being able to deal with clutter areas.</li> </ul>	[43][44][45][46]
2	Part Based	-----	<ul style="list-style-type: none"> <li>More robust in comparison to monolithic method as whole body is observable.</li> </ul>	<ul style="list-style-type: none"> <li>Unable to provide precise detection solely based on head region.</li> </ul>	[47]
3	Shape Matching	-----	<ul style="list-style-type: none"> <li>It provides the info about pose as well as count and location of each individuals.</li> </ul>	-----	[48][49]
4	Multi-Sensor	<ul style="list-style-type: none"> <li>Accessible to multiple camera where the uncertainties caused by overlap among individuals can be dissolved via one camera.</li> </ul>	-----	<ul style="list-style-type: none"> <li>Not applicable if multiple camera is not available.</li> </ul>	[50][51]
5	Transfer Learning	<ul style="list-style-type: none"> <li>Detecting the individuals in new scenes without controlling by human.</li> </ul>	-----	<ul style="list-style-type: none"> <li>Not applicable if scene contains low resolution, varying in viewpoints and illumination.</li> </ul>	[45][52][53]

(CNN) is known as the best technique for extracting the features [30]. However, the CNN method cannot efficiently tackle three difficulties of distributions of non-uniform density, scale changes, and drastic scale changes [30] and also causes the resolution of the density maps to decrease which has no discernible effect on the development of the precise individual counting, even though it prevents from localizing the individual satisfactorily [36]. By considering the mentioned difficulties as a preventive factor from attaining higher precisions, the CNN schemes particularly handle these difficulties through multi-column or multi-resolution network [36]. Though these approaches exhibit their strength to non-uniform distribution density, scale changes, and drastic scale variation, they still pose some limitation to the size that is applied during training and thus their proficiency delimited for learning better-common approach [39]. In comparison to regression techniques, the density maps have been shown to be very efficient at solving the issue of individual counting, especially where the scene contains high inter-occlusion, with low-resolution surveillance videos [22]-[25][7] and also, these methods are providing spatial information about the crowd.

### III. EXPERIMENTAL RESULTS

This research aims to cater a review based on recent counting approaches and work is to specify the effectiveness of CNN used on popular individuals counting approaches for attaining higher precision results. In the field of computer vision, individual counting is considered as a fundamental task. There are several techniques such as counting by clustering, detection, and regression that exist to resolve the issue of counting individuals. However, these techniques fall short of overcoming the difficulties such as overlapping objects, the difference in illumination, unstable weather, overpopulated scene.

These techniques, albeit, have shown astonishing performance while incorporated with CNN to attain useful data from the image /video frames which play an important role in both ROI and LOI [8]. Through utilizing the advantage from CNN for representation of the image, the individuals counting has shown great achievements in contrast to shallow approaches to tackle the issue of the populated scene. Table III shows some certain individual counting techniques mentioning the quantitative outcomes from the corresponding reference.

### IV. DISCUSSION

There are several techniques for individual counting namely, detection, clustering, and regression. These methods were employed to overcome the difficulties of counting, which have dissimilar reasons under different conditions in an overpopulated area. The detection-based approaches specify the total number of individuals in the crowded scene by assigning a detector for recognizing each individual [73]. These approaches are only practical to a specific scene and will fail for some certain real-life applications [73][74]. These approaches have difficulty to obtain precise outcome under various conditions such as unstable weather or individuals with opposite flow path [15]. Besides, they are time-consuming as the detector scans every frame of the scene [39]. The clustering methods obtain the total number of persons in a populated scene via employing tracking systems [73]. These approaches also take a long time to process and suffer from computation difficulty as they are closely dependent on the tracking algorithms. Moreover, their performance degrades significantly when employed to scene with variation in illumination, unclear resolution, and motion imaging platforms, which causes the tracking algorithm to be unsteady. The regression-based methods obtain the total value of

TABLE II. THE ADVANTAGES AND DISADVANTAGES OF INDIVIDUALS REGRESSION-BASED METHODS

No	Detection Approach	Function	Pros	Cons	Paper
1	Linear Regression	-----	<ul style="list-style-type: none"> <li>• Yields high performance in sparse scene where the crowds are small and there is less occlusion among the objects.</li> </ul>	<ul style="list-style-type: none"> <li>• Some objects are not beneficial for counting estimation.</li> </ul>	[68]
2	Partial Least Square Error (PLSR)	-----	<ul style="list-style-type: none"> <li>• Produces good performance under various crowdedness levels and unseen density.</li> <li>• Being capable to overcome the collinearity issue with fewer factors.</li> <li>• Needs less computations.</li> <li>• Converges very fast.</li> </ul>	<ul style="list-style-type: none"> <li>• Sensitive to the proportion of positive and negative class in training data.</li> </ul>	[69]
3	Kernel Ridge Regression (KRR)	<ul style="list-style-type: none"> <li>• Its extended nonlinear form of ridge regression which obtained from kernel tricks.</li> </ul>	<ul style="list-style-type: none"> <li>• Diminish the issue of multiple colinear.</li> </ul>	-----	[47]
4	Support Vector Regression (SVR)	-----	<ul style="list-style-type: none"> <li>• Requires less time testing to approximate the solution.</li> </ul>	-----	[70]
5	Gaussian Process regression (GPR)	<ul style="list-style-type: none"> <li>• Most prevalent regression-based counting technique.</li> </ul>	-----	<ul style="list-style-type: none"> <li>• Not reliable to deal with big datasets.</li> </ul>	[69]
6	Random Forest regression (RFR)	<ul style="list-style-type: none"> <li>• -Attain to nonlinear scalable regression.</li> </ul>	<ul style="list-style-type: none"> <li>• Less susceptible to the parameters.</li> <li>• Being to be scaled to big dataset.</li> </ul>	<ul style="list-style-type: none"> <li>• Disable to work with the points that are out of the range of target value.</li> </ul>	[71]



TABLE III. SOME CERTAIN INDIVIDUAL COUNTING TECHNIQUES WITH QUANTITATIVE OUTCOMES

Year	Paper	Description	Method	Challenges	Datasets	Results
2016	[1]	Count the people on basis of head detection techniques by composition of Adaboost algorithm and the CNN.	Detection based methods.	Low resolution data, body occlusion and not limited imaging viewpoints.	Real classroom surveillance.	<b>Stage=15</b> <b>Recall=0.86</b> <b>Precision=0.33</b>
2015	[8]	Proposed a new technique to approximate the overall value of arriving and leaving crowd flow with three CNN methods.	Line of feature (LOI).	High flow density crowd, various illuminations and different mal-weather.	Large dataset that consists of various real videos of public gates.	<b>Precision= 95.06%</b>
2017	[59]	Introduces passenger counting system using CNN and Spatio-temporal Context.	Region of interest (ROI).	Complex low-resolution scene for public transportation.	Public bus transportation in China.	<b>Recall= 94.215</b> <b>Precision= 92.486</b>
2017	[60]	Evaluation of produced density maps thru density estimation techniques on different crowd analysis tasks, such as counting, detecting, and tracking.	Density map based.	Reduction of density map resolution while using CNN based method which can degrade the performance of localization tasks (detection and counting).	UCSD [62] UCF CC 50 [18] WorldExpo'10 [7] TRANCOS [40]	(CNN- Pixel) Error Distance (ED)= $3.61 \pm 0.72$ (CNN- Pixel) Error Difference Distance= $2.90 \pm 0.83$  (FCNN-Skip) Error Distance (ED)= $3.61 \pm 0.72$ (FCNN-Skip) Error Difference Distance= $3.38 \pm 1.01$  <b>UCSD MAE= 1.02</b> <b>UCSD MSE= 1.18</b>  <b>MALL MAE= 1.98</b> <b>MALL MSE= 5.68</b>
2018	[30]	Proposed multi-column multi-task convolutional neural network (MMCNN) for counting the crowd.	Density estimation.	Drastic size change and unsteady density distribution.	UCSD [63] UCF CC 50 [18] WorldExpo'10 [7] Shanghai Tech [36] MALL [73]	<b>UCF CC 50 MAE= 320.6</b> <b>UCF CC 50 MSE= 323.8</b>  <b>WorldExpo'10 MAE= 9.1</b> <b>WorldExpo'10 MSE= 18.7</b>  <b>Shanghai Tech Part A MAE= 91.2</b> <b>Shanghai Tech Part A MSE= 128.6</b>  <b>ShanghaiTech Part B MAE= 18.5</b> <b>ShanghaiTech Part B MSE= 29.3</b>

individuals in the scene through mapping amongst low-level features [73]. Such techniques generally achieve higher performance for the overcrowded scene. However, they lack in providing the information of individuals location, hence not being used for object localization [66] [7]. As the goal of this study is to introduce the most suitable counting methods, the density map estimation has shown encouraging results in comparison to the other techniques, which can provide the location of the individuals in a very crowded scene. These approaches are very potential in tackling the individuals counting issue where heavy inter occlusion and unclear scene exist among the objects. However, they are unable to provide the precise individuals counting where only heads of individuals are obvious. The effectiveness of individual counting methods is indicated in Table IV.

## V. CONCLUSION

This study endeavors to provide a revision on the existing individuals counting techniques namely, clustering, detection, regression, and density map-based methods. Among these techniques, the regression-

based ones show great performance under overcrowded area. The regression-based approaches are capable to count the individuals in a crowded environment, although they are not applicable for localization task. The density map estimation approaches are very promising among other counting methods since they preserve the beneficial spatial data for two tasks of counting and localization. The density map technique is very efficient to resolve the issue of individuals counting under different circumstances such as the massive overlapping among the objects and unclear scene in consequence frames. This scheme is proper for defining individual's density distribution since it focuses on both details on location and spatial data but it might fail once the heads of individuals appear. While the functioning of this method relies on the type of features, the convolutional neural networks (CNN) are capable to extract the useful information. The CNN based counting approaches are not being able to tackle three problems, i.e., distributions of non-uniform density, a variation of scale. Furthermore, applying CNN-based approaches to density estimation techniques can reduce the resolution of these techniques rendering it ineffective for localization tasks. By taking these problems into account as a restrictive

TABLE IV. THE EFFECTIVENESS OF INDIVIDUAL COUNTING METHODS

Methods	Task	Processing Period	Computational Complication	Performance					Scene Type	
				Illumination Alteration	Low Resolution	Occlusion	Overcrowded area	Small Object Size	Consecutive	Fixed
<b>Clustering</b>	Counting by tracking techniques	High	High	Low	Low	Low	Low	Low	Applicable	Inapplicable
<b>Corresponding Authors</b>	[54]	[54]	[54]	[58][59]	[54][57]	[58][59]	[73]	[29]	[18]	[18]
<b>Detection</b>	Count by scheming detector to identify individuals	High	-----	-----	Low	-----	Low	Low	Applicable	Applicable
<b>Corresponding Authors</b>	[31]	[31]	-----	-----	[19]	-----	[73]	[29]	[18]	[18]
<b>Regression</b>	Approximate density of crowd on the basis of holistic and collective description of pattern	-----	-----	-----	-----	High	High	High	Applicable	Applicable
<b>Corresponding Authors</b>	[55]	-----	-----	-----	-----	[39]	[65]	[29]	[18]	[18]
<b>Density Map</b>	Preserves spatial info which makes it useful for both counting and localization tasks	-----	-----	-----	High	High	High	High	Applicable	Applicable
<b>Corresponding Authors</b>	[30]	-----	-----	-----	[22]-[25], [7]	[22]-[25], [7]	[29]	[29]	[18]	[18]

clause to accomplish superior precisions, particular CNN techniques accurately resolve the three mentioned difficulties via multi-column or multi-resolution net. Though these approaches showed robustness to distributions of non-uniform congestion and size variations, they are yet restricted to the size of the dataset used for training which makes their ability to be limited to achieved better methods. Based on the experimental results of previous work we found that the crowd density approach is more beneficial in comparison to other counting methods as it provides information about the location of the crowd. However, the performance of this technique is highly depending on the types of features in which the usage of best deep learning technique (CNN) can be very valuable for this method to extract the important feature from each video frame. The usage of (CNN) with a density map approach will help the future research in scheming more precise counting techniques to approximate density maps with high quality for both tasks of counting and localization.

#### ACKNOWLEDGMENT

This research is based on two research grants with code of DIP-2014-039 and AP2017005/2.

#### REFERENCES

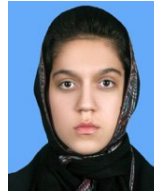
- [1] Ch. Gao, P. Li, Y. Zhang, J. Liu, and L. Wang, People counting based on head detection combining Adaboost and CNN in crowded surveillance environment, *Journal of Neurocomputing*, Vol. 208, no.C, pp.108–116, 2016.
- [2] B. Zhou, X. Tang, H. Zhang, and X. Wang, Measuring crowd collectiveness, In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol.36, pp. 1586–1599, 2014.
- [3] B. Zhou, X. Tang, and X. Wang, Coherentfiltering: detecting coherent motions from crowd clutters, In: *European Conference on computer Vision, LNCS*, vol. 7573, pp. 857-87, 2012.
- [4] M. Rodriguez, J. Sivic, I. Laptev, and J. Y. Audibert, Data driven crowd analysis in videos, In: *International Conference on Computer Vision, Barcelona, Spain*, 2011.
- [5] F. Zhu, X. Wang, and N. Yu, Crowd tracking with dynamic evolution of group structures, In: *European Conference on computer Vision, LNCS*, vol. 8694, pp. 139-154, 2014.
- [6] K. Kang, and X. Wang, Fully convolutional neural networks for crowd segmentation, *Computer Vision and Pattern Recognition*, 2014.
- [7] C. Zhang, H. Li, X. Wang, and X. Yang, Cross-scene crowd counting via deep convolutional neural networks, In: *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 07, pp. 833–841, 2015.
- [8] L. Cao, X. Zhang, W. Ren, and K. Huang, Large scale crowd analysis based on convolutional neural network, *Pattern Recognition*, vol. 48,

- no.10, pp. 3016–3024, 2015.
- [9] V. Rabaud, and S. Belongie, Counting crowded moving objects, In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 705–711, 2006.
  - [10] P. Dollár, C. Wojek, B. Schiele, and P. Perona, Pedestrian detection: An evaluation of the state of the art, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 34, pp. 743–761, 2012.
  - [11] A. B. Chan, and N. Vasconcelos, Counting people with low-level features and bayesian regression, In: IEEE Transactions on Image Processing, vol. 21, pp. 2160–2177, 2012.
  - [12] A. C. Davies, J. H. Yin, and S. A. Velastin, Crowd monitoring using image processing, Electronics & Communications Engineering Journal, vol. 7, no.1, pp. 37–47, 1995.
  - [13] C. Zhang, H. Li, X. Wang, and X. Yang, Cross-scene crowd counting via deep convolutional neural networks, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 833–841, 2015.
  - [14] C. Castellano, S. Fortunato, and A. Loreto, Statistical physics of social dynamics, Journal of Reviews of modern physics, vol. 81, no. 2, pp. 81–591, 2009.
  - [15] H. Chat'é, F. Ginelli, G. G'egoire, and F. Raynaud, Collective motion of self-propelled particles interacting without cohesion, Journal of Reviews of modern physics, vol. 77, no.4, 2008.
  - [16] J. Shao, C. C. Loy, and X. Wang, Scene-independent group profiling in crowd, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014.
  - [17] J. Shao, K. Kang, Ch.Ch. Loy, and X. Wang, Deeply Learned Attributes for Crowded Scene Understanding, In: IEEE Conference on computer vision and pattern recognition (CVPR), Boston, MA, USA, 2015.
  - [18] Ch, Ch. Loy, K. Chen, Sh. Gong, and T. Xiang, Crowd Counting and Profiling: Methodology and Evaluation, In: IEEE Transactions on Information Technology in Biomedicine, vol. 2, 2013.
  - [19] C. Wang, H. Zhang, L. Yang, S. Liu, and X. Cao, Deep People Counting in Extremely Dense Crowds, In: Proceedings of the 23rd ACM international conference on Multimedia, pp. 1299–1302, 2015.
  - [20] Ö. Yeniay, and A. Götas, A Comparison of Partial Least Squares Regression with Other, In: Journal of Mathematics and Statistics, vol. 31, pp. 99–111, 2002.
  - [21] R.D. Cramer III, Partial Least Squares (PLS): Its strengths and limitations, In: Perspectives in drug Discovery and Design, vol.1, pp. 269–278, 1993.
  - [22] C. Arteta, V. Lempitsky, J. A. Noble, and A. Zisserman, Interactive object counting, In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, vol. 8691, pp. 504–518, 2014.
  - [23] L., Nair, R., Koethe, U., F. A. Hamprecht, Learning to Count with Regression Forest and Structured Labels, In: International conference on pattern recognition (ICPR), pp. 2685–2688, 2012.
  - [24] V. Lempitsky, and A. Zisserman, Learning To Count Objects in Images, In: Advances in Neural Information Processing System, pp. 1324–1332, 2010.
  - [25] V. Q. Pham, T. Kozakaya, O. Yamaguchi, and R. Okada, COUNT forest: Co-voting uncertain number of targets using random forest for crowd density estimation, In: Proceedings of the IEEE International Conference on Computer Vision, vol. 2015, pp. 3253–3261, 2015.
  - [26] Sh. Pu, T. Song, Y. Zhang, and D. Xie, Estimation of crowd density in surveillance scenes based on deep convolutional neural network convolutional neural network, In: 8th International Conference on Advances in Information Technology, IAIT2016, pp. 154–159, Macau, China, 2016.
  - [27] F. Xiong, X. Shi, and D. Yeung, Spatiotemporal Modeling for Crowd Counting in Videos, In: International Computer Vision (ICCV), 2017.
  - [28] N. Bouchra, A. Aouatif, N. Mohammed, and H. Nabil, Deep Belief Network and Auto-Encoder for Face Classification, International Journal of Interactive Multimedia and Artificial Intelligence, <http://dx.doi.org/10.9781/ijimai.2018.06.004>, 2018.
  - [29] D. Kang, Z. Ma, and A. B. Chan, Beyond Counting: Comparisons of Density Maps for Crowd Analysis Tasks - Counting, Detection, and Tracking, In: IEEE Transactions on Circuits and Systems for Video Technology (TCSVT), pp. 1–14, 2017.
  - [30] B. Yang, J. Cao, N. Wang, Y. Zhang, and L. Zou, Counting challenging crowds robustly using a multi-column multi-task convolutional neural network, Journal of signal processing: image communication, vol. 64, pp. 118–129, 2018.
  - [31] O. Sidla, Y. Lypetsky, N. Brändle, and S. Seer, Pedestrian detection and tracking for counting applications in crowded situations, In: Proceedings IEEE International Conference on Video and Signal Based Surveillance 2006, AVSS'06, 2006.
  - [32] D. Conte, P. Foggia, G. Percannella, F. Tufano, and M. Vento, A method for counting people in crowded scenes, In: Proceedings - IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2010, pp. 225–232, 2010.
  - [33] W. Li, X. Wu, H. A. Zhao, New techniques of foreground detection, segmentation and density estimation for crowded objects motion analysis, Journal of Information and Media Technologies, vol. 6, no. 2, pp. 528–538, 2011.
  - [34] T. Van Oosterhout, S. Bakkes, and B. Krse, Head Detection in Stereo Data for People Counting and Segmentation, In: Proceedings of the International Conference on Computer Vision Theory and Applications, pp.620–625, 2011.
  - [35] H. Fradi, and J. L. Dugelay, Low level crowd analysis using frame-wise normalized feature for people counting, In: Proceedings of the 2012 IEEE International Workshop on Information Forensics and Security (WIFS 2012), pp.246–251, 2012.
  - [36] Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma, Single-Image Crowd Counting via Multi-Column Convolutional Neural Network, In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 589–597, 2016.
  - [37] H. Foroughi, N. Ray, and H. Zhang, Robust people counting using sparse representation and random projection, Journal of Pattern Recognition, vol. 48, no.10, pp.3038–3052, 2015.
  - [38] L. Zeng, X. Xu, B. Cai, S. Qiu, and T. Zhang, Multi-scale Convolutional Neural Networks for Crowd Counting, Retrieved from <http://arxiv.org/abs/1702.02359>, 2017.
  - [39] B. Xu, and G. Qiu, Crowd density estimation based on rich features and random projection forest, In: IEEE Winter Conference on Applications of Computer Vision(WACV 2016), 2016.
  - [40] K. Chen, C. C. Loy, S. Gong, and T. Xiang, Feature mining for localised crowd counting, In: Proceedings of the British Machine Vision Conference 2012, vol.1, pp. 1–11, 2012.
  - [41] Y. L. Hou, and G. K. H. Pang, People counting and human detection in a challenging situation, In: IEEE Transactions on Systems, Man, and Cybernetics Part A:Systems and Humans, vol. 41, pp. 24–33, 2011.
  - [42] Z. Q. H. Al-Zaydi, D. L. Ndzi, Y. Yang, and M. L. Kamarudin, An adaptive people counting system with dynamic features selection and occlusion handling, Journal of Visual Communication and Image Representation, vol. 39, pp. 218–225, 2016.
  - [43] O. Tuzel, F. Porikli, and P. Meer, Pedestrian detection via classification on Riemannian Mani folds, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 30, no.10, pp. 1713–1727, 2008.
  - [44] P. Sabzmejdani, and G. Mori, Detecting pedestrians by learning shapelet features, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8, 2007.
  - [45] P. Dollar, C. Wojek, B. Schiele, and P. Perona, Pedestrian detection: An evaluation of the state of the art, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 99, pp. 1–1, 2011.
  - [46] M. Pa tzold, R. Evangelio, and T. Sikora, Counting people in crowded environments by fusion of shape and motion information, In: IEEE International Conference on Advanced Video and Signal based Surveillance, pp. 157–164, 2010.
  - [47] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, no. 9, pp. 1627–1645, 2010.
  - [48] C. Lampert, Kernel methods in computer vision, vol. 4. Now Publishers Inc, 2009.
  - [49] W. Ge, R. Collins, Marked point processes for crowd counting, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 2913–2920, 2009.
  - [50] W. Ge, and R. Collins, Crowd detection with a multi view sampler, European Conference on Computer Vision, pp. 324–337, 2010.
  - [51] R. Benenson, M. Mathias, R. Timofte, and L.V. Gool, Pedestrian detection at 100 frames per second, In: IEEE Conference Computer Vision and Pattern Recognition, 2012.
  - [52] M. Wang, and W. Li, X. Wang, Transferring a generic pedestrian detector

towards specific scenes, In: IEEE Conference Computer Vision and Pattern Recognition, 2012.

- [53] M. Wang, and X. Wang, Automatic adaptation of a generic pedestrian detector to a specific traffic scene, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 3401–3408, 2011.
- [54] I. S. Topkaya, H. Erdogan, and F. Porikli, Counting people by clustering person detector outputs, In: 11th IEEE International Conference on Advanced Video and Signal-Based Surveillance, AVSS 2014, pp. 313–318, 2014.
- [55] A. M. Cheriyyadat, B. L. Bhaduri, and R. J. Radke, Detecting multiple moving objects in crowded environments with coherent motion regions, In: 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops, 2008.
- [56] Z. Ma, and A. B. Chan, Crossing the line: Crowd counting by integer programming with local features, In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2539–2546, 2013.
- [57] G. Antonini, and J. P. Thiran, Counting pedestrians in video sequences using trajectory clustering, In: IEEE Transactions on Circuits and Systems for Video Technology, vol.16, pp.1008–1020, 2006.
- [58] Y. Cong, H. Gong, S. C. Zhu, and Y. Tang, Flow mosaicking: Real-time pedestrian counting without Scene-specific learning, In: 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2009, pp. 1093–1100, 2009.
- [59] G. Liu, Z. Yin, Y. Jia, and Y. Xie, Passenger Flow Estimation Based on Convolutional Neural Network in Public Transportation System, Journal of Knowledge-Based Systems, vol. 123, pp. 102–115, 2017.
- [60] R. Shbib, S. Zhou, D. Ndzi, and K. Al-kadhimi, Distributed Monitoring System Based On Weighted Data Fusing Model, American Journal of Social Issues and Humanities, pp. 53–62, 2013.
- [61] A. E. Hoerl, and R. W. Kennard, Ridge Regression: Applications to Nonorthogonal Problems. Technometrics, vol. 12, no. 1, pp. 69–82, 1970.
- [62] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, Privacy preserving crowd monitoring: Counting people without people models or tracking, In: 26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR, 2008.
- [63] K. Chen, C. C. Loy, S. Gong, and T. Xiang, Feature Mining for Localised Crowd Counting, In: Proceedings of the British Machine Vision Conference, 2012.
- [64] D. Ryan, S. Denman, C. Fookes, and S. Sridharan, Crowd counting using multiple local features, In: DICTA 2009 - Digital Image Computing: Techniques and Applications, pp. 81–88, 2009.
- [65] A. Adegbeye, G. Hancke, and G. H. Jr, Single-pixel approach for fast people counting and direction estimation, In: Southern Africa Telecommunication Networks and Applications, 2012.
- [66] X. Zeng, W. Ouyang, M. Wang, and X. Wang, Deep learning of scene-specific classifier for pedestrian detection, In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), LNCS, Vol. 8691, pp. 472–487, 2014.
- [67] H. Idrees, K. Soomro, and M. Shah, Detecting humans in dense crowds using locally-consistent scale prior and global occlusion reasoning, In: IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 37, pp. 1986–1998, 2015.
- [68] K. De Brabanter, J. De Brabanter, J. Suykens, and B. De Moor, Approximate confidence and prediction intervals for least squares support vector regression, IEEE Transactions on Neural Networks, vol. 99, pp. 1–11, 2011.
- [69] N. Dalal, and B. Triggs, Histogram so oriented gradients for human detection, In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 886–893, 2005.
- [70] R. Haralick, K. Shanmugam, and I. Dinstein, Textural features for image classification, IEEE Transactions on Systems, Man and Cybernetics, vol. 3, no.6, pp.610–621, 1973.
- [71] S. Cohen, Background estimation as a labeling problem, In: IEEE International Conference on Computer Vision, vol. 2, pp. 1034–1041, 2005.
- [72] D. Hal, Frustratingly Easy Domain Adaptation, Journal of ACL, 2009.
- [73] Z. Zhang, H. Gunes, and M. Piccardi, Head detection for video surveillance based on categorical hair and skin colour models, In: Proceedings - International Conference on Image Processing, ICIP, pp. 1137–1140, 2009.
- [74] D. Merad, K. E. Aziz, and N. Thome, Fast people counting using head detection from skeleton graph, In: Proceedings - IEEE International

Conference on Advanced Video and Signal Based Surveillance, AVSS 2010, pp. 233–240, 2010.



Anahita Ghazvini

PhD candidate in computer science at Universiti Kebangsaan Malaysia (UKM). Received the bachelor degree with honours in information technology (computer science) at Universiti Kebangsaan Malaysia (UKM) in 2013. Received a master degree in information technology (artificial intelligence) at University Kebangsaan Malaysia (UKM) in 2016.



Siti Norul Huda Sheikh Abdullah

Received her first Degree in Computing at University of Manchester Institute of Science and Technology, United Kingdom. She furthered her master study in the area of Artificial Intelligence in Universiti Kebangsaan Malaysia. Later, she continued her Phd Study in the area of Computer Vision at Faculty of Electrical Engineering, Universiti Teknologi Malaysia. Starting the career, she involved in conducting national and international activities such as Royal Police Malaysia, Cyber Security Malaysia, Cyber Security Academia Malaysia, Federation of International Robot Soccer Association (FIRA), Asian Foundation, Global Ace Professional Certification Scheme, MIAMI, MACE and IDB Alumni. She is now holding a post as the Chairperson of Center for Cyber Security. Her research focuses are Digital Forensics, Pattern Recognition, and Computer Vision Surveillance System. She has published two books entitled “Pencegaman Pola” or “Pattern Recognition” and “Computational Intelligence for Data Science Application” and more than 50 and 100 of journal and conferences manuscripts correspondingly.



Masri Ayob

Prof Dr Masri Ayob is a lecturer in the Faculty of Information Science and Technology at the Universiti Kebangsaan Malaysia (UKM) since 1997. She has obtained her PhD in Computer Science at The University of Nottingham in 2005. Her main research areas include meta-heuristics, hyper-heuristics, scheduling and timetabling, especially educational timetabling, healthcare personnel scheduling and routing problems, and Internet of Things. She has published more than 100 papers at international journals and at peer-reviewed international conferences. She has been served as a programme committee for more than 50 international conferences and reviewers for high impact journals. She was a member of ASAP research group at the University of Nottingham. Currently, she is a principle researcher in Data Mining and Optimisation Research Group (DMO), Centre for Artificial Intelligent (CAIT), UKM.



# Detecting Image Brush Editing Using the Discarded Coefficients and Intentions

Fernando López Hernández\*, Luis de-la-Fuente Valentín, Íñigo Sarriá Martínez de Mendivil

Technology and Engineering Department of UNIR (Universidad Internacional de La Rioja), Logroño (Spain)

Received 8 June 2018 | Accepted 6 August 2018 | Published 14 August 2018



## ABSTRACT

This paper describes a quick and simple method to detect brush editing in JPEG images. The novelty of the proposed method is based on detecting the discarded coefficients during the quantization of the image. Another novelty of this paper is the development of a subjective metric named intentions. The method directly analyzes the allegedly tampered image and generates a forgery mask indicating forgery evidence for each image block. The experiments show that our method works especially well in detecting brush strokes, and it works reasonably well with added captions and image splicing. However, the method is less effective detecting copy-moved and blurred regions. This means that our method can effectively contribute to implementing a complete image-tampering detection tool. The editing operations for which our method is less effective can be complemented with methods more adequate to detect them.

## KEYWORDS

Image, Forgery, Brush, Stroke, Caption.

DOI: 10.9781/ijimai.2018.08.003

## I. INTRODUCTION

CURRENTLY, an important level of research has emerged for detecting forgery in digital images [1]. The detection of forged images has applications ranging from tampered handwriting [2] to insurance claims [3]. Forgers use different image editing tools [4]. This paper focuses on detecting scene editing with one of the currently least researched tool: the digital brush.

There are two contributions in this paper. The main contribution is the design of a new approach to detect brush editing along with the algorithm of the filter that detects this editing (see Section IV.D). As further described in Section III, there are few approaches designed to detect brush strokes, compared to other image modification techniques, such as cloning or image composition. The second contribution of this paper is the introduction of intentions as a subjective metric, along with its assessment application to forgery detection, in contrast with the more classical objective forgery metrics (see Section III.B).

## II. RELATED WORK

This section reviews the state of the art of forged image detection.

### A. Types of Digital Image Forgeries

Often authors [1][4][5][6] classify the forgery to be detected into five categories:

1. *Copy-moving* (or *cloning*). A region of the image is selected and then copy-pasted to a different region of the same image. This is the most popular form of forgery due to its simplicity to conceal

unwanted portions of the image, and effectiveness in leaving no visible traits of manipulation. Although the texture, color and noise of the pasted region are compatible with the rest of the image, there exists a wide variety of techniques to detect it (see, for instance, the survey in [6]).

2. *Image splicing* (or *image composites*). Image editing software usually allows for combining image fragments (typically represented as layers) from different images. One difference with copy-move forgery is that in composite image forgery there are no duplicate regions to be identified. Another difference is that if the forgers want to create a realistic image, they often have to apply geometric transformations (rotation, scaling, skew, etc.) to the spliced regions before pasting them into the target image [7][8]. That is, the size or orientation of the spliced regions in the source and target image usually does not match without these geometric transformations. Logically, geometric transformation in copy-move forgery detection has also been researched (e.g. [9][10]).

To identify the edited regions, often inconsistencies in region features are identified. For instance, [11] identifies JPEG compression features inconsistencies, the authors of [12][13][14] identify noise discrepancies in regions, [15] detects sharp changes surrounding the spliced region, and [16] identifies inconsistencies in shadow boundaries.

3. *Blurring and sharpening*. Blurring is an effective operation to remove traces in forged regions, especially at the edges of the manipulated regions. Fortunately, there are robust detection techniques to this attack (e.g. [17]). Sharpening is used to enhance the appearance of objects in an image, which is another form of forging. Cao et al. have studied sharpening detection in depth. They propose both, a method to detect unsharp masking (a popular sharpening operation) [18], and a method to detect sharpening in general through histogram aberration and ringing artifacts [19].

\* Corresponding author.

E-mail address: fernando.lopez@unir.net

4. *Image painting*. This category includes image tampering by painting and drawing. [20]. Cutzu et al [21] have proposed a method to discriminate between drawn images and genuine photo images by detecting changes in the hue, edge and texture features. Elgammal et al. [22] have developed a method to analyze forged strokes in paintings by characterizing personal strokes in drawings. Farid [23] has modeled brush detection as a segmentation problem, using a graph-cut algorithm to detect changes in intensity or texture. Lin and Huang [24] have detected air-brush and brush strokes by: (1) using the expectation-maximization (EM) algorithm in the JPEG coefficients, (2) generating a probability map in the frequency domain and (3) segmenting the periodicity in the probability map.
5. *Image retouching*. This category groups more subtle changes in the image that enhance or reduce certain features. For instance, Sutthiwan et al. [25] have proposed a method to detect changes in clarity or color of the texture. Mahalakshmi et al. [26] have proposed a method to detect affine transformations (rotation, scaling, etc.) by analyzing changes in the texture of the transformed region.

### B. Current Approaches, Strategies and Features

Forged image detection techniques have been broadly divided into active and passive (or blind) *approaches* [4][5][6][27][28]. *Active approaches* usually watermark or sign the image in order to detect future changes. *Passive approaches* use only the received image to assess if the image has suffered some kind of post-processing. The rest of this paper focuses on passive approaches.

H. Farid [27] (2009) classified forensic *strategies* into five categories. Ali Qureshi and Deriche [1] (2015) propose similar categories, but refer to these strategies as *tools*. In particular, these categories are:

1. *Pixel-based* strategies detect spatial irregularities in the pixel distribution properties. These strategies include, for instance, changes in noise level [12] or inter-block correlation [29]. These strategies have proved to be especially effective in identifying edited regions.
2. *Compression-based* strategies detect traces of forgery in the transformed domain, i.e., they are mainly designed for forensic analysis of JPEG images. These techniques can detect effects such as compression with a specific JPEG quantization table [30][31] or the quantization with two different quantization tables [32].
3. *Camera-based* strategies detect alterations in the characteristic artifacts that a specific camera model introduces. An example of these artifacts are the characteristic camera noise [33], or the remaining color after sensor interpolation (demosaicing) [34]. This means that they cannot be applied to analyze any image, since they only apply in certain camera models.
4. *Lighting-based* strategies detect inconsistencies in the 3D real world lighting effects, specular lighting or highlights in the surface geometry [35]. These techniques often require manual intervention to identify and analyze possible inconsistencies.
5. *Perspective-based* strategies detect when constraints are not met in the perspective of objects with respect to the camera, because the object has undergone a geometric transformation [36][37]. Although these strategies are named geometric-based in [1][27], we call them perspective-based, to distinguish them from the detection of geometric transformations in sliced regions (e.g. [7]).

Regardless of the strategy, most forgery detection methods are based on the general concept of *features*: the information extracted from the image to detect forgeries. These methods usually have two stages: 1) *Feature extraction* measures relevant characteristics of the image, and 2) *Feature matching* searches regions of the image with

similar features. The existence of regions with similar features is an indicator that one region may have been cloned from the other.

The extracted features can in turn be divided into three main types:

1. *Block-based features* are extracted from (overlapping or non-overlapping) rectangular blocks. The most typical features are the frequency representation, such as the histogram (e.g. [38]), or the Discrete Cosine Transform (DCT) (e.g. [39][40]) of the blocks. Other features are the texture of the blocks (e.g. [41]) or the *moment invariant* features, which are block features invariant to rotation and scaling [42].
2. *Keypoint-based features* are extracted from distinctive parts of the image such as corners, edges, or textures [43]. With these features, [44] identifies three issues to address: the non-uniform distribution of the keypoints, the threshold to select keypoints with low contrast, and how to cluster forged areas. For instance, [17] uses a Gabor filter for keypoint texture retrieval. Most of these features tend to be more robust to affine transformations. SIFT (Scale Invariant Feature Transform) is the most popular affine transform invariant keypoint feature. SURF [45] is an improvement on SIFT to reduce the dimension of the features and the computational time. The authors of [46] combine a point of interest detector with SIFT to extract more features points.
3. *Multi-scale features* allow for analyzing the image at different levels to achieve better detection results. The authors of [47] analyze textures at different levels to find copy-moved regions. The authors of [48] use multi-scale representation to cluster regions based on geometric constraints. The authors of [14] use multi-scale variation in noise to detect spliced regions.

As indicated above, feature matching searches for similarities (copy-move) or dissimilarities (spliced regions, blurring, retouching) between image features. An example of an effective matching method is clustering: the search space is divided into regions with similar features-vector distributions (e.g. [38]). Another popular feature matching is sorting. For instance, the authors of [49] use as features a histogram of oriented gradients that are lexicographically sorted to find duplicated blocks.

### III. PROPOSED APPROACH

As we have described in Section II.A, there is an extensive bibliography addressing the detection of forgery techniques of copy-move (cloning), image splicing (composites), blurring and sharpening. However, although graphic designers use the brush on a daily basis, its detection has not received the same level of attention. In the forth category in Section II.A we have described, to the best of our knowledge, the current research in brush painting forgeries.

Disturbances in the JPEG compression coefficients have already been successfully used to detect spliced regions [15][50] or double-compression [29][51]. In this work, we have hypothesized that brush editing also alters the distribution of these coefficients. However, the metric we use to detect these disturbances is different. Specifically, we first cancel the effect of lighting, and then assess the number of normalized coefficients that the JPEG compressor has discarded.

#### A. Forgery Localization and Forgery Mask

There are two granularity levels to represent image forgery *localization*:

1. *Image-based* localization classifies the entire image. Binary classification determines whether the image is forged or not. In this case, the true / false and positive / negative rates of the classifier are evaluated (see, for instance [40]). One way to represent the fuzzyness in the classification decision is to assign



a forgery probability to the image. To evaluate this probability, it may be useful to have a confidence interval, rather than a single point estimate. For this reason, authors such as [52] study the confidence intervals of the probabilistic classification.

2. *Region-based localization* is used when the application requires identifying the parts of the scene that have been modified. This occurs when a change in the image modifies the semantics of the scene. For example, a change in the light of a traffic light might eliminate the traffic offense of the scene. The *forgery mask* is a tool to represent these areas of the image with high probability of falsification. Our experiments use this forgery mask to highlight tampering. For example, Fig. 1(c) shows the forgery mask for an unedited image, and Fig. 1(d) shows the forgery mask after sharpening the monkey's body. In particular, in our forgery mask dark pixels indicate high probability of alteration.

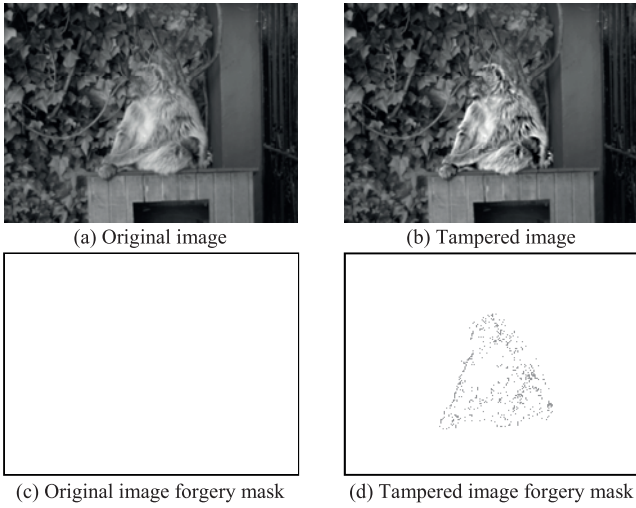


Fig. 1. Region based forgery localization for a sharpened monkey.

### B. Intentions and Interactive Forgery Mask

Typically, research in image forensics evaluates classification performance at either image or region level using an *objective metric*. For binary classification, they frequently use two metrics: *sensitivity*, i.e., the percentage of forgery correctly identified, and *specificity*, i.e., the percentage of unedited image correctly identified. Other alternative metrics, from the field of information retrieval, are *precision*, *recall* or the *F-score* (e.g., in [9][40]).

A drawback of these objective metrics is that a result like the one shown in our experiment in Fig. 2(d) has a relatively low sensitivity rate  $Se=0.5622$  (percentage of tampered pixels correctly detected). However, a visual inspection allows concluding that the image is forged. This is because a human is able to detect the intention of the forger without having to resort to soft computing techniques [53].

A second drawback of the objective metrics is their dependency on a threshold parameter. However, there is no general guideline to obtain this threshold, because usually each image has a threshold for maximum detection performance [54]. A human operator can effectively use semantics to effectively address this problem by means of an interactive gauge that allows the operator to visualize the forgery mask with different thresholds. We will refer to this gauge as the *interactive forgery mask*. Table I shows the sensitivity  $Se$  and specificity  $Sp$  for different threshold values with Fig. 2(b). Fig. 3 shows the corresponding forgery masks. Note that for a human operator the interactive mask is more helpful than the objective metrics.

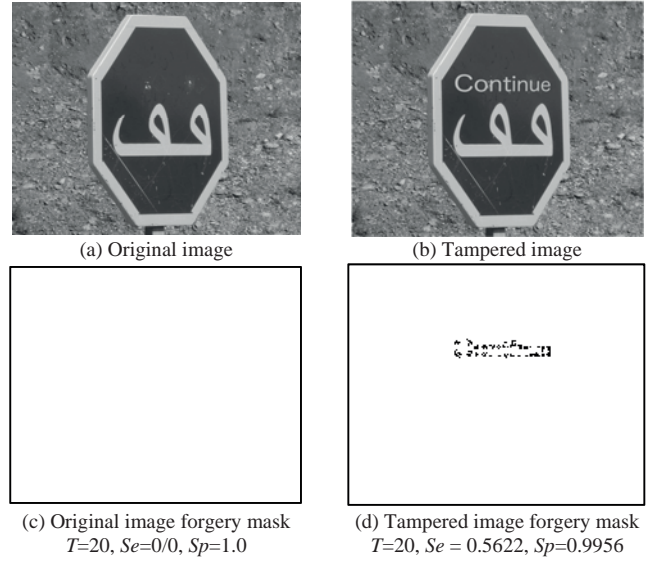


Fig. 2. First experiment (added caption).

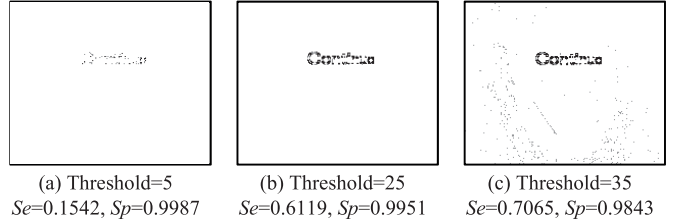


Fig. 3. Interactive forgery mask with three threshold levels in Fig. 2 (b).

TABLE I.  
SENSIBILITY AND SPECIFICITY FOR DIFFERENT THRESHOLD VALUES IN FIG. 2(B)

Threshold	Sensibility $Se$	Specificity $Sp$
5	0.1542	0.9987
10	0.3781	0.9973
20	0.5622	0.9956
25	0.6119	0.9951
30	0.6517	0.9947
35	0.7065	0.9843
40	0.7811	0.7924

A third drawback of the objective metrics is that editing does not prove tampering. For instance, the experiments in [55] reported that frequently the double-compression artifact was merely due to an image resaved with a different quality level. A human operator can semantically interpret the image to effectively decide if the marked area corresponds to an intentional forgery. This operator can also use an interactive forgery mask to best "focus" each image at its appropriate threshold level.

In spite of this, in order to facilitate the comparison, we have added to the figures of the reported experiments the objective sensitivity  $Se$  and specificity  $Sp$ .

## IV. METHOD

The method does not require the *original image* (untampered image). It requires only an *analyzed image* (potentially tampered image) in JPEG format in order to calculate the editing evidence of

each block. There is a twofold output:

1. An *objective metric* with the sensibility  $Se$  and specificity  $Sp$  of the classification.
2. A *subjective metric* with the *interactive forgery mask* indicating the probability of edition of each block (see Fig. 3).

#### A. Detected Effect

JPEG image compression uses the DCT to concentrate each block's energy in the low frequency coefficients, and high frequency coefficients are often reduced to zero. As described in [56], these high frequencies correspond to excessively sharp changes, which are the least noticeable for the human eye. The DCT tends to assign a low magnitude to these coefficients, and subsequently the JPEG compressor tends to round them to zero. In the rest of this paper we will refer to these zeroed coefficients as *discarded coefficients*.

Our working hypothesis is that brush strokes regenerate these unnoticeable sharp changes in the edited blocks. Consequently, 1) edited blocks will concentrate these high frequency coefficients, and 2) as a whole, the unedited blocks possesses a fewer count, in contrast with edited blocks.

Therefore, during recompression the JPEG compressor will need to discard a fewer number of coefficients to achieve the same compression ratio as the original image (as indicated in Section V, in our experiment we have used Adobe Photoshop default JPEG quality level:  $Q=12$ ). This effect would be even more prominent when the forger saves the image in a lossless format with the intention of preventing detection by other methods (e.g. double-compression [55][57][58]). As a consequence, edited blocks will have fewer discarded coefficients.

Note that, we are not indicating that brush painting necessarily increases the number of high coefficients than are in a natural photo (e.g. there are blurring brushes). What we are hypothesizing is that a higher number of coefficients will remain in the tampered area because during recompression the compressor discards a fewer number of them.

#### B. Tools

The tools to search for the abovementioned effect are the following:

1. *Counting discarded coefficients*. In our preliminary experiments we have observed that brush edited and recompressed blocks (especially those in the borders) keep a larger number of high coefficients, i.e., have a fewer number of discarded coefficients. Therefore, we use this count to gauge the editing probability of each block.
2. *Normalized energy*. In our preliminary experiments we found that lighting also influences the count of discarded coefficients. In particular, lighted areas (original or tampered) yield a lower count of discarded coefficients. So, before counting discarded coefficients, we need to *normalize the energy* of the analyzed image to eliminate the bias in the coefficients due the effect produced by lighting. After normalization, the count of discarded coefficients will not depend on the lighting of the blocks. The following section describes this normalization in more detail.

#### C. Canceling the Effect of Differences in Lighting

The magnitudes of the DCT coefficients indicate the energy of the block: lighter blocks will have larger DCT coefficients, and so fewer of them will be discarded. We cancel the effect of lighting in the magnitude of the coefficients by means of normalization.

To normalize the energy of the coefficients faster, we avoid converting them to their spatial representation, using the Parseval relationship. This relationship states that the mean energy of the spatial signal  $x[n, m]$  is equal to the mean energy of the coefficients in the frequency domain  $X[p, q]$ . In particular, given an  $N \times N$  block, Parseval

relationship states that:

$$\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} |x[m, n]|^2 = \sum_{p=0}^{N-1} \sum_{q=0}^{N-1} |X[p, q]|^2 \quad (1)$$

Where  $m, n$  are indexing the spatial block,  $p, q$  are indexing the coefficient of the corresponding block, and  $|\cdot|$  refers to the absolute value of the samples. Note that the spatial samples  $x[n, m]$  are integers in the range 0..255, while the coefficients  $X[p, q]$  are, in general, complex numbers.

Therefore we accomplish normalization in two steps:

1. Squaring the coefficients to measure energy and eliminate negative values:

$$E[p, q] = |X[p, q]|^2 \quad (2)$$

2. Scaling the  $E[p, q]$  values to the JPEG compressor storage range (i.e., 0..255). We can obtain these values using the following formula:

$$N[p, q] = 255 \frac{E[p, q] - \min\{E[p, q]\}}{\max\{E[p, q]\} - \min\{E[p, q]\}} \quad (3)$$

Where  $\min\{\}$  and  $\max\{\}$  refer to the minimum and maximum value in the block.

#### D. Filter Algorithm

The proposed filter algorithm is as follows:

1. Divide the image into non-overlapping blocks of  $8 \times 8$  pixels each. We propose using  $N=8$ , as this is the block size that the JPEG encoder typically uses.
2. Calculate the DCT coefficients of each block.
3. Normalize the energy of the blocks as described in Section IV.C.
4. Calculate the forgery evidence for each block as the sum of discarded coefficients  $S$  in the block (i.e., zeroed coefficients). In our implementation, for a given threshold  $T$ , we calculate forgery evidence  $E$  for each block with the following rule:  
if  $S < T$  then  
     $E=1.0$   
else if  $S < 2T$  then  
     $E=0.5$   
else  
     $E=0.0$
5. Create the forgery mask representing forgery evidence by assigning a grayscale level to each block. In our implementation a black pixel means definitely edited ( $E=1.0$ ), a white pixel unedited ( $E=0.0$ ), and a gray pixel that there is doubt ( $E=0.5$ ).

Fig. 1(d) and Fig. 3 are examples of the result of applying this algorithm. Note that in our implementation the forgery mask uses 3 gray levels to visually show the forgery evidence for each block. It is always possible to increase the number of gray levels, but we believe that, in general, it is difficult for the user to visually interpret more than 3 levels.

#### V. VALIDATION METHODOLOGY

This section demonstrates and assesses the proposed filter with different forgery techniques. For this purpose, we have surveyed the ability of the tool to detect intentions according to the purposes described in Section III.B. In addition, we are adding the sensibility

$Se$  and specificity  $Sp$  to the figure of each experiment. Due to space limitations in this section we only show a representative experiment of each type of analyzed forgery. All reported experiments have been performed with grayscale images. In the case of RGB images, the described procedure can be repeated in each channel of the JPEG image.

#### A. Detection of Brush Editing

For the reported experiments we have used Adobe Photoshop and done our best to create semantically realistic forgeries without sharp borders or any other forgery sign. We have saved the forged images with the default quality level of Adobe Photoshop ( $Q=12$ ), assuming that this default value is the more likely to be used by a forger. The figure of each experiment indicates a threshold  $T$  manually chosen for the interactive forgery mask.

##### 1) Experiment 1: Added Caption

The first experiment was made by adding a caption with perspective and blended border to the original image in Fig. 2(a). The forged image is shown in Fig. 2(b). Fig. 2(c) shows the forgery mask that the filter produces with threshold  $T=20$  on the original image.

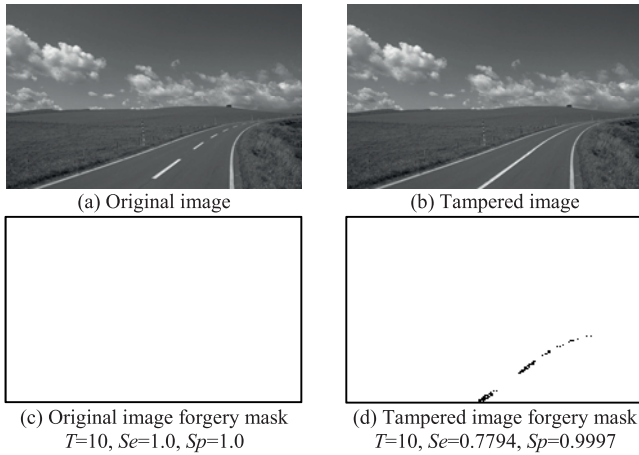


Fig. 4. Second experiment (brush painting).

The forgery mask in Fig. 2(c) does not indicate signs of forgery in any block ( $Se=0/0$ ,  $Sp=1.0$ ). Fig. 2(d) shows the forgery mask with  $T=20$  on the forged image. The signs of forgery are evident, and a human can easily detect the semantic intention. However, the objective metric is reporting relatively low sensibility  $Se=0.5622$ , i.e., the proportion of forged blocks that are correctly identified as such is 56.22%. Fig. 3 shows the result of the analysis of the same image with three different threshold values.

##### 2) Experiment 2: Brush Painting

For the second experiment in Fig. 4 we have used an 80% solid brush to turn the broken lines into a solid line. Note that the forgery mask in Fig. 4 correctly detects the edited blocks without leaving doubt about the forger's intention. In addition, the forgery mask reaches a high objective detection score:  $Se=0.7794$ ,  $Sp=0.9997$ .

#### B. Detection of Other Forgeries

Our experiments have revealed promising results with the detection of other types of forgeries, although without reaching the same level of precision. Therefore, we are demonstrating below the results that we are obtaining with these other types of forgeries.

##### 1) Experiment 3: Copy-move (Cloning)

The third experiment is for copy-move forgery. Fig. 5(b) shows a

copy-moved cat from the original photo in Fig. 5(a). The forgery mask gives some evidence of forgery, but mainly detects forgery in the edges of the forged region. Note that while the forgery mask enables us to perceive the forgery, the objective metric indicates a very low detection rate  $Se=0.0319$ .

##### 2) Experiment 4: Splicing (Composite)

The fourth experiment is for image splicing. In Fig. 6(b) a duck has been added to the lake. The forgery mask in Fig. 6(c) shows some sign of editing in the original image. We have downloaded the original lake image from the Internet, so we do not have access to the original photo. However, we think that the vegetation of the lower right corner has been edited (possibly with a contrast enhancement filter). Also the shadow over the water seems to have been artificially generated.

Regarding the forgery mask in Fig. 6(d), it gives some evidence of forgery, mainly in the edges of the spliced region.

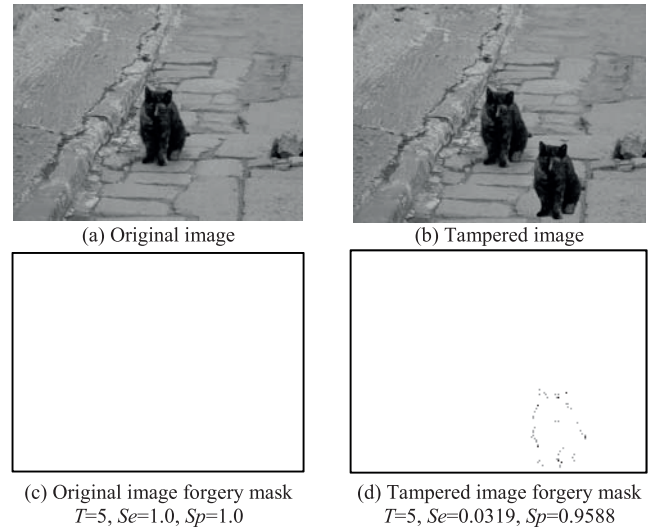


Fig. 5. Third experiment (copy-move).

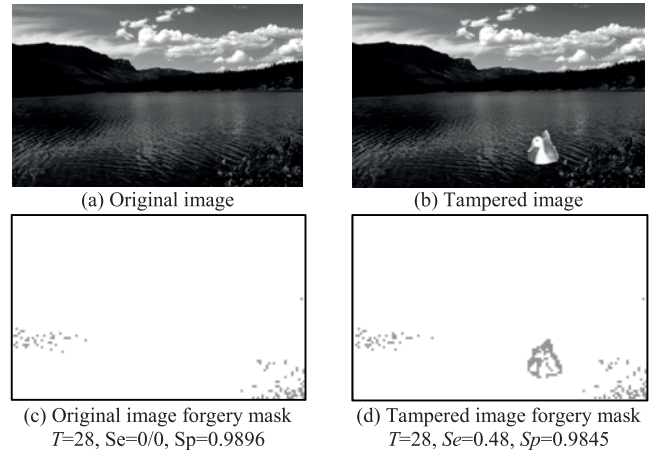


Fig. 6. Fourth experiment (splicing).

Original image source: <http://www.northamericatouring.com/images/10>

##### 3) Experiment 5: Blurring

For the fifth experiment we have spliced a bottle image in Fig. 7(b) obtaining the forgery mask in Fig. 7(c), which has traces of forgery on the edges. Then we have applied a Gaussian Blur filter with radius 3 to the tampered image in Fig. 7(b), and then recalculated the forgery mask in Fig. 7(d). These results show that our method loses its effectiveness



when the forger applies a blurring filter to the edited image.

## VI. CONCLUSIONS AND FUTURE WORK

We have observed that the recompression of an edited image block leaves a significant amount of undiscarded high frequency coefficients, and we have identified that the compressor is the responsible for it. In particular, this effect occurs because the first compression of the original JPEG image removes a large portion of these coefficients, and so the compressor is not as greedy for high coefficients when recompressing. The effect is more noticeable when the forger saves the image in a lossless format, but it is enough if the forger saves the image in a lossy format, as is the case in the reported experiments. The experiments also show that this effect is more prominent with brush-edited images, but is also able to detect other forgeries.

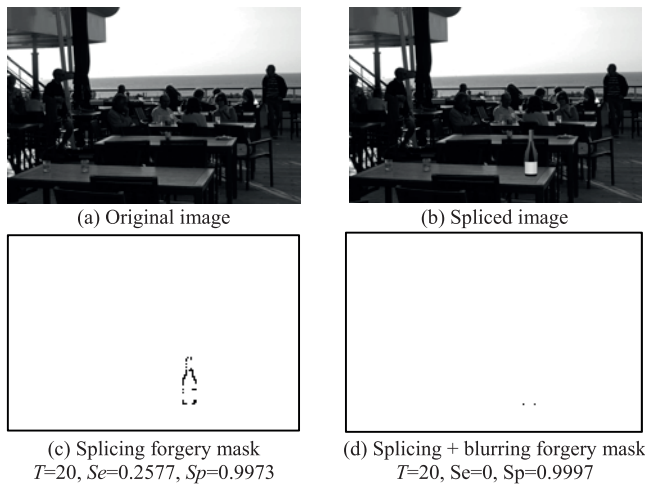


Fig. 7. Fifth experiment (splicing and blurring).

The experiments also reveal that the clustering of potentially modified blocks in semantically noticeable areas (intentions) has two benefits. 1) It makes the subjective human evaluation best determine forged areas that with a classical objective classification rate. For example, the forgery mask in the traffic infraction photo in Fig. 4 serves to provide enough evidence to identify an intentionally tampered image. 2) The interactive forgery mask (see Fig. 3) eases the determination of a suitable threshold with the help of an human operator, compared to using an optimal automatic threshold search algorithm (e.g., [54]).

The major limitation of our method is that the forger can easily erase the effect that we search for by blurring the edited region. This means that to detect blurred edited areas, our method should be combined with other methods, such as [18][59].

### A. Future Work

The first pending future work is to assess and compare intention recognition in alternative state of the art methods. The second future work is the execution of the implemented method with a standard forgery image database, such as [60][61].

## REFERENCES

- [1] M. A. Qureshi and M. Deriche, "A bibliography of pixel-based blind image forgery detection techniques," *Signal Process. Image Commun.*, vol. 39, pp. 46–74, Nov. 2015.
- [2] Y. Bouldid, A. Souhar, and M. E. Elketani, "Multi-agent Systems for Arabic Handwriting Recognition," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. Regular Issue, 2017.
- [3] W. Lin *et al.*, "Survey on blind image forgery detection," *IET Image Process.*, vol. 7, no. 7, pp. 660–670, Oct. 2013.
- [4] G. K. Birajdar and V. H. Mankar, "Digital image forgery detection using passive techniques: A survey," *Digit. Investig.*, vol. 10, no. 3, pp. 226–245, 2013.
- [5] C.-Z. X. Tanzeela Qazi, Khizar Hayat, Samee U. Khan, Sajjad A. Madani, Imran A. Khan, Joanna Kolodziej, Hongxiang Li, Weiyao Lin, Kin Choong Yow *et al.*, "Survey on blind image forgery detection," *IET Image Process.*, vol. 7, no. 7, pp. 660–670, Oct. 2013.
- [6] N. Choo *et al.*, "Copy-move forgery detection: Survey, challenges and future directions," *J. Netw. Comput. Appl.*, vol. 75, pp. 259–278, Nov. 2016.
- [7] M. Jaber, G. Bebis, M. Hussain, and G. Muhammad, "Accurate and robust localization of duplicated region in copy-move image forgery," *Mach. Vis. Appl.*, vol. 25, no. 2, pp. 451–475, Feb. 2014.
- [8] Seung-Jin Ryu, M. Kirchner, Min-Jeong Lee, and Heung-Kyu Lee, "Rotation Invariant Localization of Duplicated Image Regions Based on Zernike Moments," *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 8, pp. 1355–1370, Aug. 2013.
- [9] N. B. Abd.Warif, A. W. AbdulWahab, M. Y. IdnaIdris, RosliSalleh, and FazidahOthman, "SIFT-Symmetry: A robust detection method for copy-move forgery with reflection attack," *J. Vis. Commun. Image Represent.*, vol. 46, pp. 219–232, Jul. 2017.
- [10] D. Cozzolino, G. Poggi, and L. Verdoliva, "Splicebuster: A new blind image splicing detector," in *2015 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2015, pp. 1–6.
- [11] Juxian Zuo, Shengjun Pan, Benyong Liu, and Xiang Liao, "Tampering detection for composite images based on re-sampling and JPEG compression," in *The First Asian Conference on Pattern Recognition*, 2011, pp. 169–173.
- [12] W. C. Hu, J. S. Dai, and J. S. Jian, "Effective composite image detection method based on feature inconsistency of image components," *Digit. Signal Process. A Rev. J.*, vol. 39, 2015.
- [13] W. Wang, J. Dong, and T. Tan, "Tampered Region Localization of Digital Color Images Based on JPEG Compression Noise," Springer, Berlin, Heidelberg, 2011, pp. 120–133.
- [14] C.-M. Pun, B. Liu, and X.-C. Yuan, "Multi-scale noise estimation for image splicing forgery detection," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 195–206, Jul. 2016.
- [15] Z. Fang, S. Wang, and X. Zhang, "Image Splicing Detection Using Color Edge Inconsistency," in *2010 International Conference on Multimedia Information Networking and Security*, 2010, pp. 923–926.
- [16] Qiguang Liu, Xiaochun Cao, Chao Deng, and Xiaojie Guo, "Identifying Image Composites Through Shadow Matte Consistency," *IEEE Trans. Inf. Forensics Secur.*, vol. 6, no. 3, pp. 1111–1122, Sep. 2011.
- [17] G. Muzaffer, O. Makul, B. Ustubioglu, and G. Ulutas, "Copy Move Forgery Detection Using Gabor Filter and ORB," *Proc. 2016 International Conf. Image Process. Prod. Comput. Sci.*, pp. 23–29, 2016.
- [18] G. Cao, Y. Zhao, R. Ni, and A. C. Kot, "Unsharp Masking Sharpening Detection via Overshoot Artifacts Analysis," *IEEE Signal Process. Lett.*, vol. 18, no. 10, pp. 603–606, Oct. 2011.
- [19] Gang Cao, Yao Zhao, and Rongrong Ni, "Detection of image sharpening based on histogram aberration and ringing artifacts," in *2009 IEEE International Conference on Multimedia and Expo*, 2009, pp. 1026–1029.
- [20] E. Verdú, C. P. G. Bustelo, M. Á. M. Sánchez, and R. G. Crespo, "A System to Generate SignWriting for Video Tracks Enhancing Accessibility of Deaf People," *Int. J. Interact. Multimed. Artif. Intell.*, vol. 4, no. Regular Issue, 2017.
- [21] F. Cutzu, R. Hammoud, and A. Leykin, "Distinguishing paintings from photographs," *Comput. Vis. Image Underst.*, vol. 100, no. 3, pp. 249–273, 2005.
- [22] A. Elgammal, Y. Kang, and M. Den Leeuw, "Picasso, Matisse, or a Fake? Automated Analysis of Drawings at the Stroke Level for Attribution and Authentication," *CoRR*, vol. abs/1711.0, 2018.
- [23] H. Farid, "Exposing Digital Forgeries in Scientific Images," in *Proceedings of the 8th Workshop on Multimedia and Security*, 2006, pp. 29–36.
- [24] T. Lin and C.-L. Huang, "Digital Image Forensics Using EM Algorithm," in *PCM*, 2009.
- [25] P. Sutthiwan, Y. Q. Shi, W. Su, and T.-T. Ng, "Rake transform and edge statistics for image forgery detection," in *2010 IEEE International*

*Conference on Multimedia and Expo*, 2010, pp. 1463–1468.

- [26] S. Devi Mahalakshmi, K. Vijayalakshmi, and S. Priyadharsini, “Digital image forgery detection and estimation by exploring basic image manipulations,” *Digit. Investig.*, vol. 8, no. 3, pp. 215–225, 2012.
- [27] H. Farid, “Image Forgery Detection – A survey,” 2009.
- [28] S. V. Ashima Gupta, Nisheeth Saxena, “Detecting Copy move Forgery using DCT,” *Int. J. Sci. Res. Publ.*, vol. 3, no. 5, 2013.
- [29] Wei Wang, Jing Dong, and Tieniu Tan, “Exploring DCT Coefficient Quantization Effects for Local Tampering Detection,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 10, pp. 1653–1666, Oct. 2014.
- [30] J. D. Kornblum and J. D., “Using JPEG quantization tables to identify imagery processed by software,” *Digit. Investig.*, vol. 5, pp. S21–S25, Sep. 2008.
- [31] S. Q. Fu D, Shi YQ, “A generalized Benford’s law for JPEG coefficients and its applications in image forensics,” in *Proc. SPIE Electronic Imaging, Security and Watermarking of Multimedia Contents*, 2007.
- [32] Bin Li, Tian-Tsong Ng, Xiaolong Li, Shunquan Tan, and Jiwu Huang, “Statistical Model of JPEG Noises and Its Application in Quantization Step Estimation,” *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1471–1484, May 2015.
- [33] J. Fan, H. Cao, and A. C. Kot, “Estimating EXIF Parameters Based on Noise Features for Image Manipulation Detection,” *IEEE Trans. Inf. Forensics Secur.*, vol. 8, no. 4, pp. 608–618, Apr. 2013.
- [34] A. E. Dirik and N. Memon, “Image tamper detection based on demosaicing artifacts,” in *2009 16th IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 1497–1500.
- [35] E. Kee and H. Farid, “Exposing digital forgeries from 3-D lighting environments,” in *2010 IEEE International Workshop on Information Forensics and Security*, 2010, pp. 1–6.
- [36] H. Yao, S. Wang, Y. Zhao, and X. Zhang, “Detecting Image Forgery Using Perspective Constraints,” *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 123–126, Mar. 2012.
- [37] A. Pacheco, H. B. Barón, R. G. Crespo, and J. Pascual-Espada, “Reconstruction of High Resolution 3D Objects from Incomplete Images and 3D Information,” *Int. J. Interact. Multimed. Artif. Intell.*, vol. 2, no. Regular Issue, 2014.
- [38] H. Zhou, Y. Shen, X. Zhu, B. Liu, Z. Fu, and N. Fan, “Digital image modification detection using color information and its histograms,” *Forensic Sci. Int.*, vol. 266, pp. 379–388, Sep. 2016.
- [39] Rohini.A.Maing, A. Khade, and D.K.Chitre, “Image Copy Move Forgery Detection using Block Representing Method,” *Int. J. Soft Comput. Eng.*, vol. 4, no. 2, p. 49.53, 2014.
- [40] A. Alahmadi, M. Hussain, H. Aboalsamh, G. Muhammad, G. Bebis, and H. Mathkour, “Passive detection of image forgery using DCT and local binary pattern,” *Signal, Image Video Process.*, pp. 1–8, Apr. 2016.
- [41] E. Ardizzone, A. Bruno, and G. Mazzola, “Copy-move forgery detection via texture description,” in *Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence - MiFor '10*, 2010, p. 59.
- [42] J. Zhong, Y. Gan, J. Young, and P. Lin, “Copy Move Forgery Image Detection via Discrete Radon and Polar Complex Exponential Transform-Based Moment Invariant Features,” *Int. J. Pattern Recognit. Artif. Intell.*, vol. 31, no. 02, p. 1754005, Feb. 2017.
- [43] A. D. Warbhe, R. V. Dharaskar, and V. M. Thakare, “A Survey on Keypoint Based Copy-paste Forgery Detection Techniques,” *Procedia Comput. Sci.*, vol. 78, pp. 61–67, Jan. 2016.
- [44] G. Jin and X. Wan, “An improved method for SIFT-based copy–move forgery detection using non-maximum value suppression and optimized J-Linkage,” *Signal Process. Image Commun.*, vol. 57, pp. 113–125, Sep. 2017.
- [45] V. T. Manu and B. M. Mehtre, “Detection of copy-move forgery in images using segmentation and SURF,” in *Advances in Signal Processing and Intelligent Recognition Systems*, Springer, 2016, pp. 645–654.
- [46] F. Yang, J. Li, W. Lu, and J. Weng, “Copy-move forgery detection based on hybrid features,” *Eng. Appl. Artif. Intell.*, vol. 59, pp. 73–83, Mar. 2017.
- [47] X. Bi, C.-M. Pun, and X.-C. Yuan, “Multi-Level Dense Descriptor and Hierarchical Feature Matching for Copy–Move Forgery Detection,” *Inf. Sci. (Ny)*, vol. 345, pp. 226–242, Jun. 2016.
- [48] E. Silva, T. Carvalho, A. Ferreira, and A. Rocha, “Going deeper into copy-move forgery detection: Exploring image telltales via multi-scale analysis and voting processes,” *J. Vis. Commun. Image Represent.*, vol. 29, pp. 16–32, May 2015.
- [49] J.-C. Lee, C.-P. Chang, and W.-K. Chen, “Detection of copy–move image forgery using histogram of orientated gradients,” *Inf. Sci. (Ny)*, vol. 321, pp. 250–262, Nov. 2015.
- [50] J. He, Z. Lin, L. Wang, and X. Tang, “Detecting Doctored JPEG Images Via DCT Coefficient Analysis,” Springer Berlin Heidelberg, 2006, pp. 423–435.
- [51] I. Amerini, R. Becarelli, R. Caldelli, and A. Del Mastio, “Splicing forgeries localization through the use of first digit features,” in *2014 IEEE International Workshop on Information Forensics and Security (WIFS)*, 2014, pp. 143–148.
- [52] A. S. Alfraih, J. A. Briffa, and S. Wesemeyer, “Forgery Localization Based on Image Chroma Feature Extraction,” in *5th International Conference on Imaging for Crime Detection and Prevention (ICDP 2013)*, 2013, p. 2.11-2.11.
- [53] F. López Hernández, E. Giménez de Ory, S. Ríos Aguilar, and R. González Crespo, “Residue properties for the arithmetical estimation of the image quantization table,” *Appl. Soft Comput.*, vol. 67, pp. 309–321, Jun. 2018.
- [54] B. Ustubioglu, G. Ulutas, M. Ulutas, and V. V. Nabyev, “A new copy move forgery detection technique with automatic threshold determination,” *AEU - Int. J. Electron. Commun.*, vol. 70, no. 8, pp. 1076–1087, Aug. 2016.
- [55] A. Taimori, F. Razzazi, A. Behrad, A. Ahmadi, and M. Babaie-Zadeh, “A novel forensic image analysis tool for discovering double JPEG compression clues,” *Multimed. Tools Appl.*, vol. 76, no. 6, pp. 7749–7783, 2017.
- [56] G. K. Wallace, “The JPEG still picture compression standard,” *Commun. ACM*, pp. 30–44, 1991.
- [57] J. Yang, J. Xie, G. Zhu, S. Kwong, and Y.-Q. Shi, “An Effective Method for Detecting Double JPEG Compression With the Same Quantization Matrix,” *IEEE Trans. Inf. Forensics Secur.*, vol. 9, no. 11, pp. 1933–1942, Nov. 2014.
- [58] H. Farid, *Photo Forensics*. MIT Press, 2016.
- [59] B. Y. and B. LIU, “Feature Fusion for Blurring Detection in Image Forensics,” *IEICE Trans. Inf. Syst.*, vol. E97.D, no. 6, pp. 1690–1693, 2014.
- [60] J. Dong, W. Wang, and T. Tan, “CASIA Image Tampering Detection Evaluation Database,” in *2013 IEEE China Summit and International Conference on Signal and Information Processing*, 2013, pp. 422–426.
- [61] M. G. Dijana Tralic, Ivan Zupancic, Sonja Grgic, “CoMoFoD -New Database for Copy-Move Forgery Detection,” in *Proceedings ELMAR-2013 : 55th International Symposium ELMAR-2013, 25-27 September 2013, Zadar, Croatia*, 2013.



Fernando López Hernández

He is a full-time associate professor at UNIR. His current research interests lie in image and video processing, data-driven science, machine learning, and programming languages.



Luis de la Fuente Valentín

He is a full-time associate professor at UNIR. His current research interest is on data analysis and data mining, and data visualization mainly in the educational field.



Íñigo Sarría Martínez de Mendivil

He is a full-time professor at UNIR. His research focuses on mathematical modeling in Banach spaces, convergence of iterative methods and their dynamics.

# Deep Belief Network and Auto-Encoder for Face Classification

Nassih Bouchra\*, Amine Aouatif, Ngadi Mohammed, Hmina Nabil

LGS, National School of Applied Sciences, Ibn Tofail University, B.P. 241, university campus, Kenitra (Morocco)

Received 27 February 2018 | Accepted 14 May 2018 | Published 22 June 2018



## ABSTRACT

The Deep Learning models have drawn ever-increasing research interest owing to their intrinsic capability of overcoming the drawback of traditional algorithm. Hence, we have adopted the representative Deep Learning methods which are Deep Belief Network (DBN) and Stacked Auto-Encoder (SAE), to initialize deep supervised Neural Networks (NN), besides of Back Propagation Neural Networks (BPNN) applied to face classification task. Moreover, our contribution is to extract hierarchical representations of face image based on the Deep Learning models which are: DBN, SAE and BPNN. Then, the extracted feature vectors of each model are used as input of NN classifier. Next, to test our approach and evaluate its performance, a simulation series of experiments were performed on two facial databases: BOSS and MIT. Our proposed approach which is (DBN,NN) has a significant improvement on the classification error rate compared to (SAE,NN) and BPNN which we get 1.14% and 1.96% in terms of error rate with BOSS and MIT respectively.

## KEYWORDS

Deep Learning, Deep Belief Network, Facial Recognition, Neural Network, Stacked Auto-Encoder.

DOI: 10.9781/ijimai.2018.06.004

## I. INTRODUCTION

RECENTLY, there have been a lot of applications which rely on the face classification. For instance, a security system that allows access only to a people who are members of a certain group or a surveillance system that can give an alert to law enforcement agencies of the presence of a person who belongs or has a link with an international terrorist group. Each of these applications relies on the integration of a face classification system. This article is devoted to the problem of face classification based on feature extraction and classification task [1] [32]. Therefore, the objective of the feature extraction procedure is to extract invariant features representing the face information. Many different kinds of features have been used in previous works, such as hand-crafted features presented by the Local Binary Pattern (LBP) [2] [3] [31], Discrete Cosine Transform (DCT) [4] and Discrete Wavelet Transform (DWT) [5], etc. With the extracted features, general machine learning algorithms such as Support Vector Machine (SVM) [6] and Artificial Neural Network [36] can be used to discriminate between faces and non-faces.

Nowadays, the Deep Learning becomes a most powerful subfield of machine learning that focuses on learning deep hierarchical models which widely applied in a variety of areas such as speech recognition, pattern recognition and computer vision. Many Deep Learning algorithms were proposed and successfully applied as Convolutional Neural Network (CNN) which used it to extract distinctive features for face patterns and classification [7] [8]. The Restricted Boltzmann Machines (RBMs) are usually used as feature extractors for another algorithm or to keep a good initialization for deep feed-forward neural

network classifiers. Deep Belief Network (DBN) and Deep Boltzmann Machines (DBM), which are powerful techniques in pattern recognition task [9] [10]. In the latest progress of Deep Learning, researchers have achieved new records in face verification by exploiting different CNN structures [11], [8], [12], [13], [14].

In this paper, we intend to perform a new contribution based on the Deep Learning models for the face classification task by investigating two different kinds of Deep Learning architectures which consist of DBN and Stacked Auto-Encoder (SAE) besides of Back Propagation Neural Networks (BPNN) in order to explicitly capture various latent facial features. These representations offer several advantages over those obtained through hand-crafted descriptors, because, they can capture higher-order statistics such as corners and contours.

The rest of the paper is organized as follows: In Section II, we describe briefly some Deep Learning architectures, which are DBN and SAE, besides of BPNN. The proposed approach is presented in Section III. The experimental results of the proposed contribution along with comparative analysis are discussed in Section IV. Finally, we draw conclusions and give avenues for future work in Section V.

## II. DEEP LEARNING MODELS

### A. Deep Learning

Deep Learning is one of the most powerful parts of the wider machine learning field focused on learning representations and abstraction of data. The idea behind the Deep Learning advanced from the study of Artificial Neural Networks (ANNs) [15]. This network operates a large number of successive transformations, making it possible to discover more and more abstract representations. These transformations make it possible to represent the data at different levels of abstraction, for example, the first layer corresponds to the raw data, the last layer corresponds to the

\* Corresponding author.

E-mail address: nassih.bouchra@univ-ibntofail.ac.ma



outputs and each successive layer uses the output from the previous layer as input. Then, the core philosophy within the Deep Learning framework is to let the network directly learn the useful feature representations and at the same time train with the prediction tasks from end to end, which helps the Deep Learning technique to set a new records for many vision tasks. Further, there are various Deep Learning architectures such as, CNNs, DBN, and SAE. These architectures are oftentimes built with a ravening layer-by-layer technique.

## B. Feature Learning

Feature learning [16] is a set of methods that permits a system to spontaneously learn the representations required for feature detection or classification of raw data which influences the final result. This technique replaces the hand-crafted descriptors as LBP [2] [3], DCT [4] and DWT [5], it permits a machine to learn the features in cooperation and use them to carry out a particular task as classification, detection or recognition. We generally can get a satisfactory result based on these feature learning, which yields a good classification rate for face classification. Therefore, the most important function of Deep Learning is to pick up the most pertinent features by means of a great number of training data to build a machine learning model that has an interesting number of hidden layers in order to improve the accuracy of the classification.

### 1) Back-Propagation Neural Network

A Back Propagation Neural Network (BPNN) is a multilayered, feed forward Neural Network [23]. It is a kind of Neural Network (NN) architecture trained with Back Propagation (BP) algorithm. This last is an efficient gradient descent algorithm which represents a big importance in NN. In fact, the BP is a popular technique that has been known for its exactness, because, it enables itself to learn and enhancing itself, for this reason, it can accomplish a higher precision [17]. Therefore, the principle of this algorithm is to optimize the parameters of the NN. Further, when a result is obtained, the classification error is calculated. Subsequently, this error is propagated back from one layer to the other starting from the output layer until the minimal Mean Squared Error (MSE) is achieved, so that the weights can be modified according to the obtained error. In brief, the BP is a training algorithm which includes two processes [19] [28]. The first one is feed forward the values to generate output. The second one is to calculate the error between the output and the target output, then, propagate it back to the earlier layers in order to update weights and biases. The BPNN is used as a classifier basically due to its capacity to produce complex ruling limits in the characteristic features space [29]. There are different activation functions which can be utilized in BPNN. Among the more prevalent is the sigmoid, expressed by (1):

$$f(x) = \frac{1}{1+e^{-x}} \quad (1)$$

The BPNN comprises of three layers, which are input layer, hidden layer and output layer. Then, each layer has a specific number of neurons depending upon the complexity of the problem. Fig. 1 illustrates the structure of standard BPNN.

To adjust the weight of one neuron  $j$ , we used the following equations:

$$W_j = W_j + \Delta W_j \quad (2)$$

$$\Delta W_j = \alpha E_j Y_j \quad (3)$$

With:

$E_j$ : The output error for the neuron  $j$  in the output layer defined by (4).

$Y_j$ : The output of neuron  $j$  presented by (5).

$\alpha$ : The learning rate.

$$E_j = Y_j(1 - Y_j)(T_j - Y_j) \quad (4)$$

$$Y_j = \frac{1}{1+e^{-x_j}} \quad (5)$$

With:

$T_j$ : The target output.

We took only an example of one neuron  $j$ , but, the same procedure is done for all used neurons.

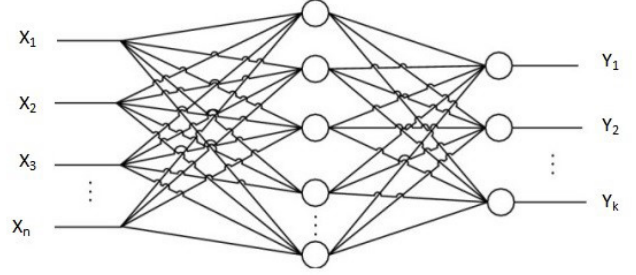


Fig.1. An example of a three-layer BPNN [30].

According to the experiments, the optimal parameters of the BPNN are reported in Table I.

TABLE I. THE ADOPTED PARAMETERS OF BPNN

Number of layers	3
Activation function	Sigmoid
Learning Rate	1
Input Zero Masked Fraction	0.5
Number of epochs	5

### 2) Deep Belief Network

A DBN is a deep structured learning architecture introduced by Hinton since 2006 [20]. It is a generative model interpretable as a stack of the Restricted Boltzmann Machine (RBM). The DBN is composed of multiple layers of stochastic and latent variables and can be regarded as a special form of the Bayesian probabilistic generative model. When we used the DBN for classification task, it is treated as a MultiLayer Perceptron (MLP), by adding a logistic regression layer on top, but DBN is more effective than MLP. To develop the pliability of DBN, a novel model of Convolutional Deep Belief Networks (CDBNs) was introduced by Arel [21]. The first step in the training of DBN is to learn a layer of features from the visible units, using Contrastive Divergence (CD) algorithm [18], the following step is to treat the activations of previously trained features as visible unites and learn features in a second hidden layer. At last, the whole DBN is trained when the learning for the last hidden layer is getting at. According to the experiments, the optimal parameters of the DBN are depicted in Table II.

TABLE II. THE ADOPTED PARAMETERS OF DBN

Number of layers	4
Activation function	Sigmoid
Learning Rate	1
Input Zero Masked Fraction	0.5
Number of epochs	5

Fig. 2 and 3 present the visualization of the output response of the DBN model on both databases, BOSS and MIT respectively. Regarding BOSS, most features extracted contain shape information which is illustrated by the facial contours as related in Fig. 2. While for MIT database, the shape information is not clear compared to BOSS database.

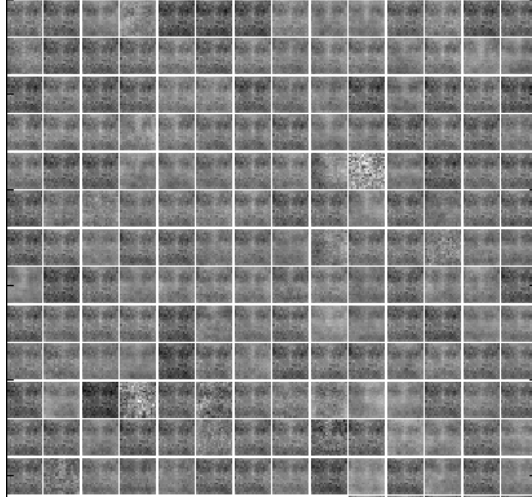


Fig. 2. Weight visualization using DBN on BOSS database.

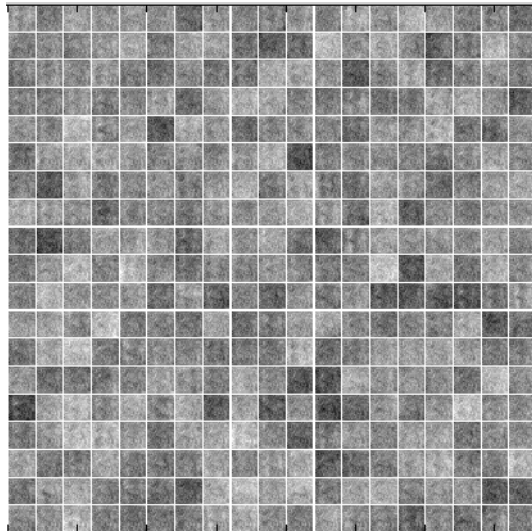


Fig. 3. Weight visualization using DBN on MIT database.

### 3) Stacked Auto-Encoder

A SAE is a Neural Network consisting of multiple layers of Auto-Encoders in which the outputs of each layer are wired to the inputs of the successive layer. The SAE enjoys all the benefits of any deep network of greater expressive power. A SAE is a building of deep network architecture by hierarchically training layers of Auto-Encoders.

The Auto-Encoders have been proposed in [22] [23], as methods for dimensionality reduction. In addition, it often captures a “hierarchical grouping” or a useful “partial-total decomposition” of the input. The first layer of a SAE tends to learn first-rate features in the raw input (such as the edges of an image). The second layer tends to learn second-order characteristics corresponding to patterns in the appearance of the first order features. Upper layers of SAE tend to learn higher order characteristics in order to form a good representation of its input.

The SAE model is characterized by 5 parameters. A series of tests is performed to determine these optimal parameters, as depicted in Table III.

TABLE III. THE ADOPTED PARAMETERS OF SAE

Number of layers	3
Activation function	Sigmoid
Learning Rate	1
Input Zero Masked Fraction	0.5
Number of epochs	5

In brief, the SAE is based on unsupervised layer learning, which drives each Auto-Encoder independently, like the case of DBN, in which each neuron has a weight vector associated with it, which will be set to respond to a particular visual characteristic. Fig. 4 and 5 illustrate the representation of these facial characteristics for both databases BOSS and MIT respectively, based on the SAE algorithm. But, according to Fig. 2, 3, 4 and 5, the DBN learns useful representations compared to SAE model which demonstrates the robustness of our proposed approach for face classification, especially on BOSS database.

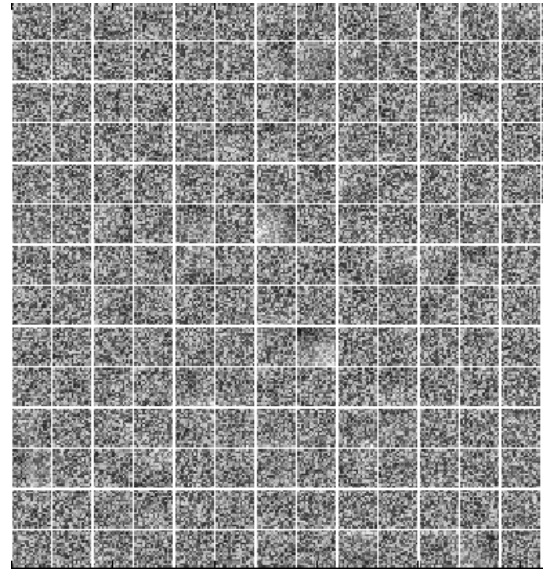


Fig. 4. Weight visualization using SAE on BOSS database.

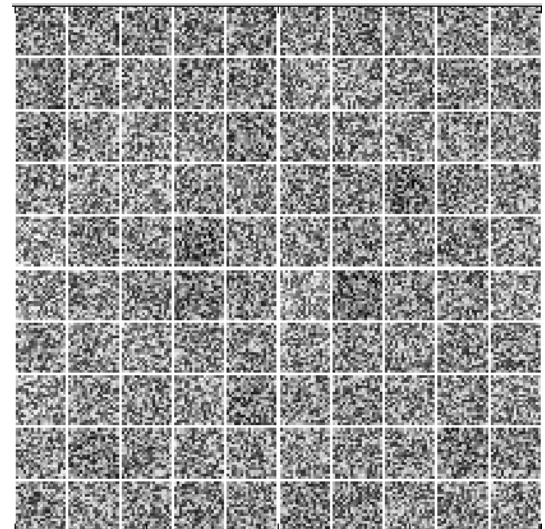


Fig. 5. Weight visualization using SAE on MIT database.

### III. PROPOSED APPROACH

Among the important stages of the face classification system is to characterize all face images as a feature vector. Subsequently, we can apply many learning algorithms, for instance, SVM [24] [25] in order to perform the classification task. Thus, the efficiency of face classification algorithm primarily focuses on the useful features. Such as, for feature extraction step, many comparative studies suggested varied hand-crafted features as LBP and Scale Invariant Feature Transform (SIFT), but in our work, we exploit a robust feature learning method to discriminate between faces and non-faces in order to ensure an efficient classification. Further, the Deep Learning models are adopted to replace and overcome the limitations of the hand-crafted methods, which enhance the features by the high level presentation of each face image.

The main objective of our work is to test and compare the effectiveness and the robustness of our proposed approach of face classification, so, it consists to adopt the Deep Learning as a technique of feature extraction and classification based on its representative models, which are DBN, SAE and BPNN.

After the preprocessing step, including the normalization of our data, the Deep Learning models, which are, DBN, SAE, and BPNN, are applied on two facial databases, BOSS and MIT, in order to extract the high level features. Thus, the different feature vectors that result from the different operations of extracting the deep features will be used as inputs of the NN classifier as presented in Fig. 6, 7 and 8, which describe the adopted procedure in detailed way.

Oftentimes to assess the performance, there are diverse particular elements that show the system classification performance. For instance, the accuracy (Acc) and the error (Er). The equations of these metrics are defined by (6) and (7) as follows:

$$\text{Acc} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (6)$$

$$\text{Er} = \frac{\text{FP} + \text{FN}}{\text{TP} + \text{FP} + \text{TN} + \text{FN}} \quad (7)$$

With:

TP: True positive.

TN: True negative.

FP: False positive.

FN: False negative.

The training holds the key to an accurate solution, so, the criterion to stop training must be more described. In general, it is known that a network with enough weights will always learn the training set better as the number of iteration is increased [33], for this reason, we stopped at 5 epochs when the error stopped decreasing.

The weights of the system are continually adjusted to reduce the difference between the output of the system and the desired response, the difference is referred to as the error and can be measured in a different way. The Training Error (TE) is the error that we get when we run trained model back on the training data in order to update the weights. Then, the MSE is the most common measurement which computes the average squared difference between the desired output and each predicted output. Besides of MSE, there are many kinds of errors, including *Root-Mean-Square Error (RMSE)*, Sum of Squared Errors (SSE), Mean Absolute Percent Error (MAPE), which can be used for measuring the error of Neural Network [33][34].

### IV. RESULT AND ANALYSIS

#### A. Description of BOSS Database

BOSS is a new database of faces and non-faces. Currently, our database is private and can be delivered upon request. Most face images were captured in uncontrolled environments and situations, such as, illumination changes, facial expressions (neutral expression, anger, scream, sad, sleepy, surprised, wink, frontal smile, frontal smile with teeth, open / closed eyes,), head pose variations, contrast, sharpness and occlusion. Besides, the majority of individuals is between 18-20 years old, but some older individuals are also present with distinct appearance, hair style, adorns and wearing a scarf. The database was created to provide more diversity of lighting, age, and ethnicity than currently available landmarked 2D face databases. All images were taken with 26 ZOOM CMOS digital camera of full HD characteristics. The majority of images were frontal, nearly frontal or upright. Fig. 9 shows some people images of BOSS database. We detect people faces in our BOSS database by using the cascade detected of Viola-Jones algorithm. All the faces are scaled to the size 30\*30 pixels. This database contains 9619 with 2431 training images (with 771 faces and 1660 non-faces) and 7188 test images (178 faces and 7010 non-faces). The face images are stored in JPG format. Fig. 10 presents some detected face images of BOSS database.

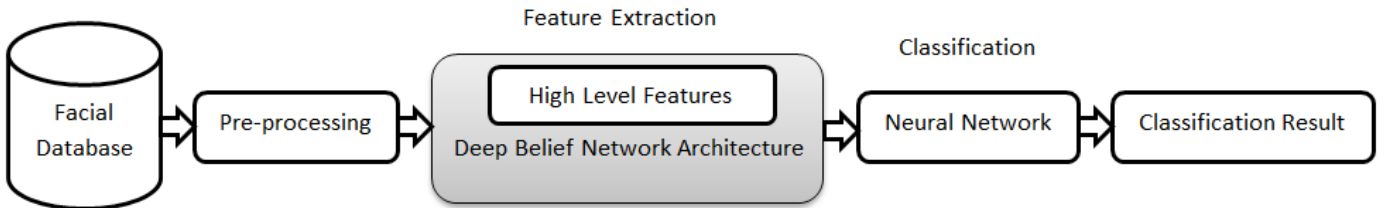


Fig. 6. The proposed model of face classification system based on DBN and NN.

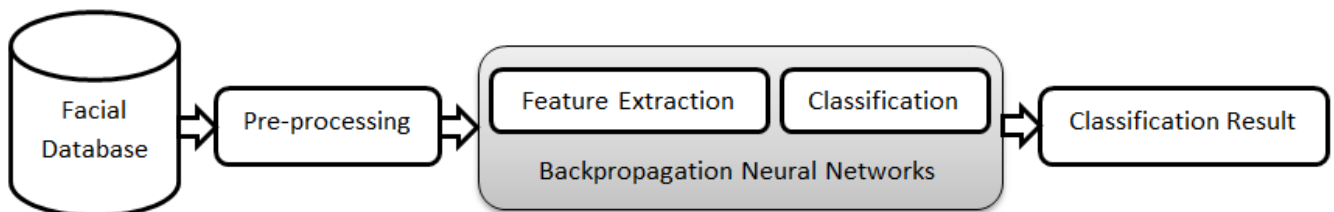


Fig. 7. The proposed model of face classification system based on BPNN.



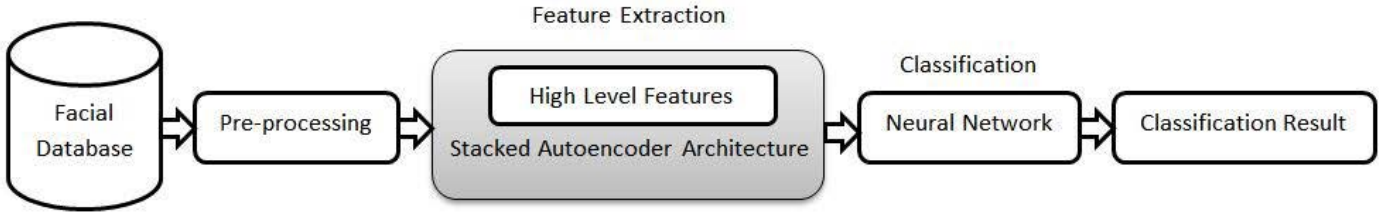


Fig. 8. The proposed model of face classification system based on SAE and NN.



Fig. 9. Some people images of BOSS database.

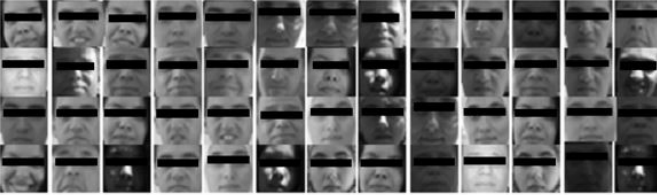


Fig. 10. Some detected face images of BOSS database.

### B. Description of MIT Database

The proposed approach has been tested on the famous MIT database. This database is available publicly [35]. It is a database of faces and non-faces that has been used extensively at the Center for Biological and Computational Learning at MIT. It has 6977 training images (with 2429 faces and 4548 non-faces) and 24045 test images (472 faces and 23573 non-faces). The face images are stored in PGM format. The size of each image is 19\*19 pixels, with 256 level gray scales. An example of some samples of MIT database is presented in Fig. 11.



Fig. 11. Some images of MIT facial database.

### C. Results and Discussion

Among the preprocessing of our data is the normalization of training and testing sets, then, we standardize all values, and Randomize data within batches to ensure class stratification, and Data augmentations.

We train typically several epochs, but we choose only 5 epochs when the validation error is no longer decreasing. Therefore, this problem is dependent of the number of data samples, and the batch size.

We have carried out several experiments on both databases. BOSS and MIT to evaluate the impact of the high level features on the face classification and to prove the effectiveness of our proposed approach.

Fig. 12 and 13 expose the performance of the Deep Learning models based on MSE and TE of both databases, BOSS and MIT respectively. We can observe that the performance of the three methods including (SAE,NN), (DBN,NN) and BPNN during 5 epochs is slightly different. Although, for BOSS, we can find that the BPNN outperforms the combination of (SAE,NN) and (DBN,NN) during the 5 epochs, in the same time, the methods of (SAE,NN) and (DBN,NN) depict a more competitive performance. In addition, as regards MIT database, we notice that the combination of (DBN,NN) performs better than the methods (SAE,NN) and BPNN during the 5 iterations in terms of MSE and TE.

### D. Comparison of MIT and BOSS Facial Databases

In this section, we made a comparative study between two facial databases. BOSS and MIT, based on the three Deep Learning models, which are DBN, SAE and BPNN. Fig. 14, 15 and 16 illustrate the comparison between BOSS and MIT in terms of their robustness with respect to the previous Deep Learning models. The exam of the Fig. 14, 15 and 16 reveals that the results obtained as regards BPNN, DBN and SAE respectively in terms of MSE and TE, are always better during the 5 iterations by using the BOSS database, which more demonstrates the robustness of BOSS compared to MIT.

To better demonstrate the performance of the three Deep Learning models on BOSS and MIT, we used the classification error rate of the previous models as an evaluation criterion. The Table IV presents the performance of BPNN, (DBN,NN) and (SAE,NN) applied on the BOSS database to discriminate faces and non-faces. In this table we can find that (DBN,NN) with two hidden layers outperforms BPNN and (SAE,NN), getting 1.14% in terms of the classification error rate.

The results on MIT are reported in Table V, which illustrates the performance of BPNN, (DBN,NN) and (SAE,NN). Therefore, the superiority of (DBN,NN) is clearly demonstrated compared to BPNN and (SAE,NN) obtaining 1.96% in terms of error rate.

In brief, the combination of (DBN, NN) achieves always the best result in terms of classification error rate based only on two hidden layers.



TABLE IV. PERFORMANCE OF THE DEEP LEARNING MODELS APPLIED ON BOSS DATABASE

Methods	Error	Number of neurons per hidden layer	Number of hidden layer
BBNN	1.49	100	1
<b>(DBN,NN)</b>	<b>1.14</b>	<b>100</b>	<b>2</b>
(SAE,NN)	2.49	100	1

TABLE V. PERFORMANCE OF THE DEEP LEARNING MODELS APPLIED ON MIT DATABASE

Methods	Error	Number of neurons per hidden layer	Number of hidden layer
BBNN	2.49	100	1
<b>(DBN,NN)</b>	<b>1.96</b>	<b>100</b>	<b>2</b>
(SAE,NN)	1.96	200	1

### E. Comparison with the State-of-the-Art Methods

This section provides a comparative study of our approach and multiple well-known techniques recently published regarding face classification by using both facial databases, BOSS and MIT. Referring to Table VI, the result clearly shows that the classification accuracy is good for the majority of algorithms. For fair comparisons, it is clear that our contributions of (DBN,NN) produced the best classification accuracy compared to other published methods which we get 98.86% and 98.04% in terms of classification rate on BOSS and MIT respectively. According to this result, our proposed approach yields a significant improvement over the state-of-the-art methods.

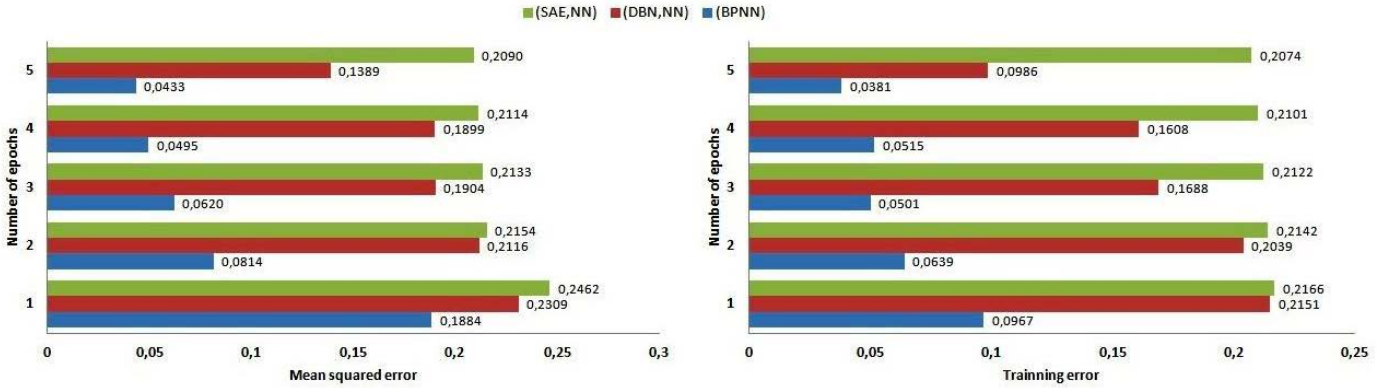


Fig. 12. MSE and TE of the Deep Learning models on BOSS database.

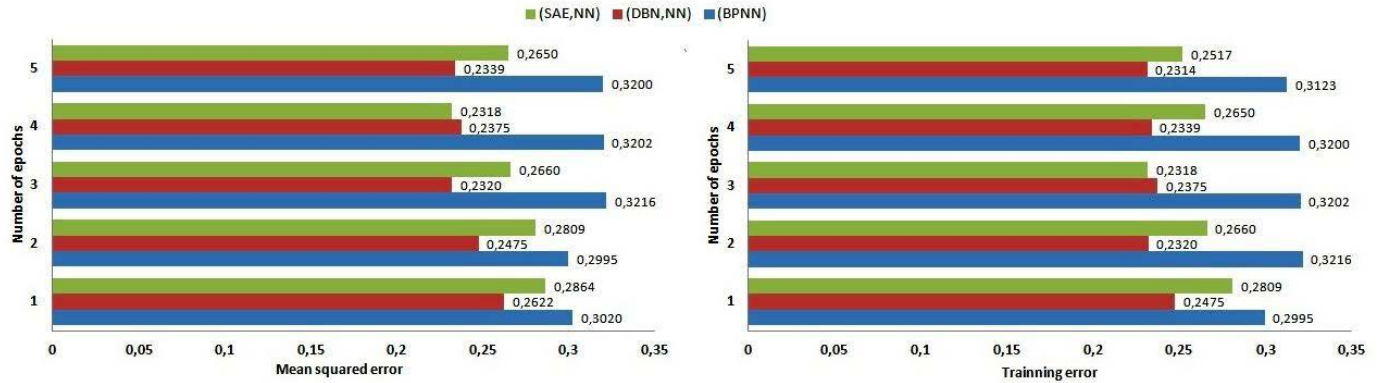


Fig. 13. MSE and TE of the Deep Learning models on MIT database.

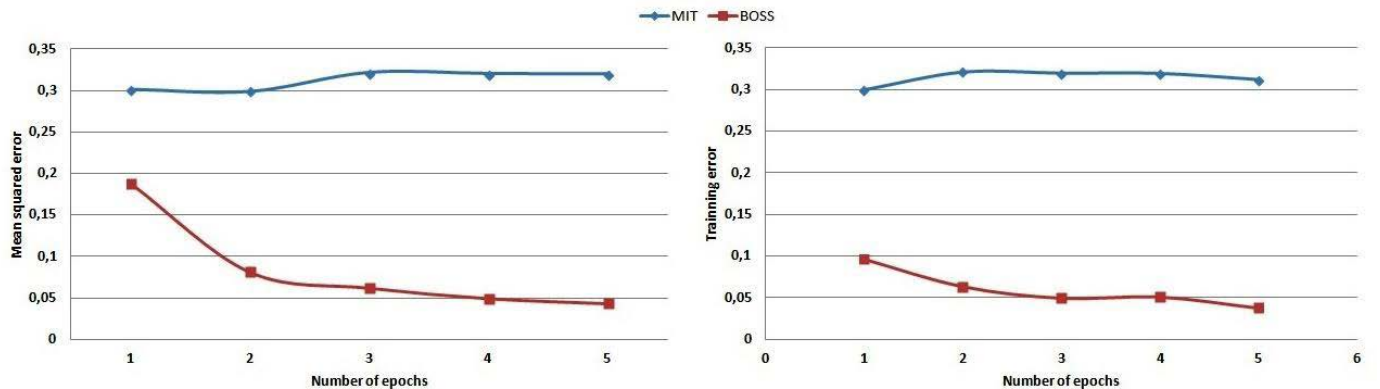


Fig. 14. Comparison of BOSS and MIT databases based on the performance of BPNN.

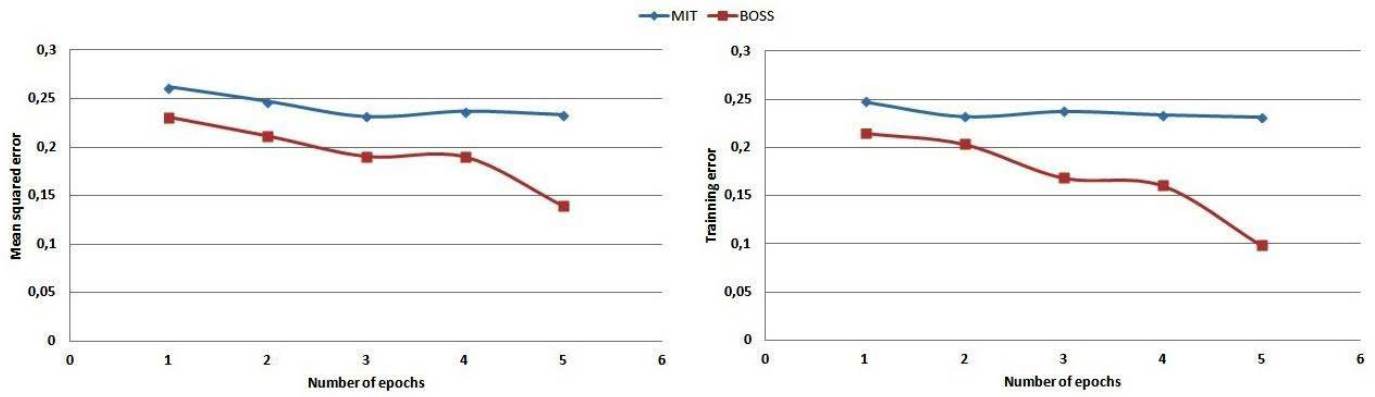


Fig. 15. Comparison of BOSS and MIT databases based on the performance of DBN combined with NN.

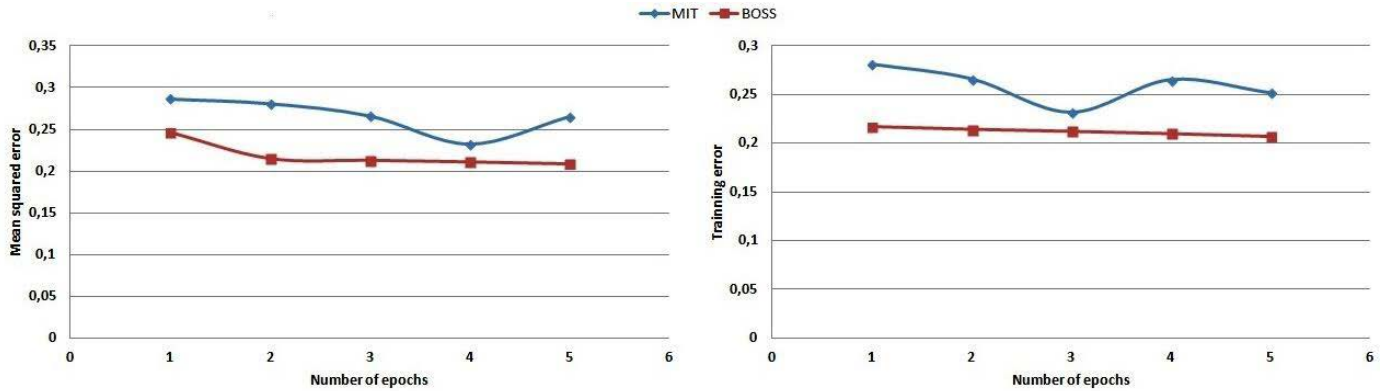


Fig. 16. Comparison of BOSS and MIT databases concerning the performance of SAE combined with NN.

TABLE VI. THE PERFORMANCE OF THE OTHER METHODS IN LITERATURE USING MIT AND BOSS DATABASE

Methods	Accuracy (%)	Research group	Database
DWT	95.5	[26]	MIT
LDCT+DWT	95.5	[26]	MIT
GDCT+DWT	95.5	[26]	MIT
GDCT	96.04	[26]	MIT
LDCT	96	[26]	MIT
Autoencoder+SVM	98.04	[27]	MIT
PCA+SVM	98.04	[27]	MIT
Autoencoder+SVM	97.52	[27]	BOSS
PCA+SVM	96.98	[27]	BOSS
<b>(DBN,NN)</b>	<b>98.04</b>	<b>Our approach</b>	<b>MIT</b>
<b>(DBN,NN)</b>	<b>98.86</b>	<b>Our approach</b>	<b>BOSS</b>

## V. CONCLUSION

In this paper, we investigate the Deep Learning models based on the face classification task. We have progressively improved its performance by investigating three different kinds of Deep Learning models, which are DBN, SAE and BPNN using sigmoid activation function. The Deep Learning algorithms are used to extract high-level features in order to use them as inputs of NN classifier for the purpose of discriminating faces and non-faces. We perform an experimental evaluation on two facial databases, BOSS and MIT. The results indicate the robustness and the efficiency of our proposed approach (DBN,NN) for face classification, obtaining 1.14% and 1.96% in terms of error rate with BOSS and MIT respectively.

## ACKNOWLEDGMENT

This research work is supported by the SAFEROAD project under contract No: 24/2017.

## REFERENCES

- [1] S. Zafeiriou, C. Zhang, and Z. Zhang, "A survey on face detection in the wild: past, present and future," *Computer Vision and Image Understanding*, vol. 138, pp. 1-24, 2015.
- [2] J. Li, Z. Chen and C. Liu, "Low-Resolution Face Recognition of Multi-Scale Blocking CS-LBP and Weighted PCA," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 30, no. 8, 2016.
- [3] T. Ojala, M. Pietikainen, T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971-987, 2002.
- [4] A. Amine, A. el Akadi, M. Rziza, D. Aboutajdine, "GA-SVM and Mutual Information based Frequency Feature Selection for Face Recognition," *infocomp Journal of Computer Science*, vol. 8, no. 1, pp. 20-29, 2009.
- [5] S. Sunitha, S. Katamaneni, P. M. Latha, "Efficient Gender Classification Using DCT and DWT," *International Journal of Innovations in Engineering and Technology*, vol. 7, no. 1, 2016.
- [6] D. K. Srivastava, L. Bhambu, "Data Classification Using Support Vector Machine," *Journal of Theoretical and Applied Information Technology*, vol. 12, no. 1, 2009.
- [7] S. Zhan, Q-Q. Tao and X-H. Li, "Face Detection using Representation Learning," *Neurocomputing*, vol. 187, pp. 19-26, 2016.
- [8] Y. Sun, X. Wang, X. Tang, "Hybrid deep learning for face verification," in: *Proceedings of the IEEE International Conference on Computer Vision*, 2013.
- [9] Y. Chen, T. Huang, H. Liu, D. Zhan, "Multi-pose face ensemble classification aided by Gabor features and deep belief nets," *Optik International Journal for Light and Electron Optics*, vol. 127, no. 2, pp.

- 946-954, 2016.
- [10] R. Salakhutdinov, G. Hinton, "Deep Boltzman machines," in: Proceedings of the International Conference on Artificial Intelligence and Statistics, pp. 448-455, 2009.
  - [11] G.B. Huang, H. Lee, E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," in: Proceedings of the IEEE International Computer Vision and Pattern Recognition, pp. 2518-2525, 2012.
  - [12] Y. Taigman, M. Yang, M.A. Ranzato, L. Wolf, "DeepFace: closing the gap to human-level performance in face verification," in: Proceedings of the IEEE International Computer Vision and Pattern Recognition, 2014.
  - [13] E. Zhou, Z. Cao, and Q. Yin, "Naïve-deep face recognition: touching the limit of LFW benchmark or not," <http://arxiv.org/abs/1501.04690>, 2015.
  - [14] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, M.S. Lew, "Deep learning for visual understanding: a review," *Neurocomputing*, vol. 184, pp. 27-84, 2016.
  - [15] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504-507, 2006.
  - [16] Y. Bengio, A. Courville, P. Vincent, "Representation Learning: A Review and New Perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (special issue Learning Deep Architectures), vol. 35, no. 8, pp. 1798-1828, 2013.
  - [17] H. Yeremia, N. A. Yuwono, P. Raymond, W. Budiharto, "Genetic Algorithm and Neural Network For Optical Character Recognition," *Journal of Computer Science*, vol. 9, no. 11, 2013.
  - [18] H. S. Bhatt, S. Bharadwaj, R. Singh, M. Vatsa, "Recognizing Surgically Altered Face Images Using Multiobjective Evolutionary Algorithm," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 89-100, 2013.
  - [19] R. Rojas, "Neural networks: a systematic introduction," Chapter 7, Springer, 1996.
  - [20] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural computation*, vol. 18, no. 7, pp. 1527-1554, 2006.
  - [21] I. Arel, D. C. Rose, and T. P. Karnowski, "Deep machine learning-a new frontier in artificial intelligence research [research frontier]," *IEEE Computational Intelligence Magazine*, vol. 5, no. 4, pp. 13-18, 2010.
  - [22] D.E. Rumelhart, G.E. Hinton, R.J. Williams, "Learning representations by backpropagating errors," *Nature*, vol. 323, pp. 533-536, 1986.
  - [23] V.F. Rodriguez-Galiano, M. Sanchez-Castillo, M. Chica-Olmo, M. Chica-Rivas, "Machine learning predictive models for mineral prospectivity: An evaluation of neural networks, random forest, regression trees and support vector machines," *Ore Geology Reviews*, vol. 71, pp. 804-818, 2015.
  - [24] Abedi, M., Norouzi, G.H., Bahroudi, A. "Support vector machine for multi-classification of mineral prospectivity areas," *Computers and Geosciences*, vol. 46, pp. 272-283, 2012.
  - [25] P. Baldi, K. Hornik, "Neural networks and principal component analysis: learning from examples, without local minima," *Neural Network.*, vol. 2, pp. 53-58, 1989.
  - [26] B. Nassih, A. Amine, M. Ngadi, S. Tayb, N. Hmina, "A New Face Decomposition for Face Detection using SVM Classifier," *International Journal of Imaging and Robotics*, vol. 17, no. 3, 2017.
  - [27] B. Nassih, A. Amine, M. Ngadi, N. Hmina, "Non-Linear Dimensional Reduction Based Deep Learning Algorithm for Face Classification," *International Journal of Tomography and Simulation*, vol. 30, no. 4, 2017.
  - [28] B. Shah, B. Trivedi, "Optimizing Back Propagation Parameters For Anomaly Detection," *IEEE - International Conference on Research and Development Prospectus on Engineering and Technology*, 2013.
  - [29] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359-366, 1989.
  - [30] Y. Liu, W. Jing, and L. Xu, "Parallelizing Backpropagation Neural Network Using MapReduce and Cascading Model," *Computational Intelligence and Neuroscience*, 2016.
  - [31] Y. Hbali, L. Ballihi, M. Sadgal, A. El Fazziki, "Face Detection for Augmented Reality Application Using Boosting-based Techniques," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 2, pp. 22-28, 2016.
  - [32] M. Magdin, M. Turčani, and L. Hudec, "Evaluating the Emotional State of a User Using a Webcam," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 1, pp. 61-68, 2016.
  - [33] S. Haykin, "Neural Networks: A Comprehensive Foundation," New York: Macmillan, 1994.
  - [34] I. A. Basheer, M. Hajmeer, "Artificial neural networks: fundamentals, computing design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3-31, 2000.
  - [35] Center for Biological and Computational Learning at MIT and MIT. CBCL Face Data. <http://cbcl.mit.edu/cbcl/softwaredatasets/FaceData2.html>
  - [36] M. A. Mansor, M. S. M. Kasihmuddin, S. Sathasivam, "Enhanced Hopfield network for pattern satisfiability optimization," *International Journal of Intelligent Systems and Applications*, vol. 8, no. 11, pp. 27-33, 2016.



Nassih Bouchra



Amine Aouatif



Ngadi Mohammed



Hmina Nabil

# A Novel Approach on Visual Question Answering by Parameter Prediction using Faster Region Based Convolutional Neural Network

Sudan Jha<sup>1</sup>, Anirban Dey<sup>1</sup>, Raghvendra Kumar<sup>2</sup>, Vijender Kumar-Solanki<sup>3\*</sup>

<sup>1</sup> Kalinga Institute of Industrial Technology, Bhubaneswar (India)

<sup>2</sup> LNCT College, Jabalpur (India)

<sup>3</sup> CMR Institute of Technology, (Autonomous), Hyderabad, TS (India)

Received 16 April 2018 | Accepted 31 July 2018 | Published 31 August 2018

**unir**  
LA UNIVERSIDAD  
EN INTERNET

## ABSTRACT

Visual Question Answering (VQA) is a stimulating process in the field of Natural Language Processing (NLP) and Computer Vision (CV). In this process machine can find an answer to a natural language question which is related to an image. Question can be open-ended or multiple choice. Datasets of VQA contain mainly three components; questions, images and answers. Researchers overcome the VQA problem with deep learning based architecture that jointly combines both of two networks i.e. Convolution Neural Network (CNN) for visual (image) representation and Recurrent Neural Network (RNN) with Long Short Time Memory (LSTM) for textual (question) representation and trained the combined network end to end to generate the answer. Those models are able to answer the common and simple questions that are directly related to the image's content. But different types of questions need different level of understanding to produce correct answers. To solve this problem, we use faster Region based-CNN (R-CNN) for extracting image features with an extra fully connected layer whose weights are dynamically obtained by LSTMs cell according to the question. We claim in this paper that a single R-CNN architecture can solve the problems related to VQA by modifying weights in the parameter prediction layer. Authors trained the network end to end by Stochastic Gradient Descent (SGD) using pre-trained faster R-CNN and LSTM and tested it on benchmark datasets of VQA.

## KEYWORDS

Natural Language Processing, Computer Vision, Convolutional Neural Network, Recurrent Neural Network, Long Short Term Memory, Faster Region Based Convolutional Neural Network, Stochastic Gradient Descent.

DOI: 10.9781/ijimai.2018.08.004

## I. INTRODUCTION

**U**NDERSTANDING an image by the help of computer vision or image processing technique is a complex procedure studied in the two last eras. Then the scientists introduced compatible circumstances between Image Processing and Natural Language Processing (NLP) to solve the problem of image understanding. Traditionally the researchers all over the world applied the process of image understanding to solve the problem of Visual Question Answering by the machine learning. Recently, deep learning architectures constructed by knowledge artificial neural networks have enhanced visual image understanding [1, 2, 3]. Object recognition from an image is done by Convolutional Neural Network (CNN). Recurrent Neural Network (RNN) with Long Short Term Memory (LSTM) cell has outstretched the bar on sequence prediction jobs as well as machine translation [18, 19]. CNN performs feature representation of the image and LSTMs process the representation of question and answer. The researchers directly combined both networks and trained end to end to generate the answer [20, 21]. But this kind of approach is able to answer the common and simple questions that are related to the image's content i.e. 'What is the

shape ...?' or 'How many?'. To find a correct answer for the different types of questions, understanding of an image should be different [4, 5].

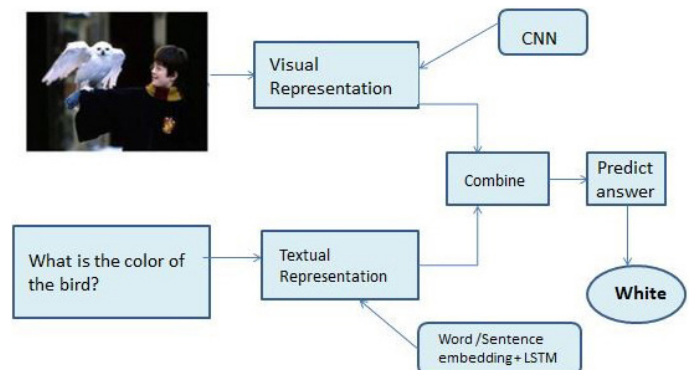


Fig. 1. Visual Question Answering: Common Approach.

Fig.1 indicates the steps of VQA approach at a glance. First image and question need the visual and textual representations. CNN provides the visual representation (feature vector of image) and Word or sentence embedding followed by RNN +LSTM provides textual representation (feature vector of question). Then both features are combined together for predicting the answer [6, 7, 8].

\* Corresponding author.

E-mail address: spesinfo@yahoo.com



There are various kinds of questions and answers which are used to train the VQA system. Questions can be mainly of these types: fine grained recognition (e.g. “What kind of animal is in the ground?”), Object detection (e.g. “How many birds are there?”), activity recognition (e.g. is this boy smiling?”), knowledge based reasoning (e.g. “which animal is the reptile in this picture?”), commonsense reasoning (e.g. “Is the boy expecting a job?”). Answer format can be of two categories: yes/no (binary) or multiple-choice. For open-ended, formless questions, answer has not any particular form it can be one word or several words. Images can be mostly two types: Real (real world images like photograph of nature, human, object, etc.) and abstract (clip-arts, animated scene). That is why one type of object recognition for the image is not sufficient for visual question answering problem; selection of the appropriate recognition task is also significant [9, 10].

Computer vision and NLP community released different types of approaches among those techniques, actually the main methodology is to combine the image and question features to create deep learning architecture for classifying jointly combined features to produce a correct answer. Some approaches use CNN for extracting image features and bag of words (BOW), for obtaining question features [11, 12].

VQA is a complete AI finish task and it is multi-disciplinary and heterogeneous in nature. The visual/ image question answering model can solve the complex real life problems (discussed in future work). Another point is that it tackled different sub problems in a single task i.e. image processing, question answering in NLP and sometimes knowledge representation. Motivation from this research problem is a variety of standpoints like new datasets and techniques in computer vision and image processing, integration of vision, language and common sense and merging group of methods of machine learning and deep learning. It has the ability to train the machine how to learn and answer the question like human being and also learn about the pictorial world [13]. The VQA system can be able to learn what do need to learn and what do not need. VQA system is relevant to various types of social applications that include situational information from visual content, making decisions from huge amount of investigation data and, last but not the least, interaction with robots. This research has the purpose to improve the lives of visually-impaired people and transfigure the society through interaction between human beings and visual information [14].

VQA needs various types of understanding of an image not only caption generation or image entity recognition but visual scene understanding and reasoning of knowledge about to the image. The approaches that we discuss in related work took the VQA problem as a classification task, in addition to some network architectures designed to solve the above mentioned problem in joint embedding manner. To overcome the previous problem, we use Faster Region based CNN (R-CNN) for feature extraction of image and updating weights of that CNN according to the question. For weights (parameter) updating we use different types of functions into the weights matrix of the main network. Actually we are trying to generate the answer of free-form and open ended question from the image.

In the section II, we discuss the recent existing approaches. It divided the approaches into four different types; without deep learning based approaches [1, 2], deep learning based approaches [3, 5, 6, 15, 21], deep learning based approaches with attention [15, 16], other different approaches [17].

In the section III we discuss the basic concepts behind the deep learning tools CNN, RNN, LSTM and Word embedding which are applied to build model to solve VQA problem.

In the section IV & V we describe our proposed model and

methodologies to solve the problems over VQA in field of CV and NLP. Our key contributions to this research are that we describe a different methodology that depends on the question part. We have applied faster R-CNN (Region based CNN) [7] with a prediction of parameter layer where the weights are updated according to the question. We apply the word embedding method [12] to question and build a parameter prediction network with a series of LSTM cells. In the section VI we conclude proposed work in details. It uses pre-trained faster R-CNN [7] and LSTM [9] to initialize the weights of entire network by using Stochastic Gradient Descent [22] for training the whole network.

## II. RELATED WORK

Malinowski et al. [2] proposed a model that can answer the probability of a given question according to an image. This model gives a single word answer. Authors proposed another model which produces multi word answer. This model is trained and tested on DAQUAR dataset.

Kaffle et. al. [1] proposed a model that is based on Bayes’ rule of probability. They trained on COCO-QA data set that contains four types of answers such as counting, object, location; color. The model evaluates the probability of answer dependent on answer type.

Malinowski et al. [3] proposed a method called Neural image QA where they jointly fed image and question features into LSTM encoder-decoder to generate answer. Image features are extracted by CNN and question embedding is obtained from LSTM. The main limitation of the model is that it can be applied on large dataset only. They also propose two metrics i.e. “Average Consensus” which defines human disagreement and “Min Consensus” which defines the disagreement in question answering of a human.

Ren et al. [6] proposed a method that uses RNN and visual semantic embedding in the middle phases of image segmentation and object detection. Their main contribution is the question generation algorithm that transforms image descriptive dataset into question - answer format. They treated the VQA problem as classification rather than answer generation.

Gao et al. [21] proposed a little different procedure that uses LSTMs to translate the question and create the answer by two diverse forms. First LSTM is shared weight mechanism into LSTM encoder –decoder architecture that focuses on grammatical structure of the sentence. Second LSTM extracts features from image by CNN that are not directly served into encoder at each time step.

Noh et al. [5] derived a new approach named “DPPnet” of VQA which learns by CNN with dynamic parameter prediction layer. Here weight matrix of the CNN is dynamically predicted by a parameter predictive network. Parameters were predicted by Gated Recurrent Unit (GRU), which takes the asking, question as an input and produce candidate weights as an output. They used hashing technique for the arrangement of final weight matrix of CNN. This method improved the accuracy on the benchmark data set.

Lin et al. [14] did not utilize any RNN for encoding question. They utilize 3 kinds of CNN for answering the question from an image: CNN for image, CNN for sentence and extra multimodal CNN. In the sentence CNN they utilize 3 layers of convolution and max pooling for creating the sentence portrayal. Catching the relations amongst feature vector of image and question are ended by multimodal convolution step. They added two extra fully connected layer; one for multimodal convolution i.e. they embedded one image feature in between two consecutive semantic components of question side and other one is softmax classification layer to predict answer.

Shih et al. [15] proposed an attention-based model henceforth referred to as Where To Look (WTL). The authors use VGG net to

extract the image features. They average the each word vector in the question to obtain the question vector.

Then they compute attention vector over the set of the image features to choose which region to be focused on. The final image representation is the summation of attention weighted regions of the image. After that they concatenated between question features and summed result and fed through full connected softmax layer to predict answer.

Yang et al. [16] proposed a model called Stacked Attention Networks (SAN) encode question using either CNN or LSTM. Then the question encoding is fed to attend over image. Then question encoding and the attention weight are concatenated to calculate attention over the unique image. This model increases the accuracy of VQA system for use of attention.

Andreas et al. [17] proposed a new approach called Neural Memory Network (NMN) that is based on semantic parsing of the question. Parse tree is converted to module to answer the question. Modules are composable and independent. The question parsing is performed to identify the grammatical relation between parts of the sentence. They use ad-hoc handwritten rules to convert the parse tree in structured queries. The main problem in this model is parsing of a question due to fixed network structure. Error cannot be recoverable.

### III. PRELIMINARIES

#### A. Convolutional Neural Network (CNN)

CNN is used to extract the feature vector from an image. There are two phases in CNN: feature extraction and output prediction. There are two layers in feature extraction: convolution layer and sub sampling (pooling) layer. And one fully connected layer for classification or output prediction. After convolution layer the obtained feature goes through an activation function. In the convolutional layer there are series of matrix multiplications followed by summation operation.

#### B. Recurrent Neural Network (RNN)

Recurrent Neural Network [4, 24, 25] is used for learning the sequence of data like series of video frame, text, music etc. The main difference of RNN over feed forward ANN is that we add another weight matrix, that matrix comes from previous hidden state. We just give input of the hidden state in every time step, and keep repeating that. Main problem with the RNN is vanishing gradient problem i.e. whenever we do back propagating in the neural network gradient (product of partial derivative of error) tends to vanish. The mathematical formulation RNN is the following:

$$(O)^t = f(h^t; \theta_o) \quad (1)$$

$$h^t = f(h^{t-1}; x^t; \theta_h) \quad (2)$$

Where  $o^t$  is the output of the RNN and  $x^t$  is the input at time  $t$ ,  $h^t$  is the state of the hidden layer(s) at time  $t$ .  $\theta$  is encapsulated parameter of weight and bias. Fig. 2 defines a simple graphical model to explain the relation between these three variables in an RNN computation graphs.

In Fig. 2, the values  $\theta_i$ ,  $\theta_h$ ,  $\theta_o$  represent the parameters associated with the inputs, previous hidden layer states, and outputs, respectively.

#### C. Long Short Term Memory (LSTM)

LSTM [9] resolves the problem of vanishing gradient and getting successes in natural language processing applications like machine translation, speech recognition etc. LSTM cell consists mainly three gates i.e. input gate, output gate and forget gate and a cell state. Actually LSTM says what is the relevant part of that a network has learned and what to forget.

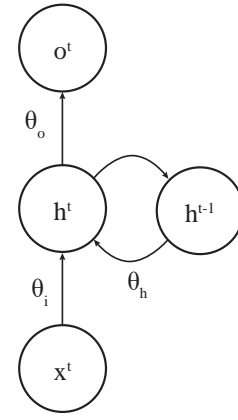


Fig. 2. A graphical model for an RNN.

#### D. Long Short Term Memory (LSTM)

LSTM [9] resolves the problem of vanishing gradient and getting successes in natural language processing applications like machine translation, speech recognition etc. LSTM cell consists mainly three gates i.e. input gate, output gate and forget gate and a cell state. Actually LSTM says what is the relevant part of that a network has learned and what to forget.

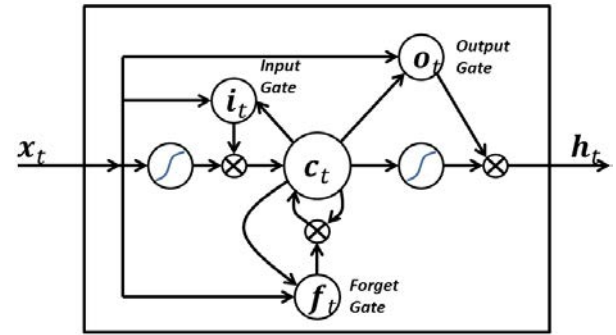


Fig. 3. State diagram of LSTM cell.

In Fig. 3,  $f_t$ ,  $i_t$ ,  $o_t$  are variables; forget gate, input gate and output gate, respectively.  $W_f$ ,  $W_i$ ,  $W_o$  are the corresponding weight vectors and  $b_f$ ,  $b_i$ ,  $b_o$  are the bias.  $C_{t-1}$ ,  $h_{t-1}$  are values of previous cell and previous hidden state respectively at time  $t-1$ .  $c_t$  is the value of cell state at time  $t$ ;  $h_t$  is the output vector of LSTM cell at time  $t$ .

Cell state is long-term memory, it represents all the learning at over time, and hidden state (is like a current memory). Forget gate is called remember vector that learns what to forget and what to remember. Input gate is also called the save vector that determines number of inputs getting into the cell state, and output gate is also called focus vector that is akin to an attention mechanism (what part of data should be focused on). Actually these gates of LSTM are perception i.e. single layer neural network. So LSTM functionality is mainly based on forgetting, remembering and paying attention of data. The mathematical formulation LSTM is the following:

$$f_t = \sigma(W_f \cdot [C_{t-1}, h_{t-1}, x_t] + b_f) \quad (3)$$

$$i_t = \sigma(W_i \cdot [C_{t-1}, h_{t-1}, x_t] + b_i) \quad (4)$$

$$C_t = \text{Tanh}(W_c \cdot [h_{t-1}, x_t] + b_c) \text{ "Memory cell state"} \quad (5)$$

$$c_t = f_t \cdot [C_{t-1}] + [i_t] \cdot C_t \quad (6)$$

$$o_t = \sigma(W_o \cdot [C_{t-1}, h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t \cdot [\text{Tanh}(C_t)] \quad (8)$$

In the equation (3), the output of the forget gate ( $f_t$ ) denotes the cell state that what to forget by multiplying 0 to the particular position in the LSTM matrix. The information is remembered if  $f_t$  is equal to 1. Here sigmoid activation function  $\sigma$  is applied to the weighted input and previous hidden state.

In the equation (4, 5), the output of input gate ( $i_t$ ) is a sigmoid function ranged of [0, 1]. That is why sigmoid function is not capable to forget the information of the cell state. The output of input modulation gate ( $C_t$ ) is  $\tanh$  activation function ranged of [-1, 1]. It permits the cell state to forget the information.

Equation (6) is called cell state equation. The previous cell state  $C_{t-1}$  forgets information multiplying by output of forget gate  $f_t$  and adds new information through the output of input gate ( $i_t$ ).

Equation (7) is called output gate equation. The output of the output gate equation ( $O_t$ ) defines all possible values from LSTM matrix which must be moving forward to the next hidden state.

In the equation (8),  $h_t$  is output of hidden state equation that defines what information we should take for next sequence.

### E. Word Embedding

Words can be represented as a vector of real valued numbers. In the vector representations words with similar vectors should be semantically the same, that is, they have to represent related concepts. Sometime for practical purposes these vector can have the dimension of 10 or 100 as compared to the vocabulary size which is at least dimension 10 or 1000. Word2vec, GloVe, Skip thought [12] are some methods of word embedding.

## IV. PROPOSED WORK

In this section we discuss each tool that we use in the model and how they work together for solving this heterogeneous problem.

### A. Image Features

In our method, we use faster R-CNN pretrained model VGG-16 [7] (16 layer CNN in faster R-CNN framework).

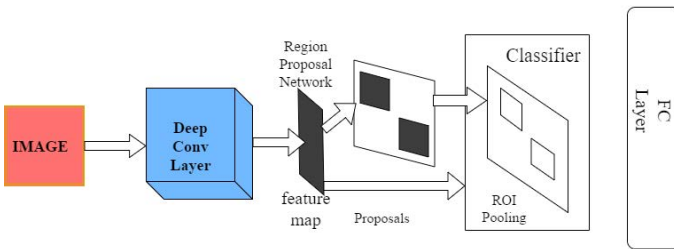


Fig. 4. Faster R-CNN block architecture.

The input image is fed through the Faster R-CNN [7] to get a vector representation. The R-CNN network is trained on Visual Genome [13] dataset for the focusing element of images with the help of image annotation. We vary the value of K (number of image regions) according to the variety of each image. After the convolution layer of faster R-CNN, there is a region proposal network (Fig. 4 and 5) in which some part of the feature pixel matrix of input image is called proposal. Proposal is divided in some regions and selection of maximum feature value from each region is called ROI (Region of Interest) pooling. After the ROI pooling the required image feature vector is generated. We remove the last layer which is fully connected and attached three fully connected layers. The second last fully-connected layer is the parameter prediction layer whose weights are updated according to the question. The dimension of final output vector is equal to the number of possible answers. In the final layer (fully connected) softmax classifier

[23] is applied to the output vector for calculating the probabilities of each answer. Output of the classification network is denoted by  $O = [o_1, \dots, o_n]^T$  is

$$O = W_r(Q) \times I + b^r \quad (9)$$

Where  $I$  is the input vector of parameter prediction layer  $I = [i_1, \dots, i_n]^T$ ,  $W_r(Q) \in R^{M \times N}$  represents the matrix constructed by LSTM network given to the question  $Q$ .  $b$  is bias.

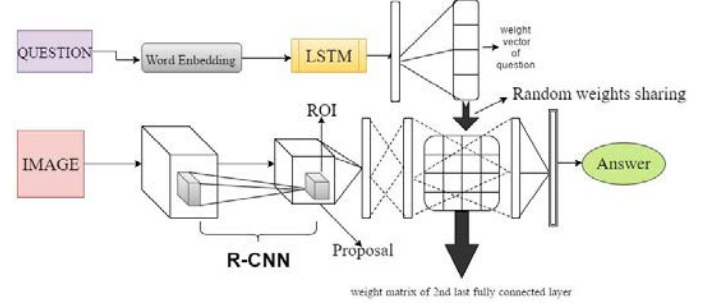


Fig. 5. Network architecture for VQA system.

### B. Question Features

First we tokenize the question sentence means splitting the words with space and punctuation and any number is considered as words also. Questions are clipped into 14 words sentence and extra words are thrown out. Then we use pretrained GloVe word embedding (Global Vector for Word Representation). Each word and their corresponding vector have dimension of 300. We also use padding scheme (vectors of zeros) for the question which length is shorter than 14 words. After word embedding we fed the output sequence into LSTM. Output of the LSTM ( $Q_i$ ) is multiplied by weight vector ( $W_i$ ) of last fully connected layer of LSTM for getting the candidate weight vector of LSTM network. Candidate weight vector denoted by  $L = [l_1, \dots, l_n]$  is

$$L = W_i \cdot Q_i \quad (10)$$

Where  $W_i$  is the weight matrix of a fully connected layer of LSTM network and  $Q_i$  is final output vector of question.

### C. Parameter Prediction

We update the weight (parameter) matrix of 2<sup>nd</sup> last layer of R-CNN on the basis of the candidate vector of LSTM network. Here we have used parameter reduction techniques to optimize the network. Here the weight matrix of 2<sup>nd</sup> last fully connected layer of faster R-CNN is denoted  $W_r(Q) = [w_{11}^r, \dots, w_{kl}^r]$ .

Now,  $w_{kl}^r$  is a weight of the weight matrix  $W_r(Q)$ , it is the corresponding weight between  $k^{\text{th}}$  output and  $l^{\text{th}}$  input neuron.  $\mathcal{E}(k, l)$  is a function to map a key  $(k, l)$  to a natural number in  $\{1, \dots, Z\}$ , where  $Z$  is the dimensionality of  $l$ . The final value is given by:

$$W_{k-1}^f = l_{\mathcal{E}(k-1)} \cdot \omega(k, l) \quad (11)$$

Where  $\omega(k, l) \in \mathbb{N} \times \mathbb{N} \rightarrow \{+1, -1\}$  is another function to remove the bias. By these two functions we update the weight matrix  $W_r(Q)$  according to the question  $Q$ .

#### Algorithm: Proposed Algorithm

Input:  $I, Q$  ( $I$  is input image,  $Q$  is input question)

Output:  $A$  ( $A$  is predicted answer of the question  $Q$  about the image  $I$ )

1. Extract the image features by FasterR-CNN, output of feature vector  $V$  size of  $(k \times 2048)$ .  $k$  = number of image location
2. Resize the feature maps  $(14 \times 14)$  by ROI pooling operation into  $(7 \times 7)$ .
3. Embed the question  $Q$  by GloVe word embedding into  $(14 \times 300)$  vector  $Q_e$ .



4.  $Q_e$  is fed to the LSTM which internal state dimension is 512 and number of hidden layer MLP is 3.
5. After question embedding output is final vector obtained from series of LSTM cell  $Q_r$ .
6. Now the weight vector of the LSTM network  $l=W_i \cdot Q_r$ ; where  $W_i$  is the weight matrix of a fully connected layer of LSTM network and  $Q_r$  is final output vector of question.
7. The weight matrix of 2nd fully connected layer of R-CNN is denoted by  $W_r(Q)$  and we update the  $W_r(Q)$  with respect to  $l$  (candidate weight vector obtained from LSTM network) by random weight sharing method. In other words, the weight matrix corresponding to the layer is parameterized by a function of the input question  $Q$ .
8. In the final layer softmax classifier is applied to the output vector for calculating the probabilities of each answer.

#### D. Training

Stochastic gradient is used to train the network. Questions are preprocessed by tokenization method [10] and transformed the upper case letter to lower case before training. We do the same thing for answers. To prevent over fitting we identify the epoch which gives the best performance on VQA dataset. Then the training is repeatedly running for the similar number of epochs. Optimization of weight updating layer of R-CNN changes in each batch. For clip gradient of LSTM networks we identify the range of gradient norm and then scale down the exceeding gradient range. We reuse the hyper parameters to normalize the loss. When training time of LSTM goes long then we stop the training because loss of the training does not improve for several epochs. When LSTM network starts to over fit, we stop tuning LSTM but training on the other parts of the network is going on.

#### E. Training Error by Stochastic Gradient Descent (SGD)

The use of SGD in the neural network is driven by the high cost of running back propagation over the full training set. SGD can overcome this cost and still lead to fast convergence. Back propagation gives you the gradients, but not how to use them. SGD optimizes the gradient and computes the objective function; the standard gradient descent algorithm updates the parameters  $\theta$  of the objective function  $J(\theta)$  as,

$$\theta = \theta - \alpha \nabla \theta E \times J(\theta) \quad (12)$$

The above equation (12) is approximated by evaluating the cost and gradient over the full training set. Stochastic Gradient Descent (SGD) simply does away with the expectation in the update and computes the gradient of the parameters using only a single or a few training examples.

### V. EXPERIMENT

We tested our model in various VQA datasets like COCO-QA, VQA, DAQUAR-all and DAQUAR-reduced [13] and tested before fine-tuning of LSTM and RCNN and after fine-tuning. We also describe the differences between existing methods and our method. And test results in terms of image and question also is displayed in this section.

TABLE I. EMPIRICAL COMPARISON BETWEEN EXISTING MODELS AND OUR MODEL

Models	Methods	Answer Type	Image Features
Neural-Image-QA [3]	Joint embedding	Generation	GoogLeNet
VIS+LSTM [6]	Joint embedding	Classification	VGG-Net
Multimodal-CNN [14]	Joint embedding	Generation	GoogLeNet
LSTM+Attention [20]	Attention method	Classification	VGG-Net
StackAttn network [16]	Attention method	Classification	VGG-Net
iBowing [18]	Joint embedding	Classification	GoogLeNet
Bayesian [1]	Joint embedding	Classification	VGG-Net
NMN [17]	Compositional model	Classification	VGG-Net
Attribute LSTM [19]	Knowledge base	Generation	VGG-Net
<b>Proposed Work</b>	<b>Parameter prediction</b>	<b>Classification</b>	<b>VGG-Net (Faster R-CNN)</b>

Table I indicates the differences between previous approaches and our approach on basis of the method used, answer type and image features (pre-trained CNN model name). Joint embedding means combining both image and question representations together. Attention method means focusing the particular regions of the image or questions rather the whole image or question. Knowledge base signifies facts about the world. It has an inference engine for reasoning about the facts by using rules and logic, and deduces the new facts.

#### A. Datasets

Generally Datasets are prepared of triplets (question, answer, and image). Some number of datasets contains some kind of additional notation. Most of the answers are single words or phrases.

Table II includes the description dataset in terms of number of images, number of questions, number of question types (number questions, color questions, object question and location questions), and number of training and testing questions, question collection and evaluation metrics.

#### B. Evaluation Metrics

In DAQUAR and COCO-QA dataset, there are two types of evaluation metrics; one is classification accuracy and another one is WUPS (Wu-Palmer similarity [6]) based on WordNet [5] categorization to calculate the semantic similarity between words. It signifies the similarity between machines generated answer and human answer.

**Accuracy:** The accuracy of classification is calculated by the following equation:

$$\text{Accuracy} = \frac{\text{Item Classified Correctly}}{\text{All items classified}} \quad (13)$$

**WUPS (Wu-Palmer similarity):** The Wu & Palmer calculates relatedness by considering the depths of the two synsets in the WordNet taxonomies, along with the depth of the LCS (Longest Common Subsequence).

TABLE II. BRIEF DESCRIPTION ABOUT SOME WELL-KNOWN DATASET OF VQA

Dataset	No. of images	No. of questions	No. of question types	Question collection	No. of training questions	No. of testing question	Evaluation metrics
DAQUAR[13]	1449	12468	4	Human	3,876	297	Acc. & WUPS
COCO-QA[13]	117,684	117,684	4	Automatic	78736	38948	Acc. & WUPS
VQA[13]	204,721	614163	20+	Human	248349	244302	Acc. against 10 humans
Visual genome [13]	108,000	1,445,322	7	Human	-	-	Acc.



$$WUPS = \frac{1}{N} \sum_{t=1}^n \text{Min} \left( \prod_{a \in A^t} \text{Max}_{t \in T^t} \mu(a, t), \prod_{t \in T^t} \text{Max}_{a \in A^t} \mu(a, t) \right) \quad (14)$$

Where  $\mu(a, t)$  defines the threshold Wu-Palmer similarity between prediction (a) and ground-truth (t). 0.9 and 0.0 are two threshold values in our evaluation. This means that  $0 \leq \text{score} < 1$ .

### C. Results and Discussion

In this section, results of the proposed model are tested by using four types of datasets i.e. DAQUAR –all, DAQUAR-reduced, COCO-QA, VQA test-dev. It describes the comparative analysis with the existing methods with respect to accuracy and WUPS score.

TABLE III. RESULT ON DAQUAR- ALL DATASET

Models	Accuracy (%)	WUPS@ 0.0	WUPS@ 0.9
Neural QA[3]	19.43	25.25	62.00
3-CNN[14]	23.40	29.59	62.95
Attribute LSTM[19]	24.27	30.41	62.29
DPPNet[5]	28.98	34.80	67.82
Bayesian[1]	28.96	34.74	67.33
<b>Proposed Work</b>	<b>29.92</b>	<b>35.56</b>	<b>72.79</b>

TABLE IV.RESULT ON DAQUAR- REDUCED DATASET

Models	Accuracy (%)	WPUS @ 0.0	WPUS @ 0.9
GUESS	18.24	29.65	77.59
VIS +BOW[6]	34.17	44.99	81.48
VIS+LSTM[6]	34.41	46.05	82.48
Neural QA[3]	34.68	40.76	79.54
3-CNN[14]	39.66	44.86	83.06
Attribute LSTM[19]	40.07	45.43	82.67
DPPNet[5]	44.48	49.56	83.95
Bayesian[1]	45.17	49.74	83.95
<b>Proposed Work</b>	<b>47.28</b>	<b>51.76</b>	<b>88.56</b>

TABLE V. RESULTS ON COCO-QA DATASET

Model name	Accuracy (%)	WUPS @0.0	WUPS @0.9
GUESS	6.65	17.42	73.42
VIS +LSTM[6]	53.31	63.91	88.58
3-CNN[14]	54.95	65.36	88.58
VIS+BOW[6]	55.95	66.78	88.99
Attribute LSTM[19]	61.38	71.15	91.58
DPPNet[5]	61.19	70.84	90.61
Bayesian	63.18	73.14	91.32
<b>Proposed Work</b>	<b>64.71</b>	<b>74.87</b>	<b>92.56</b>

Table III, IV and V show comparison of the results obtained by our proposed model and other existing models “Neural QA, 3-CNN, Attribute LSTM, DPPNet, Bayesian” in terms of both metrics accuracy and WPUS@0.0 and WUPS@0.9 score. The proposed algorithm beats all existing approaches reliably in all benchmarks.

TABLE VI. RESULT ON VQA (TEST\_DEV) DATASET

Models	Open ended questions (Accuracy %)	Multiple choice questions (Accuracy %)
Question	43.09	58.68
Image	28.15	30.53
Q+I	52.64	58.97
LSTM Q	48.76	54.75
LSTM Q+I	53.74	57.17
DPPNet [5]	57.22	62.48
<b>Proposed Work</b>	<b>61.99</b>	<b>67.13</b>

Table VI contains the results of VQA (test-dev) on the open-ended and multiple-choice (M.C.) settings. Here we see that our model performs well in open-ended question as well as in multiple choice questions. The model is tested on only question, only image, question and image both and reported the accuracy.

Now some test images are displayed in Fig. 6 and 7 and shows how the proposed model performs accurately in these different type of questions.



Fig. 6. Result of our proposed model in the terms of Question and Image on COCO-QA and DAQUAR Dataset.

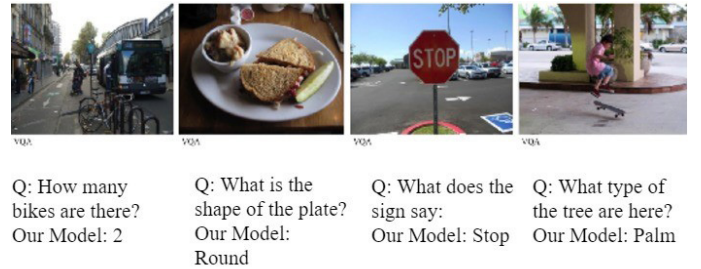


Fig. 7. Result of our proposed model in terms of Question and Image on VQA dataset.

The qualitative results of the proposed model are introduced in Fig. 6. As a rule, the proposed method is effective to deal with different sorts of inquiries that need diverse levels of semantic understanding. It demonstrates that the system is able to recognize the task relying upon questions. On the other hand, in Fig.7, the proposed model is compelling to find the answer of different questions on various images.

Some major advantages of our proposed model are discussed in the following:

- Accuracy of this model is better than existing previous VQA models. Because we use single faster RCNN pre-trained model for obtaining visual also.
- We solved the VQA problem not only as a joint embedding approach but we update the weight (parameters) of the network (RCNN), that is why in the last classification layer the number of probabilities (classes) is less than in the other approaches.
- We generate the answer of free-form and open ended question from the image.

CNN entirely loses all their inner data about the position and the placement of the object and they route all the info to the same neurons that may not be able to deal with this kind of information. A CNN makes estimations by looking at an image and then testing to see if assured modules are present in that image or not. If they are, then it classifies that image accordingly. Fine tuning of hyperparameters of our proposed model is non-trivial and we need a large dataset for training the model. We identify that when we trained our model in DAQUAR reduced dataset loss after several epochs are increased drastically. These are the complexities of using a CNN architecture.

## VI. CONCLUSION

This paper focuses on a deep learning based model for both open ended and multiple choice questions. It describes several types of experiments and the contributions of each design. It also provides the importance of several mechanisms of a VQA model. It shows how Faster R-CNN magnifies the ability of object detection in an image fastly, increasing overall accuracy by 13% more than other image recognition models "Neural QA, 3-CNN, Attribute LSTM, Bayesian" do. The paper describes how our model updates the weight matrix of fully connected layer of faster R-CNN according to candidate weights of questions. It has been shown that VQA problem can be solved by a single R-CNN model. The following further directions should be investigated by using capsule Network for the object detection part of VQA system.

## REFERENCES

- [1] Kafle, K., and Kanan, C. (2016) Answer-type prediction for visual question answering, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 4976–4984.
- [2] Malinowski, M., and Fritz, M. (2014) A multi-world approach to question answering about real-world scenes based on uncertain input, In Advances in Neural Information Processing Systems, 1682–1690.
- [3] Malinowski, M., Rohrbach, M., and Fritz, M. (2015) Ask your neurons: A neural based approach to answering questions about images, In Proceedings of the IEEE international conference on computer vision, 1–9.
- [4] Mikolov, T., Karafi'at, M., Burget, L., Cernock'y, J., and Khudanpur, S. (2010) Recurrent neural network based language model. In Inter speech, 3-12.
- [5] Noh, H., HongsuckSeo, P., and Han, B. (2016) Image question answering using convolutional neural network with dynamic parameter prediction, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 30–38.
- [6] Ren, M., Kiros, R., and Zemel, R. (2015) Exploring models and data for image question answering, In Advances in neural information processing systems, 2953–2961.
- [7] Ren, S., He, K., Girshick, R., and Sun, J. (2015) Faster r-cnn: Towards real-time object detection with region proposal networks, In Advances in neural information processing systems, 91–99.
- [8] Sabour, S., Frosst, N., and Hinton, G. E. (2017) Dynamic routing between capsules, In Advances in Neural Information Processing Systems, 3857–3867.
- [9] Sundermeyer, M., Schlüter, R., and Ney, H. (2012) LSTM neural networks for language modeling, In Thirteenth Annual Conference of the International Speech Communication Association, 147-156.
- [10] Teney, D., Anderson, P., He, X., and Hengel, (2017) A. v. d. Tips and tricks for visual question answering: Learnings from the 2017 challenge, 24-31.
- [11] Yang, Z., He, X., Gao, J., Deng, L., and Smola, A. (2016) Stacked attention networks for image question answering, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 21–29.
- [12] Zou, W. Y., Socher, R., Cer, D., and Manning, C. D. (2013) Bilingual word embeddings for phrase-based machine translation, In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing, 1393–1398.
- [13] Antol, S., A. Agrawal, J. Lu, M. Mitchell, D. Batra, C. L. Zitnick, and D. Parikh (2015) VQA: visual question answering. In ICCV, 24-35.
- [14] Ma, Lin, Zhengdong Lu, and Hang Li. (2016) Learning to Answer Questions from Image Using Convolutional Neural Network. AAAI. 3(7), 12-21.
- [15] Shih, K. J., Singh, S., & Hoiem, D. (2016). Where to look: Focus regions for visual question answering, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4613-4621).
- [16] Yang, Z., He, X., GAO, J., Deng, L., & Smola, A. (2016). Stacked attention networks for image question answering, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 21-29).
- [17] Andreas, J., Rohrbach, M., Darrell, T., & Klein, D. (2016). Neural module networks, In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 39-48).
- [18] Wang, P., Wu, Q., Shen, C., & Hengel, A. V. D. (2016). The VQA-Machine: Learning How to Use Existing Vision Algorithms to Answer New Questions. ArXiv preprint arXiv: 1612.05386, 147-154.
- [19] Yao, T., Pan, Y., Li, Y., Qiu, Z., & Mei, T. (2016). Boosting image captioning with attributes. arXiv preprint arXiv:1611.01646, 21-31.
- [20] Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., & Bengio, Y. (2015, June). Show, attend and tell: Neural image caption generation with visual attention, In International Conference on Machine Learning (pp. 2048-2057).
- [21] GAO, H., Mao, J., Zhou, J., Huang, Z., Wang, L., & Xu, W. (2015). Are you talking to a machine? Dataset and methods for multilingual image question, In Advances in Neural Information Processing Systems (pp. 2296-2304).
- [22] Acharya, U. R., Fujita, H., Lih, O. S., Hagiwara, Y., Tan, J. H., & Adam, M. (2017). Automated detection of arrhythmias using different intervals of tachycardia ECG segments with convolutional neural network. Information sciences, 405, 81-90.
- [23] Liu, Y., Chen, X., Peng, H., & Wang, Z. (2017). Multi-focus image fusion with a deep convolutional neural network. Information Fusion, 36, 191-207.
- [24] Mansor, M. A., Kasihmuddin, M. S. M., & Sathasivam, S. (2016). Enhanced Hopfield network for pattern satisfiability optimization. International Journal of Intelligent Systems and Applications, 8(11), 27.
- [25] Mansor, M. A. B., Kasihmuddin, M. S. B. M., & Sathasivam, S. (2017). Robust Artificial Immune System in the Hopfield network for Maximum k-Satisfiability. International Journal of Interactive Multimedia and Artificial Intelligence 4(4), 63-71.



Sudan Jha

Sudan Jha holds the Ph.D. in CSE, working as professor in School of Computer Engineering, KIIT University, India with 18 years of experience in Academics, Administration as a Principal in four Colleges in India and Nepal, Industrial Research and Software Project Development & Management; As a team leader executed three government funded projects; Board of directors and Technical Advisor in Nepal's National Television (NTV); Chief IT Consultant in Nepal Telecom Authority – the regulating body of telecommunication of Nepal; published so far 31 International Journal transaction papers (SCOPUS Indexed-3, UGC Approved-8, Others-20); 60+ conference papers; 8 major software developed as Team Leader including Govt and non Govt. organization. Editorial board member in several Journals, International Conferences, Books. Provisional issuance is under process for his second Ph.D. in Computer Engineering which is already under the final verge. He has been an invitee for ROBOTRONICSUSA twice, World Youth Festival Russia as a VIP Guest Speaker, Teacher's Education Club, Harvard University USA as an active audience. He has visited 14 countries and delivered talks in various topics related with Network Systems, Embedded Systems, Parallel & Distributed Systems, Operating System, Distributed Operating System, Programming Skills like Programming in C, Programming in C++, Object Oriented Software Engineering, Emerging Subjects like High Speed Downlink Packet Access Networks (HSDPA), Universal Mobile Telecommunication Systems (UMTS), Cell throughput evaluation of HSDPA Networks, Fundamentals of Software Engineering, Data Mining and Warehousing, Wired and wireless networking, Software Engineering, Data Mining and Warehousing, UMTS based HSDPA. He also holds standing on consultant for the Nepal Telecommunication Authority, the regulatory board for Government of Nepal and member of board director of Nepal Television.



Anirban Dey

Anirban Dey pursuing M.Tech in computer science and engineering from KIIT University and a research scholar. He is doing researches in the field of computer vision, Natural Language Processing and Deep Learning. His some papers were published in IEEE-approved and Springer conference.



Raghvendra Kumar

Raghvendra Kumar is working as Assistant Professor in Computer Science and Engineering Department at L.N.C.T Group of College Jabalpur, M.P. India. He received B. Tech. in Computer Science and Engineering from SRM University Chennai (Tamil Nadu), India, M. Tech. in Computer Science and Engineering from KIIT University, Bhubaneswar, (Odisha) India and Ph.D. in Computer Science and Engineering from Jodhpur National University, Jodhpur(Rajasthan), India. He has published 86 research papers in international / National journal and conferences including IEEE, Springer and ACM as well as serve as session chair, Co-chair, Technical program Committee members in many international and national conferences and serve as guest editors in many special issues from reputed journals (Indexed By: Scopus, ESCI). He also received best paper award in IEEE Conference 2013 and Young Achiever Award-2016 by IEAE Association for his research work in the field of distributed database. His research areas are Computer Networks, Data Mining, cloud computing and Secure Multiparty Computations, Theory of Computer Science and Design of Algorithms. He authored 12 computer science books in field of Data Mining, Robotics, Graph Theory, and Turing Machine by IGI Global Publication, USA, IOS Press Netherland, Lambert Publication, Scholar Press, S. Chand Publication and Laxmi Publication.



Vijender Kumar-Solanki

Vijender Kumar Solanki, Ph.D. is an Associate Professor in Computer Science & Engineering, CMR Institute of Technology (Autonomous), Hyderabad, TS, India. He has more than 10 years of academic experience in network security, IoT, Big Data, Smart City and IT. Prior to his current role, he was associated with Apeejay Institute of Technology, Greater Noida, UP, KSRCE(Autonomous) Institution, Tamilnadu, India & Institute of Technology & Science, Ghaziabad, UP, India. He has attended an orientation program at UGC-Academic Staff College, University of Kerala, Thiruvananthapuram, Kerala & Refresher course at Indian Institute of Information Technology, Allahabad, UP, India. He has authored or co-authored more than 25 research articles that are published in journals, books and conference proceedings. He has edited or co-edited 4 books in the area of Information Technology. He teaches graduate & post graduate level courses in IT. He received Ph.D in Computer Science and Engineering from Anna University, Chennai, India in 2017 and ME, MCA from Maharishi Dayanand University, Rohtak, Haryana, India in 2007 and 2004, respectively and a bachelor's degree in Science from JLN Government College, Faridabad Haryana, India in 2001. He is Editor in International Journal of Machine Learning and Networked Collaborative Engineering (IJMLNCE) ISSN 2581-3242, Associate Editor in International Journal of Information Retrieval Research (IJIRR), IGI-GLOBAL, USA, ISSN: 2155-6377 | E-ISSN: 2155-6385. He is guest editor with IGI-Global, USA, InderScience & Many more publishers. He can be contacted [spesinfo@yahoo.com](mailto:spesinfo@yahoo.com).

# Contribution to the Association Rules Visualization for Decision Support: A Combined Use Between Boolean Modeling and the Colored 2D Matrix

Fatima Zohra Benhacine\*, Baghdad Atmani, Fawzia Zohra Abdelouhab

Laboratoire d'Informatique d'Oran (LIO), University of Oran 1 Ahmed Benbella (Algeria)

Received 22 February 2018 | Accepted 14 September 2018 | Published 24 September 2018



## ABSTRACT

In the present paper we aim to study the visual decision support based on Cellular machine CASI (Cellular Automata for Symbolic Induction). The purpose is to improve the visualization of large sets of association rules, in order to perform Clinical decision support system and decrease doctors' cognitive charge. One of the major problems in processing association rules is the exponential growth of generated rules volume which impacts doctor's adaptation. In order to clarify it, many approaches meant to represent this set of association rules under visual context have been suggested. In this article we suggest to use jointly the CASI cellular machine and the colored 2D matrices to improve the visualization of association rules. Our approach has been divided into four important phases: (1) Data preparation, (2) Extracting association rules, (3) Boolean modeling of the rules base (4) 2D visualization colored by Boolean inferences.

## KEYWORDS

Association Rules, Boolean Modeling, Cellular Automaton, Clinical Decision Support System (CDSS), Colored 2D Matrix, Data Mining.

DOI: 10.9781/ijimai.2018.09.002

## I. INTRODUCTION

THE decision support is a wide area. Since the 90's the decision making became a major activity that requires the establishment of effective dedicated systems. The Decision support is intended to assist the decision maker and to support him on his understanding of the decision situation by providing him the reason of the selected choices, allowing him to assess the risks when he adopts any strategy [1]. Decision support can be encountered in many application domains as economy, industry, agriculture and medicine. In medicine, the decision is regarded as the center of the medical act. The medical decision process is to make a diagnosis, and also propose a treatment. So, a large number of decision support systems (DSS) have been developed in this area. These DSS are destined to support health workers in their decision making [2]. The clinical decision support system (CDSS) is defined as an information technology tool whose main objective is to assist clinicians to organize, store, extract, and exploit medical knowledge [3].

We contribute to an interactive approach to clinical decision support. The objective of such CDSS is to provide interactive help to users which face similar medical decision problems.

Research in the field of CDSS has resulted in the appearance of new technologies on storage, processing and analysis of data and information required for decision process. The development of such systems requires a real knowledge of the application area, which gave rise to Knowledge Discovery from Databases approach [3]. The decision value is in data history. The purpose of clinical decision support system (CDSS) based on Knowledge Discovery from Databases (KDD) is to capture medical data that can provide decision

support at the time of the treatment. Using the techniques of KDD, the large databases become potentially rich and reliable information sources for knowledge generation and validation. The Data Mining represents the center phase of KDD and to apply the machine-learning algorithms on data to extract models [3].

In our work, we will focus on association rules, constituting one of the powerful models in Data Mining. Association rules are used to treat and discovering interesting rules from large collection of data as  $X \rightarrow Y$  where X and Y are set of items [4]. Association rules have been successfully used in many areas as economic planning support [5], diagnosis assistance and medical research [6]. They are able to detect the trends and hidden relationships and exploring correlations from data. However the significant number of rules generated by this method makes it difficult for human eye. Given these large volumes, the text representation mode is not suitable and the interpretation is impossible. To overcome this problem various visualization techniques [7], [8], [9], [10], [11], [12], [13] have been proposed to improve interpretation of extracted knowledge, clearly and precisely.

The objective of our work is to improve visualization of association rules to increase performance of CDSS and reduce the physician's cognitive loading using 2D colored matrices [14] and Cellular Automaton for Symbolic Induction CASI [15]. We propose a solution that provides a reasoning and essentially guarantees, storage space optimization and execution time.

Our contribution deals mainly with two aspects:

- Firstly, we propose an Intelligent Clinical Decision Support System for immunization using Knowledge Discovery from Databases to provide interactive help to physicians.
- Secondary, we use jointly the CASI cellular machine and the colored 2D matrix to improve visualization of association rules to increase performance of ICDSS and reduce the physician's

\* Corresponding author.

E-mail address: benhacine.fatima@gmail.com



cognitive loading.

This article is structured as follows: In Section II, an overview of CDSS, association rules and techniques for their visualization are presented. In Section III, the proposed approach is described in detail. The results of the approach are explained in Section IV. Finally, we present our conclusions and our future research perspectives in section V.

## II. RELATED WORK

In a patient environment, the medical decision process is to choose an investigative mode. Thence, medical decision support is defined as information management techniques helping partially or completely, the physician in decision processes [16]. The clinical decision support system (CDSS), are playing increasingly important roles in medical practice by helping physicians or other medical professionals making clinical decisions. CDSS are having a greater influence about the care process. They are expected to improve the medical care quality; their impact should intensify due to increasing capacity for more efficient data processing [17]. So, great numbers of decision support systems have been developed in this domain [18], [19]. These applications are made to support health services in their deciding making [2].

There are three main categories of CDSS, the first category is indirect decision support systems or documentary assistance systems whose objective is to facilitate access to relevant information rapidly; but these systems have no reasoning method [20]. The second category is about the automatic reminder systems that prevent doctors from making mistakes or remind important elements to be taken into account for the decision. Assistance provided is not a reasoning help but rather a reminder providing useful information in an easy defined situation [21]. These systems, like the previous ones do not reason [20]. Finally, the third category is consulting systems that provide the user with reasoned conclusions according the reasoning methods used. The conception is more satisfactory than precedents. Developers are primarily interested in this category [20]. This category includes medical expert systems [21] and systems based on Data Mining [22]. Our contribution is part of the CDSS based on Data Mining.

Data Mining uses a variety of methods to process large amounts of data and information to discover useful knowledge for decision. It is a decision support tool in various sectors like health sector facing increasing pressures to improve the quality of health care while reducing costs. It is therefore not surprising that health organizations have been interested in data mining to improve physician practices, disease management and the resource utilization. Hence the progressive use of data mining in the medical domain.

Throughout our study, we identify different data mining techniques [23], [24],[25] such as association rules [23], [24], neural networks and naive bayes [24], Decision Tree and Case-Based Reasoning. The competing objective leading us to propose association rules is their simplicity and unlike other data analysis methods, they provide simple and easily interpretable results.

Since its introduction in [4] the task of association rules has received a great deal of attention. Today the association rules are still one of the most popular pattern discoveries in KDD. An association rule is an expression  $R:X \rightarrow Y$  where  $X$  and  $Y$  are sets of items.  $X$  is called antecedent or left-hand-side (LHS) and  $Y$ , consequent or right-hand-side (RHS). The association rules search expresses, from data contained in a relational database, implicit trends between attributes of antecedent and consequent. In order to select interesting rules from the set of all possible rules; the best-known indicators are support and confidence [4].

One of the recurring problems for association rules is the exponential growth of generated rules volume which impacts doctor's

adaptation. The expert must be able to move as quickly as possible to the essentials during an evaluation. That is why we are interested in graphical representation techniques of association rules. In fact, some approaches are pretty clear for very small data sets; but less so when these quantities increase. In addition, if an expert searches particular information, his eye has a high chance to mask some important information for his analysis [12].

There are many works that deal with the association rules visualization in Data Mining [7], [8], [9], [10], [11], [12], [13].

Different applications have been developed to visualize association rules as tables [9]. Each row corresponds to a rule, and each column represents a rule characteristic whose first column element represents the antecedent, the second column represents the consequent and the following columns represent quality measures of rules. These applications have a rule filtering interface, by specifying the items or setting the support and/or confidence. These applications allow you to classify the rules in ascending or descending order using support or confidence measure [26]. The main limitation of its application is the textual presentation, which does not suit studying large quantities of rules.

On the other hand, the graph visualization [27] is used to represent relationships between items. They are therefore adapted to visualize association rules connecting their antecedent and consequent with an arc. The graph may be directed or undirected. In the undirected graph, in [28], authors used two colors to identify the antecedent and consequent. If an itemset is at the same time the antecedent of one rule and the consequent of another rule, its node has two colors. Three other colors are used to express rule support; light color for low supports, a dark color for high supports and a black color for supports with value 1. The rule confidence is represented by arc length. High confidence is represented by a long arc [26]. An association graph can quickly turn into a tangled display with as few rules. Mosaic plot was introduced by [7] to visualize contingency tables and have been arranged for association rules in [29]. Each rule is represented by a rectangle, whose width is the support and height is the confidence. The main limit to this method is that it can only represent a section of the association rules that have the same consequent and have the same attributes in [26].

In [8], author introduced parallel coordinates as a multidimensional data representation method and they are adapted by [28] to visualize association rules. Variables are represented by parallel axes on which items are distributed and each rule is represented by a broken line which cuts the parallel axes at the level of the items it contains [26].

Matrices are certainly the most used means to represent rules. We identify two types of matrices: Two-Dimensional Matrix (2D) [14], [10], [11], [12]) and three-dimensional (3D) matrices [10], [11].

The work of [14] proposes a Two-Dimensional Matrix (2D). A rule is represented by a cell, the antecedent is displayed on row and consequent is displayed in column or inversely. Visualization of association rule with 3D matrices was introduced with MineSet tools [30]. These matrices known as «item to item» are composed of three axes whose two are used to display items of antecedent or consequent. The third axis is used to represent indicators as support or confidence. A rule is represented by a bar whose gradual color expresses its support and the height corresponds to his confidence [26]. This visualization technique has been improved in matrices «rule-to-item» where each row represents a consequent item and each column has an antecedent item. A rule between two items is represented in the intersection box by a 2D/3D object that indicates the presence or not of the item in the rule. The color of these objects indicates the presence of the item in antecedent or consequent. The matrix is completed by two lines that indicate support and confidence by the height of the bars in three-dimensional. The main disadvantage of this representation is that it

reaches considerable sizes for important sets of rules and become illegible.

The majority of visualization methods are not adapted to represent large sets of patterns. They become unusable; when the pattern number is too important and few methods give an overview of the pattern sets. Finally, no method is adapted to explain how to deduce some rules or the presence of other rules. We are interested in our work on the Two-Dimensional Matrix (2D) for the simplicity to represent the rules on two colors, blue for antecedent and red for consequent.

Nonetheless, in the presence of a significant number of association rules this representation also becomes illegible, and the rules overlap. Occultation problems of such representation become inevitable. To resolve this problem, we have opted for rule optimization using boolean modeling offered by the Cellular Automaton (CASI) [15], and its inference engine to explain the reasoning of some deductions.

In our approach we based on some aspects of the CASI Cellular Automaton: its simplicity to express knowledge in rules and facts, its efficiency in optimizing storage space and execution time. The latter is a particular model of dynamic and discrete systems able to acquire, represent and process extracted knowledge in Boolean form. On the other hand, the machine exploitation as a Cellular Automaton, in the CDSS area, is a novel idea within the team and also at the scientific community. The Cellular Automaton showed its evidence in several research studies in Data Mining: [31] proposed an approach using cellular automaton for the regulation and the reconfiguration of urban transportation systems, [32] proposed a new approach based on cellular automata to reduce the classification time of knearest neighbors algorithm, and [33] proposed a new mapping approach based on the Boolean modeling of critical domain knowledge and also, on the use of different data sources via the data mining technique for the purpose of improving the process of acquiring knowledge explicitly. The problem addressed by [34] is data mining of biological Mycobacterium Tuberculosis responsible for tuberculosis. The author proposed a process of data-enough to generate new knowledge that will be profitable for extraction of particular patterns in the rules of association and they are modeled by the Boolean principle adopted by the cellular machine CASI (Cellular Automaton for Symbolic Induction). [35] proposed a new text categorization framework using concepts lattice and cellular automata. Finally, [36] developed a new approach to boolean fusion ontologies using the CASI cellular machine to optimize the complexity of classical fusion algorithms. The originality of our system is essentially in a proposal of a Intelligent Clinical Decision Support System for immunization using Knowledge Discovery from Databases and the combination

between the 2D matrix and the CASI Cellular Automaton to prove its effectiveness in a new field which is visualization.

### III. PROPOSED APPROACH

In order to handle the large volume of generated rules by an extraction algorithms, our approach operates as a real Knowledge Discovery Process with an important focus on the post-treatment phase. Indeed, our contribution is attempting to provide an effective means for the association rules analysis by graphical visualization.

Our objective is the involvement of visualization techniques in the medical decision-making process guided by data mining and Boolean modeling of CASI. Our contribution is a CDSS based on Data Mining named VisuelAR (Visuel Association Rules). The latter has an important role in medical practice, by helping physicians or other medical professionals making clinical decisions, using visualization techniques. Our process is divided into four important phases 1) Data preparation, (2) Extracting association rules, (3) Boolean modeling of the rule base (4) 2D visualization colored by Boolean inferences. (Fig. 1).

#### A. Preprocessing Module

Preprocessing is an crucial step in the Knowledge Discovery from Databases (KDD). Results obtained at the end of this phase depend mostly on the quality of the data used. Pre-treatment steps relate to data access to build tables named individual-variable tables, grouping observations (explicit data). Depending on data type (numeric, symbolic), pre-treatment methods structure data, clean them, process missing data, and select the attributes when they are numerous: selection of the most informative attributes then sampling. This phase is very important because it will condition the model quality established during Data Mining. Finally, these choices are intended to emerge information contained in the data set [37].

#### B. Data Mining Module

Data mining module extracts association rules using the APRIORI algorithm [4]. The user can modify configuration settings such as support and confidence as needed. At the output of this extraction algorithm, the association rule sets are simple text lists.

#### C. Boolean Modeling Module

CASI [15] is a cellular automaton that simulates basic functioning of an inference engine within an expert system, considering a cellular automaton made of two finite arbitrary long layers of finite-

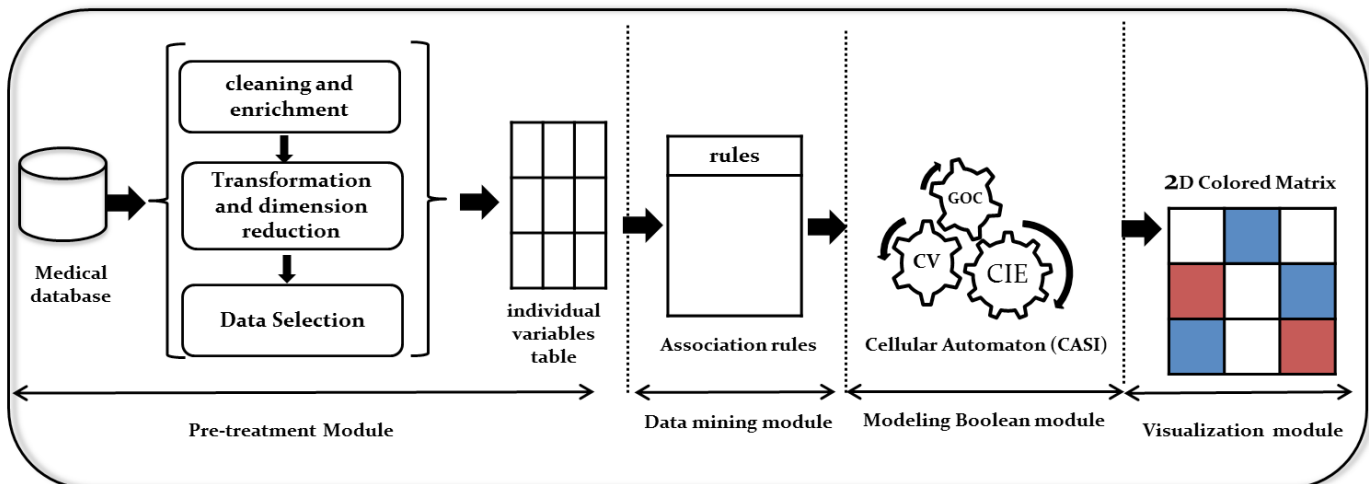


Fig. 1. VisualAR System Architecture.

state machines (cells), all identical. The operation of the system is synchronous, and the state of each cell at time  $t+1$  depends only on the state of its vicinity cells, and on its own state at time  $t$ . The behavior of a knowledge base can be represented by such a cellular automaton with two layers. A first layer known as CELFACT is the fact base, and a second layer, known as CELRULE, is the rule base. In each layer, the cell contents determine whether and how it participates in each inference step: at every iteration, a cell can be active or passive, can participate in the inference or not. We suppose that there are  $l$  cells in the layer CELFACT, and  $r$  cells in the layer CELRULE. The cell states are composed of three parts: the input EF, internal state IF and output parts SF of the CELFACT cells and ER, IR and SR which are the input, internal state and output parts of the CELRULE cells,  $R_E$  and  $R_S$  are the input and the output incidence matrices, respectively:

- The input relation, noted RE, is formulated as follows:  $\forall i \in [1, l], \forall j \in [1, r]$ , if (fact  $i \in$  antecedent of rule  $j$ ) then  $RE(i, j) \leftarrow 1$ .
- The output relation, noted RS, is formulated as follows:  $\forall i \in [1, l], \forall j \in [1, r]$ , if (fact  $i \in$  consequent of rule  $j$ ) then  $RS(i, j) \leftarrow 1$ .

The incidence matrices  $R_E$  and  $R_S$  represent the input/output relation of the facts and are used in forward chaining. One can also use  $R_E$  as output relation and  $R_S$  as input relation for backward chaining. Finally, since there are  $l$  cells in the layer CELFACT, EF, IF and SF will be considered as  $l$ -dimensional vectors ( $EF, IF, SF \in \{0, 1\}$ ). Similarly, since there are  $r$  cells in the layer CELRULE, ER, IR and SR will be considered as  $r$ -dimensional vectors ( $ER, IR, SR \in \{0, 1\}$ ). The cellular automaton dynamics implements the CIE module as a cycle of an inference engine made up of two local transitions  $\delta_{fact}$  and  $\delta_{rules}$  where  $\delta_{fact}$  corresponds to the evaluation, selection and filtering phases, and  $\delta_{rules}$  corresponds to the execution phase.

$\delta_{fact}$  transition function

$$(EF, IF, SF, ER, IR, SR) \xrightarrow{\delta_{fact}} (E, F, E, R + (R_E^t \bullet E) R, R) \quad (1)$$

$\delta_{rules}$  transition function

$$(EF, IF, SF, ER, IR, SR) \xrightarrow{\delta_{rules}} E + (RS.ER), IF, SF, ER, IR, ^{ER} \quad (2)$$

Where the matrix  $R_E^t$  is the transpose of  $R_E$ .

We consider  $G_0$  as the initial cellular automaton configuration and the  $\Delta = \delta_{fact} \circ \delta_{rules}$  as a global transition function:  $(G_0) = G_1$  if  $G_0 \xrightarrow{\delta_{fact}} G'_0$  and  $G'_0 \xrightarrow{\delta_{rules}} G_1$

$\Delta(G_0) = G_1$  si  $G_0 \xrightarrow{\delta_{fact}} G'_0$  et  $G'_0 \xrightarrow{\delta_{rules}} G_1$  Let  $G \{G_0, G_1, G_2, \dots, G_q\}$  be the configuration set of our cellular automaton. The automaton evolution in discrete time steps from one generation to the next is defined by the configuration sequence  $G_0, G_1, \dots, G_q$   $G_{i+1} = \Delta(G_i)$

Integration of an inference engine in 2D colored matrix gives it the possibility to justify the presence of some rules and facilitates expert decision-making.

#### D. Visualization Module

This module uses a 2D matrix visualization technique «rule-to-item», where each row represents a consequent item and each column has an antecedent item. The cell color indicates the itemset presence in the  $R_E$  or  $R_S$  matrix of CASI. The presence in the  $R_E$  matrix is represented by the blue and its presence in the  $R_S$  matrix is represented by the red. This technique reduces the physician's cognitive loading and improving their visual perception results.

## IV. RESULTS AND DISCUSSION

In our approach, we are interested in a preventive medicine area which concerns more than half of the world, namely vaccination.

In the world in general, and particularly in Africa, immunization became, through the different Expanded Program on Immunization, one of the most effective strategies for controlling viral and bacterial diseases. In the late 1970s, it was erected as a health program by the World Health Organization (WHO). This program is to protect children aged 0 to 11 months against the six (6) most deadly diseases. These include tuberculosis, diphtheria, tetanus, whooping cough, polio and measles. Indeed, these diseases are responsible for more than two (2) million deaths per year globally, according to WHO. Yet effective ways to protect children against these diseases have been popularized around the world. Unfortunately, a related problem is that all children presented at the first immunization one week after their birth do not receive all of the vaccines after 11 months for many reasons. They are the lost ones. A true problem for the country's health authorities who are always looking to reduce the causes of lost ones problems.

Through the various surveys funded by the WHO in African countries such as Cameroon, Niger, Burkina Faso, Benin [38] [39], [40], [41]: one of the causes of this low immunization coverage is the poor quality of the service offered. This quality of service can be appreciated from the knowledge viewpoint, staff attitudes and practices, the resource availability, the services organization, satisfaction and the beneficiaries' knowledge. Several studies have revealed that immunization service quality is often poor: it is characterized by the administration age of the different vaccines, high wait times, the many missed occasions, poor reception of providers and the appearance of abscesses after vaccination.

Several studies have shown the correlation between the levels of knowledge mothers have about vaccination and the lost ones [41]. In this context, we are interested in this work to implement our approach on the detection of lost ones and abundance causes in relation to the mother's socio-economic characteristics.

#### A. The questionnaire

To collect the data about the mother's socio-economic characteristics we have developed an online questionnaire: <https://goo.gl/forms/mqDTqjXjs659AlvG3> with 9 questions in total about the mothers' socio-economic profile using the Google Forms platform. We opted for a majority of single-answer multichotomous questions (7 of 9 questions) in order to have a rapid response and data exploitation. The questionnaire results obtained have been saved in an Excel file (.csv).

#### Study Population Characteristics

In Table I, we present the socio-economic characteristics of the mothers interviewed. A total of 368 mothers replied to the questionnaire. Mothers aged above 30 represented the largest age group with 64.2% of the total. The university level represents 61.6% of the education level; the main occupation was state function work with 42.5%. 43.7% of these mothers had only 1 child. The most common means of travel for the immunization session was personal transportation in 72% of cases, with travel time less than 30 minutes for 62.3% of mothers. 66.4% of women are dissatisfied with the start time of the vaccination session and 76.1% of women respect vaccination dates.

TABLE I. SAMPLE DESCRIPTION

	EFFECT	Pourcentage
The number of mothers interviewed	268	100%
Distribution by Age		
Between 18 and 30 years old	96	35.8%
More than 30 years old	172	64.2 %
Education level		
Primary	3	1.1 %
Medium	4	1.5 %
Secondary	22	8.2 %
Academic	165	61.6 %
Post-Graduation	74	27.6 %
Occupation		
Student	21	7.8 %
Housewife	91	34 %
State function	114	42.5 %
Liberal function	42	15.7 %
Number of children		
1 only	117	43.7 %
2	85	31.7 %
3	44	16.4 %
4	16	6 %
5 and more	6	2.2 %
Means of travel for vaccination		
Walking	58	21.6 %
Public transport	6	2.2 %
A taxi	11	4.1 %
Personal transportation	193	72 %
Travel time for vaccination		
Less than half an hour	167	62.3 %
Between 30 minutes and 1 hour	68	25.4 %
More than an hour	33	12.3 %
Satisfaction of the immunization session		
Yes	147	54.9 %
No	121	45.1 %
Satisfaction of the start time for the immunization session?		
Yes	178	66.4 %
No	90	33.6 %
Do you respect vaccination dates and times?		
Yes, at each session	204	76.1 %
Yes, occasionally	54	20.1 %
No, never	10	3.7 %

To ensure data quality data pretreatment has been developed:

- *Data selection*

We can select data now that already exist in our initial file; we have selected the following attributes: Breakdown by age, occupation, number of children, education level, Means of travel, Travel time, Respect for vaccination.

- *Data cleansing and enrichment*

Data cleaning starts immediately after data selection [42]. If enrichment is required a second cleaning step is essential. In cleaning, we find the processing of missing data.

- *Data transformation*

This step consists of transforming one attribute A to another A' [42], in our case, to simplify the nomenclature we have used indexation to designate the different answers obtained and prepare the individual/variable tables (Table II).

TABLE II. UNITS FOR MAGNETIC PROPERTIES

ATTRIBUT	INDEXATION
Distribution by Age	Age
Between 18 and 30 years old	18-30
More than 30 years old	>30
Education level	NvEtud
Primary	Pr
Medium	Moy
Secondary	Sec
Academic	Univ
Post-Graduation	PG
Occupation	Occup
Student	Etud
Housewife	FmFoy
State function	FctEt
Liberal function	FctLib
Number of children	NbEnf
1 only	1
2	2
3	3
4	4
5 and more	<=5
Means of travel for vaccination	MoyDep
Walking	Marche
Public transport	TransPb
A taxi	Taxi
Personal transportation	TransPer
Travel time for vaccination	TempDep
Less than half an hour	<30mn
Between 30 minutes and 1 hour	30-60mn
More than an hour	>60mn
Do you respect vaccination dates and times?	RespVacc
Yes	Oui
No	Non

## B. Extraction of Association Rules

We implemented the Apriori algorithm in JAVA language on a Windows platform. We conducted a series of experiments on our mother's socio-economic database to choose the right model (Table III). We have fixed a suitable and sufficient number for our experience equal to 500 rules. An excerpt from the rules is given in Fig. 2.

TABLE III. ASSOCIATION RULES SERIES

Test	Min Support	Min Confidence	Rules number	Execution time
Series 1	10%	90%	15	0.0s
Series 2	10%	80%	87	0.2s
Series 3	9%	80%	110	0.2s
Series 4	7%	80%	359	0.5s
Series 5	5%	80%	500	1s

```

R1={NbEnf=4, Occup=FctLib}---->{RespVacc=Non}
R2={NbEnf=4, NvEtud=Univ, Occup=FctLib}---->{RespVacc=Non}
R3={NbEnf=4, Occup=FctLib}---->{NvEtud=Univ, RespVacc=Non}
R4={MoyDep=TransPer, NbEnf=4, Occup=FctLib}---->{RespVacc=Non}
R5={NbEnf=4, Occup=FctLib}---->{MoyDep=TransPer, RespVacc=Non}
R6={NbEnf=4, Occup=FctLib}---->{MoyDep=TransPer, NvEtud=Univ}
R7={NbEnf=4, Occup=FctLib}---->{RespVacc=Oui}
R8={MoyDep=Marche, NbEnf=3}---->{RespVacc=Oui}
R9={MoyDep=TransPer, NbEnf=3, TempDep=<30mn}---->{RespVacc=Oui}
R10={NbEnf=4, NvEtud=Univ, RespVacc=Oui}---->{Occup=FctLib}

```

Fig. 2. Extract from the Association Rules of the 5th Series.



### C. CASI and the 2D Colored Matrix

The integration of an inference engine in the 2D color matrix gives it the opportunity to justify the presence of certain rules and facilitate expert decision-making. We will use the sample of the preceding association rules to explain how the cellular Automaton CASI works especially backward chaining. The expert can launch CASI on all iterations and visualize all rules that he/she can deduce from a certain fact. Based on formal mathematical reasoning, the resulting rules can be automatically validated. On the other hand, the expert can visualize the rules obtained, step by step, and stop the system any time. This shows the interactivity between the expert and the visualization matrix which is the expected contribution of our approach.

#### Inference by Purpose

If inference by Purpose, the user can interact with the inference engine so that he can stop it at any iteration level following a specific purpose (Fig. 3, Fig. 4, Fig. 5):

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	0	1	0	R1	0	1	1
Occup=FctLib	0	1	0	R2	0	1	1
MoyDep=TransPer	0	1	0	R3	0	1	1
NvEtud=Univ	0	1	0	R4	0	1	1
RespVacc=Non	0	1	0	R5	0	1	1
RespVacc=Oui	0	1	0	R6	0	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	0	1	1
CELFACT				CELRULE			

Fig. 3. Initial cellular automaton configuration CELFACT/CELRULE.

R <sub>E</sub>	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	R <sub>7</sub>	R <sub>8</sub>	R <sub>9</sub>	R <sub>10</sub>
NbEnf=4	1	1	1	1	1	1				1
Occup=FctLib	1	1	1	1	1					
MoyDep=TransPer				1				1	1	
NvEtud=Univ		1								1
RespVacc=Non										
RespVacc=Oui										1
MoyDep=Marche							1			
NbEnf=3							1	1	1	
Occup=FctEt						1				
TempDep=<30mn								1	1	

R <sub>s</sub>	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	R <sub>7</sub>	R <sub>8</sub>	R <sub>9</sub>	R <sub>10</sub>
NbEnf=4										
Occup=FctLib										1
MoyDep=TransPer					1	1				
NvEtud=Univ			1			1				
RespVacc=Non	1	1	1	1	1					
RespVacc=Oui							1	1	1	
MoyDep=Marche										
NbEnf=3										
Occup=FctEt										
TempDep=<30mn										

Fig. 4. Input/output incidences matrices.

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	0	1	0	R1	1	1	1
Occup=FctLib	0	1	0	R2	1	1	1
MoyDep=TransPer	0	1	0	R3	0	1	1
NvEtud=Univ	0	1	0	R4	1	1	1
RespVacc=Non	1	1	1	R5	0	1	1
RespVacc=Oui	0	1	0	R6	0	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	0	1	1
CELFACT				CELRULE			

Fig. 5. Configuration obtained with  $\delta_{fact}$ .

#### Application of $\delta_{fact}$

The 2D matrix displays the applicable rules step by step, we note that the rules applicable in this step are R1, R2 R4 (Fig. 6.). Fig. 7 and Fig. 8 represent the 2<sup>nd</sup> configuration G2. The visualization of this configuration is represented by Fig. 9.

	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	R <sub>4</sub>	R <sub>5</sub>	R <sub>6</sub>	R <sub>7</sub>	R <sub>8</sub>	R <sub>9</sub>	R <sub>10</sub>
NbEnf=4										
Occup=FctLib										
MoyDep=TransPer										
NvEtud=Univ										
RespVacc=Non										
RespVacc=Oui										
MoyDep=Marche										
NbEnf=3										
Occup=FctEt										
TempDep=<30mn										

Fig. 6. Visualization of the applicable rules.

#### Application de $\delta_{rules}$

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	1	1	0	R1	1	1	0
Occup=FctLib	1	1	0	R2	1	1	0
MoyDep=TransPer	1	1	0	R3	0	1	1
NvEtud=Univ	1	1	0	R4	1	1	0
RespVacc=Non	1	1	1	R5	0	1	1
RespVacc=Oui	0	1	0	R6	0	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	0	1	1
CELFACT				CELRULE			

Fig. 7. Configuration G1 =  $\Delta(G_0)$  obtained with  $\delta_{fact}(G_0)$  and  $\delta_{rules}(G_0)$ .

Application of  $\delta_{fact}$ 

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	1	1	1	R1	1	1	0
Occup=FctLib	1	1	1	R2	1	1	0
MoyDep=TransPer	1	1	1	R3	1	1	1
NvEtud=Univ	1	1	1	R4	1	1	0
RespVacc=Non	1	1	1	R5	1	1	1
RespVacc=Oui	1	1	1	R6	1	1	1
MoyDep=Marche	0	1	0	R7	0	1	1
NbEnf=3	0	1	0	R8	0	1	1
Occup=FctEt	0	1	0	R9	0	1	1
TempDep=<30mn	0	1	0	R10	1	1	1
CELFACT				CELRULE			

Fig. 8. The 2nd configuration G2.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
NbEnf=4										
Occup=FctLib										
MoyDep=TransPer										
NvEtud=Univ										
RespVacc=Non										
RespVacc=Oui										
MoyDep=Marche										
NbEnf=3										
Occup=FctEt										
TempDep=<30mn										

Fig.9. Visualization of the 2nd iteration.

## Automatic inference

In automatic inference case, the user does not interfere in the inference engine process so that iterations are started automatically until no rules are applicable (Fig. 10 and Fig. 11).

Fact	EF	IF	SF	Rule	ER	IR	SR
NbEnf=4	1	1	1	R1	1	1	0
Occup=FctLib	1	1	1	R2	1	1	0
MoyDep=TransPer	1	1	1	R3	1	1	0
NvEtud=Univ	1	1	1	R4	1	1	0
RespVacc=Non	1	1	1	R5	1	1	0
RespVacc=Oui	1	1	1	R6	1	1	0
MoyDep=Marche	1	1	1	R7	1	1	0
NbEnf=3	1	1	1	R8	1	1	0
Occup=FctEt	1	1	1	R9	1	1	0
TempDep=<30mn	1	1	1	R10	1	1	0
CELFACT				CELRULE			

Fig. 10. Final iteration.

	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
NbEnf=4										
Occup=FctLib										
MoyDep=TransPer										
NvEtud=Univ										
RespVacc=Non										
RespVacc=Oui										
MoyDep=Marche										
NbEnf=3										
Occup=FctEt										
TempDep=<30mn										



	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10
NbEnf=4										
Occup=FctLib										
MoyDep=TransPer										
NvEtud=Univ										
RespVacc=Non										
RespVacc=Oui										
MoyDep=Marche										
NbEnf=3										
Occup=FctEt										
TempDep=<30mn										

Fig. 11. Visualization of the last iteration.

## V. DISCUSSION AND VALIDATION

For this experiment we used the 5th series (Table III) with 500 interesting rules (Fig. 12) which we cannot all interpret. In order to retrieve the rules that can meet our objective which is the detection of the causes of non-vaccination we will use the Boolean modeling and colored 2D matrix visualization.

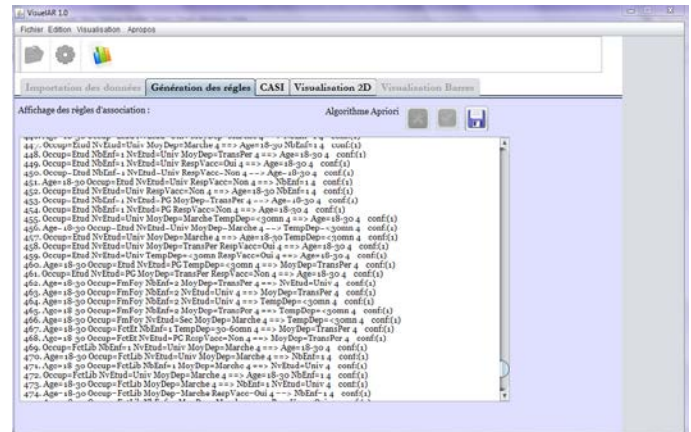


Fig. 12. Textual rules list.

In our case the fact (RespVacc=non) represents the initial fact to be determined (Fig. 13 and Fig. 14).

CELFACT	CELRULE	RE	RS
FACT	EF	IF	SF
NbEnf=2	0	1	0
Age=18-30	0	1	0
NbEnf=1	0	1	0
Occup=Etd	0	1	0
Occup=FmFoy	0	1	0
RespVacc=Non	1	1	0
TempDep=30-60mn	0	1	0
TempDep=>60mn	0	1	0
MoyDep=Marche	0	1	0
NbEnf=4	0	1	0

Fig. 13. Initial cellular automaton configuration CELFACT.

CELRULE	RE	RS
RULE	ER	IR
R421	0	1
R422	0	1
R423	0	1
R424	0	1
R425	0	1
R426	0	1
R427	0	1
R428	0	1
R429	0	1
R430	0	1

Fig. 14. Initial cellular automaton configuration CELRULE.

The incidence matrices RE (Fig. 15) and RS (Fig. 16) represent the input/output relation of the facts and are used in forward chaining from antecedent to consequent. One can also use in transition function  $\delta_{fact}$  and  $\delta_{rules}$ , RS as input relation and RE as output relation for backward chaining, from consequent to antecedent. In our case we will choose a backward chaining to detect the facts that are responsible for the non-vaccination.

RE	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14	R15	R16	R17	R1
NbEnf=3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
RespVacc=Oui	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TempDep=<30mn	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Age=>30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
MoyDep=TransPer	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NvEtud=Univ	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Occup=FctLib	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
NvEtud=PG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Occup=FctEt	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Fig. 15. Input incidence matrix  $R_E$ .

RS	R25	R26	R27	R28	R29	R30	R31	R32	R33	R34
24	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0	0

Fig. 16. Output incidence matrix  $R_S$ .

Fig. 17 and Fig.18 illustrate configuration  $G1 = \Delta(G_0)$  obtained with  $\delta_{fact}$  and  $\delta_{rules}$  (CELFACT)

CELFACT	CELRULE	RE	RS
FACT	EF	IF	SF
NbEnf=2	0	1	0
Age=18-30	0	1	0
NbEnf=1	0	1	0
Occup=Etd	0	1	0
Occup=FmFoy	0	1	0
RespVacc=Non	1	1	0
TempDep=30-60mn	0	1	0
TempDep=>60mn	0	1	0
MoyDep=Marche	0	1	0
NbEnf=4	0	1	0

Fig. 17. Configuration  $G1 = \Delta(G_0)$  obtained with  $\delta_{fact}$  and  $\delta_{rules}$  (CELFACT).

CELRULE	RE	RS
RULE	ER	IR
R427	0	1
R428	0	1
R429	1	1
R430	1	1
R431	0	1
R432	0	1
R433	1	1
R434	1	1
R435	0	1

Fig. 18. Configuration  $G1 = \Delta(G_0)$  obtained with  $\delta_{fact}$  and  $\delta_{rules}$  (CELRULE).

	R378	R429	R430	R433	R434
NbEnf=3	.	.	.	.	.
RespVacc=Oui	.	.	.	.	.
TempDep=<30mn	.	.	.	.	.
Age=>30	.	.	.	.	.
MoyDep=TransPer	.	.	.	.	.
NvEtud=Univ	.	.	.	.	.
Occup=FctLib	.	.	.	.	.
NvEtud=PG	.	.	.	.	.
Occup=FctEt	.	.	.	.	.
NbEnf=2	.	.	.	.	.
Age=18-30	.	.	.	.	.
NbEnf=1	.	.	.	.	.
Occup=Etd	.	.	.	.	.
Occup=FmFoy	.	.	.	.	.
RespVacc=Non	.	.	.	.	.
TempDep=30-60mn	.	.	.	.	.
TempDep=>60mn	.	.	.	.	.
MoyDep=Marche	.	.	.	.	.
NbEnf=4	.	.	.	.	.
NvEtud=Sec	.	.	.	.	.
NbEnf=>5	.	.	.	.	.
MoyDep=Taxi	.	.	.	.	.
NvEtud=Moy	.	.	.	.	.

Fig. 19. Visualization of the applicable rules.

The CASI cellular Automaton gave us the opportunity to retrieve 5 important rules from our series of 500 rules (Fig. 19). The rules, presented in Fig. 19, are R378, R429, R430, R433, R434.

These rules were presented to an expert who helped us interpret them.

The rule 378:  $\{NbEnf=4, Occup=FctLib\} \rightarrow \{RespVacc=Non\}$  indicates that women who have at least 4 children and exercise a liberal profession (presented by a blue color), do not respect the vaccination date (presented by a red color).

The rule 429:  $\{NbEnf=4, NvEtud=Univ, Occup=FctLib\} \rightarrow \{RespVacc=Non\}$  indicates that women who have at least 4 children and exercise a liberal profession; have a university education (presented by a blue color), do not respect the vaccination date (presented by a red color).

The rule 433:  $\{MoyDep=TransPer, NbEnf=4, Occup=FctLib\} \rightarrow \{RespVacc=Non\}$  indicates that women who have personal transportation, at least 4 children, exercise a liberal profession, have a university education (presented by a blue color), do not respect the vaccination date (presented by a red color).

The rest of the rules can be interpreted in the same way and it is clear that this result depends on the population and sample size of the study.

#### System validation

The complexity of every system is calculated by two factors: the temporal factor, it must go faster as possible, and a spatial factor, it must consume less of memory.

The incidence matrix, RE, facilitates the rules transformation into boolean expressions and makes it possible, to use elementary boolean algebra to test different simplifications. Boolean modeling proves that this can be achieved by reducing the amount of storage and execution time. Indeed, this is due to the use of the Boolean representation of the RE and RS matrices, and the Boolean multiplication used by the transition functions  $\delta_{fact}$  and  $\delta_{rules}$ . The two intensive computation processes in CASI are the storage and Boolean multiplication of the RE and RS incidence matrices.

- Storage in  $R_E$  and  $R_S$  memory; the latter, being Boolean matrices, can be expressed as follows: two vectors of various binary sequences (in Hexa). The amount of memory necessary to store Boolean matrices is in the order of  $O(q)$  when using  $q$  sequences of  $r$  bits or  $O(r)$  when  $r$  sequences of  $q$  bits are used. Such matrices can be processed in  $q \times r$  steps. On the other hand, these matrices are in all the hollow iterations (contain a lot of zeros), it is enough to store only the other values equal to 1.
- The standard algorithm ( $R_E^t$ , EF), for example, used by the transition function  $\delta_{fact}$  can be expressed by a sequential algorithm of boolean vector-matrix multiplication executed in a time  $O(rq)$  where  $q$  is the dimension of the vector EF and  $r \times q$  the matrix  $R_E^t$  dimension. The multiplication of  $R_E^t$  and EF can be executed using the boolean matrix vectorization technique, in a time  $\approx O(r \log q)$ , where the internal product of a REt. line with the vector EF is reduced to the product of the parity bit (bit wise And). We can therefore conclude that boolean modeling by cellular Automaton, can be a powerful tool for exploration of research spaces efficiently and effectively. It represents an algorithmic alternative of less complexity that facilitates scaling.

## VI. CONCLUSION

This paper presents a new solution to the complex problem of decision visualization. We are interested in improving the visualization of large sets of association rules to increase system performance while reducing the user's cognitive load. In this article we have essentially tried to highlight the combined use of visualization and data mining techniques in an intelligent clinical decision support system. We proposed an approach combining a colored 2D matrix with Boolean modeling, this brings a deduction strategy that makes it expressive. Indeed, the integration of an inference engine on the 2D colored matrix can justify the presence of certain rules and facilitate the expert's decision-making. The result obtained illustrates this intention. The expert can choose to run the Automaton on all its iterations and see all of the rules it can infer from a certain fact. Using formal mathematical reasoning, the resulting rules can be automatically validated. On the other hand, the expert can visualize the rules obtained step by step and stop the system at will. This reflects the interactivity between the expert and the visualization matrix, which is the contribution of our approach. However, some improvements can be envisaged to make our approach more complete: set grayscale to colors based on confidence and support rules; submit to experts, sets of rules from medical areas to validate our approach. The results presented in this paper provide the basis for future research in several areas. Firstly, we propose to evaluate our approach in other application areas and with other visualization

techniques. Secondly, one future direction of our work is to evaluate association rules with the different measures [43].

## REFERENCES

- [1] A. Akharraz, J. Montmain, G. Mauris, "Elucidation and risk expressions of a movie recommendation based on a multi-criteria aggregation with a Choquet integral". In IPMU 2004, 10th International Conference on Information processing and Management of uncertainty in Knowledge-Based Systems, 2004.
- [2] S. Benbelkacem, B. Atmani, A. Mansoul, "Planification guidée par Raisonnement à base de cas et Datamining : Remémoration des cas par Arbre de décision", Atelier aide à la Décision à tous les Etages (Aide@EGC2012), 2012.
- [3] H. Ltfi, C. Kolski, M. B. Ayed & A. M. Alimi, "A human-centred design approach for developing dynamic decision support system based on knowledge discovery in database". Journal of Decision Systems, 22(2), 69-96, 2013.
- [4] R. Agrawal, T. Imielinski, A. N. Swami, "Mining association rules between sets of items in large databases". In Peter Buneman and Sushil Jajodia, editors, Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207-216, Washington, D.C., 1993.
- [5] D. Dahmani, "Data Mining Using Association Rules Conducted By Pre-selected Rules Patterns and Ontology", ICCESSE 2010
- [6] N. Pasquier, "Extraction de bases pour les Règles d'Association à partir des Itemsets Fermés Fréquents", INFORSID'2000 Congress, (pp. 56-77), 2005.
- [7] J. Hartigan, B. Kleiner "Mosaics for contingency". Proceedings of the 13th symposium on the interface. Springer US, 1981.
- [8] A. Inselberg, "N-dimensional graphics, Part I—lines and hyperplanes", Scientific Center Report G320-2711, IBM, Los Angeles, CA, 1981.
- [9] S. Chakravarthy, H. Zhang. "Visualization of association rules over relational dbms", SAC, 2003.
- [10] S. Ben Yahia, E. Nguifo, "Emulating a cooperative behavior in a generic association rule visualization tool with Artificial Intelligence". ICTAI 2004. 16th IEEE International Conference 2004.
- [11] K. Zhao, B. Liu, M. Thomas, Tirpak, X. Weimin, "Opportunity map: a visualization framework for fast identification of actionable knowledge. In CIKM'05": Proceedings of the 14th ACM international conference on Information and knowledge management. ACM Press, New York, NY, USA, 2005.
- [12] O. Couturier, J. Rouillard, V. Chevrin, "Une approche hybride pour une meilleure visualisation de grands ensembles de règles d'association", ERGOIA'06, 2006.
- [13] M. Ounifi, S. Amdouni, H. Elhoussine, R. Slimane. "New 3D Visualization and Validation Tool for Displaying Association Rules and Their Associated Classifiers". In Information Visualization (IV). 20th International Conference (pp. 152-158). IEEE, 2016.
- [14] J. Boulicaut, P. Marcel, C. Rigotti, "Query driven knowledge discovery in multidimensional data", DOLAP '99: Proceedings of the 2nd ACM international workshop on Data warehousing and OLAP, pp. 87-93, New York, NY, USA. ACM Press 1999.
- [15] B. Atmani, B. Beldjilali, "Knowledge Discovery in Database: Induction Graph and Cellular Automaton", Computing and Informatics Journal, 26(2), 171-197, 2007.
- [16] S. Darmoni, "Titres et travaux, Informatique de santé", Sciences et Technologies de l'Information et de la Communication 2003.
- [17] G. Leroy, H. Chen. "Introduction to the special issue on decision support in medicine", Decision Support Systems, 43 (4), 1203-1206, 2007.
- [18] C. R. Chen, Y. H. Huang, C. T. Bau, S. M. Chen, "A recommendation system based on domain ontology and SWRL for anti-diabetic drugs selection". Expert Systems with Applications, 39(4), 3995-4006, 2012.
- [19] A. Khdega Yosef Galala, "An Expert System for Diagnosis of Ear Problems in Children, International Journal of Advanced Research in Computer and Communication Engineering 4(7), 2015.
- [20] M. Cleret, P. Le Beux, F. Le Duff, "Les systèmes d'aide à la décision médicale", Les Cahiers du numérique, 2(2/2001), 125-154, 2001.
- [21] P. Degoulet M. Fieschi: "Traitement de l'information médicale: méthodes et applications hospitalières", Masson, 1991.
- [22] González-Ferrer, A., Seara, G., Cháfer, J., & Mayol, J. "Generating Big Data Sets from Knowledge-based Decision Support Systems to Pursue



Value-based Healthcare.” International Journal of Interactive Multimedia and Artificial Intelligence, Special Issue on Big Data and e-Health 4(7), 42-46, 2018.

- [23] V. Chevrin, , O. Couturier, , E. Mephu Nguifo, , and J. Rouillard, “Recherche anthropocentrée de règles d’association pour l’aide à la décision. Revue d’Interaction Homme-Machine, 8(2), 2007.
- [24] J. Soni, , U. Ansari, , D. Sharma, , S. Soni, “Predictive data mining for medical diagnosis: An overview of heart disease prediction”. International Journal of Computer Applications, 17(8), 43-48, 2011.
- [25] B. diri, & A. Napoli, Découverte de règles d’association pour l’aide à la prévision des accidents maritimes. In Revue des Nouvelles Technologies de l’Information (No. RNTI-E-23, pp. pages-243). Hermann Editio, 2012, January.
- [26] A. Achouri . “Extraction de relations d’associations maximales dans les textes : représentation graphique”, PhD thesis, University of Québec Trois-Rivières, 2012.
- [27] J. Han, Y. Fu, W. Wang, J. Chiang, W. Gong, K. Koperski, D. Li, Y. Lu, A. Rajan, N. Stefanovic, B. Xia, O. Zaïane , “Dbminer : A system for mining knowledge in large relational databases ”. In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 250–255, 1996.
- [28] D. Bruzzese, P. Buono, “Combining Visual Techniques for Association Rules Exploration”, Proceedings of the International Conference Advances Visual Interfaces, Gallipoli, Italy, May 25-28, 2004.
- [29] H. Hofmann, A. Wilhelm, “Visual comparison of association rules”, Computational Statistics 16(3), 399-415, 2001.
- [30] C. Brunk, J. Kelly, and R. Kohavi, “Mineset : An integrated system for data mining ”. KDD, (135-138), 1997.
- [31] F. Amrani, K. Bouamrane, B. Atmani, and D. Hamdadou, “Une nouvelle approche pour la régulation et la reconfiguration spatiale d’un réseau de transport urbain collectif”. Journal of Decision Systems 20(2), 207-239, 2011.
- [32] F. Barigou, , B. Atmani, , Y. Bouziane, and N. Barigou, “Accélération de la méthode des K plus proches voisins pour la catégorisation de textes ”. EGC’2013 , 241-246, 2013.
- [33] M. Brahmi, B. Atmani, N. Matta, “Dynamic knowledge mapping guided by data mining : Application on Healthcare ”. JIPS 9(1), 1-30, 2013.
- [34] A. Mansoul, B. Atmani, S. Benbelkacem, “A hybrid decision support system: application on healthcare”. CoRR abs/1311.4086, 2013
- [35] H. Benfriha, F. Barigou , B. Atmani, “A text categorisation framework based on concept lattice and cellular automata ”. Int. J. of Data Science, 1(3), 227 – 246, 2016.
- [36] F. Abdelouhab, and B. Atmani, “Fusion cellulaire des ontologies”. Journal of Decision Systems ISSN: 1246-0125 (Print) 2116-7052 (Online) Journal homepage: <http://www.tandfonline.com/loi/tjds20>, 2016.
- [37] D.A Zighed, R.RaKotomalala, “Graphs of induction, Training and datamining”, Hermes science publication (Edition Hermé), 2000.
- [38] S. Savadogo. Etude des causes des abandons de vaccination des enfants de 0 – 11 mois, DIPLOME INTERUNIVERSITAIRE (DIU) Organisation et Management des systèmes Publics de Prévention Vaccinale dans les Pays en Développement, 2003.
- [39] N. Ganda, Etude des causes de non vaccination des enfants de 12-23 mois en zone semi nomade dans le district sanitaire de N’Guigmi au Niger en 2007, DIPLOME INTERUNIVERSITAIRE (DIU) 3ème Cycle, Organisation et Management des Systèmes Publics de prévention vaccinale dans les Pays en Développement. 2009
- [40] V. D Sodjinou. Etude des déterminants de non vaccination en milieu holli dans la zone sanitaire de pobe au Bénin, Organisation et Management des Systèmes Publics de prévention vaccinale dans les Pays en Développement, 2008.
- [41] J.R. Balo Evaluation de la qualité des prestations de vaccination dans le district de sante de Mbankomo, Cameroun, Master Economie et Management de la Santé Publique dans les Pays en Développement, Université de Cocody Abidjan et Université de Paris Dauphine, 2010.
- [42] U G., Fayyad, Piatetsky-Shapiro et P. Smyth. Knowledge Discovery and Data Mining: Towards a Unifying Framework», Proc. 2nd International Conference on KDD & DM, Simoudis E. and Han J. (Eds.), AAAI Press, Menlo Park CA, P. 82-87, 1996.
- [43] S. Lallich, & O. Teytaud. Évaluation et validation de l’intérêt des règles d’association. Revue des Nouvelles Technologies de l’Information, 1(2), 193-218, 2004.



Fatima Zohra Benhacine

She is currently a PhD candidate at the University of Oran 1 Ahmed Ben Bella and affiliated researcher in Laboratoire d’Informatique d’Oran, Algeria. She received her Master on ID-IHM Degree from university of Oran 1 Ahmed BenBella, Algeria in 2013. Her research topics include data mining, Visualization, medical decision support systems, machine learning.



Baghdad Atmani

He is a Professor of Computer Science at the University of Oran 1 Ahmed Benbella. His field of interests is Data Mining and Machine Learning Tools. His research is based on Knowledge Representation, Knowledge-based Systems and CBR, Data and Information Integration and Modelling, Data Mining Algorithms, Expert Systems and Decision Support Systems. His research is guided and evaluated through various applications in the field of control systems, scheduling, production, maintenance information retrieval, simulation, data integration and spatial data mining.



Fawzia Zohra Abdelouhab

She is a Master of Conference and teacher at the department of computer Science, University of Oran1, Ahmed Ben Bella, Algeria. She received her Magister in Computer Science from Oran University. She is currently a Ph.D. in the Computer Science Department at the same university. Her research interests focus on Knowledge Engineering, Ontologies Matching, and information retrieval areas.

# Optimal Performance of Doubly Fed Induction Generator Wind Farm Using Multi-Objective Genetic Algorithm

Ahmed H. A. Elkasem<sup>1</sup>, Salah Kamel<sup>1,2</sup>, Ahmed Rashad<sup>3</sup>, Francisco Jurado<sup>4\*</sup>

<sup>1</sup> Department of Electrical Engineering, Faculty of Engineering, Aswan University, 81542 Aswan (Egypt)

<sup>2</sup> State Key Laboratory of Power Transmission Equipment & System Security and New Technology, Chongqing University, Chongqing, 400030 (China)

<sup>3</sup> Upper Egypt Electricity Distribution Company, Qena Rural Electrification Sector (Egypt)

<sup>4</sup> Department of Electrical Engineering, University of Jaén, 23700 EPS Linares, Jaén (Spain)

Received 19 October 2018 | Accepted 18 March 2019 | Published 29 March 2019



## ABSTRACT

The main purpose of this paper is allowing doubly fed induction generator wind farms (DFIG), which are connected to power system, to effectively participate in feeding electrical loads. The oscillation in power system is one of the challenges of the interconnection of wind farms to the grid. The model of DFIG contains several gains which need to be achieved with optimal values. This aim can be accomplished using an optimization algorithm in order to obtain the best performance. The multi-objective optimization algorithm is used to determine the optimal control system gains under several objectives. In this paper, a multi-objective genetic algorithm is applied to the DFIG model to determine the optimal values of the gains of DFIG control system. In order to point out the contribution of this work; the performance of optimized DFIG model is compared with the non-optimized model of DFIG. The results show that the optimized model of DFIG has better performance over the non-optimized DFIG model.

## KEYWORDS

DFIG, SCIG, Multi-Objective Genetic Algorithm (MOGA).

DOI: 10.9781/ijimai.2019.03.007

## I. INTRODUCTION

THE use of renewable energy in these days became necessary. Wind energy is one of renewable energy resources. Wind energy is a clean energy hence, it is a friend of the environment and human. This will decrease the releasing of harmful gases such as (CO<sub>2</sub>). These harmful gases are emitted from the conventional generation stations. Wind energy has several advantages such as cost-effective, a clean fuel source, does not pollute the air like power plants which rely on combustion of fossil fuels, such as coal or natural gas, which emit particulate matter, nitrogen oxides, and sulfur dioxide causing human health problems and economic damages, also does not produce atmospheric emissions that cause acid rain and smog [1]. The first appearance of wind farms was based on the squirrel cage induction generator (SCIG). The second appearance of wind farms was based on DFIG. The DFIG is a wound rotor induction machine hence it can operate in super-synchronous, sub-synchronous speed and unity power factor manner [2]. The active power from DFIG is produced from the captured mechanical energy which obtained from wind energy [3].

The benefits of the DFIG over SCIG are to improve power quality and to reduce mechanical stress. The optimization algorithm can be single objective or multi-objective. All single-objective optimization methods can find the solution of one specific control problem. This solution differs according to the change in the objective of the same control problem [4, 5].

The occurrence of power system oscillation could cause instability of DFIG wind farm and hence, of connected grid. According to the new optimization techniques that applied to achieve the optimal controller parameters, system stability and good performance are achieved. The improvement of the performance of DFIG wind farms using single-objective optimization technique has been presented in [6, 7]. The improvement of the performance of DFIG wind farms using multi-objective genetic algorithm was investigated in [8]. Many researchers have achieved for enhancing the generated active power [9, 10]. In [11], the dynamic performance of DFIG has been well improved by achieving the optimal controller parameters. The output power from DFIG has been enhanced to help in system stability [12]. This paper aims to enhance the generated active power of DFIG and effective load frequency control with different power system areas. The optimized active power that connected to grid is produced via optimized controller parameters in the studied system. Hence, the oscillation and short settling time will be reduced. This paper represents the basis of DFIG working with power systems which are used in different applications. The control parameters of DFIG are investigated using MOGA. The simulation is carried out in two cases. Firstly, DFIG model performance without MOGA. Secondly, the performance of DFIG model with MOGA.

## II. MULTI-OBJECTIVE OPTIMIZATION

### A. Genetic Algorithm (GA)

The evolutionary Theory is the basis of GA working method. The process of finding the optimal solution using GA is composed of

\* Corresponding author.

E-mail address: fjurado@ujaen.es

three stages. Selection or reproduction represents the first stage and it can be defined as the process of choosing the sets of parents from the population. Crossover is the second stage and it can be defined as the process of choosing two parent solutions from the selected sets. Mutation is the last stage of GA; in this process of generating the next generation from the selected parents [13, 14]. There are many papers studied the enhancement of DFIG based on single-objective optimization like using craziness-based particle swarm optimization (CRPSO) [13]. The energy of wind turbine and power factor have been enhanced using evolutionary computation algorithm (ECA) [15].

### B. Multi-Objective Optimization Technique (MOO)

This optimization technique includes a set of several objectives that need to be achieved. All objectives are gained without overlapping between them. These objectives can be written in the form of functions; these functions can be written as shown in (1).

$$F(x) = [f_1(x), f_2(x), f_3(x) \cdots f_i(x)] \quad (1)$$

The variables of the objective function are denoted by (x) where  $x = (x_1, x_2, \dots, x_n)^T \in X$  and the solution is represented by the vector of X. The role of multi-objective genetic algorithm is obtaining the fit solution. This solution must be suitable to a specific set of objectives and it is not requisite to be fit with single object [16]. The main objectives presented in this work are reducing the oscillation in active power wave form and reducing settling time to enhance performance of power system application.

## III. WIND TURBINE MODEL

In this study a GE 3.6 MW is modeled using MATLAB Simulink program, all data are according to Ref [17]. The block diagram of wind turbine containing modeling of induction generator is shown in Fig. 1.

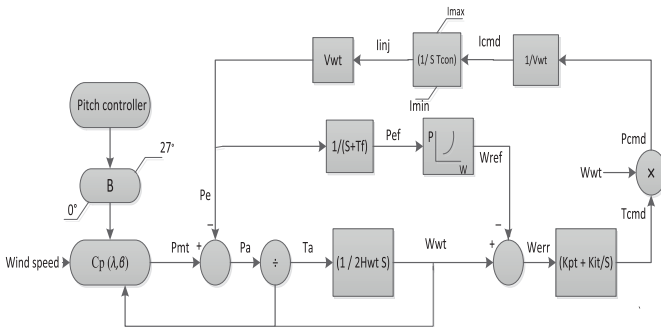


Fig.1. Variable speed wind turbine block diagram.

The captured power by the rotor of wind turbine of doubly fed induction generator from the wind is given by (2) [18] :

$$\left\{ \begin{array}{l} P_{wt} = \frac{1}{2} \pi r^2 \rho v^3 C_p(\lambda, \beta) \\ C_p(\lambda, \beta) = 0.517 \left( \frac{116}{\lambda_i} - 0.4\beta - 5 \right) e^{\frac{-0.0068}{\lambda_i}} + 0.0068\lambda \\ \frac{116}{\lambda_i} = \frac{1}{\lambda + 0.08\beta} - \frac{0.035}{\beta^3 + 1} \\ \lambda = \omega_0 r \frac{\omega_{wt}}{v} \end{array} \right. \quad (2)$$

$P_{wt}$  is referring to power in watts,  $\rho$  is referring to the density

of air,  $r$  is the radius of the turbine rotor (blade-length that equals to 52m),  $v$  is referring to the wind speed.  $C_p$  represents the power coefficient (betz coefficient). Betz coefficient reaches maximum value = 0.5173, in fact the obtainable power coefficient is in the range of 45% according to pitch angle  $\beta$  and the tip speed ratio  $\lambda$ . The optimal tip speed ratio is equal to  $\lambda_{opt} = 8.76 \omega_{wt}$ , where  $\omega_{wt}$  is referring to the rotor speed in pu.  $\omega_0$  is the rotor based speed in Rad/s, its value is equal to 1.33 rad/s. ( $\omega_{ref}$ ) is referring to reference speed associated with maximum power tracking in pu. It depends on the measured electric power ( $P_{ef}$ );  $\omega_{ref}$  can be computed according to the following equation:

$$\omega_{ref} = -0.67 p_{ref}^2 + 1.42 p_{ref} + 0.51 \quad (3)$$

where  $P_{ref}$  is the measured electrical power in pu.

The reference speed calculated by (3) is around 1.2 pu when the generated power levels are above 0.75 pu. In this limit (above 0.75 pu), the pitch controller begins to reset and adjust the speed in order to generate the maximum power 1.0 pu. The pitch angle control is used to keep the generated power within the normal limit. The pitch angle control is applied to DFIG as shown in Fig. 2.

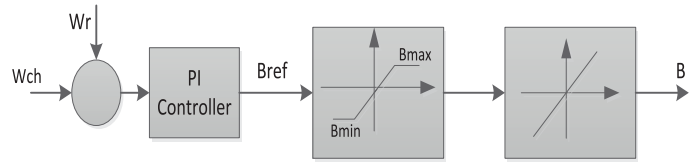


Fig. 2. Strategy of pitch angle control system.

With the increase in wind speed above its rated value, the pitch angle control system role will start increasing in  $\beta$  to its maximum value  $\beta_{max}$ . Thereby, power capture from the wind return to reference output power [19].

Fig. 3 shows the schematic diagram of DFIG connected to grid [20]. The converters of the DFIG are connected between the rotor and the electrical grid. The stator of DFIG is attached to the power system directly to feed grid with generated power.

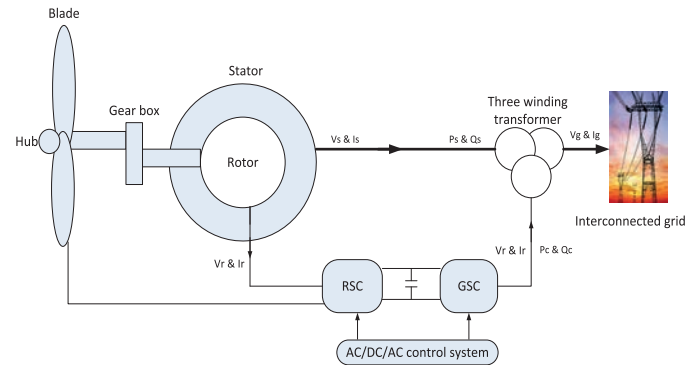
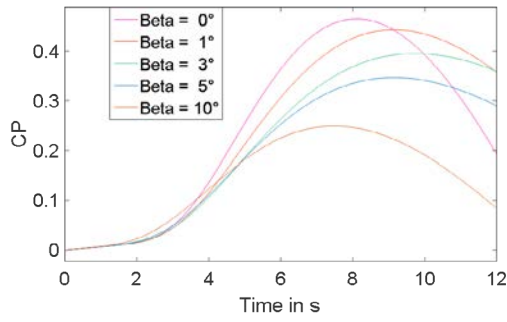


Fig. 3. Schematic diagram of DFIG.

The curves of  $C_p$  of the turbine according to (2) are plotted for five values of  $\beta$  [ $\beta = (0, 1, 3, 5, 10)$ ]. The power coefficient reaches maximum value at pitch angle equal to zero  $\beta=0$  whenever. With the increase in the pitch angle value, the power coefficient value will be decreased as shown in Fig. 4.


 Fig. 4.  $C_p \rightarrow \lambda$  Curve of GE wind turbine for different pitch angles.

#### IV. MODEL OF DFIG STUDIED SYSTEM

As it has been mentioned before, the first case represents the DFIG model without using a MOGA. This model has been modeled according to Ref [21] and it is shown in Fig.5. This studied model includes non-optimized controller parameters such as  $T_f$  which gives the power reference,  $P_{ef}$  extracts the speed reference  $W_{ef}$ . The controller parameter for obtaining the injected current to rotor of DFIG is  $T_{con}$ . The speed controller gains are divided into proportional gain  $K_{pt}$  and integral gain  $K_{it}$ .

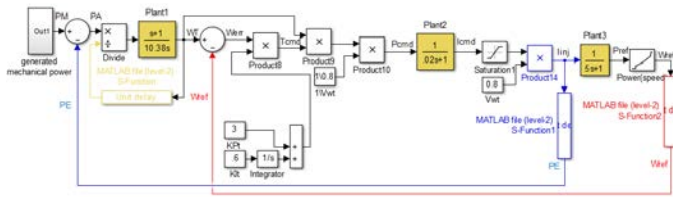


Fig. 5. The model of DFIG without using MOGA technique.

The gains of the non-optimized DFIG wind turbine model are given in Table I.

TABLE I. GAINS OF NON-OPTIMIZED DFIG MODEL (WITHOUT USING MOGA)

Symbol	Quantity	Values
$H_{wt}$	The inertia of WT	5.19 s
$T_f$	WT control parameter for obtaining $P_{ef}$	5 s
$T_{con}$	WT control parameter for obtaining $i_{inj}$	0.02 s
$K_{pt}$	Proportional gain	3
$K_{it}$	Integral gain	0.6

The second case represents the model of DFIG with MOGA. The model of DFIG with MOGA is shown in Fig. 6.

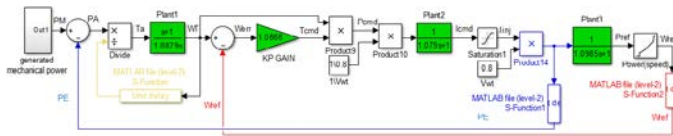


Fig. 6. The model of DFIG using MOGA technique.

The optimized studied model contains transfer functions which are responsible for obtaining a good performance with less perturbations when achieving the best control parameters.

Fig. 7 shows the flowchart of MOGA which is applied on the DFIG model.

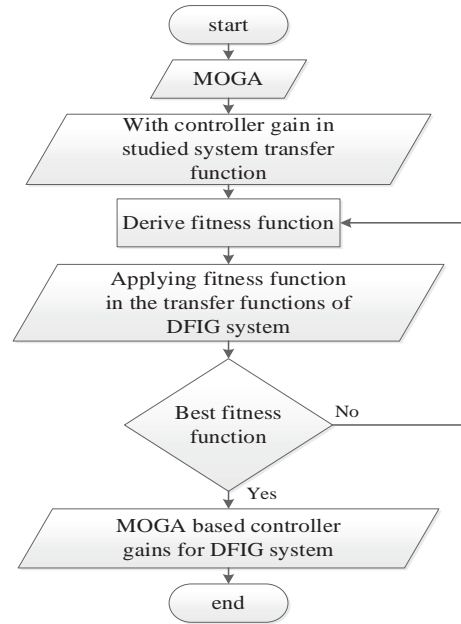


Fig. 7. Flow chart of MOGA based controller gain.

The MOGA is based on the fitness function shown in the next equation:

$$F = \begin{cases} f_1 = \int_0^t (\Delta\omega)^2 dt \\ f_2 = \int_0^t (u_{out})^2 dt \end{cases} \quad (4)$$

Where  $F$  consists of two objectives  $f_1$  and  $f_2$ .  $f_1$  is minimizing the error signal of speed deviation ( $\Delta\omega$ ) from 0 time to the duration time of simulation ( $t$ ).  $f_2$  is maximizing the output signal of DFIG from 0 time to the duration time of simulation ( $t$ ). The included control parameters in studied model can be restricted to certain limits by specifying simple bound constraints to the constrained optimizer function. For `fmincon`, the command is

`x = fmincon (@objfun, x0, [], [], [], [], lb, ub, @confun, options);`

which limits  $x$  to be within the range  $lb \leq x \leq ub$ . To restrict  $x$  (control parameters) to be greater than zero (i.e.,  $x_1 \geq 0$ ,  $x_2 \geq 0$ ,  $x_3 \geq 0$ ,  $x_4 \geq 0$ , ...), these commands are used [22]:

- $X0 = [0.1 \ 0.1 \ 0.1 \ 0.1]$  this command makes a starting guess at the solution.
- $Lb = [0 \ 0 \ 0 \ 0]$  this order includes the lower boundaries of controller gains in studied model.
- $Ub = []$  there are not any upper boundaries constraints.

The main steps which are related to MOGA technique for achieving the best controller parameters are mentioned as:

Step 1: write an M-file name at MATLAB script:

- Function  $F = \text{tracklsq}(x1, x2, x3, \dots)$  where  $F$  is the function which presents the object, `tracklsq` is the name of M-file that will be stored and  $(x1, x2, x3, \dots)$  are the variables of the function.
- To move variables into model parameters, this order must be added `kp1=x (1), kp2=x (2), kp3=x (3)`.
- For choosing solver and set model workspace to this function, the next command needs to be written as mentioned: -  
`opt = simset('solver','ode5','SrcWorkspace','Current');`
- For running Simulink model, this command must be added: -  
`[tout, xout, yout] = sim('optsim', [0 40], opt);` where `tout`, `xout` and `yout`



are the optimized objectives, (sim) is MATLAB command used to run the simulation, (optsim) is the name of the Simulink model and [0 40] is the running time of simulation.

e) At the end, determination of what is the function such as: -

$F = [\text{tout}, \text{xout}, \text{yout}]$ , where tout, xout, yout are the objective functions that need to be modified for a good performance.

Step 2: doing the optimization M-file in new MATLAB script and detecting the best method for minimizing output errors and enhancing the performance of studied model by obtaining best controller parameters. The optimal values of gains of the model of DFIG wind turbine which have been obtained by MOGA are given in Table II.

TABLE II. GAINS OF NON-OPTIMIZED DFIG MODEL (WITH USING MOGA)

Symbol	Quantity	Values
$H_{wt}$	The inertia of WT	1.0879 s
$T_f$	WT control parameter for obtaining $P_{ef}$	1.0985 s
$T_{con}$	WT control parameter for obtaining $i_{inj}$	1.0791 s
$k_p$	Proportional gain	1.0866

## V. SIMULATION RESULTS

The simulation results are carried out in two cases. The first case represents the effect of the change in wind speed on the DFIG wind turbine model without using optimization. The second case represents the effect of the change in wind speed on the model of DFIG with MOGA. The applied wind speed is changed from (0.1 m/s) to (16 m/s) as shown in Fig. 8.

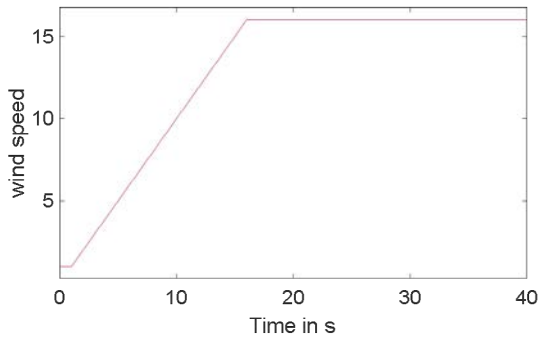


Fig. 8. The variation of the applied wind speed.

### 1) DFIG Model Results Without MOGA Method

Fig. 9 shows the impact of the applied wind speed on the pitch angle. As it can be observed from Fig. 9, the pitch angle has varied from (0 degree) to (27 degree).

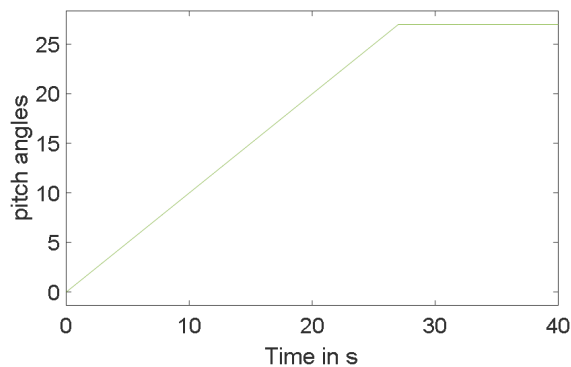


Fig. 9. The pitch angle during the change in wind speed.

Fig. 10 (a) represents captured mechanical power. As shown in Fig. 10 (a), there is no oscillation in  $p_m$ . Fig 10(b) shows the electrical output power ( $p_e$ ). Fig 10 (b) shows that the electrical output power wave has an oscillation at the beginning. This oscillation takes almost 2 seconds (unstable in  $P_e$  wave form). Also, the value of oscillation is large and its duration is long. According to the oscillation in the extracted active power of DFIG, the power system performance will not be stable. System instability is summarized in fluctuations in system frequency which it supposed to be in steady state case. Fig.10 (c) shows the value of ( $P_a$ ) where  $p_a$  is the difference between  $p_m$  and  $p_e$ . This oscillation occurs due to the absence of optimized controller gains values.

In Fig. 11 the reference power curve takes more than 20 s to reach the steady state value. Fig. 12 shows the reference wind speed. As shown in Fig. 12 the reference wind speed takes a period to reach the reference value (1 pu). This means that the non-optimized model of DFIG takes a long time to reach stability.

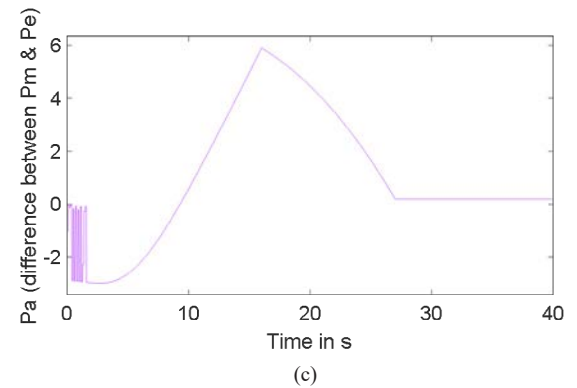
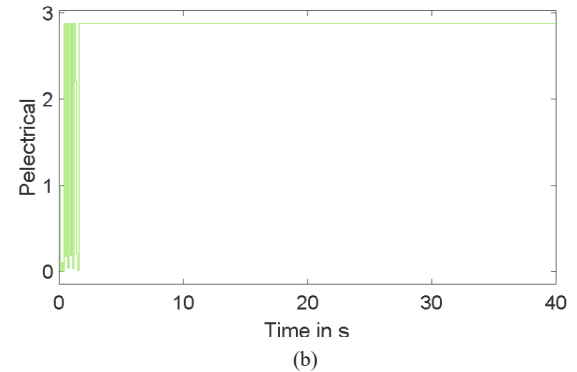
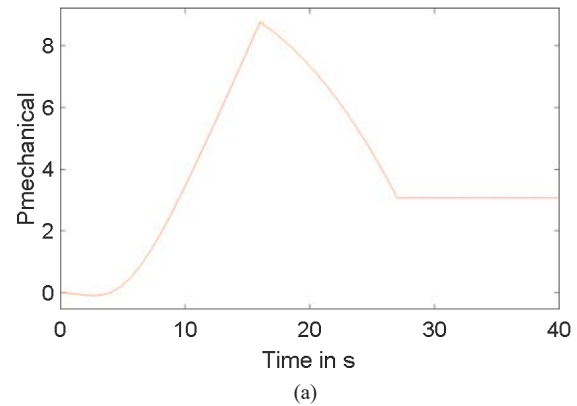


Fig. 10. Power of Non-optimized DFIG model (a)  $p_m$ , (b)  $p_e$  and (c)  $p_a$ .

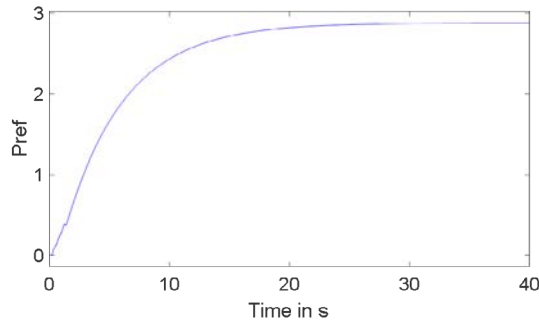


Fig. 11. The reference power  $P_{ref}$ .

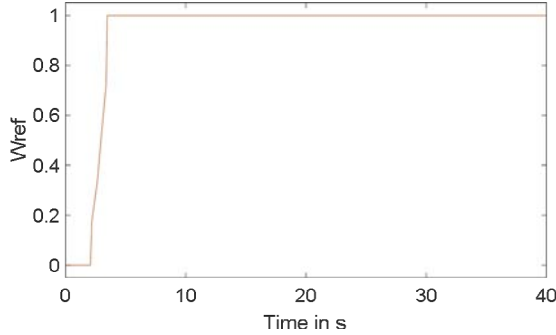


Fig. 12. The reference wind speed  $W_{ref}$ .

## 2) (DFIG) Model Results with (MOO) Method

Fig. 13 shows the power of the model of DFIG with MOGA:  $p_m$ ,  $p_e$  and  $p_a$ . As shown in this figure, oscillation in  $p_e$  and  $p_a$  are decreased to be very small value compared to non-optimized model. Not only that, the time of oscillation occurring became very small. This will lead to more stability of model and more reliability.

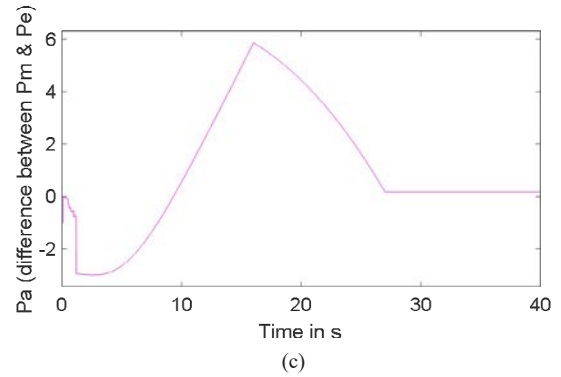
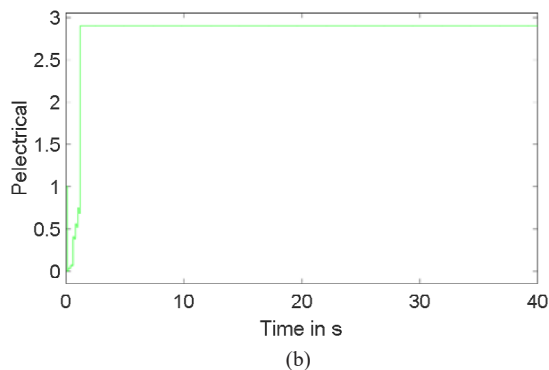
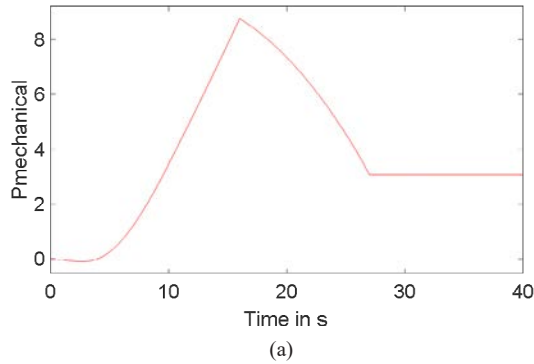


Fig. 13. The power of model of DFIG with MOGA: (a)  $p_m$ , (b)  $p_e$  and (c)  $p_a$ .

There is an effective performance of power system operation with optimized generated active power participation. In the case of DFIG participation to enhance the frequency of power system, good results are achieved with enhanced active power which are characterized by lowest perturbations with smallest settling time.

Fig. 14 shows the reference power  $p_{ref}$  of the model of DFIG with MOGA. Fig. 15 shows the reference wind speed of the model of DFIG with MOGA.

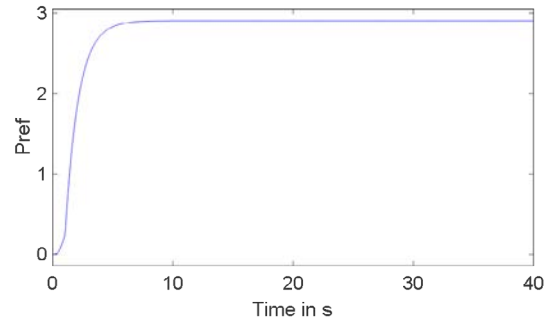


Fig. 14. The reference power  $p_{ref}$  of the model of DFIG with MOGA.

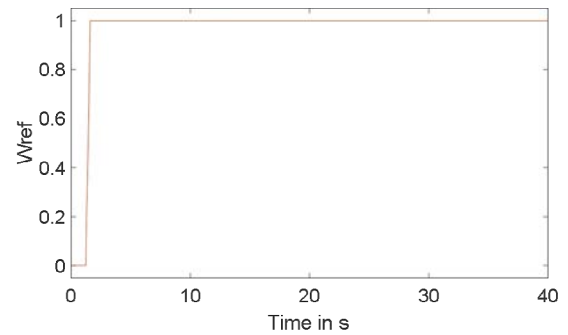


Fig. 15. The reference wind speed of model of DFIG with MOGA.

By comparing Fig. 14 with Fig. 11, it can be observed that  $p_{ref}$  of the model of DFIG takes lower than 2s to reach the steady state value while in case of non-optimized DFIG it takes more than 20 s to reach the steady state value. Also, by comparing Fig. 15 with Fig. 12, it can be observed that in case of DFIG with MOGA the reference wind speed has reached its steady state much faster than in case of non- optimized DFIG.

## VI. CONCLUSION

In this paper, the model of DFIG wind turbines has been based on the optimal value of control gains of DFIG. The optimal value of control gains of DFIG has been investigated using MOGA. The performance

of the model (DFIG with MOGA) has been examined during wind speed variation as ramp function. The performance of the model of DFIG with MOGA has been compared with the performance of non-optimized model of DFIG.

From simulation results, it can be concluded that the model of DFIG with MOGA has much lower oscillation in electrical power and both the reference power and reference wind speed reach their steady state level much faster than non-optimized model of DFIG. This will lead to make DFIG wind turbines more stable and more reliable.

## REFERENCES

- [1] O. Ellabban, H. Abu-Rub, and F. Blaabjerg, "Renewable energy resources: Current status, future prospects and their enabling technology," *Renewable and Sustainable Energy Reviews*, vol. 39, pp. 748-764, 2014.
- [2] Y. Zou, M. Elbuluk, and Y. Sozer, "A complete modeling and simulation of induction generator wind power systems," in *Industry Applications Society Annual Meeting (IAS)*, 2010 IEEE, 2010, pp. 1-8.
- [3] S. Muller, M. Deicke, and R. W. De Doncker, "Doubly fed induction generator systems for wind turbines," *IEEE Industry applications magazine*, vol. 8, pp. 26-33, 2002.
- [4] C. A. C. Coello, "A comprehensive survey of evolutionary-based multiobjective optimization techniques," *Knowledge and Information systems*, vol. 1, pp. 269-308, 1999.
- [5] K. Deb and M.-O. O. U. E. Algorithms, "vol. 16," ed: John Wiley & Sons, 2001.
- [6] Y. Tang, P. Ju, H. He, C. Qin, and F. Wu, "Optimized Control of DFIG-Based Wind Generation Using Sensitivity Analysis and Particle Swarm Optimization," *IEEE Trans. Smart Grid*, vol. 4, pp. 509-520, 2013.
- [7] O. P. Bharti, R. Saket, and S. Nagar, "Controller Design of DFIG Based Wind Turbine by Using Evolutionary Soft Computational Techniques," *Engineering, Technology & Applied Science Research*, vol. 7, pp. 1732-1736, 2017.
- [8] J. P. Vieira, M. V. Nunes, and U. H. Bezerra, "Using Genetic Algorithm to Obtain Optimal Controllers for the DFIG Converters to Enhance Power System Operational Security," in *Wind Turbines*, ed: InTech, 2011.
- [9] N. A. Janssens, G. Lambin, and N. Bragard, "Active power control strategies of DFIG wind turbines," in *Power Tech, 2007 IEEE Lausanne*, 2007, pp. 516-521.
- [10] L. Xu, "Enhanced control and operation of DFIG-based wind farms during network unbalance," *IEEE Transactions on Energy Conversion*, vol. 23, pp. 1073-1081, 2008.
- [11] L. Yang, G.-Y. Yang, Z. Xu, Z. Y. Dong, K. P. Wong, and X. Ma, "Optimal controller design of a doubly-fed induction generator wind turbine system for small signal stability enhancement," *IET generation, transmission & distribution*, vol. 4, pp. 579-597, 2010.
- [12] F. Wu, X.-P. Zhang, P. Ju, and M. J. Sterling, "Decentralized nonlinear control of wind turbine with doubly fed induction generator," *IEEE Transactions on Power Systems*, vol. 23, pp. 613-621, 2008.
- [13] P. Bhatt, R. Roy, and S. Ghoshal, "Dynamic participation of doubly fed induction generator in automatic generation control," *Renewable Energy*, vol. 36, pp. 1203-1213, 2011.
- [14] P. Sravanthi, K. R. Rani, J. Amarnath, and S. Kamakshiah, "Critical clearing time and transient stability analysis of SCIG based wind farm with STATCOM," in *Smart Electric Grid (ISEG), 2014 International Conference on*, 2014, pp. 1-8.
- [15] A. Kusiak and H. Zheng, "Optimization of wind turbine energy and power factor with an evolutionary computation algorithm," *Energy*, vol. 35, pp. 1324-1332, 2010.
- [16] R. Khezri and H. Bevrani, "Voltage performance enhancement of DFIG-based wind farms integrated in large-scale power systems: Coordinated AVR and PSS," *International Journal of Electrical Power & Energy Systems*, vol. 73, pp. 400-410, 2015.
- [17] N. W. Miller, W. W. Price, and J. J. Sanchez-Gasca, "Dynamic modeling of GE 1.5 and 3.6 wind turbine-generators," *GE-Power systems energy consulting*, 2003.
- [18] B. Wu, Y. Lang, N. Zargari, and S. Kouro, *Power conversion and control of wind energy systems* vol. 76: John Wiley & Sons, 2011.
- [19] A. M. Rashad and S. Kamel, "Enhancement of Hybrid Wind Farm

performance using tuned SSSC based on Multi-Objective Genetic Algorithm," in *Power Systems Conference (MEPCON), 2016 Eighteenth International Middle East*, 2016, pp. 786-791.

- [20] В. Дьяконов, *Simulink. Самоучитель*: Litres, 2017.
- [21] N. R. Ullah, T. Thiringer, and D. Karlsson, "Temporary primary frequency control support by variable speed wind turbines—Potential and applications," *IEEE Transactions on Power Systems*, vol. 23, pp. 601-612, 2008.
- [22] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optimization methods and software*, vol. 11, pp. 625-653, 1999.



Ahmed. H. A. Elkasem

Ahmed H. A. Elkasem received the B.Eng. from Faculty of Energy Engineering, Aswan University, Egypt in 2014. He is currently pursuing his MSc degree in Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University. His research activities include renewable energy and power system optimization.



Salah Kamel

Salah Kamel received the international PhD degree from University of Jaen, Spain (Main) and Aalborg University, Denmark (Host) in Jan. 2014. He is an Assistant Professor in Electrical Engineering Department, Aswan University. Also, He is a Leader for power systems research group in the Advanced Power Systems Research Laboratory (APSR Lab), Aswan, Egypt. He is currently a Postdoctoral

Research Fellow in State Key Laboratory of Power Transmission Equipment and System Security and New Technology, School of Electrical Engineering, Chongqing University, Chongqing, China. His research activities include power system modeling, analysis and simulation, and applications of power electronics to power systems and power quality.



Ahmed Rashad

Ahmed Rashad received the B.Eng. from Faculty of Energy Engineering, Aswan University, Egypt and M.Sc. degree in electrical power engineering from Faculty of Engineering, South Valley University, Egypt in 2013. He received the jointly-supervised PhD degree in Department of Electrical Engineering, Aswan Faculty of Engineering, Aswan University, Egypt and University of Jaen, Spain in 2018.

His research activities include Power system modeling, analysis and simulation, renewable energy and smart grid technologies, applications of power electronics to power systems and power quality.



Francisco Jurado

Francisco Jurado obtained the MSc and PhD degrees from the UNED, Madrid, Spain, in 1995 and 1999 respectively. He is Full Professor at the Department of Electrical Engineering of the University of Jaén, Spain. His research activities have focused on two topics: power systems and renewable energy.

# User Identification and Verification from a Pair of Simultaneous EEG Channels Using Transform Based Features

Loay E. George<sup>1</sup>, Hend A. Hadi<sup>2</sup> \*

<sup>1</sup> College of Science, Remote Sensing and GIS, University of Baghdad, Baghdad (Iraq)

<sup>2</sup> Ministry of Education, General Directorate of Education-Karkh III, Baghdad (Iraq)

Received 17 June 2018 | Accepted 18 October 2018 | Published 28 December 2018



## ABSTRACT

In this study, the approach of combined features from two simultaneous Electroencephalogram (EEG) channels when a user is performing a certain mental task is discussed to increase the discrimination degree among subject classes, hence the visibility of using sets of features extracted from a single channel was investigated in previously published articles. The feature sets considered in previous studies is utilized to establish a combined set of features extracted from two channels. The first feature set is the energy density of power spectra of Discrete Fourier Transform (DFT) or Discrete Cosine Transform; the second one is the set of statistical moments of Discrete Wavelet Transform (DWT). Euclidean distance metric is used to accomplish feature set matching task. The combinations of features from two EEG channels showed high accuracy for the identification system, and competitive results for the verification system. The best achieved identification accuracy is (100%) for all proposed feature sets. For verification mode the best achieved Half Total Error Rate (HTER) is (0.88) with accuracy (99.12%) on Colorado State University (CSU) dataset, and (0.26) with accuracy (99.97%) on Motor Movement/Imagery (MMI) dataset.

## KEYWORDS

EEG, Wavelet Transforms, DCT, DFT, Energy Features, Statistical Moments, Euclidean Measure.

DOI: 10.9781/ijimai.2018.12.008

## I. INTRODUCTION

**A** USER identification process attempts to determine who an individual is based on the information provided, and the verification process involves confirming the claimed identity of an individual. All existing security access methodologies consider one or more of the following three basic factors: Knowledge (e.g. password), Possessions (e.g. ATM card), and biological traits (e.g. fingerprint, voice, and retina) [1].

These methodologies are subject to be lost, theft or forgery; Electroencephalogram (EEG) signals are relatively new biological traits which are recently explored for user identification and verification processes due to its robustness against forgery and theft unlike traditional biometrics [2], [3].

There have been various proposed approaches for EEG based user recognition system, Palaniappan in [4] and [5] proposed to use EEG signals for user identification and verification, respectively. The authors proposed features of Autoregressive (AR) coefficients, channel spectral powers, inter-hemispheric channel spectral power differences, linear and the non-linear complexity of inter-hemispheric channel from 6 channels, Linear Discriminant Classifier (LDC) was used to achieve best identification result (i.e., 100%) with features combined from Rotation, Math., Letter, Baseline tasks, whereas for verification mode the author proposed a two stage authentication approach with same features in [4] and used Manhattan distance. The best achieved result

was obtained with False Accepted Rate (FAR) and False Rejected Rate (FRR) equal to zero using 6 channels belong to single task, testing five subjects from CSU dataset [4]. The author used multiple feature types, more than one channel and mental tasks to identify the subject or verify the input identity. In this paper, for each proposed system one type of features is used, two channels and one mental task are used for identification and verification purposes.

Altahat et al. [4] discussed the reduction of required EEG channels of the EEG based verification system. Signal Power Spectral Density (PSD) was considered as features; the best achieved HETR was (14.69%) using 8 channels when the system was tested on (106) subjects from MMI dataset. In this paper, the energy density of DFT or DCT power spectra are proposed as features from two channels only.

Bajwa and Dantu [5] suggested features for EEG authentication system using statistical information of Daubechies (db8) wavelet transform sub-bands after applying the Fast Fourier Transform (FFT) on the EEG signal. Two types of classifiers were tested: Support Vector Machine (SVM) and Bayesian network; and they achieved the best accuracy rate (100%) when the system was tested on 7 subjects from CSU dataset. In this study, Daubechies (db4) with some statistical moments are considered as features, and the energy distribution of DFT spectra is considered as another feature set.

For identification mode, Yang et al. (2012) [6] proposed the use of wavelet transform for extracting features from raw EEG signals by calculating the mean and standard deviation of 5 sub-bands for 8 channels generating a feature vector with 80 features. In this paper, seven types of statistical moments are calculated for six sub-bands and for two channels generating a feature pool that contains 84 features. They proposed two classifiers Support Vector Machine (SVM),

\* Corresponding author.

E-mail addresses: loayedwar57@yahoo.com (L. E. George), hend\_amir2007@yahoo.com (H. A. Hadi).



and k-nearest neighbor classifier to recognize the individuals. A comparison of two kinds of tasks was conducted: motor movement and motor imagery. Their study indicated that imagery tasks show better performance than motor movement tasks. The system was tested on 18 subjects from the Motor/Movement Imagery dataset, the best achieved recognition rate when using SVM classifier was 97.4%.

Daria La Rocca et al. [7] produced an approach based on the fusion of spectral coherence-based connectivity; they fused features from two channels for a single task. They proposed Power Spectral Density (PSD) and Spectral Coherence Connectivity as features; they used a Mahalanobis distance-based classifier to classify (108) subjects from Motor Movement/Imagery dataset, the best achieved accuracy was (100%).

Kumari and Vaish [8] discussed in their paper the fusion of features that were extracted from different mental tasks using canonical correlation analysis from two mental tasks and 6 channels. They proposed to use Empirical Mode Decomposition (EMD), Information Theoretic Measure (ITM) and statistical measurement to extract features. They classified 7 subjects of CSU dataset using Learning Vector Quantization Neural Network (LVQNN) and its extension (LVQ2); they achieved an accuracy of (96.05%). One type of features from two channels and single task are fused in this study to generate the feature vector.

Kumari and Vaish [9] focused in their study on the comparison between the motor movement task and the imagery task. They proposed different methods of Daubechies wavelet transform and different energy methods as features, and then they used Artificial Neural Network (ANN) to classify 5 subjects from Motor Movement/Imagery dataset, achieving True Accepted Rate (TAR) of (95%). In this paper, the Daubechies (db4) wavelet transform was considered with some of the statistical moments as features from two channels belonging to a single task, the statistical moments are applied on each sub-band, and the statistical distance measure was adopted for matching stage.

Yang et al. [10] discussed the sensitivity of EEG-based recognition system to the type of mental tasks; they proposed Daubechies (db4) packet decomposition and calculated the standard deviation of each sub-band as features. Features from different tasks and electrodes (9 electrodes) were fused to generate the final features vector, then they fed to Linear Discriminant Analysis (LDA) classifier to classify (108) subjects from MMI dataset. The best achieved CRR was (99%) for identification mode whereas they achieved a best verification result with Equal Error Rate (EER=4.5).

However, to make EEG-based user identification and verification system applicable, fast, and accurate, the system must go through few and uncomplicated stages. Also, the acquisition process should be easy and simple so as not to disturb the user. Therefore, the least number of electrodes (or channels) must be attached to the user's scalp, and a minimum number of mental tasks must be asked to be performed by the user. These main problems are discussed in this study in which simple, fast, and different methods are proposed using only two EEG channels when the user is performing one mental task, in order to reduce system complexity while maintaining high system accuracy.

In previous work, [11], [12] and [13] the approach of extracting features from single EEG channels when the user is performing certain mental task was discussed, to keep the complexity of the recognition system as less as possible; competitive results were achieved. In this work, the approach of extracting features from two simultaneous EEG channels when the user is performing one task is discussed, to increase the discrimination ratio of the classes and enhance the performance of the recognition system, using the same feature types proposed in the above mentioned previous works.

This paper is organized as follows: Section II presents the description of used datasets and the proposed methods, Section III discusses the

experiments result, Section VI discusses previous works related to this paper, and Section V presents conclusions.

## II. METHODOLOGY

In this study, The EEG-based identification and verification system is based on the approach of combined features from two simultaneous EEG channels through the following main stages: (i) Feature extraction stage which in turn comprises three steps; the first step is aimed to transfer the input EEG signal to either frequency domain or scale-shift domain, whereas the second one is aimed to extract the main features from the transformed signal. The third step is feature analysis and combination stage which is aimed to select and combine the more related and discriminated features from two EEG channels belong to the same task to prepare the final feature vector to be the input to the matching stage to make the final decision.

The main problem facing the automatic EEG identification and verification system is the suitable selection of discriminative features from the EEG signal. Extraction of EEG features is conducted in different domains such as the time domain or the frequency domain. The most used feature extraction methods for EEG biometric systems are AR modeling, Power Spectral Density (PSD), the energy of EEG channels and wavelet packet decomposition (WPD) [3].

### A. Datasets

Two public and free datasets which are used and described in [11] and [13] are also tested in this study. The first one is the Colorado State University dataset which is a small dataset that consists of the recordings of 7 healthy volunteers, collected by Keirn and Aunon [14], whereas the second one is Motor Movement/Imagery dataset which is a relatively large dataset that consists of EEG recordings of 109 healthy volunteers; it was described in [15]. The number of samples in each class in CSU dataset is shown in Table I (Note: class 4 has 9 samples for the letter-composing task because of the error occurred in the dataset and mentioned in [8], [14]), whereas in MMI datasets each class consists of 3 samples.

TABLE I. THE NUMBER OF SAMPLES FOR EACH CLASS IN CSU DATASET

Class No.	No. of Samples	Class No.	No. of Samples
1	10	5	15
2	5	6	10
3	10	7	5
4	10 (only 9 for composing task)		

### B. Proposed System

The proposed methods in [12], [11], and [13] which worked under the approach (using a single channel and a single task) are tested in this study under the second approach (combining features from two channels belonging to the same task).

#### 1) Proposed Features

Two separate sets of features were used for the identification and verification system in previous studies; these are the energy distribution features and statistical moments features:

##### a) Energy Distribution Features

This proposed set of features includes the use of transforms: (1) DFT which is defined by (1), or (2) DCT which is defined by (2); they are used to transform the input signal to the frequency domain, and their output is used to calculate the energy distribution. Equation (3) is used to calculate the energy distribution to the sliced power spectra (i.e. AC components) [16], [3], and [17].

$$F(u) = \frac{1}{N} \sum_{i=0}^{N-1} s(i) \left[ \cos\left(\frac{2\pi i u}{N}\right) - j \sin\left(\frac{2\pi i u}{N}\right) \right] \quad (1)$$

$$C(u) = \alpha(u) \sum_{i=0}^{N-1} s(i) \cos\left(\frac{u\pi(2i+1)}{2N}\right) \quad (2)$$

$$\text{where, } \alpha(u) = \begin{cases} \sqrt{1/N} & \text{if } u = 0 \\ \sqrt{2/N} & \text{if } u \neq 0 \end{cases}$$

Where  $C(u)$  and  $F(u)$  are the  $u^{\text{th}}$  AC coefficient of the DCT and DFT, respectively, and  $s()$  is the input EEG signal.

$$en(j) = \frac{1}{L} \sum_{i=jL+1}^{jL+L} |T(i)|^2, \quad (3)$$

Where  $T(i)$  represents the  $F(u)$  or  $C(u)$  coefficients array;  $en(j)$  is the energy average of  $j^{\text{th}}$  band;  $L$  is the number of coefficients belonging to each band;  $j=0 \dots (N-1)/L$  which is the total number of bands. The array  $en()$  is considered the feature vector.

### b) Statistical Moments Features

The second set of features is the statistical moments of Discrete Wavelet Transforms sub bands. Three types of DWT were proposed to use in previous works: Haar Wavelet Transform which is described by (4) and (5) [18], Daubechies Wavelet Transform (db4) which is described by (6) and (7) [19], and Bi-orthogonal (Tap9/7) Transform which is described by (8-13) [20]:

$$L(i) = s(2i) + s(2i+1) \quad (4)$$

$$H(i+N/2) = s(2i) - s(2i+1) \quad (5)$$

Where  $i=0 \dots N/2$ ;  $N$  is the length of the input signal;  $L()$  is the approximation coefficients;  $h()$  is the detailed coefficient.

$$L(i) = \sum_{k=0}^{N/2} \alpha(k) s(j+k) \quad (6)$$

$$H(i+N/2) = \sum_{k=0}^{N/2} \beta(k) s(k+j) \quad (7)$$

Where,  $i \in \{0, \dots, (N/2)-1\}$ ,  $j \in \{0, \dots, N-3\}$ , and  $k \in \{0, \dots, 3\}$ . The scale values ( $\alpha$ ) and wavelets ( $\beta$ ) are given below:

$$\alpha_1 = (1+\sqrt{3})/(4\sqrt{2}), \quad \alpha_2 = (3+\sqrt{3})/(4\sqrt{2})$$

$$\alpha_3 = (3-\sqrt{3})/(4\sqrt{2}), \quad \alpha_4 = (1-\sqrt{3})/(4\sqrt{2})$$

$$\beta_1 = \alpha_4, \quad \beta_2 = -\alpha_3$$

$$\beta_3 = \alpha_2, \quad \beta_4 = -\alpha_1$$

The bi-orthogonal (Tap9/7) wavelet transform is applied through three consecutive phases: (i) split phase (ii) lifting phase which is described by (8-11) (iii) scaling phase which is described by (12) and (13) [20]:

$$Y(2n+1) = s(2n+1) + a[s(2n) + s(2n+1)] \quad (8)$$

$$Y(2n) = s(2n) + b[s(2n-1) + s(2n+1)] \quad (9)$$

$$Y(2n+1) = Y(2n+1) + c[Y(2n) + Y(2n+2)] \quad (10)$$

$$Y(2n) = Y(2n) + d[Y(2n-1) + Y(2n+1)] \quad (11)$$

$$Y(2n) = Y(2n) / k \quad (12)$$

$$Y(2n+1) = -k \times Y(2n) \quad (13)$$

Where  $a = -1.586134342$ ,  $b = -0.052980118$ ,  $c = 0.8829110762$ ,  $d = 0.4435068522$ , and  $k = 1.230174105$

Following equations describe the two sets of statistical moment which are proposed to be applied on the extracted sub-bands.

The 1<sup>st</sup> Statistical Moments Set is described by (14) whereas a 2<sup>nd</sup> set is described by (16):

$$Mom(n) = \frac{1}{k} \sum_{i=0}^{p-1} [S(i) - \bar{S}]^n \quad (14)$$

Where  $S(i)$  is the  $i^{\text{th}}$  coefficient of the sub-band,  $k$  is the sub-band length, and  $\bar{S}$  is the mean which is determined as:

$$\bar{S} = \frac{1}{k} \sum_{i=0}^{p-1} S(i) \quad (15)$$

$$Mom(n) = \frac{1}{k} \sum_{i=0}^{p-2} [\Delta S(i) - \overline{\Delta S}]^n \quad (16)$$

Where  $\Delta S(i) = S(i) - S(i+1)$  for  $(i=0, \dots, p-2)$ , and  $\bar{S}$  is the average of  $\Delta S(i)$  as described by (15), and the power  $n$  is taken (0.5, 0.75, 1, 2, and 3).

### 2) Two Simultaneous EEG Channels Feature Analysis and Combination

In this stage the features from two simultaneous EEG channels when the user is performing a certain task are combined to make one feature pool, then the pool size is reduced by applying feature analysis and combination by selecting the most related and discriminated features with lowest within distance and highest between variations to make a final feature vector which led to best recognition and verification accuracy [21] [17].

### 3) Matching Stage

In this stage, the normalized Euclidean distance measure (nMSD) is proposed to calculate the distance between the input pattern and the stored templates(s) to make the final decision which either to identify the user identity in identification mode or to verify the claimed identity based on similarity distance threshold in verification mode. nMSD is given by (17) [22]:

$$nMSD(S_i, T_j) = \sum_{k=0}^{p-1} \left( \frac{s_i(k) - t_j(k)}{\sigma_j(k)} \right)^2 \quad (17)$$

Where  $S_i$  is the samples belonging to the  $i^{\text{th}}$  class,  $T_j$  is the template feature vector of the  $j^{\text{th}}$  class, and  $\sigma_j$  is the standard deviation vector of the  $j^{\text{th}}$  template.

## III. EXPERIMENTAL RESULTS

The experimental study of the second approach (combined features from two EEG channels) was conducted on both considered datasets, and the accuracy of verification and identification system with all proposed features was tested. The second adopted approach enhanced the performance of all suggested features for the identification mode, whereas, for the verification mode, the performance of some methods with the first approach (using one EEG channel) is better than with the second approach.

### A. Identification and Verification Experimental Results

Correct Recognition Rate (CRR); that is given by (18); is used to

check the identification system accuracy [17]. The system was partially trained with 67.66% of total samples of each class for CSU dataset, whereas for MMI dataset each class has three samples; so two samples are used for training in which each one is considered as a template, and one sample is used to test the system.

$$CRR = \frac{\text{no. of correctly classified samples}}{\text{total no. of tested samples}} * 100\% \quad (18)$$

The Receiver Operating Characteristic (ROC) Curve is the most used statistical tool for describing the verification system behavior by plotting the False Accepted Rate (FAR) which is given by (19) and measures the average of accepted imposter patterns, against the False Rejected Rate (FRR) which is given by (20) and measures the average of rejected genuine patterns, at various threshold settings to obtain the intersection point between FRR and FAR so the Half Total Error Rate (HTER) can be calculated using (20) to evaluate the performance of the verification system [23], [24]:

$$FAR = \frac{\text{No. of accepted imposters}}{\text{Total No. of imposters}} \quad (19)$$

$$FRR = \frac{\text{No. of rejected genuines}}{\text{Total No. of genuines}} \quad (20)$$

$$HTER = \frac{1}{2} (FAR + FRR) \quad (21)$$

So the accuracy of the verification system can be calculated using (22):

$$\text{Accuracy} = \frac{TP + TN}{P + N} \quad (22)$$

Where P refers to the genuine patterns; and N refers to the imposter patterns [24].

### 1) Experimental Results of Energy Features of DFT Bands

The result of combining the features of DFT energy distribution extracted from two EEG channels; are shown in this section. Tables II and III show some of the identification results when the method is applied to CSU and MMI datasets, respectively. Tables IV and V show some of the verification results of the both datasets.

TABLE II. CRR OF SOME TESTED FEATURE SETS USING THE ENERGY OF DFT BANDS, CSU DATASET

#	Feat. Set	Full training	Partial training
1	C4-P3-base	100%	98.46%
2	P3-O1-rot	100%	98.46%
3	P4-O2-rot	100%	98.46%
4	P4-O1-rot	100%	98.46%
5	C3-O1-Math	100%	98.46%
6	P3-O2-Math	100%	98.46%
7	C4-O2-base	100%	98.46%
8	C3-P4-rot	100%	98.46%
9	C3-P4-base	100%	98.46%
10	P3-C4-base	100%	98.46%

TABLE III. CRR OF SOME TESTED FEATURE SETS USING THE ENERGY OF DFT BANDS, MMI DATASET

#	Feat. Set	Full Training	Partial Training
1	Cz-O1-Task1	100%	100%
2	Fc1-O1-Task1	100%	100%

3	Cp2-Af4-Task4	100%	100%
4	F4-P6-Task4	100%	100%
5	Fcz-O2-Task1	100%	99.69%
6	P7-Oz-Task1	100%	99.69%
7	Fc2-Iz-Task1	100%	99.69%
8	F1-P5-Task1	100%	99.69%
9	Fz-Po3-Task1	100%	99.69%
10	Fc1-Fpz-Task1	100%	99.69%

TABLE IV. FRR, FAR, ACCURACY, AND HTER OF SOME TESTED FEATURE SETS USING THE ENERGY OF DFT BANDS, CSU DATASET

Feat. Set	Thr.	FRR	FAR	Acc.	HTER
C4-P3-base	11.1	1.43	1.56	98.46%	1.49
P3-O1-rot	10.1	1.43	1.58	98.46%	1.51
P4-O2-rot	9.7	2.38	2.47	97.58%	2.42
P4-O1-rot	7.5	2.86	2.39	97.58%	2.62
C3-O1-Math	10.3	2.86	3.84	96.48%	3.35

TABLE V. FRR, FAR, ACCURACY, AND HTER OF SOME TESTED FEATURE SETS USING THE ENERGY OF DFT BANDS, MMI DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
Af4-O2-Task4	21.3	0.31	0.21	99.79%	0.26
P7-Oz-Task1	20.3	0.31	0.27	99.73%	0.29
C2-T9-Task1	24.1	0.31	0.30	99.70%	0.30
Af3-Oz-Task4	20.6	0.31	0.34	99.66%	0.32
Cz-O1-Task1	24.4	0.31	0.35	99.65%	0.33
F1-P5-Task1	22.8	0.31	0.38	99.62%	0.34
Fz-Po3-Task1	22.4	0.31	0.39	99.62%	0.35
Fc2-Iz-Task1	22.7	0.31	0.39	99.61%	0.35
Fpz-P3-Task4	21.2	0.31	0.42	99.58%	0.36
Fc1-O1-Task1	25.1	0.31	0.44	99.56%	0.38

### 2) Experimental Results of Energy Features of DCT Bands

Tables VI and VII show some of the achieved identification results for the DCT bands energy distribution features when the proposed system is applied on both datasets, whereas Tables VIII and IX show some the achieved verification results.

TABLE VI. CRR OF SOME TESTED FEATURE SETS USING THE ENERGY OF DCT BANDS, CSU DATASET

#	Feat. Set	Full training	Partial training
1	C3-P4-Rot	100%	100%
2	C4-O2-Base	100%	98.46%
3	P3-O1-Math	100%	98.46%
4	P3-O2-Math	100%	98.46%
5	P4-O2-Lett	100%	98.44%
6	C4-O2-Rot	100%	98.44%
7	P3-P4-Rot	100%	98.44%
8	P3-O1-Rot	100%	98.44%
9	P3-O2-Rot	100%	98.44%
10	P4-O2-Rot	100%	98.44%

TABLE VII. CRR OF SOME TESTED FEATURE SETS USING THE ENERGY OF DCT BANDS, MMI DATASET

#	Feat. Set	Full training	Partial training
1	Afz-P7-Task4	100%	100%
2	F7-O2-Task4	100%	100%
3	Fc6-P4-Task4	100%	100%
4	F7-P6-Task4	100%	100%
5	Fp2-P6-Task4	100%	100%
6	Cp1-Iz-Task4	100%	100%
7	Cpz-Oz-Task4	100%	100%
8	Fcz-F6-Task4	100%	99.69%
9	C1-Af3-Task4	100%	99.69%
10	Cz-Fp1-Task4	100%	99.69%

TABLE VIII. FRR, FAR, ACCURACY, AND HTER USING THE ENERGY OF DCT BANDS, CSU DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
P3-O2-Math	9.6	0.95	1.12	98.90	1.03
C3-P4-Rot	17.0	2.38	2.39	97.58	2.39
P4-O2-Lett	17.1	2.38	2.39	97.58	2.39
C4-O2-Base	9.7	2.86	2.63	97.36	2.74
P3-O1-Math	9.3	2.38	2.73	97.36	2.55

TABLE IX. FRR, FAR, ACCURACY, AND HTER OF THE ENERGY OF SLICED DCT SPECTRA FEATURES, MMI DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
Cp1-Iz-Task4	19.6	0.31	0.22	99.78%	0.26
Fp2-T9-Task4	19.3	0.31	0.25	99.75%	0.28
Cpz-Oz-Task4	18.8	0.31	0.25	99.74%	0.28
T7-O2-Task1	19.2	0.31	0.28	99.72%	0.29
Fz-Po7-Task1	20.0	0.31	0.30	99.70%	0.30
Fp1-Pz-Task4	20.1	0.31	0.31	99.69%	0.31
Po4-Iz-Task1	19.2	0.31	0.31	99.69%	0.31
Cpz-Iz-Task1	19.8	0.31	0.31	99.69%	0.31
F7-O2-Task4	20.3	0.31	0.36	99.64%	0.33
Af3-O2-Task1	20.1	0.31	0.36	99.64%	0.33

### 3) Experimental Results of Statistical Moments Features of Haar Wavelet Transform

Tables X and XI present the results of some conducted tests of the introduced identification system using Haar wavelet transform with the 2<sup>nd</sup> set of statistical moments that is applied to CSU and MMI datasets, respectively. Tables XII and XIII show some conducted verification results on both datasets.

TABLE X. CRR OF SOME TESTED FEATURE SETS USING THE STATISTICAL MOMENTS OF HWT BANDS, CSU DATASET

#	Feat. Set	Full training	Partial training
1	C3-O1-Rot	100%	100%
2	C4-O1-Rot	100%	100%
3	C4-O2-Rot	100%	100%
4	P3-O2-Rot	100%	100%
5	C3-O1-Base	100%	98.46%
6	C3-O2-Base	100%	98.46%
7	C4-O1-Base	100%	98.46%
8	P4-O2-Rot	100%	98.46%
9	P4-O2-Math	100%	98.46%
10	C3-O2-Rot	100%	98.46%

TABLE XI. CRR OF SOME TESTED FEATURE SETS USING THE 2<sup>nd</sup> STATISTICAL MOMENTS SET OF HWT BANDS, MMI DATASET

#	Feat. Set	Full training	Partial training
1	F7- O1 -Task4	100%	99.39%
2	Fz-Po4 -Task4	100%	99.08%
3	F6-Oz -Task4	100%	99.08%
4	Fc2-F8-Task4	100%	99.39%
5	F4-Po4-Task4	100%	99.08%
6	Cz-Af4-Task4	100%	98.78%
7	Afz-O1-Task4	100%	98.78%
8	Fp2-Po8-Task4	100%	98.47%
9	Fp1-Oz-Task4	100%	98.78%
10	Cp1-F8-Task4	100%	98.47%

TABLE XII. FRR, FAR, ACCURACY AND HTER USING THE 2<sup>nd</sup> STATISTICAL MOMENTS SET OF HWT BANDS, CSU DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
C3-O1-Rot	10.2	1.4	1.3	98.7	1.4
C4-O1-Rot	12.1	1.4	1.6	98.5	1.5
C4-O2-Rot	10.0	1.4	1.6	98.5	1.5
P3-O2-Rot	10.0	2.9	2.9	97.4	2.9
C3-O1-Base	11.2	2.4	2.4	97.6	2.4

TABLE XIII. FRR, FAR, ACCURACY AND HTER USING THE 2<sup>nd</sup> STATISTICAL MOMENTS SET OF HWT BANDS, MMI DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
F6-Poz-Task4	19.8	0.31	0.41	99.59%	0.36
F7-O1 -Task4	15.4	0.31	0.45	99.55%	0.38
Fp1-Iz-Task4	22.3	0.31	0.50	99.51%	0.40
Fp1-Oz-Task4	18.6	0.61	0.49	99.51%	0.55
Fp2-Po8-Task4	17.4	0.61	0.51	99.49%	0.56
Fz-Po4-Task4	20.8	0.61	0.53	99.47%	0.57
Afz-O1-Task4	18.1	0.61	0.65	99.35%	0.63
Cp1-F8-Task4	19.6	0.61	0.66	99.34%	0.63
F6-P2-Task4	19.6	0.61	0.66	99.34%	0.63
Fc2-Iz-Task4	20.5	0.61	0.77	99.23%	0.69

### 4) Experimental Results of Statistical Moments Features of Daubechies (db4) Wavelet transform

Tables XIV and XV show some of the conducted tests results of the identification system based on Daubechies wavelet transform (db4) with the 2<sup>nd</sup> set of statistical moments on CSU dataset and 1<sup>st</sup> statistical moments on MMI dataset. Tables XVI and XVII show some conducted verification results on both datasets.

TABLE XIV. CRR OF SOME TESTED FEATURE SETS USING THE STATISTICAL MOMENTS OF DB4 BANDS, CSU DATASET

#	Feat. Set	Full training	Partial training
1	C4-O1-Rot	100%	100%
2	C4-O2-Rot	100%	100%
3	P3-P4-Rot	100%	100%
4	P4-O1-Rot	100%	100%
5	P4-O2-Rot	100%	100%
6	C3-O2-Lett	100%	98.44%
7	C3-O1-Base	100%	98.46%
8	P4-O2-Base	100%	98.46%
9	C3-O1-Rot	100%	98.46%
10	P3-O1-Rot	100%	98.46%



TABLE XV. CRR OF SOME TESTED FEATURE SETS USING THE STATISTICAL MOMENTS OF DB4 BANDS, MMI DATASET

#	Feat. Set	Full training	Partial training
1	Af8-P3-Task4	100%	100%
2	Af8-Poz-Task4	100%	100%
3	Fp1-Po4-Task4	100%	99.69%
4	Fpz-Po4-Task4	100%	99.69%
5	Af8C56-Task4	100%	99.69%
6	T10-C48-Task4	100%	99.69%
7	Af8-Po4-Task4	100%	99.69%
8	Af8-Iz-Task4	100%	99.69%
9	Fc3-O1-Task4	100%	99.39%
10	Fcz-Po4-Task4	100%	99.39%

TABLE XVI. FRR, FAR, ACCURACY AND HTER USING THE 2<sup>ND</sup> STATISTICAL MOMENTS SET OF DB4 BANDS, CSU DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
C3-O2-Lett	6.9	0.95	1.04	98.88	1.00
C4-O2-Rot	8.0	1.43	1.04	0.99	1.23
C4-O1-Rot	10.6	1.43	1.32	98.68	1.38
P4-O2-Rot	10.3	1.43	1.35	98.68	1.39
C3-O1-Base	11.5	1.43	1.58	98.46	1.51

TABLE XVII. FRR, FAR, ACCURACY AND HTER USING THE 2<sup>ND</sup> STATISTICAL MOMENTS SET OF DB4 BANDS, CSU DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
Af8-Poz-Task4	23.1	0.31	0.31	99.69%	0.31
Af8-Po4-Task4	18	0.31	0.38	99.62%	0.34
Fp2-P8-Task4	19.4	0.31	0.38	99.62%	0.34
F2-Iz-Task1	24.6	0.31	0.37	99.63%	0.34
F7-Po4-Task4	19.1	0.31	0.44	99.56%	0.37
Fp1-Po4-Task4	22.5	0.31	0.49	99.51%	0.4
Af8-Iz-Task4	19.6	0.31	0.49	99.51%	0.4
Fcz-Po4-Task4	19.8	0.31	0.54	99.46%	0.42
Af3-Po3-Task4	22	0.61	0.47	99.53%	0.54
Fpz-Po4-Task4	22	0.61	0.48	99.52%	0.55

### 5) Experimental Results of Statistical Moments Features of TAP 9/7 Wavelet Transform

The best identification results of the system based on Statistical Moments of Tap9/7 Sub-bands are in Tables XVIII and XIX, whereas for verification system in Tables XX and XXI; for both datasets.

TABLE XVIII. CRR OF SOME TESTED FEATURE SETS USING THE STATISTICAL MOMENTS OF TAP9/7 BANDS, CSU DATASET

#	Feat. Set	Full training	Partial training
1	P4-O2-Rot	100%	100%
2	P3-O2-Rot	100%	100%
3	C3-O2-Base	100%	100%
4	C3-O2-Rot	100%	100%
5	P3-O1-Rot	100%	100%
6	C4-O1-Rot	100%	98.46%
7	P4-O1-Rot	100%	98.46%
8	P4-O1-Math	100%	98.46%
9	C3-O2-Math	100%	98.46%
10	C3-P4-Rot	100%	98.46%

TABLE XIX. CRR OF SOME TESTED FEATURE SETS USING THE STATISTICAL MOMENTS OF TAP9/7 BANDS, MMI DATASET

#	Feat. Set	Full training	Partial training
1	T10-P4-Task4	100%	100%
2	Fc1-Af8-Task4	100%	100%
3	Fp1-O2-Task4	100%	99.69%
4	Fpz-Po8-Task4	100%	99.69%
5	Fc3-P8-Task4	100%	99.69%
6	Fcz-O1-Task4	100%	99.69%
7	Fc2-Po7-Task4	100%	99.69%
8	C6-Po8-Task4	100%	99.69%
9	Cp3-Af8-Task4	100%	99.69%
10	Cp6-Af4-Task4	100%	99.69%

TABLE XX. FRR, FAR, ACCURACY AND HTER USING THE 2<sup>ND</sup> STATISTICAL MOMENTS SET OF TAP9/7 BANDS, CSU DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
C4-O1-Rot	15.8	0.95	0.81	99.12	0.88
P4-O1-Rot	8.2	0.95	1.30	98.68	1.13
P4-O2-Rot	12.4	1.43	1.23	98.68	1.33
P3-O2-Rot	10.2	1.43	1.30	98.68	1.37
P4-O1-Math	8.1	1.43	1.52	98.46	1.47

TABLE XXI. FRR, FAR, ACCURACY AND HTER USING THE 2<sup>ND</sup> STATISTICAL MOMENTS SET OF TAP9/7 BANDS, MMI DATASET

Feat. Set	Thr.	FRR	FAR	Accuracy	HTER
F6-Oz-Task4	22.2	0.31	0.24	99.76%	0.27
Fpz-Po8-Task4	17.9	0.31	0.30	99.70%	0.3
Cp3-Af8-Task4	19.5	0.31	0.31	99.69%	0.31
Fp1-O2-Task4	17.7	0.31	0.36	99.64%	0.33
Po4-Iz-Task1	22.4	0.31	0.37	99.63%	0.34
C6-Po8-Task4	18.8	0.31	0.38	99.62%	0.34
Af8-Po8-Task4	19.2	0.31	0.39	99.61%	0.35
Fcz-O1-Task4	20.3	0.31	0.40	99.60%	0.35
Fc2-Po7-Task4	18.3	0.31	0.42	99.58%	0.36
Af8-Poz-Task4	20.1	0.31	0.46	99.55%	0.38

### B. Execution Time

The specification of the computer lap top that was used in the conducted tests is Intel® Core™ i5-2450M CPU with (4GB) RAM, the operating system is windows7 (64bit), and the development programming language is Microsoft Visual C#. Table XXII shows the average elapsed time, (in milliseconds) of the proposed methods on both datasets for one signal only. Taking into consideration the recording time for each sample of CSU dataset is (10 sec) with the sampling rate (250 Hz), and the recoding time for MMI dataset is (1 minute) with the sampling rate (160 Hz).

TABLE XXII. THE AVERAGE PROCESSING TIME RESULTS (IN MSEC) FOR CSU AND MMI DATASETS

CSU dataset		MMI dataset	
Proposed Method	Feature Extraction Time	Proposed Method	Feature Extraction Time
DFT	13.359	DFT	217.364
DCT	22.0991	DCT	323.864
HWT	1.7495	HWT	3.559
Daub4	0.95797	Daub4	3.389
Tap9/7	1.007	Tap9/7	3.719

#### IV. COMPARISON WITH RELATED WORKS

In this section, the comparison between the two adopted approaches (using two EEG channels and using one EEG channel) and the comparison with the other recent related works are shown. Tables XXIII and XXIV show the comparisons of the results of the adopted (as 2<sup>nd</sup>) approach in this paper with the (1<sup>st</sup>) adopted approach in previously proposed work. The second approach combines the features from two channels belonging to the same task, so this approach also keep the complexity of the system low because the user is asked to perform only one mental task in the acquisition stage.

The comparison of the findings of this work and other related works for identification and verification modes are shown in Tables XXV and XXVI.

TABLE XXIII. THE RESULTS OF THE PROPOSED METHODS UNDER THE TWO APPROACHES, IDENTIFICATION MODE

Proposed Method	First approach		Second approach	
	<i>Full training</i>	<i>Partial training</i>	<i>Full training</i>	<i>Partial training</i>
CSU dataset				
DFT	100%	100%	100%	98.46
DCT	100%	100%	100%	100%
HWT	100%	98.46%	100%	100%
Db4	100%	100%	100%	100%
Tap9/7	100%	98.46%	100%	100%
MMI dataset				
DFT	100%	99.08%	100%	100%
DCT	100%	99.08%	100%	100%
HWT	100%	97.25%	100%	99.39%
Db4	100%	97.55%	100%	100%
Tap9/7	100%	98.78%	100%	100%

TABLE XXIV. THE RESULTS OF THE PROPOSED METHODS UNDER THE TWO APPROACHES, VERIFICATION MODE

Proposed Method	First approach		Second approach	
	Accuracy	HTER	Accuracy	HTER
CSU dataset				
DFT	99.56%	0.26	98.46%	1.49
DCT	99.34%	0.4	98.90%	1.03
HWT	98.64%	0.95	98.70%	1.4
Db4	97.58%	2.39	98.88%	1
Tap9/7	99.34%	0.39	99.12%	0.88
MMI dataset				
DFT	99.69%	0.16	99.79%	0.26
DCT	99.61%	0.35	99.78%	0.26
HWT	99.37%	0.62	99.59%	0.36
Db4	99.33%	0.64	99.69%	0.31
Tap9/7	99.40%	0.61	99.76%	0.27

#### V. CONCLUSIONS AND FUTURE WORK

In this paper an extended approach to extract features from user EEG signal is adopted, The features which are proposed in previously conducted studies are tested in this study to check the discriminative degree of this features when they are combined from two simulated EEG channels to generate one feature pool. This approach has improved the performance of the identification system, but for the verification system the performance of the first approach for most types of features is better than the second approach.

This approach also keeps the computational complexity low, and the user performs a single task to take his EEG features. After completing this study, the findings showed that one or two EEG channels are enough to extract discriminate features and recognize the individuals when the proposed methods were tested on the available datasets.

Wider Daubechies wavelet methods such as (db8, db10, ..) and a new type of statistical moments are recommended as new features for EEG based user identification and verification system.

TABLE XXV. COMPARISONS WITH OTHER PUBLISHED WORKS ON CSU DATASET AND MMI DATASET BASED ON THE NUMBER OF SUBJECTS, THE NUMBER OF USED CHANNELS AND TASKS, IDENTIFICATION

CSU dataset					
Author	#subject	#Ch.	#Task	# Features	Accuracy (%)
[25]	5	6	4	4	CRR=100%
[8]	7	6	2	3	CRR=96.05%
Proposed DFT	7	2	1	1	CRR=100%
Proposed DCT	7	2	1	1	CRR=100%
Proposed HWT	7	2	1	1	CRR=100%
Proposed Daub4	7	2	1	1	CRR=100%
Proposed Tap9/7	7	2	1	1	CRR=100%
Motor Movement/Imagery dataset					
[6]	18	8	1	2	CRR=97.4%
[7]	108	2	1	1	CRR=100%
[9]	5	1	1 for train 1 for test	4	TAR=95%
[10]	108	9	4	1	CRR=99%
Proposed DFT	109	2	1	1	CRR=100%
Proposed DCT	109	2	1	1	CRR=100%
Proposed HWT	109	2	1	1	CRR=100%
Proposed Db4	109	2	1	1	CRR=100%
Proposed Tap9/7	109	2	1	1	CRR=100%

TABLE XXVI. COMPARISONS WITH OTHER PUBLISHED WORKS ON CSU DATASET AND MMI DATASET BASED ON THE NUMBER OF SUBJECTS, THE NUMBER OF USED CHANNELS AND TASKS, VERIFICATION

Author	#Subject	# Ch.	# Tasks	# Features	Accuracy (%)
CSU Dataset					
[26]	5	6	1	4	FAR=0, FRR=0
[5]	7	6	1	1	Acc.=100%
Proposed DFT	7	2	1	1	Acc.=98.46%, FAR=1.43, FRR=1.56
Proposed DCT	7	2	1	1	Acc.=98.90%, FAR=0.95, FRR=1.03
Proposed HWT	7	2	1	1	Acc.=98.70%, FAR=1.4, FRR=1.3
Proposed Db4	7	2	1	1	Acc.= 98.88%,FAR=0.95, FRR=1.04
Proposed Tap9/7	7	2	1	1	Acc.= 99.12%,FAR=0.95, FRR=0.81
MMI Dataset					
[10]	108	9	4	1	EER=0.045
[4]	106	8	1	1	HTER=14.64
Proposed DFT	7	2	1	1	Acc.= 99.79%,FAR=0.31, FRR=0.21
Proposed DCT	7	2	1	1	Acc.= 99.78%,FAR=0.31, FRR=0.22
Proposed HWT	7	2	1	1	Acc.= 99.59%,FAR=0.31, FRR=0.41
Proposed Daub4	7	2	1	1	Acc.= 99.69%,FAR=0.31, FRR=0.31
Proposed Tap9/7	7	2	1	1	Acc.= 99.76%,FAR=0.31, FRR=0.24

## REFERENCES

- [1] C. Ashby, A. Bhatia, F. Tenoreb, and J. Vogelsteina, "Low-Cost Electroencephalogram (EEG) based Authentication," in *15th International IEEE EMBS Conference on Neural Engineering*, Cancun, Mexico, 2011, pp. 422-445.
- [2] P. Campisi and D. La Rocca, "Brain Waves for Automatic Biometric-based User Recognition," *IEEE transactions on information forensics and security*, vol. 9, no. 5, pp. 782-800, 2014.
- [3] M. Abo-Zahhad, S. M. Ahmed, and S., N. Abbas, "State-of-the-art Methods and Future Perspectives for Personal Recognition Based on Electroencephalogram Signals," *IET Biometrics*, vol. 4, no. 3, pp. 179–190, 2015.
- [4] S. Altahat, M. Wagner, and E., M. Marroquin, "Robust Electroencephalogram Channel Set for Person Authentication," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015, pp. 997-1001.
- [5] G. Bajwa and R. Dantu, "Neurokey: Towards a New Paradigm of Cancelable Biometrics-based Key Generation using Electroencephalograms," *Computers & Security*, vol. 62, pp. 95-113, 2016.
- [6] S. Yang and F. Deravi, "On the effectiveness of EEG signals as a source of biometric information," in *2012 Third International Conference on Emerging Security Technologies (EST)*, 2012, pp. 49-52.
- [7] D. La Rocca, P. Campisi, B. Vegso, P. Cserti, and G. Kozmann, "Human Brain Distinctiveness Based on EEG Spectral Coherence Connectivity," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 9, pp. 2406-2412, 2014.
- [8] P. Kumari and A. Vaish, "Feature-level Fusion of Mental Task's Brain Signal for an Efficient Identification System," *Neural Computing and Applications*, vol. 27, no. 3, pp. 659-669, 2015.
- [9] P. K. Sharma and A. Vaish, "Individual Identification Based on Neuro-Signal using Motor Movement and Imaginary Cognitive Process," *Optik-International Journal for Light and Electron Optics*, vol. 127, no. 4, pp. 2143-2148, 2015.
- [10] S. Yang, F. Deravi, and S. Hoque, "Task Sensitivity in EEG Biometric Recognition," *Pattern Analysis and Applications*, pp. 1-13, 2016.
- [11] H., A. Hadi and L., E. George, "EEG Based User Identification Methods Using Two Separate Sets of Features Based on DCT and Wavelet," sent for publication to *Journal of Theoretical and Applied Information Technology*, vol. xx, no. x, pp. xx-xx, 2017.
- [12] H., A. Hadi and L., E. George, "EEG Based User Identification and Verification Using the Energy of Sliced DFT Spectra," *International Journal of Science and Research (IJSR)*, vol. 6, no. 9, pp. 46-51, 2017.
- [13] L. E. George and H. A. Hadi, "Brainwaves for User Verification using Two Separate Sets of Features based on DCT and Wavelet," *International Journal of Advanced Computer Science and Applications (IJACSA)*, no. 9, pp. 240-246, 2018.
- [14] Z., A. Keirn and J., I. Aunon, "A new Mode of Communication between Man and his Surroundings," *IEEE transactions on biomedical engineering*, vol. 37, no. 12, pp. 1209-1214, 1990.
- [15] G. Schalk, D., J. McFarland, T. Hinterberger, N. Birbaumer, and J., R. Wolpaw, "BCI2000: a General-Purpose Brain-Computer Interface (BCI) System," *IEEE Transactions on biomedical engineering*, vol. 51, no. 6, pp. 1034-1043, 2004.
- [16] N. AHMED, T. Natarajan, and K., R. RAO, "Discrete Cosine Transform," *IEEE transactions on Computers*, vol. 100, no. 1, pp. 90-93, 1974.
- [17] A., M., J. Abbas and L., E. George, "Palm Vein Identification and Verification System Based on Spatial Energy Distribution of Wavelet Sub-Bands," *International Journal of Emerging Technology and Advanced Engineering*, vol. 4, no. 5, pp. 727-734, may 2014.
- [18] J. S. Walker, *A primer on Wavelets and Their Scientific Applications*.: CRC press, 2008.
- [19] J. Shen and G. Strang, "Asymptotics of Daubechies Filters, Scaling Functions, and Wavelets," *Applied and Computational Harmonic Analysis*, vol. 5, no. 3, pp. 312-331, 1998.
- [20] M. Beladgham, A. Bessaid, A., M. Lakhdar, and A., T. Ahmed, "Improving Quality of Medical Image Compression using Biorthogonal CDF Wavelet Based on Lifting Scheme and SPIHT Coding," *Serbian Journal of Electrical Engineering*, vol. 8, no. 2, pp. 163-179, 2011.
- [21] S. N. Mohammed and L. E. George, "Subject Independent Facial Emotion Classification Using Geometric Based Features," *Research Journal of Applied Sciences, Engineering and Technology*, vol. 11, no. 9, pp. 1030-1035, 2015.
- [22] W. K. Pratt, *Digital Image Processing*. A Wiley-Inter Science Publication, 2001.
- [23] R. Fluss, B. Reiser, D. Faraggi, and A. Rotnitzky, "Estimation of The ROC Curve under Verification Bias," *Biometrical Journal*, vol. 51, no. 3, pp. 475-490, 2009.
- [24] T. Fawcett, "An Introduction to ROC Analysis," *Pattern recognition letters*, vol. 27, no. 8, pp. 861-874, 2006.
- [25] R. Palaniappan, "Electroencephalogram Signals from Imagined Activities: A novel Biometric Identifier for a Small Population," in *International Conference on Intelligent Data Engineering and Automated Learning*, 2006, pp. 604-611.

- [26] R. Palaniappan, "Two-Stage Biometric Authentication Method using Thought Activity Brain Waves," *International Journal of Neural Systems*, vol. 18, no. 1, pp. 59-66, 2008.



Loay E. George

Dr. Loay E. George was born in Baghdad, Iraq 1957, PhD holder since 1979. Currently, he is a member of teaching staff in college of science/ University of Baghdad, Iraq. His main research concerns are: Digital Multimedia Processing, Coding (encryption, digital signature, data compression, representation) Pattern Recognition & Classification, Fast Strings Processing and Analysis, Biometrics, Visual Based

application.



Hend A. Hadi

Hend A. Hadi was born in Baghdad, Iraq 1986, assistant instructor at ministry of education, Baghdad, Iraq. She holds a master degree from college of science/ University of Baghdad in 2018. Her main research concerns are: Digital Signal Processing, Pattern Recognition & Classification, Database applications, Visual Based application.



# A Diversity-Accuracy Measure for Homogenous Ensemble Selection

S. Taleb Zouggar<sup>1\*</sup>, A. Adla<sup>2</sup>

<sup>1</sup> Department of Economics, University of Oran 2 (Algeria)

<sup>2</sup> Department of Computer Science, University of Oran 1 (Algeria)

Received 25 December 2017 | Accepted 30 May 2018 | Published 29 June 2018



## ABSTRACT

Several selection methods in the literature are essentially based on an evaluation function that determines whether a model  $M$  contributes positively to boost the performances of the whole ensemble. In this paper, we propose a method called Diversity and Accuracy for Ensemble Selection (DIACES) using an evaluation function based on both diversity and accuracy. The method is applied on homogenous ensembles composed of C4.5 decision trees and based on a hill climbing strategy. This allows selecting ensembles with the best compromise between maximum diversity and minimum error rate. Comparative studies show that in most cases the proposed method generates reduced size ensembles with better performances than usual ensemble simplification methods.

## KEYWORDS

Machine Learning, Classification, Decision Trees, Ensemble Methods, Bagging, Ensemble Pruning, Hill Climbing.

DOI: 10.9781/ijimai.2018.06.005

## I. INTRODUCTION

**D**ECISION trees are classification methods for generating mutually exclusive decision rules structured in trees that must be simple with maximum performances on the learning sample.

These methods have the advantage to generate intelligible rules but are less efficient than the competing methods because of a significant variance due to the trees instability [5] that increases their generalization error rate [20].

To alleviate this weakness and find a compromise between the complexity of a model and its generalization reliability, stopping criteria as well as post-pruning algorithms have been proposed. These techniques reduce the variance but deteriorate the performances (increase the value of the bias). Another solution is to improve an unstable learning algorithm by using it several times to construct a set of different models [6], [40], [15], [18]. The generated models are then aggregated in order to combine their predictions. In the case of learning by regression, the aggregation is based on individual models predictions, whereas in the case of learning by classification, it is done by a majority vote among the classes predicted by the different models.

The effectiveness of these methods is related to the construction of a collection of classifiers that are both sufficiently precise (performance) and sufficiently diverse (diversity). Diversity offers the opportunity to benefit from the complementarity of the individual models that make up the ensemble.

If each classifier is precise and does not commit errors on the same individuals as others, then the uncorrelated errors of the different

classifiers are removed using the voting process.

The decision trees ensemble methods [43], [8], [9], [28], [25], [26], [27], [1] are homogeneous ensemble methods for which the basic model used is a tree induction algorithm.

The ensemble methods are noise-resistant; do not suffer from over-learning and give good performances [41], but have the disadvantage of relying on a large number of models, which can have as consequences increased learning time, storage resources [31] and the prediction time related to the interrogation of all models in the set.

The aim of ensemble pruning methods is to improve both the efficiency (prediction time) and the prediction performances [35] because a large number of models increase the computational complexity but guarantees a great diversity within the ensemble. This diversity is represented by models with good or bad predicting performances.

Models with poor performance negatively affect the overall performance of the ensemble. Eliminating them while maintaining a large diversity of the remaining elements in the ensemble, improves performance while reducing prediction time.

We propose a new method to simplify homogeneous ensembles composed of C4.5 decision trees [39]. This method is based on a DHCEP (Directed Hill Climbing Ensemble Pruning) strategy with a multi-objective function to evaluate the relevance of an ensemble of trees. The function, used in a Hill Climbing process in Forward Selection (FS), allows selection of ensembles with the best compromise between maximum diversity and minimum error rate. The motivation behind the joint use of the two criteria is that there is a correlation between the individual performance of classifiers and their diversity. The more accurate the classifiers, the less they disagree. The use of one of the two properties is not sufficient to find the best performing ensemble.

\* Corresponding author.

E-mail address: souad.taleb@gmail.com

The proposed new multi-objective function is based on this compromise between individual performance of trees and their diversity. A comparison with UWA methods [36], Complementariness (Comp) [32], and Margin Distance Minimization (MARGIN) [33] shows, in most cases, that the proposed method allows generating ensembles that are both smaller in size and more efficient, than those of the methods cited above. This reduced number of trees allows a gain in memory space and computing time which can be very significant for large samples.

The paper is organized as follows: In Section 2 we present a state-of-the-art on ensembles selection methods with more details on UWA [36], Complementariness [32], and Margin Distance minimization [33].

These methods allow simplifying ensembles and will serve as a basis for comparison (The source code for these different methods is available at <http://mlkd.csd.auth.gr/ensemblepruning.html> [36]). In Section 3, we present the DIACES method, detailing the proposed new function as well as the path strategy. Section 4 contains all the experiments and analysis of the obtained results, whereas in the last section we conclude and propose some insights on future work.

## II. BAGGING, AGGREGATION, AND HILL CLIMBING

In this section, we highlight the basic elements used in this paper, namely, the bagging method used for the generation of the initial ensemble, aggregation by unweighted and weighted vote.

### A. Diversification by Bagging

Bagging Bootstrap Aggregating is a resampling method introduced by Breiman in 1996 [6]. Given a learning sample  $\Omega_L$  and a learning method which generates a predictor  $\hat{h}(\cdot, \Omega_L)$  using  $\Omega_L$ . The principle of bagging is to draw several bootstrap samples ( $\Omega_L^{o1}, \dots, \Omega_L^{on}$ ) and generate for each one a collection of predictors ( $\hat{h}(\cdot, \Omega_L^{o1}), \dots, \hat{h}(\cdot, \Omega_L^{on})$ ) using the base learning method for finally aggregating them.

A bootstrap sample  $\Omega_L^{o1}$  is obtained by randomly drawing  $n$  observations in the starting sample  $\Omega_L$ . Each observation has the probability of  $1/n$  of being shot;  $|\Omega_L| = n$ , the random variable  $\Theta_i$  represents the random drawing.

Initially, Bagging was introduced with a decision tree as basic rule. But the schema is general and can be apply to other basic rules. In Fig. 1 is presented the principle of Bagging.

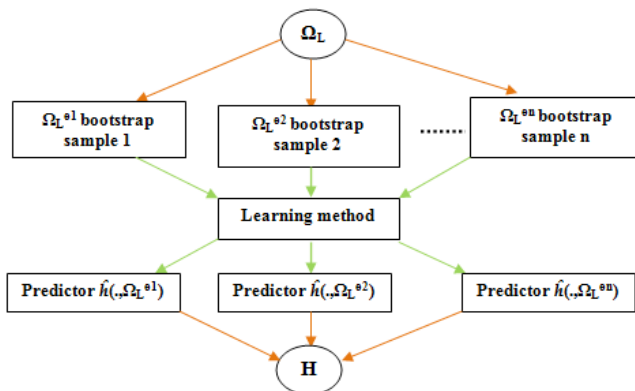


Fig. 1. Representative diagram of Bagging.

### B. Aggregation (Unweighted Vote, Weighted Vote)

The unweighted and weighted voting are the most used methods for combining (aggregating) whether homogenous or heterogeneous models. In ensemble methods each model, for an instance, gives a class value, a probability, and the class with most votes, highest average

probability is assigned to the instance by the ensemble.

In weighted vote, the classification models are associated with weights assigned relatively to their classification accuracy. Formally this can be written [36]:

Let  $x$  be an instance and  $m_{i, i=1..k}$  a set of models that output a probability distribution  $m_i(x, c_j)$  for each class  $c_j, j=1..n$ . The output of the (weighted) voting method  $y(x)$  for instance  $x$  is given by the following mathematical expression:

$$y(x) = \operatorname{argmax}_{c_j} \sum_{i=1}^k w_i m_i(x, c_j) \quad (1)$$

### C. Hill Climbing

Hill climbing is an optimization technique belonging to the family of local search. The algorithm starts with any solution to a problem, then tries iteratively to find a better solution by changing one element of the solution. If the change produces a better solution (maximize or minimize the evaluation function used for the course), an incremental change is made to the new solution. The process is repeated until no improvements can be found (the function reached the maximum or the minimum).

Hill climbing attempts to maximize (or minimize) a target function  $f(X)$  where  $X$  is a vector of continuous and/or discrete values. Each iteration, hill climbing will adjust a single element in  $X$  and determine if the change improves the value of  $f(X)$ . Any change improving the function  $f(X)$  is accepted, the process continues until no amelioration of the function can be found.

For ensemble selection, DHCEP (Directed Hill Climbing Ensemble Pruning) is used, in this case the vector  $X$  is composed of classifiers or predictors.

The course can be realized either in backward elimination or in forward selection, in the first case the whole ensemble is considered as a solution and then repeatedly elements not improving the evaluation function are eliminated one by one, in the second case we initialize with an element randomly and we add the elements that improve the evaluation function one by one. The elements to be added or removed are part of the neighborhood of the current solution. In Fig. 2, a hill climbing diagram for an ensemble composed of four models is presented.

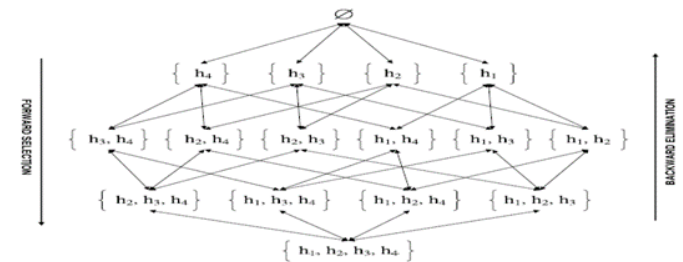


Fig. 2. Hill climbing search for selection in an ensemble composed of four classifiers [36].

## III. RELATED WORK

Several methods and heuristics have been proposed to reduce the size of a set of classifiers. They are categorized in four main classes: 1) ordering-based methods [31], [47], [33], [34], 2) clusters based methods [16], [21], [29], [19], 3) optimization based methods [49], [50], and 4) the others [31], [17], [10], [32], [3].

Among these methods, a large number for ensemble pruning based on a hill climbing research process have recently been proposed [31], [17], [10], [32], [3]. The methods differ from each other by the adopted

research directions, the different evaluation measures or the evaluation ensembles. It will be noted that some methods use the learning sample for evaluation while others promote the use of a separate validation set. The latter depends mainly on the availability of the data.

A first type of approaches uses performance measures. Fan et al. [17] propose a profit-based evaluation function and propose dynamic scheduling to accelerate the prediction process. For the reduction of the ensemble size, the total benefit is used as selection criterion in conjunction with a greedy search algorithm with and without back fitting. The path begins with the model with the greatest benefit. A set of instances  $x$  is considered, each instance  $x$  can be positive or negative,  $B(x)$  denotes the benefit of predicting  $x$  as positive and the total benefit  $BT = \sum_x B(x)$ , the authors choose the sub ensembles which maximize the total benefit.

Empirical evaluations of several data sets have revealed that the profit-driven greedy approach with or without back fitting eliminates 90% of the size of an ensemble maintaining or sometimes exceeding the total benefit on the test sample of the original ensemble. The authors have also studied the possibility of combining diversity and total benefit, but the experimental results have shown that the total benefit is a good criterion by itself.

Caruana et al. [10] use several performance metrics and a hill climbing strategy for building ensembles of models from libraries of thousands of models. Model libraries are generated using different learning algorithms. The Forward Stepwise selection consists to add models that maximize its performance.

The second type of approaches uses diversity-based measures. Martínez-Muñoz et al. [32] use diversity with a hill climbing forward selection process. The diversity measure called Complementariness is defined as the complementarity of a model  $h_k$  with respect to the current sub set and a set of instances of the evaluation sample  $Eval = \{(x_i, y_i)\}$  such that  $|Eval| = n$ , the measure COM is calculated as follows:

$$COM_{Eval}(h_k, Sub) = \sum_{i=1}^n I(y_i = h_k(x_i) \text{ and } y_i \neq Sub(x_i)) \quad (2)$$

Where  $I(\text{True}) = 1$ ,  $I(\text{False}) = 0$ ,  $Sub(x_i)$  is the classification of the instance  $x_i$  by the sub-set  $Sub$ . The measurement principle is to add to  $Sub$ , the model which allows classifying correctly the examples misclassified by the subset.

Martínez-Muñoz et al. [33] propose the minimization of the marginal distance which allows calculating the diversity by associating to each classifier  $h_i$  a vector  $c_i$  whose dimension is equal to the number of individuals of the evaluation sample. An element  $c_i(i)$  takes the value 1 if  $h_i$  properly classifies the individual  $i$  and -1 otherwise. An average vector  $C_{Sub}$ , associated with a subset  $Sub$  is calculated as  $C_{Sub} = \frac{1}{|Sub|} \sum_{t=1}^{|Sub|} c_t$ . The objective is to reduce the Euclidean distance  $d(o, C_{Sub})$  where  $o$  is a predefined vector. The measure that represents the margin is written:

$$MAR_{Eval}(h_k, Sub) = d(o, \frac{1}{|Sub|+1} (c_k + C_{Sub})) \quad (3)$$

Partalas et al. [35], [36] propose the Uncertainty Weighted Accuracy measure (UWA) in (4) that considers four cases when adding a model to a sub ensemble and using justified weights to distinguish favor cases from others.

$$UWA_{Eval}(h_k, Sub) = \sum_{i=1}^{|Eval|} (\alpha * I(y_i = h_k(x_i) \text{ ET } y_i \neq Sub(x_i)) - \beta * I(y_i \neq h_k(x_i) \text{ ET } y_i = Sub(x_i)) + \beta * I(y_i = h_k(x_i) \text{ ET } y_i = Sub(x_i)) - \alpha * I(y_i \neq h_k(x_i) \text{ ET } y_i \neq Sub(x_i))) \quad (4)$$

Where the parameters  $\alpha$ ,  $\beta$  represent respectively the number of models in the sub-set  $Sub$  correctly classifying the instance  $(x_i, y_i)$  and the number of models incorrectly classifying the same instance.

More recent related work [30] theoretically deals with the effect of diversity on voting generalization performance using Probably Approximately Correct (PAC) learning. It is revealed that diversity is closely related to the space complexity hypothesis, and strengthening it can be achieved by applying regularization to ensemble methods. Based on this analysis, the authors apply an explicit regularization of the diversity for the selection of ensembles.

Dai [12] proposes an improvement of the ensemble selection method of the same authors. This method uses backtracking in depth, which is perfectly adapted to systematically seek solutions to combinatorial problems of great magnitude. This improvement concerns the response time of this method, which has been considerably improved in this study.

Zhou et al. [51] propose a new algorithm based on frequent item learning that links data and the simplified ensemble to a transactional database whose transactions are instances and items are classifiers. A Boolean classification matrix is used for each model of the pruned ensemble. Using this matrix, several candidate ensembles are obtained by iterative and incremental extraction of basic classifiers with the best performances.

Bhatnagar et al. [4] perform ensemble selection using a performance-based and diversity-based function that considers the individual performance of classifiers as well as the diversity between pairs of classifiers. A bottom-up search is performed to generate the sub ensembles by adding various pairs of classifiers with high performance.

To simplify a set of classifiers usually involves reducing the number of trees while maximizing performance. Qian et al. [38] adopt the Pareto diagram to solve this two-goal problem using an evolutionary optimization method of Pareto combined with a local search operator. The method is applied in the field of mobile human activity recognition.

Based on the approximate ensembles, Guo et al. [22] propose a new framework for ensemble selection. In this context, the relationship between attributes in an approximate space is considered a priori as well as their degree of maximum dependence. This effectively reduces the search space and increases the diversity of selected sub-ensembles. Finally, to choose the appropriate sub-ensemble, an evaluation function that balances diversity and precision is used. The proposed method allows repetitively changing the search space of the relevant sub-ensembles and selecting the next sub-ensemble from a new search space.

Cavalcanti et al. [11] combine in pairs different matrices of diversity using a genetic algorithm. The combined diversity matrix is then used to group similar (not very diverse) models; they must not belong to the same ensemble. To generate candidate ensembles, the combined diversity matrix is transformed into one or more graphs and then a graph coloring technique is applied.

Guo et al. [23] propose a new metric using the margin (instances) and the diversity (of classifiers) to explicitly evaluate the importance of individual classifiers. By adding the models to the ensemble in decreasing order of the metric, the user can choose the first  $T$  models to form a sub-ensemble.

Dai et al. [13] emphasize the utility of optimizing predictive performance together with diversity, which are two indispensable and inseparable parameters for ensemble selection. There have been three measures proposed to simplify ensembles using a greedy algorithm: 1) The first measure simultaneously considers the difference (diversity) between the current subset and the candidate classifier and the performance of each one; 2) The second allows evaluating the diversity within the ensemble and; 3) the last measure reinforces the concern about the accuracy of the resulting sub-ensemble. Experimental results confirm the interest of the three measures which is illustrated by the improvement of performances.

#### IV. THE DIACES PROPOSED MEASURE

Our goal is to construct a distribution of the number of errors associated with each case and to calculate the diversity of this distribution. Our goal is to minimize diversity while maintaining good performance for each classifier.

The set of data  $\Omega$  is divided into two sub samples  $\Omega_L$  (generally 80% of  $\Omega$ ) for learning and pruning and  $\Omega_T$  (generally 20% of  $\Omega$ ) for testing. A bagging ensemble BE of  $t$  C4.5 trees is constructed,  $BE = \{T_1, \dots, T_t\}$ , using  $\Omega_L$  with  $|\Omega_L| = n$ . Each tree  $T_i$  is represented by a vector  $(x_{i1}, x_{i2}, \dots, x_{ij}, \dots, x_{in})^T$ . We have the following notations:

- $x_{ij}$ : Result of classification of the individual  $i$  by the tree  $j$ ,  $x_{ij} = 1$  if the individual  $i$  is misclassified by the tree  $T_j$  and  $x_{ij} = 0$  otherwise,
- $x_{i+}$ : The total number of errors committed for the individual  $i$ :  $x_{i+} = \sum_{j=1}^t x_{ij}$
- $X$ : The total number of errors committed by the set:  $X = \sum_{i=1}^n \sum_{j=1}^t x_{ij}$
- $(\theta_{ix_{i+}})$ : The relative distribution of the error frequencies associated with the different cases:  $\theta_i = \frac{x_{i+}}{X}, i = 1, n$
- $x_{+j}$ : The number of errors committed by the classifier  $T_j$  over all the individuals:  $x_{+j} = \sum_{i=1}^n x_{ij}$
- $e_j$ : The error rate associated with the tree  $T_j$ :  $e_j = \frac{x_{+j}}{n}$

The evaluation function to optimize noted  $S$  connects diversity  $\theta_i$  and the error rate  $e_j$ :  $S = \sum_{i=1}^n \theta_i^2 + \sum_{j=1,t} e_j^2$ .

The component  $C = \sum_{i=1}^n \theta_i^2$  is a concentration index of error distribution which is derived from the quadratic entropy (or Gini index). The smallest value of  $C$  is  $1/n$ , where all  $x_{i+}$  have the same value. This situation is best conditioned by the value of  $X$ . The minimization of  $C+E$ , using two metrics error rates and concentration index of errors distribution, allows having a good compromise between the diversity of the trees and their average performance.

$C^*$  represents the normalized coefficient  $\Rightarrow C^* = \frac{nx - x^2}{nk - x}$

$E^*$  represents the normalized coefficient  $\Rightarrow E^* = \frac{kn^2 E - x^2}{xkn - x}$

Finally the function to minimize  $S = C^* + \alpha E^*$  ( $\alpha$  is a parameter determined empirically by the user).

Algorithm 1 presents the proposed method DIACES in a pseudo code:

Algorithm1 DIACES;

Input

$BE = \{T_1, \dots, T_t\}$ ;

$\Omega_L$ : selection set;

Neighborhood( $\psi_j$ ): Function that returns the subsets of models obtained from  $\psi_j$  by adding a classifier (tree);

Output

Sub ensemble  $\psi_0$  of BE;

Begin

Initialize( $\psi_0$ );

1. Calculate  $S(\psi_0, \Omega_L)$ ;

if  $\exists \psi_j$  such as  $S(\psi_j, \Omega_L) < S(\psi_0, \Omega_L)$  where  $\psi_j \in \text{Neighborhood}$

( $\psi_0$ ) Then  $\psi_0 = \arg\min_{\psi_j} (S(\psi_j, \Omega_L))$ ;

Goto 1;

End.

The algorithm complexity consists in calculating the hill climbing path method complexity which is  $O(k^2)$  where  $k$  is the number of classifiers of the ensemble. The function  $S$  is computed from a matrix

composed of  $n$  rows (number of individuals in the validation set), its complexity is  $O(n)$ , the calculation of the function is repeated  $k'$  times where  $k'$  is the number of ensembles traveled in the hill climbing scheme. The complexity of the proposed method is  $O(n * k' * k^2)$ .

#### A. Initialization and Path

The path strategy used by our method is a hill climbing strategy whose principle is simple [44]. It consists in reducing the number of ensembles generated in the case of an exhaustive path in which  $2^k$  of ensembles are explored ( $k$  is the number of classifiers).

The hill climbing allows obtaining a sub-optimal solution by going through  $\frac{k(k+1)}{2}$  subsets, considering a set of states and selecting the next state to be visited from the neighborhood of the current state. In this case, the states are the different ensembles of models and the neighborhood of a subset Sub of BE (set of all hypotheses) is composed of ensembles constructed by adding (forward selection) or deleting (backward elimination) a model of Sub. The method goes across the search space (all subsets of models) from one end to the other; one of the two ends is composed of the empty set and the other of the set of all the models. The complexity of hill climbing is  $O(k^2)$ .

In our case, we go across the set in forward selection and for the initialization we choose a tree that is not very good and not very bad. The fact of not choosing the most precise tree, as is the case for many methods that adopt a hill climbing method for selection in a set, is due to the fact that in some cases the most accurate tree can be perfect (does not make any error) on the evaluation sample, consequently a subset is produced with a single tree (the initialization tree) which can be very bad in generalization.

The proposed solution consists in choosing a tree with “average” performances on the evaluation sample. First the  $k$  trees of the initial set BE are ordered using performance on the evaluation sample in ascending or decreasing way, the ordered set B is then decomposed in 3 subsets BE1, BE2, and BE3, finally the initialization tree is randomly selected from the subset BE2.

According to experimental studies that we have carried out, the proposed multi-objective function directly affects the error rate in generalization (on the test sample  $\Omega_T$ ). In Fig. 3, the curve shows the correlation between the function and the error rate for the Ionosphere dataset.

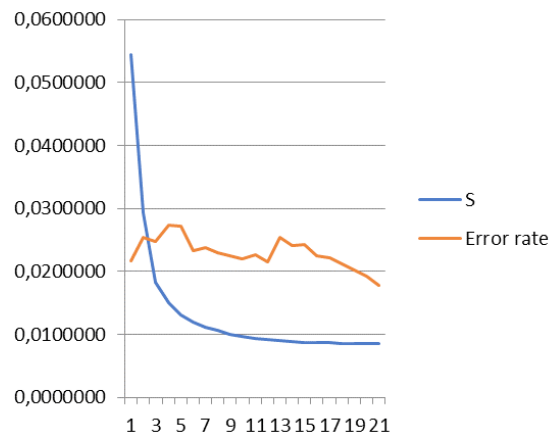


Fig. 3. Correlation between the function  $S$  and the error rate on  $\Omega_T$  for Ionosphere dataset.

#### V. EXPERIMENTS

Experiments consist in building homogeneous sets by sampling the starting sample and using the C4.5 decision tree generation algorithm as a basic rule.



## A. Materials and Methods

The Weka platform [46] is used as a source for the C4.5 learning algorithm and validation. For this purpose, we consider 24 Benchmarks of the UCI Repository [2] which are described in Table I, the table is composed of five columns:

ID: Identifier of the data set, NI: Number of Instances, ND: Number of Descriptors, CM: Class Modalities,  $\alpha$ : the value of the parameter  $\alpha$  for each data set.

TABLE I. DESCRIPTION OF THE DATA SETS USED FOR THE COMPARISON

	id	NI	ND	CM	$\alpha$
Audiology	b1	226	69	24	181
Breast-Cancer	b2	286	9	2	229
Breast w	b3	699	9	2	559
CMC	b4	1473	9	3	1178
Diabetes	b5	768	8	2	614
Credit-a	b6	690	15	2	552
Credit-g	b7	1000	20	2	800
Heart-Statlog	b8	270	13	2	216
Anneal	b9	798	38	6	638
Balance-Scale	b10	625	4	3	500
Colic	b11	368	22	2	294
Haberman	b12	306	3	2	245
Titanic	b13	390	13	2	312
Primarytumor	b14	339	17	2	271
Sonar	b15	195	60	2	156
Soybean	b16	683	35	19	546
Vehicle	b17	946	18	4	757
Vote	b18	435	16	2	348
Vowel	b19	990	13	11	792
Autos	b20	205	25	6	164
Glass	b21	214	9	6	171
Hearth-h	b22	294	13	5	235
Ionosphere	b23	351	34	2	281
Lymph	b24	148	18	4	118

The initial sample decomposition is based on the results presented in [37]. It concludes that for the case of simplification of homogeneous set, pruning on the same learning sample allows obtaining better results than using separate samples, which is also an advantage in the case of unavailability of large amounts of data.

The initial sample is subdivided into two sub-samples, one of which is composed of 80% of the individuals and will be used for learning (model generation) and evaluation (pruning), while the remaining 20% will be used for testing.

The method we propose, DIACES, is compared to a set of ensemble pruning methods based on diversity: UWA, COM, and MARGIN, detailed the state of art. For all these methods, the unweighted majority vote is used for the combination of models and the performance calculation.

The methods use a forward selection strategy in a hill climbing scheme. The stop criterion for methods in literature is the performance on the evaluation sample which generates subsets of reduced sizes compared with the usual stop criteria defined as a fixed number of models [35]. In our case we use the same function for both the path and the stop.

## B. Results and Analysis

A first criterion to compare the different pruning methods is the performance of the subsets obtained. For each method and each dataset, 10 different draws are made for which the averages of the success rates obtained for each draw are calculated.

For 24 benchmarks, DIACES shows better performances in 15 cases followed by the COM method with 6 victories, 3 victories for UWA and finally MARGIN with two victories. For the average success rate on all datasets, DIACES is ranked first with an average rate of 80.48%, exceeding the other methods with a rate of at least 0.9%.

Compared to UWA, DIACES improves performance by about 3% on the Audiology, Breast cancer, Vehicle, Primary Tumor, and Titanic datasets. Moreover, DIACES improves the performances obtained by the COM and MARGIN methods by at least 0.2%, 4% on Audiology data set.

To ensure whether the performance differences between DIACES and the other methods are significant; we will use the statistical test of the sign. The test allows ensuring the effectiveness of the proposed method for any data collection. The test considers the method achieving the best success rate and compares it to DIACES. Based on the results presented in Table II, we compare DIACES to COM which gives an average success rate of 79.87% following the next steps:

1. Calculate the performance difference D between DIACES and COM methods.
2. Calculate the p-value, using the sign test:

Success number 7

p-value 0.04

The p-value < 0.05 (0.05 being the risk value) implies that the performance difference between DIACES and COM is significant, and consequently DIACES is better than all the other methods.

Depending on the average success rates values, a rank is assigned to each method for each benchmark. In this way the different methods are compared according to their average ranks which is an appropriate criterion of comparison [14], the ranks and average ranks of the methods are presented in Table III. AVR represents the average rank of the generated sub ensembles based on success rates.

TABLE II. COMPARING SUCCESS RATES FOR DIFFERENT METHODS

	DIACES	UWA	COM	MARGIN
b1	<b>0.85432</b>	0.81773	0.80661	0.81108
b2	0.72618	<b>0.72977</b>	0.72631	0.72205
b3	<b>0.97459</b>	0.95032	0.95248	0.9496
b4	0.52106	0.52617	0.52515	<b>0.52651</b>
b5	<b>0.77519</b>	0.76469	0.76727	0.76524
b6	0.85329	0.86085	<b>0.86882</b>	0.86156
b7	<b>0.742</b>	0.73	<b>0.7365</b>	0.729
b8	0.81857	0.81848	<b>0.82034</b>	0.81478
b9	<b>0.99651</b>	0.9916	0.99048	0.99104
b10	0.8145	0.8144	<b>0.8264</b>	0.8208
b11	0.81598	0.84245	<b>0.83423</b>	0.83012
b12	<b>0.73913</b>	0.73118	0.7131	0.72131
b13	<b>0.75509</b>	0.72176	0.73075	0.73331
b14	<b>0.45732</b>	0.41787	0.41587	0.42685
b15	<b>0.79043</b>	0.78048	0.77803	0.77071
b16	<b>0.9537</b>	0.93676	0.94705	0.93897
b17	<b>0.78626</b>	0.74612	0.74198	0.74553
b18	0.95571	0.954	0.95171	<b>0.9586</b>
b19	0.89749	<b>0.92016</b>	0.91058	0.91866
b20	0.81362	<b>0.81703</b>	0.80483	0.80972
b21	<b>0.789</b>	0.7619	0.77284	0.77047
b22	<b>0.80951</b>	0.79652	0.80858	0.80687
b23	<b>0.9235</b>	0.90856	0.90142	0.91998
b24	<b>0.84</b>	0.81376	0.83789	0.82411
AVSR	0.8084	0.7980	0.7987	0.7986

TABLE III. RANKING METHODS BASED ON SUCCESS RATE

	DIACES	UWA	COM	MARGIN
b1	1	2	4	3
b2	3	1	2	4
b3	1	3	2	4
b4	4	2	3	1
b5	1	4	2	3
b6	4	3	1	2
b7	1	3	2	4
b8	2	3	1	4
b9	1	2	4	3
b10	3	4	1	2
b11	4	1	2	3
b12	1	2	4	3
b13	1	4	3	2
b14	1	3	4	2
b15	1	2	3	4
b16	1	4	2	3
b17	1	2	4	3
b18	2	3	4	1
b19	4	1	3	2
b20	2	1	4	3
b21	1	4	2	3
b22	1	4	2	3
b23	1	3	4	2
b24	1	4	2	3
AVR	<b>1.81</b>	2.81	2.71	2.67

We note that the proposed method has the best average rank with 1.81, followed by MARGIN method with 2.67, COM comes in the third position with 2.67, and finally the UWA method with 2.81.

A second comparison criterion is the size of the ensemble obtained. For each benchmark and on 10 draws, the average sizes of the subsets obtained is calculated.

Table IV shows the average sizes (over 10 iterations) of the sub sets obtained for each method. The number of models selected is reduced for all methods compared to the original size of the ensemble.

TABLE IV. SUB ENSEMBLES SIZES GENERATED USING THE DIFFERENT METHODS

	DIACES	UWA	COM	MARGIN
b1	<b>10.2</b>	14.6	14	21.7
b2	<b>11.2</b>	11.4	11.3	15.5
b3	<b>9.8</b>	11.7	13.1	14.3
b4	<b>18</b>	48.7	38.2	42.1
b5	<b>14.2</b>	20.4	26.3	36.3
b6	<b>13.3</b>	16.1	18.7	18
b7	<b>16.9</b>	23.2	25.7	33.2
b8	<b>12.7</b>	14.7	17.9	18.7
b9	<b>2.6</b>	3.4	2.9	13.8
b10	<b>14.7</b>	26.9	22.4	21.8
b11	12.2	8.2	<b>4.7</b>	7
b12	12.5	13.5	15.6	24.1
b13	<b>15.2</b>	22.4	24	34.1
b14	<b>13</b>	53	37.2	33.6
b15	10.7	<b>7</b>	8.8	15
b16	<b>12.2</b>	<b>12.2</b>	19.3	23.5
b17	<b>13.4</b>	31.6	33.8	37.7
b18	10.8	<b>4.6</b>	9.4	11
b19	<b>12.4</b>	18.8	13.1	22.7
b20	<b>11.00</b>	17	14.5	20
b21	<b>13</b>	15.9	20.6	16.2
b22	<b>13.2</b>	15.7	17.2	19.7
b23	10.6	<b>6.7</b>	7.9	12.7
b24	10.9	<b>10.6</b>	15.3	21.5
AVSZ	<b>12.28</b>	17.84	17.99	22.26

The reductions for all methods compared to the initial ensemble of all models vary between 93% and 97%.

The new method allows obtaining subsets of reduced sizes compared to the other methods for 17 data sets, 66% of cases; UWA comes in second place with 6 victories, and finally the COM method which counts one victory.

The study of algorithmic complexity makes it possible to calculate the quantity of resources (time or space) necessary for its execution. The complexities in time of the proposed method are calculated on the 24 data sets (For each one of them it is a question of calculating the average of time and space on 10 iterations) on a machine treating 109 instructions/seconds (1 Gigahertz) and a memory of 3 Gigabytes. The Hill Climbing search method is fast because it does not cover all the possible case combinations, the maximum time is calculated for cmc and credit-g data sets for which the search lasts 5.12 ms, the minimum times are 0.15 ms for the anneal data set; For other data sets the times vary between 3.7 ms and 0.5 ms. The 24 data sets have a total time of 0.82 seconds.

## VI. CONCLUSION AND FUTURE WORKS

The ensemble methods improve the performance of an unstable classifier but have the disadvantage of the loss of readability of the model provided; Composed of a large number of distinct trees and therefore more difficult to synthesize by humans.

This is why other methods have also been proposed to synthesize not the results but the structure of a tree in the form of a “consensus” from a set of classifiers of the same type [42] [45] nevertheless with a deterioration in the quality of prediction.

In this paper we presented a new evaluation function combining performance and diversity for selection in a homogeneous ensemble used in a process of climbing hill path. The method was evaluated on several benchmarks and compared to pruning homogeneous ensembles in literature.

The results show that the proposed method obtains ensembles with performances exceeding the ensembles obtained by the compared methods. In addition the interest of the function is that it can be used equally with homogeneous sets as well as heterogeneous sets (the basic rules are not the same).

We plan, in future work, to use this new pruning measure and apply it for selection in a random forest ensemble, knowing that a random forest ensemble improves the performance of a bagging [7]. We also propose to study the possibility of using another path strategy for the selection. The third contribution consists in using the method in the field of predicting student performance in education institutions and comparing the results with those obtained in [24]. In their work [48] use various classification methods (Neural networks, Kppv) separately for the diagnosis of breast cancer. We propose to use a heterogeneous ensemble composed of several classification methods for the same application. The ensemble will then be simplified by using the proposed measure.

The last point consists in finding a value for the parameter  $\alpha$  for which we have noticed during empirical research that an appropriate value, determined in a non-empirical way, could significantly improve the results as observed in our experiments, for this we will use the Pareto principle.

## REFERENCES

- [1] Y. Amit, D. Geman, “Shape quantization and recognition with randomized trees,” *Neural Computation*, (9), 1997, pp. 1545–1588.
- [2] D.N.A. Asuncion. (2007). UCI machine learning repository. Available at <http://www.ics.uci.edu/~mllearn/MLRepository>.

- [3] R.E. Banfield, L.O. Hal, K.W. Bowyer, W.P. Kegelmeyer, "Ensemble diversity measures and their application to thinning," *Information Fusion*, 6 (1), 2005, pp. 49–62.
- [4] V. Bhatnagar, M. Bhardwaj, S. Sharma, S. Haroon, "Accuracy-diversity based pruning of classifier ensembles," *Progress in Artificial Intelligence*, 2(2-3), 2014, pp. 97–111.
- [5] L. Breiman, J.H. Friedman, R.A. Olshen, C.J. Stone, *Classification and Regression Trees*, Chapman and Hall, New York, 1984.
- [6] L. Breiman, "Bagging Predictors," *Machine Learning*, 26, No. 2, 1996, 123-140.
- [7] L. Breiman, "Random Forests," *Machine Learning*, 45(1), 2001, pp. 5-32.
- [8] B.G. Buchanan, E.H. Shortliffe, "Rule Based Expert Systems," Addison-Wesley, Reading, Massachusetts, 1984, pp. 288-291.
- [9] P.L. Bogier, "Shafer-Dempster reasoning with applications to multisensor target identification," *IEEE Trans. Sys. Man. Cyb. SMC-17*, 1984, pp. 968-977.
- [10] R. Caruana, A. Niculescu-Mizil, G. Crew, A. Ksikes, "Ensemble selection from libraries of models," In *Proceedings of the 21st international conference on machine learning*, 2004, pp. 18.
- [11] G.D.C. Cavalcanti, L.S. Oliveira, T.J.M. Moura, G.V. Carvalho, "Combining diversity measures for ensemble pruning," In *Pattern Recognition Letters*, Volume 74, 2016, pp. 38-45.
- [12] Q. Dai, "An efficient ensemble pruning algorithm using One-Path and Two-Trips searching approach," In *Knowledge-Based Systems*, Volume 51, 2013, pp. 85-92.
- [13] Q. Dai, R. Ye, Z. Liu, "Considering diversity and accuracy simultaneously for ensemble pruning," In *Applied Soft Computing*, Volume 58, 2017, pp. 75-91.
- [14] J. Demsar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine Learning Research*, 7, 2006, pp. 1–30.
- [15] T. Dietterich, "Ensemble Methods in Machine Learning," *Lecture Notes in Computer Science*, 1857, 2000, pp. 1-15.
- [16] P. Domingos, "Knowledge acquisition from examples via multiple models," In: *Proc. 14th International Conference on Machine Learning*, Morgan Kaufmann, 1997, pp. 98–106.
- [17] W. Fan, F. Chu, H. Wang, P.S. Yu, "Pruning and dynamic scheduling of cost-sensitive ensembles," In *Eighteenth national conference on artificial intelligence*, American association for artificial intelligence, 2002, pp. 146–151.
- [18] Y. Freund, R.E. Schapire, "A Short Introduction to Boosting," *J. Japanese Soc. Artificial Intelligence Res.*, 14(5), 1999, pp. 771-780 (1999).
- [19] Q. Fu, S.X. Hu, S.Y. Zhao, "Clusterin-based selective neural network ensemble," *Journal of Zhejiang University SCIENCE6A*(5), 2005, pp. 387-392.
- [20] P. Geurts, "Contributions to decision tree induction: bias/variance tradeoff and time series classification," Phd. Thesis, Department of Electronical Engineering and Computer Science, Liège Univ of Liège, Belgium, 2002.
- [21] G. Giacinto, F. Roli, G. Fumera, "Design of effective multiple classifier systems by clustering of classifiers," In: *15th International Conference on Pattern Recognition*, ICPR 2000, 2000, pp. 160-163.
- [22] Y. Guo, L. Jiao, S. Wang, F. Liu, K. Rong, T. Xiong, "A novel dynamic rough subspace based selective ensemble," In *Pattern Recognition*, Volume 48, Issue 5, 2015, pp. 1638-1652.
- [23] H. Guo, H. Liu, R. Li, C. Wu, Y. Guo, M. Xu, "Margin and diversity based ordering ensemble pruning," In *Neurocomputing*, 2017; ISSN 0925-2312.
- [24] A.K. Hamoud, A. S. Hashim, and W. A. Awadh, "Predicting Student Performance in Higher Education Institutions Using Decision Tree Analysis," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. In Press, issue In Press, no. In Press, pp. 1-6, 02/2018.
- [25] T.K. Ho, J.J. Hull, S.N. Srihari, "Decision Combination in Multiple Classifier Systems," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 16, no. 1, 1994, pp. 66-75.
- [26] T.K. Ho, "Random Decision Forests," *Proc. Third Int'l Conf, Document Analysis and Recognition*, 1995, pp. 278-282.
- [27] T.K. Ho, "The random subspace method for constructing decision forests," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(8), 1998, pp. 832-844.
- [28] S. Kwok, C. Carter, "Multiple decision trees," *Uncertainty in Artificial Intelligence 4*, ed. Shachter, R., Levitt, T., Kanal, L., and Lemmer, J., North-Holland, 1990, pp. 327-335.
- [29] A. Lazarevic, Z. Obradovic, "The effective pruning of neural network classifiers," In: *2001 IEEE/INNS International Conference on Neural Networks, IJCNN 2001*, pp. 796-801.
- [30] N. Li, Y. Yu, H. Zhou, "Diversity Regularized Ensemble Pruning," In *Proceedings of the 23rd European Conference on Machine Learning and Knowledge Discovery in Databases - Volume Part I (ECMLPKDD'12)*, Peter A. Flach, Tijl De Bie, and Nello Cristianini (Eds.), Vol. Part I. Springer-Verlag, Berlin, Heidelberg, 2012, pp. 330-345.
- [31] D.D. Margineantu, T.G. Dietterich, "Pruning adaptive boosting," In: *Proc Of the 14th International Conference on Machine Learning*, 1997, pp. 211-218.
- [32] G. Martínez-Muñoz, A. Suarez, "Aggregation ordering in bagging," In *International Conference on Artificial Intelligence and Applications (IASTED)*, 2004, pp. 258–263.
- [33] G. Martínez-Muñoz, A. Suarez, "Pruning in ordered bagging ensembles," In *23rd international conference in machine learning (ICML-2006)*, 2006, pp. 609–616.
- [34] G. Martínez-Muñoz, A. Suarez, "Using boosting to prune bagging ensembles," *Pattern Recognition Letters* 28 (1), 2007, pp. 156–165.
- [35] I. Partalas, G. Tsoumakas, I. Vlahavas, "Focused ensemble selection: A diversity-based method for greedy ensemble selection," In: M. Ghallab, C.D. Spyropoulos, N. Fakotakis, N.M. Avouris (eds.) *ECAI 2008 - 18th European Conference on Artificial Intelligence*, Patras, Greece, *Proceedings, Frontiers in Artificial Intelligence and Applications*, vol. 178, 2008, pp. 117-121.
- [36] I. Partalas, G. Tsoumakas, I. Vlahavas, "An ensemble uncertainty aware measure for directed hill climbing ensemble pruning," *Machine Learning*, 81, 2010, pp. 257–282.
- [37] I. Partalas, G. Tsoumakas, I. Vlahavas, "A Study on Greedy Algorithms for Ensemble Pruning," *Technical Report TR-LPIS-360-12*, LPIS, Dept. of Informatics, Aristotle University of Thessaloniki, Greece, 2012.
- [38] C. Qian, Y. Yu, Z.H. Zhou, "Pareto ensemble pruning," In: *AAAI Conference on Artificial Intelligence*, 2015.
- [39] J.R. Quinlan, "C4.5: Programs For Machine Learning," Morgan Kaufmann Publishers, San Mateo, California, 1993.
- [40] R.E. Schapire, Y. Freund, P. Bartlett, "Lee W.S., Boosting the margin: a new explanation for the effectiveness of voting methods," In Douglas H. Fisher, editor, *Machine Learning: Proceedings of the Fourteenth International Conference (ICML'97)*, 1997, pp. 322–330, Morgan Kaufmann.
- [41] Settouti, N., M. E. A. Bechar, and M. A. Chikh, "Statistical Comparisons of the Top 10 Algorithms in Data Mining for Classification Task," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, issue Special Issue on Artificial Intelligence Underpinning, no. 1, pp. 46-51, 09/2016.
- [42] W.D. Shannon, D. Banks, "Combining classification trees using MLE," *Statist. Med.*, 18(6), 1999, pp. 727-740.41
- [43] S. Shlien, "Non parametric classification using matched binary decision trees," *Pattern Recognition Letters*, 13(2), 1992, pp. 83-88.
- [44] G. Tsoumakas, I. Partalas, I. Vlahavas, "An Ensemble Pruning Primer," *Applications of Supervised and Unsupervised Ensemble Methods* (Eds.) Okun and Valentino, 2009, pp. 1-13, Springer-Verlag.
- [45] J. T. L. Wang, K. Zhang, "Finding similar consensus between trees: an algorithm and a distance hierarchy," *Pattern Recognition*, 2001, 34:127.137.
- [46] I. H. Witten, E. Frank, *Data Mining: Practical machine learning tools and techniques*, 2nd Edition, Morgan Kaufmann, 2005.
- [47] Y. Yang, K. Korb, K. Ting, "Webb G., Ensemble selection for superparent-one-dependence estimators," In: *AI 2005: Advances in Artificial Intelligence*, 2005, pp. 102-112.
- [48] Youh, H., and G. Rumbe, "Comparative Study of Classification Techniques on Breast Cancer FNA Biopsy Data," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 1, issue A Direct Path to Intelligent Tools, no. 3, pp. 5-12, 12/2010.
- [49] H. Zhou, J. Wu, W. Tang, "Ensembling neural networks: Many could better than all," *Artificial intelligence* 137 (1-2), 2002, pp. 239-263.
- [50] H. Zhou, W. Tang, "Selective ensemble of decision trees," In: *9th International Conference on Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing*, 2003, pp. 476-483.
- [51] H. Zhou, X. Zhao, X. Wang, "An effective ensemble pruning algorithm based on frequent patterns," In *Knowledge-Based Systems*, Volume 56, 2014, pp. 79-85.



[Souad Taleb Zouggar](#)

Souad Taleb Zouggar is Assistant Professor in Computer Science at University of Oran 2, Algeria. She received a PhD in Computer Science from University of Oran in 2014. She has published papers on Machine Learning, Data mining. Her research interests focus on Knowledge Discovery in Databases, Machine Learning, Optimization, and Decision Support systems.



[Abdelkader Adla](#)

Abdelkader Adla is full Professor in Computer Science at University of Oran 1, Algeria. He received a PhD in Computer Science, Artificial Intelligence from Paul Sabatier University Toulouse III, France. He received also a State Doctorate in Computer-Aided Design and Simulation from University of Oran in 2007. He has published papers on collaborative decision making, Decision Support Systems (DSS), distributed group DSS and multi-agents DSS. His research interests focus on Group DSS, Group work facilitation, Cooperative and collaborative systems, Knowledge and Multi-agent decision support systems.



# Improved Behavior Monitoring and Classification Using Cues Parameters Extraction from Camera Array Images

Ahmad Jalal, Shaharyar Kamal\*

Department of Computer Science and Engineering, Air University, Islamabad (Pakistan)

Received 26 February 2018 | Accepted 6 July 2018 | Published 20 July 2018



## ABSTRACT

Behavior monitoring and classification is a mechanism used to automatically identify or verify individual based on their human detection, tracking and behavior recognition from video sequences captured by a depth camera. In this paper, we designed a system that precisely classifies the nature of 3D body postures obtained by Kinect using an advanced recognizer. We proposed novel features that are suitable for depth data. These features are robust to noise, invariant to translation and scaling, and capable of monitoring fast human body-parts movements. Lastly, advanced hidden Markov model is used to recognize different activities. In the extensive experiments, we have seen that our system consistently outperforms over three depth-based behavior datasets, i.e., IM-DailyDepthActivity, MSRDailyActivity3D and MSRAction3D in both posture classification and behavior recognition. Moreover, our system handles subject's body parts rotation, self-occlusion and body parts missing which significantly track complex activities and improve recognition rate. Due to easy accessible, low-cost and friendly deployment process of depth camera, the proposed system can be applied over various consumer-applications including patient-monitoring system, automatic video surveillance, smart homes/offices and 3D games.

## KEYWORDS

Activity Recognition,  
Body Posture  
Recognition System,  
Pattern Clustering,  
SmartCities.

DOI: 10.9781/ijimai.2018.07.003

## I. INTRODUCTION

**I**DENTIFICATION, monitoring, classification and recognition of human from behavior images is very necessary as it is very effective to convey subject's situation, identity, emotion, gait and gestures [1-4]. Still human identification and monitoring is not absolutely perfect in various conditions such as position changes, illumination, orientation, noise variations and dark-area places [5-8]. In spite of the research efforts and significant results in the past decade, recognition accuracy of human behavior still remains a challenge because of self-occlusion of human body parts, variation of body size and appearance, un-clear or hidden body parts behind objects and fast human movements during indoor scenes decade [9, 10]. In addition, several researchers mainly focused on recognizing activities from videos captured by conventional cameras which are less effective due to complex backgrounds, light sensitivity and motion ambiguities (i.e. color and texture variability) [11-13]. Thus, to access the high quality imaging and 3D motions, the development of low-cost and easy-processing depth cameras such as Microsoft Kinect or bumblebee, have initiated new era for a variety of image recognition tasks including human behavior recognition (BR) [14-16]. Depth images provide several opportunities to enhance BR such as additional body joints information, spatial continuity, insensitivity to lighting conditions and controlling overlapping issues of different human body parts.

A large number of methods have been designed for efficient BR method and also a lot of comparative studies were evaluated by series

of researchers over depth videos [16-18] to examine the best algorithms for recognition. These methods mainly interact with depth data using two different approaches: skeleton joints features and depth silhouette features. For example, Oreifej and Liu [19] proposed a new descriptor for behavior recognition using a histogram capturing the distribution of the surface normal orientation in the 4D space of time, depth, and spatial coordinates. To build the histogram, they created 4D projectors, which quantize the 4D space and represent the possible directions for the 4D normal. In [17], Yang et al described an effective method that project depth maps onto three orthogonal planes and accumulate global activities through entire video sequences to generate the Depth Motion Maps (DMM). Histograms of Oriented Gradients (HOG) are then computed to enhance the activity recognition results. In [20], authors proposed a behavior recognition system that deals with motion features as magnitude and directional angular features from body joints information between consecutive frames to recognize daily routine human activities. In [21], authors designed mid-level features from Kinect skeletons by considering the orientations of human body limbs connected by two skeleton joints and each orientation is encoded into different states. They employed frequent pattern mining to pick the most frequent feature values, relevant states of parts in continuous several frames and recognize different activity/actions.

However, such methods show better performance and contributions, but different factors having negative impact surrounded each method. Those methods just relied on the skeleton data which became unreliable for postures with self-occlusion. Also, some methods were depended on depth silhouettes information which causes low recognition accuracy especially in case of hidden or missing body parts, fast moving human silhouettes and large distance of subject from the source (i.e. depth camera). Therefore, we elaborate some novel features along with

\* Corresponding author.

E-mail address: shaharyar.kamal@mail.au.edu.pk

advanced HMM to overcome the above mentioned problems and improve recognition accuracy.

In this paper, we propose a novel behavior recognition framework based on cues-parameters, which has an improved accuracy over existing algorithms. At the start of the BR framework, we handle the noisy input posture and unclear background data by designing a set of reliability measurement to extract true silhouettes and tracked joint values. These true data is examined to extract human silhouette by considering spatial/temporal continuity, constraints of human motion information and frame differentiation. These data are further processed to get feature representation by considering cues-parameters including angular direction, spatiotemporal velocity and invariant features which provide compact and sufficient feature values for better BR performance. While, all feature values are mapped into codewords and recognized each behavior via advanced Hidden Markov model (HMM). We evaluate our method according to the standard experimental protocols definition on three challenging depth behavior datasets: IM-DailyDepthActivity, MSRDailyActivity3D and MSRAction3D. Our experimental results show that the proposed method is able to achieve better recognition accuracy than the state-of-the-art methods. Since our system is well-organized, affordable and easily installable, therefore, it is the preferable solution for reliable body tracking, orientation, smart environments and behavior recognition systems.

The organization of this paper is as follows. Section II elaborates the system architecture of the proposed method starting from depth image preprocessing, evaluation of feature extraction by cues-parameters and behavior training/recognition via advanced HMMs. Experimental results and comparisons between proposed and state of the art methods are described in Section III. Finally, we conclude this work in Section IV.

## II. SYSTEM ARCHITECTURE AND DESCRIPTION

The proposed behavior recognition system consists of various steps as: 1) raw depth data captured by RGB-D video sensor, 2) noisy background removal, 3) human detection, tracking and identification from the time-sequential behavior video images. 4) feature extraction based on cues-parameters techniques, clustering using Linde, Buzo, and Gray (LBG)'s algorithm and training/recognition using advanced HMM. Fig. 1 shows the overall flow of our proposed BR system.

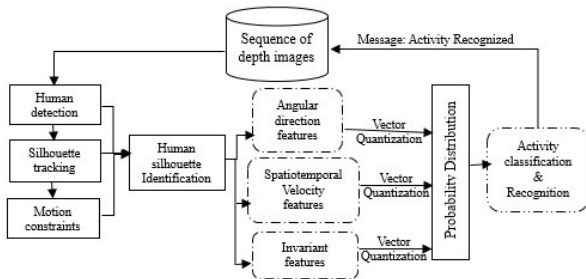


Fig. 1. Overall flow of the proposed human behavior recognition system.

### A. Depth Image Processing

For the human identification in depth sequential maps, background subtraction routine is applied which consists of least squares method for estimating the angle and center point of the floor in a real world coordinate system [22]. The depth value  $y$  in a spaced grid having least value is used to ignore floor from noisy background. Then, we have localized different objects in the scene and segmented them by computing the modified connected component labeling (CCL) method and examining degree of freedom [23, 24]. It is used to label all candidates pixels separately. Finally, we extract the human moving

silhouettes, temporal depth intensity differentiation [22] is applied as expressed in (1) to obtain the depth human silhouettes from consecutive frames.

$$D = \sqrt{(I_i^x - I_{i-1}^x)^2 + (I_i^y - I_{i-1}^y)^2 + (I_i^z - I_{i-1}^z)^2} \quad (1)$$

To properly track the entire human body, we performed disparity segmentation. Ignoring the 0's, we consider the average of the disparity values in the detected moving parts and compare neighboring pixels surrounded by the detected moving parts to add the pixels with closure disparity values that make a separate region (i.e. human silhouettes). Thus, the disparity segmentation is employed to find the target human silhouette candidates and subjects which are free to move more naturally.

Overall system's framework allows a smart environment to analysis what the user is doing from the noisy data obtained from the depth camera. We implement the depth motion database targeting at different behaviors. The database includes correctly and incorrectly performed skeleton models for different activity/action purposes, annotated posture information, as well as depth and color images obtained from the RGB-D camera. Some of the examples of the database are shown in Fig. 2.

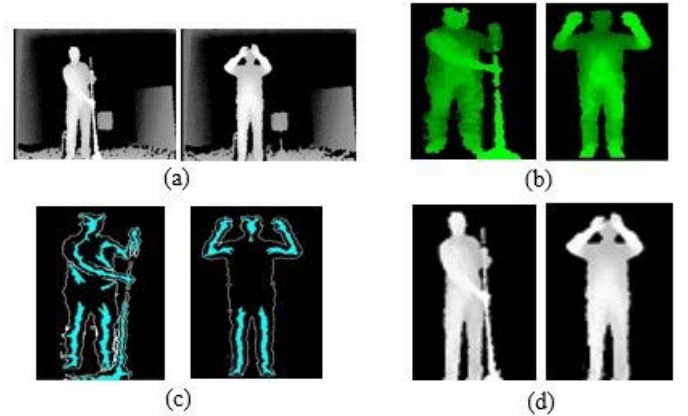


Fig. 2. Human silhouette identification. (a) Noisy background, (b) labeled silhouette, (c) ridge information and (d) depth human silhouette.

### B. Evaluation of Cues-parameters Techniques

In this section, we revive human motion and joint tracking information by considering unique cue evaluation techniques using OpenNI environment, and point out why these methods are applied efficiently and produce vital results.

For initialization, we utilized several body silhouettes and joints values cues for local body parts motions, temporal frames information, spatial/temporal depth silhouettes characteristics and speed measurements of human body shape using depth datasets. Following are the sub-sections that describe the features under our evaluation.

#### 1) Angular Direction Feature

Each posture  $P$  in the database is represented by a vector of 3D joint points as

$$P = [C_1, C_2, \dots, C_n] \quad (2)$$

The angular direction feature  $\varphi_{\cos}$  is defined as the difference of angular movements between similar joints of two different frames (i.e., consecutive frames) at time  $t_1$  and  $t_2$ . It captures the temporal movements of different body parts of human silhouette. It is defined as

$$\varphi_{\cos} = \cos^{-1} \left( \frac{C_k^{t_1} \times C_k^{t_2}}{\|C_k^{t_1}\| \|C_k^{t_2}\|} \right) \quad (3)$$

where  $C^{t_1}$  and  $C^{t_2}$  are the joints information with respect to consecutive frames and  $k$  indicates the all three coordinates axis (i.e., x, y, z) of respective joints [22]. Fig. 3 shows an example of the directional angular features with different activities. However, the joints angular values are quite effective in order to improve the performance accuracy during feature discrimination and recognition.

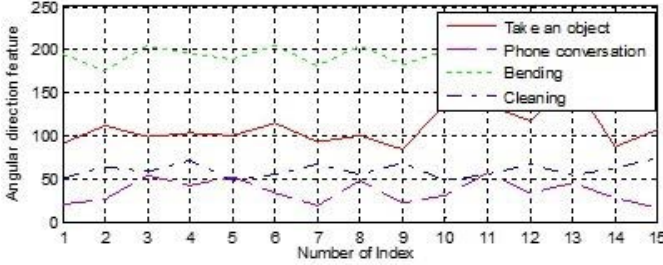


Fig. 3. Human silhouette identification. (a) Noisy background, (b) labeled silhouette, (c) ridge information.

## 2) Velocity Feature

The spatiotemporal velocity feature  $f_v$  captures the velocity of each joint in the direction of the normal vector of the plane from starting frame till ending frame. It is defined as

$$f_v(t, t+1) = \frac{1}{|t_e - t_s + 1|} \sum_{t=t_s}^{t_e} (v(j_{t+1}^k) - v(j_t^k)) \quad (4)$$

where  $t_s$  and  $t_e$  are the starting and ending frames of overall data sequence and  $j^k$  deals with the coordinate axis of human body joint information. These features deal with the intensity differentiation and spatiotemporal motion values of body parts. Fig. 4 explains the detail description of spatiotemporal velocity features using depth dataset.

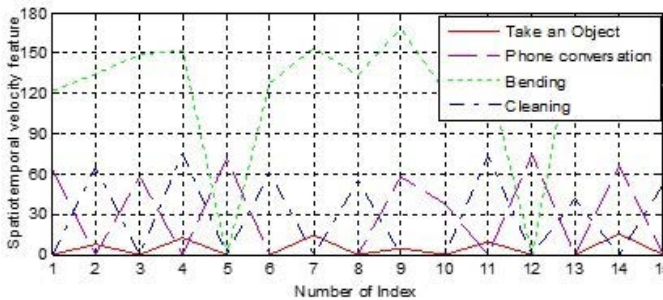


Fig. 4. 2D plot of spatiotemporal velocity features from RGB-D video dataset.

## 3) Invariant Feature

To observe the invariant characteristics of human body silhouette, we measure the integral of a function over a line as the Radon transform. Basically, it extracts some interested regions in the Euclidean plane as

$$E(\theta) = \int_{-\infty}^{\infty} R_z^2(\rho, \theta) d_\rho \quad (5)$$

where  $R_z(\rho, \theta)$  is 2D Radon map that is the line integral of depth data. During extraction of human behavior silhouettes, radon transform

is used to represent the distance and local directional movements in human body parts motion. During the computation of radon transform, the line integrals of human silhouettes amplify low frequency components which are useful in behavior recognition. It is quite suitable for BR because there is a variation in the angles of different human body parts such as feet, hands, head, hips and shoulders. Thus, it represents the maximum energy of the human behavior that appears in specific coefficients which vary considerably through time. Fig. 5 shows some human behaviors and their corresponding Radon transforms.

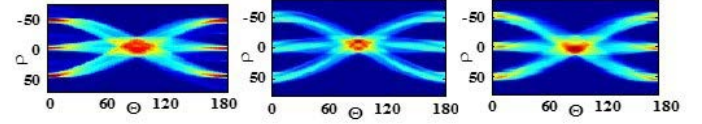


Fig. 5. Distance and local directional movements of different human behaviors using Radon transform.

## C. Features Symbolization

Now, these cues-parameters vectors are further symbolized based on vector quantization technique known as Linde, Buzo, and Gray (LBG)'s clustering algorithm [15]. Initially, LBG initializes with a codebook size of 1 and recursively splits the centroids of feature vectors (i.e., datasets) to get an optimally sized codebook. We used the optimal codebook size of 64 after experimenting over different depth datasets. In addition, the codebook size and codevector are directly related with the precision value of a locally global feature values and parallelly intact with the size of the source image. This optimization of the centroids is done to reduce the distortion. While, these code-values are generated per each behavior sequence and stored by considering buffer strategy. Fig. 6 shows the procedure of code-values generation and symbolization of proposed features.

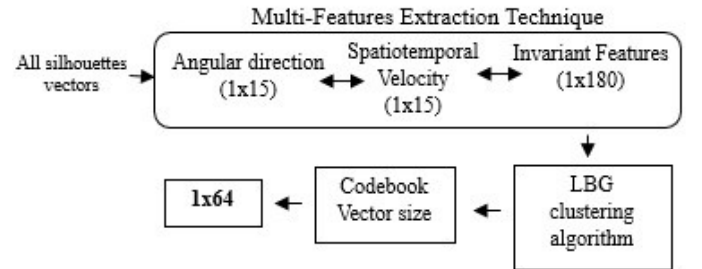


Fig. 6. Overall procedure of code-values generation and symbolization of cues-parameters.

## D. Advanced HMMs

To train and recognize different depth datasets, we modified conventional HMM into advanced HMM technique. In conventional HMM approach, each behavior makes its specific HMM based on finite states having transition and symbol observation probability [21]. Such method consists of redundant information in the form of whole body silhouette and less active moving body parts [25]. Also, HMMs are mainly dependent on the feedback of possible transitions and need the priori knowledge to manage the parameters which cause over-fitting in each class [26, 27]. Such kind of unnecessary information causes reduction at overall performance of accuracy results. Thus, advanced HMM is developed which focused on active areas of human body parts such as hands, feet, head, hips and shoulders. For training phase, we need to train  $(N+1)$  distinct HMMs for  $N$  different human activities. During testing phase, maximum likelihood [28-30] value of specific sequential data is chosen to recognize distinct behavior as expressed in (6).



$$H_a = \arg \max_k \{ P(O / \lambda_h) \} \quad (6)$$

where  $P(O / \lambda_h)$  denotes the probability of likelihood of the  $h$  behavior HMM among a number of activities. Fig. 7 shows active features regions of overall human silhouettes to calculate specific likelihood of each behavior.

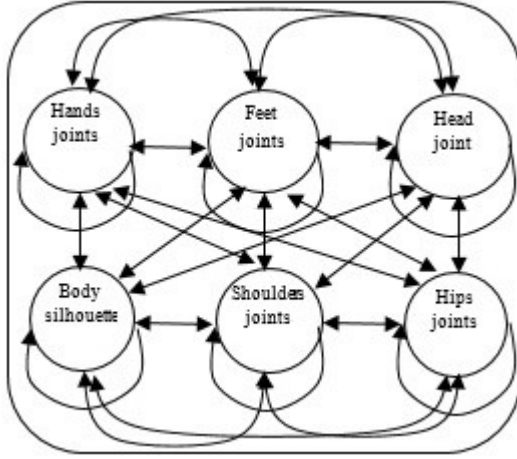


Fig. 7. Structural view of advanced hidden Markov model.

### III. EXPERIMENTAL RESULTS

In this section, we describe how our skeleton models and human postures are represented in the database, and details about what kind of behaviors are included to create the database. We evaluate the proposed cues-parameters approach on a newly collected depth-based behavior recognition datasets (i.e. IM-DailyDepthActivity) and two public datasets (i.e., MSRDailyActivity3D and MSRAction3D).

During experiment, we used the leave one subject out (LOSO) cross validation method [26].

#### A. Datasets Description

Following sub-sections are used to describe each dataset, its experimental setting and size of each dataset, respectively. We use the Microsoft Kinect to capture the skeleton data, its joints location and posture data for the database, as it is one of the known running and developed device of depth camera based motion sensors. Apart from that, we manually handled descriptions such as the natural movements of the human behavior, risky injury during captured sequence and slotting of each video sequence.

##### 1) IM-DailyDepthActivity

We capture the RGB images, depth images, labeled data and skeleton joints information of 15 different activities as: *sit down, both hands waving, phone conversation, kicking, reading an article, throwing, bending, clapping, right hand waving, take an object, exercise, eating, boxing, cleaning and stand up*, respectively. The dataset is captured in indoor environments (i.e., labs, halls and classrooms) having multiple background scenes. However, the dataset includes 45 segmented videos of each activity for training and 30 unsegmented continuous videos for testing performed by 15 different subjects.

Both of the RGB videos and depth maps have the resolution of 640 x 480 pixels and the skeleton contains 15 joints per person. Fig. 8 gives some examples of IM-DailyDepthActivity dataset.

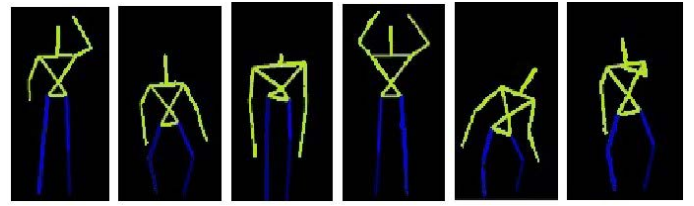


Fig. 8. Some skeleton images of annotated IM-DailyDepthActivity dataset.

##### 2) MSRDailyActivity3D

The dataset has sixteen activities including: drink, eat, read book, call cellphone, write on a paper, use laptop, use vacuum cleaner, cheer up, sit still, toss paper, play game, lay down on sofa, walk, play guitar, stand up and sit down. The total dataset includes 320 video sequences which are performed by 10 subjects. Each subject performs each activity twice, one in standing position and the other in sitting position. Fig. 9 shows some depth images of MSRDailyActivity3D dataset. In addition, most activities involve human-object interactions which make this dataset more challenging.

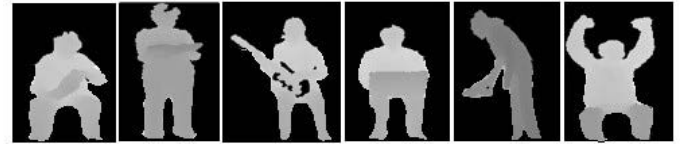


Fig. 9. Sample depth silhouettes images of MSRDailyActivity3D dataset.

##### 3) MSRAction3D

The MSRAction3D dataset [27] is an actions dataset of depth map sequences captured by a depth camera. It contains 20 different actions performed by 10 subjects. These actions include: high arm wave, horizontal arm wave, hammer, hand catch, forward punch, high throw, draw x, draw tick, draw circle, hand clap, two hand wave, side boxing, bend, forward kick, side kick, jogging, tennis swing, tennis serve, golf swing and pick up & throw. There are 567 depth videos sequences. Furthermore, the background of the dataset is clean and mostly actions involve specific body parts movements (i.e., head, arm and leg), which makes the dataset quite challenging. Fig. 10 shows different sequences of depth maps actions used in MSRAction3D dataset.



Fig. 10. Depth sequential images of hand catch and bend actions used in MSRAction3D dataset.

#### B. Recognition Results

##### 1) IM-DailyDepthActivity

Table I, II and III contain the confusion matrices obtained. Specifically, Table I shows the confusion matrix of 15 typical human activities on the annotated IM-DailyDepthActivity dataset.

In the newly collected dataset, we evaluated the proposed cues-parameters method and compared the results with the state-of-the-art methods. Table IV shows the comparison of recognition results. Here, the existing methods such as motion templates [31] proposed new methods for automatic classification and retrieval of motion capture data facilitating the identification of logically related motions in specific dataset. In [17], Yang et al. describes the depth motion maps



TABLE I. CONFUSION MATRIX 15 TYPICAL HUMAN ACTIVITIES ON THE ANNOTATED IM-DAILYDEPTHACTIVITY DATASET

**SD**= sit down, **BH**= both hands waving, **PC**= phone conversation, **KG**= kicking, **RA**= reading an article, **TG**= throwing, **BG**= bending, **CG**= clapping, **RH**= right hand waving, **TA**= take an object, **EX**= exercise, **EA**= eating, **BO**= boxing, **CL**= cleaning, **SU**= stand up

Activities	SD	BH	PC	KG	RA	TG	BG	CG	RH	TA	EX	EA	BO	CL	SU
<b>SD</b>	<b>72.5</b>	4.0	0	6.50	4.0	0	0	5.0	0	0	1.0	2.5	1.5	0	3.0
<b>BH</b>	2.5	<b>77.0</b>	3.0	0	0	1.5	4.0	0	1.5	3.5	0	0	1.5	4.5	1.0
<b>PC</b>	11.0	3.5	<b>64.5</b>	6.0	2.5	0	0	2.5	3.5	1.0	2.0	0	1.5	0	2.0
<b>KG</b>	2.0	1.5	0	<b>81.0</b>	2.5	3.0	0	2.5	0	3.5	0	1.5	0	1.0	1.5
<b>RA</b>	0	13.0	0	4.5	<b>69.5</b>	1.5	1.0	0	2.5	5.0	1.0	0	0	2.0	0
<b>TG</b>	1.5	3.5	3.0	0	0	<b>73.0</b>	3.5	4.0	3.5	0	1.5	3.0	2.0	0	1.5
<b>BG</b>	7.0	2.5	0	1.0	0	8.5	<b>5.85</b>	3.5	1.0	2.0	4.0	1.0	0	9.5	1.5
<b>CG</b>	1.5	0	1.0	2.5	0	2.0	2.5	<b>81.0</b>	2.50	1.0	0	0	3.5	0	2.5
<b>RH</b>	5.0	7.5	2.0	0	1.0	0	1.5	0	<b>6.75</b>	3.5	2.0	1.5	2.0	1.5	5.0
<b>TA</b>	3.5	1.0	4.5	5.0	0	3.0	1.0	0	3.5	<b>72.5</b>	0	1.5	1.0	0	3.5
<b>EX</b>	8.5	3.5	0	5.5	6.5	0	2.0	1.5	0	2.5	<b>55.0</b>	7.5	0	5.5	2.0
<b>EA</b>	3.0	7.5	2.0	0	11.5	2.5	0	2.0	4.0	0	0	<b>52.5</b>	8.5	0	6.5
<b>BO</b>	6.5	0	5.0	0	1.5	0	0	3.5	0	1.5	0	0	<b>70.5</b>	9.0	2.5
<b>CL</b>	3.5	12.5	4.5	3.0	0	2.0	0	0	3.5	5.5	1.0	0	0	<b>57.0</b>	7.5
<b>SU</b>	1.5	4.0	0	0	3.5	6.5	1.0	0	4.5	0	2.0	1.0	2.5	0	<b>73.5</b>
Mean Recognition Rate = 68.4%															

TABLE II. CONFUSION MATRIX 16 DIFFERENT HUMAN ACTIONS ON THE MSRDAIlyACTIVITY3D DATASET

**DK**= drink, **ET**= eat, **RB**= read book, **CC**= call cellphone, **WP**= write on a paper, **UL**= use laptop, **UV**= use vacuum cleaner, **CU**= cheer up, **SS**= sit still, **TP**= toss paper, **PG**= play game, **LD**= lay down on sofa, **WK**= walk, **PR**= play guitar, **ST**= stand up, **SI**= sit down.

Activities	DK	ET	RB	CC	WP	UL	UV	CU	SS	TP	PG	LD	WK	PR	ST	SI
DK	87.0	2.5	1.0	0	1.5	0	3.0	0	1.0	0	0	1.5	0	0	1.0	1.5
ET	0	94.5	0	0	1.0	2.5	0	0	1.0	0	0	0	0	1.0	0	0
RB	1.5	0	93.0	1.0	0	0	1.5	0	0	2.0	0	0	1.0	0	0	0
CC	2.5	4.0	0	84.5	0	4.5	0	0	1.0	0	1.0	0	0	2.5	0	0
WP	1.5	2.0	1.5	0	89.0	0	0	1.0	0	0	0	2.5	0	0	1.0	1.5
UL	0	0	0	1.0	0	98.0	0	0	0	1.0	0	0	0	0	0	0
UV	1.5	0	1.0	0	0	0	94.5	0	0	0	1.5	0	0	1.5	0	0
CU	3.5	0	0	1.5	0	0	2.0	86.5	1.5	0	1.5	0	0	2.5	0	1.0
SS	2.0	1.0	0	2.5	0	0	1.5	0	88.0	0	0	2.0	1.5	0	1.5	0
TP	0	0	1.5	0	0	1.0	0	0	0	97.5	0	0	0	0	0	0
PG	0	2.0	0	0	1.0	0	1.0	0	1.0	0	94.0	0	0	1.0	0	0
LD	1.5	0	2.0	0	0	2.5	0	1.5	0	1.0	0	89.5	0	1.0	0	1.0
WK	0	0	1.0	0	0	0	1.5	0	0	0	0	0	97.5	0	0	0
PR	0	2.0	0	1.5	0	0	3.5	0	0	1.0	0	0	0	91.0	0	1.0
ST	1.5	0	2.0	1.0	1.5	0	0	1.0	0	1.0	0	1.5	1.0	0	89.5	0
SI	3.0	0	0	1.5	0	2.5	1.5	0	1.0	0	1.5	0	0	2.0	1.0	86.0
Mean Recognition Rate = 91.2%																

as feature representation and HOG to characterize the local appearance for recognizing data. While, [32] deals with the positions of joints to locally define reference system and multi-part bag-of-poses approach is then defined, which permits the separate alignment of body parts through a nearest-neighbor for classification and recognition. In [33],

multi-modality fusion scheme is developed based on spatio-temporal interest points and motion history images features to recognize different activities. From results in Table IV, it is clearly seen that the proposed method improves the recognition results as compared to state-of-the-art methods.

TABLE III. CONFUSION MATRIX 20 DIFFERENT HUMAN ACTIONS ON THE MSRACTION3D DATASET

HW= high arm wave, HA= horizontal arm wave, HM= hammer, HC= hand catch, FP= forward punch, HT= high throw, DX= draw x, DT= draw tick, DC= draw circle, HC= hand clap, TH= two hand wave, SB= side boxing, BD= bend, FK= forward kick, SK= side kick, JO= jogging, TS= tennis swing, TE= tennis serve, GS= golf swing, PU= pick up & throw.

Actions	HW	HA	HM	HC	FP	HT	DX	DT	DC	HC	TH	SB	BD	FK	SK	JO	TS	TE	GS	PU
HW	<b>91.5</b>	1.0	0	0	2.5	0	0	2.0	0	0	1.0	0	0	0	1.0	0	0	1.0	0	0
HA	2.0	<b>86.0</b>	1.0	0	0	2.5	0	1.0	1.5	0	3.0	0	0	1.0	0	1.0	0	0	0	1.0
HM	0	0	<b>91.5</b>	1.5	0	2.0	0	0	1.0	2.0	0	1.0	0	0	0	0	1.0	0	0	0
HC	0	0	1.0	<b>97.0</b>	0	0	0	0	0	1.0	0	0	1.0	0	0	0	0	0	0	0
FP	1.0	0	0	0	<b>96.0</b>	1.0	0	0	0	0	0	0	0	0	0	1.0	0	0	1.0	0
HT	2.0	1.0	0	0	1.0	<b>88.5</b>	3.0	0	1.0	0	2.5	0	0	0	0	0	0	1.0	0	0
DX	0	0	2.0	0	1.5	0	<b>90.0</b>	0	1.5	0	2.0	1.0	1.0	0	1.0	0	0	0	0	0
DT	0	0	0	0	0	0	0	<b>97.0</b>	0	1.0	0	0	0	1.0	0	0	1.0	0	0	0
DC	0	0	1.0	0	1.5	0	0	0	<b>94.5</b>	0	0	0	0	1.0	0	0	0	1.0	0	1.0
HC	0	1.0	0	2.5	0	2.0	0	0	0	<b>91.0</b>	0	1.5	0	0	1.0	0	0	0	1.0	0
TH	2.0	0	0	0	3.5	0	1.5	0	0	1.0	<b>90.5</b>	0	1.5	0	0	0	0	0	0	0
SB	0	1.0	0	1.0	0	0	0	2.0	0	0	0	<b>96.0</b>	0	0	0	0	0	0	0	0
BD	0	0	0	0	0	1.5	0	0	2.5	0	0	0	<b>95.0</b>	0	0	0	1.0	0	0	0
FK	0	0	0	0	1.5	0	0	0	0	2.0	0	0	0	<b>93.5</b>	0	1.0	0	1.0	0	1.0
SK	1.0	0	1.0	0	0	0	1.5	0	0	0	0	2.5	0	0	<b>92.0</b>	0	1.0	0	1.0	0
JO	0	1.5	0	0	2.5	0	0	1.0	0	1.0	0	0	1.5	0	0	<b>92.5</b>	0	0	0	<b>0</b>
TS	1.0	0	0	1.0	0	0	1.0	0	0	0	1.5	0	0	0	0	0	<b>95.5</b>	0	0	<b>0</b>
TE	0	0	1.0	0	0	0	0	0	0	1.0	0	0	0	0	0	0	0	<b>97.0</b>	0	<b>1.0</b>
GS	0	1.0	0	0	2.0	1.0	0	0	1.5	0	0	0	0	0	0	0	0	0	<b>94.5</b>	<b>0</b>
PU	0	2.5	0	3.0	0	0	0	1.5	0	0	1.0	0	2.0	0	0	2.0	0	0	0	<b>88.0</b>
Mean Recognition Rate = 92.9%																				

TABLE IV. RECOGNITION RESULTS USING IM-DAILYDEPTHACTIVITY

Methods	Recognition Accuracy
Motion templates [31]	38.7
Depth motion maps [17]	42.3
Naive-Bayes-Nearest-Neighbor [32]	47.6
Color-depth fusion features [33]	51.6
<b>Proposed Cues-parameters</b>	<b>68.4</b>

## 2) MSRDailyActivity3D

We evaluate our proposed cues-parameters method using MSRDailyActivity3D dataset [34, 35]. We performed the experiment using the depth silhouettes and the joint information together in cues-parameters. We used existing methods [19, 31, 36] where [31] and [36] mainly deal with joint points information to monitor the human movements and recognize human activities/actions. While, in [10], a novel approach is proposed for human action recognition with histograms of 4D joint locations (HON4D) as a compact representation to recognize different activities. In [26], Wang et al developed the actionlet ensemble model that used 3D point cloud to model body shape and reported a recognition accuracy of 85.7%. In [38], Xia and Aggarwal proposed a novel depth cuboid similarity feature (DCSF) to describe the local 3D depth cuboid around the depth images to recognize different actions/activities.

Table II demonstrates the confusion matrix of 16 different human activities that is obtained from the proposed cues parameters method on the MSRDailyActivity3D dataset.

Besides, we compare the experimental results of our proposed method with the algorithms defined as state-of-the-arts methods and the results are shown in Table V. As the table shows, our proposed method performs much better than state-of-the-art methods. It is clearly shown that the proposed method achieved significantly better recognition accuracy as 91.2% than the state-of-the-art methods as 54.0%, 58.1%, 72.1%, 80.0%, 85.7% and 88.2%, respectively.

TABLE V.

RECOGNITION ACCURACY COMPARISON USING MSRDAILYACTIVITY3D

Methods	Recognition Accuracy
Motion templates [31]	54.0
Eigenjoints [36]	58.1
Spatiotemporal features [37]	63.7
HON4D [19]	80.0
Actionlet ensemble [26]	85.7
Multilayer perceptron [45]	87.6
Cuboid Similarity Feature [38]	88.2
Deep learning approach [46]	90.5
<b>Proposed Cues-parameters</b>	<b>91.2</b>

## 3) MSRACTION3D

The confusion matrix of 20 different human actions that is obtained from the proposed cues parameters method on the MSRACTION3D dataset is shown in Table III.

We compare the performance of our proposed method with the state of the art methods using MSRACTION3D dataset. In the first approach [39], developed by Lv and Nevatia, learning-based algorithm for automatic recognition and segmentation of 3D human actions is defined. In [27] and [36], bag of 3D model and position differences of joints as eigenjoints are analyzed for action recognition. While, actionlet ensemble [26], pose set [42] and moving pose [43] methods deal with body joints information for activity classification and recognition.

To make a fair comparison, we performed cross subject test between proposed method and the state-of-the-art methods because the cross subject is more challenging due to dynamic intra-class differences in actions among different subjects. The proposed cues-parameters method achieves the recognition accuracy of 92.9% which significantly outperforms the existing methods, as listed in Table VI.

TABLE VI.  
RECOGNITION ACCURACY COMPARISON OF OUR METHOD AND PREVIOUS  
METHODS USING MSRACTION3D.

Methods	Recognition Accuracy
Hidden Markov model [39]	63.0
Bag of 3D points [27]	74.7
Shape and motion features [40]	82.1
Eigenjoints [36]	82.3
Hybrid features [41]	83.6
Actionlet ensemble [26]	88.2
Multilayer perceptron [45]	88.7
Pose Set [32]	90.0
Deep learning approach [46]	91.3
Moving Pose [43]	91.7
<b>Proposed Cues-parameters</b>	<b>92.9</b>

#### IV. CONCLUSION

In this paper, we proposed a novel methodology having multi-joints features along with advanced HMM for BR system using depth sequences. Our proposed BR system utilizes a combination of human silhouettes identification and tracking process using modified connected component labeling method. In-addition, it includes angular direction, spatiotemporal velocity and invariant features which are used to extract local body part information, intensity differentiation and temporal variation properties to reinforce the feature classification and accuracy. Finally, these features are modeled, trained and recognized using recognizer engine. The proposed method is implemented with Matlab (R2009b) using an Intel Core i3 processor with 4 GB RAM in Windows XP platform. We evaluated the proposed method using three different datasets as IM-DailyDepthActivity, MSRDailyActivity3D and MSRAction3D. Experimental results showed some promising performance of the proposed BR method over the state of the art methods.

In the future work, we will improve the effectiveness of our system by adding biometric identification techniques [44] such as face reorganization, iris detection, gesture recognition, etc., with the help of involving multi-cameras views or RGB contents in Kinect sensors. Besides, we will also use combined cues of biometric identification cues [44] along with proposed cues-parameters techniques to cover large interventions of worldly-scenarios and act as multi-mode biometric system.

#### REFERENCES

- [1] H. Chaminda, V. Klyuev and K. Naruse, "A smart reminder system for complex human activities," in *International Conference on Advanced Communication Technology*, 2012, pp. 235-240.
- [2] V. B. Semwal, N. Gaud and G. C. Nandi, "Human Gait State Prediction Using Cellular Automata and Classification Using ELM," in *International Conference on Machine Intelligence and signal processing*, 2017.
- [3] A. Jalal and A. Shahzad, "Multiple facial feature detection using Vertex-Modeling structure," in *International Conference on Interactive Computer Aided Learning (ICL)*, 2007, pp. 1-7.
- [4] M. Mehmood A. Jalal, and H. A. Evans, "Facial Expression Recognition in Image Sequences Using 1D Transform and Gabor Wavelet Transform," in *IEEE International Conference on Applied and Engineering Mathematics*, 2018.
- [5] M. Raj, V. B. Semwal, and G. C. Nandi, "Bidirectional association of joint angle trajectories for humanoid locomotion: the restricted Boltzmann machine approach," *Neural Computing and Applications*, pp. 1-9, 2016.
- [6] A. Jalal and I. Uddin, "Security architecture for third generation (3G) using GMHS cellular network," in *IEEE Conference on Emerging Technologies*, 2007, pp. 74-79.
- [7] A. Jalal and Y. Rasheed, "Collaboration achievement along with performance maintenance in video streaming," in *Proceedings of the IEEE conference on Interactive computer aided learning*, pp. 1-8, 2007.
- [8] V. B. Semwal and G. C. Nandi, "Toward developing a computational model for bipedal push recovery—a brief," *IEEE Sensors Journal*, vol. 15, no. 4, pp. 2021-2022, 2015.
- [9] A. Jalal, and M. A. Zeb, "Security enhancement for e-learning portal," *International Journal of Computer Science and Network Security*, vol. 8, no. 3, pp. 41-45, 2008.
- [10] A. Jalal and S. Kim, "Global security using human face understanding under vision ubiquitous architecture system," *World Academy of Science, Engineering, and Technology*, vol. 13, pp. 7-11, 2006.
- [11] V. B. Semwal, J. Singha, P. K. Sharma, A. Chauhan and B. Behera, "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification," *Multimedia tools and applications*, vol. 76, no. 22, pp. 24457-24475, Nov. 2017.
- [12] A. Jalal and S. Kim, S, "Advanced performance achievement using multi-algorithmic approach of video transcoder for low bit rate wireless communication," *ICGST International Journal on graphics, vision and image processing*, vol. 5, no. 9, pp. 27-32, 2005.
- [13] A. Jalal and S. Kim, "The Mechanism of Edge Detection using the Block Matching Criteria for the Motion Estimation," *Proc. Human Computer Interaction*, pp.484-489, Jan. 2005.
- [14] G. C. Nandi, V. B Semwal, M. Raj and A. Jindal, "Modeling bipedal locomotion trajectories using hybrid automata," *Proc. IEEE Region 10 Conference (TENCON)*, pp. 1013-1018, 2016.
- [15] A. Jalal, N. Sharif, J. T. Kim and T. S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home," *Indoor and Built Environment*, vol. 22, no. 1, pp. 271-279, January, 2013.
- [16] P. Turaga, R. Chellappa, V. S. Subrahmanian and O. Udrea, "Machine recognition of human activities: A survey," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 11, pp. 1473-1488, November, 2008.
- [17] X. Yang, C. Yang and Y. Tian, "Recognizing actions using depth motion maps-based histograms of oriented gradients," in *International Conference on Multimedia (ICM)*, 2012, pp. 1057-1060.
- [18] A. Jalal, Y. Kim, and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," in *Conference on computing, communication and networking technologies*, 2014, pp. 1-6.
- [19] O. Oreifej and Z. Liu, "Hon4d: Histogram of oriented 4d normal for activity recognition from depth sequences," in *Conference on Computer Vision and Pattern Recognition*, 2013, pp. 716-723.
- [20] A. Jalal, S. Kamal and D. Kim, "A depth video sensor-based life-logging human activity recognition system for elderly care in smart indoor environments," *Sensors*, vol. 14, no. 7, pp. 11735-11759, July, 2014.
- [21] A. Jalal, Y.-H. Kim, Y.-J. Kim, S. Kamal and D. Kim, "Robust human activity recognition from depth video using spatiotemporal multi-fused features," *Pattern recognition*, vol. 61, pp. 295-308, 2017.
- [22] A. Jalal and Y. Kim, "Dense Depth Maps-based Human Pose Tracking and Recognition in Dynamic Scenes Using Ridge Data," in *Conference on Advanced Video and Signal-Based Surveillance*, 2014, pp. 119-124.
- [23] V. B. Semwal and G. C. Nandi, "Generation of joint trajectories using hybrid automate-based model: a rocking block-based approach," *IEEE Sensors Journal*, vol. 16, no. 14, pp. 5805-5816, May, 2016.
- [24] M. Raj, V. B. Semwal and G. C. Nandi, "Multiobjective optimized bipedal locomotion," *International Journal of Machine Learning and Cybernetics*, pp. 1-17, 2017.
- [25] A. Jalal, S. Y. Lee, J. T. Kim, T. S. Kim, "Human activity recognition via the features of labeled depth body parts," in *International Conference on Smart Homes and Health Telematics*, 2012, pp. 246-249.
- [26] J. Wang, Z. Liu, Y. Wu, J. Yuan, "Mining actionlet ensemble for action recognition with depth cameras," in *International Conference on Computer Vision and Pattern Recognition*, 2012, pp. 1290-1297.
- [27] W. Li, Z. Zhang and Z. Liu, "Action recognition based on a bag of 3D points," in *International Workshop on Computer Vision and Pattern Recognition*, 2010, pp. 9-14.
- [28] A. Jalal, J. T. Kim, and T.-S Kim, "Development of a life logging system via depth imaging-based human activity recognition for smart homes," in *Proceedings of the International Symposium on Sustainable Healthy*

*Buildings*, 2012, pp. 91-95.

- [29] A. Jalal, S. Kamal and D.-S. Kim, "Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System," *KSI Transactions on Internet and Information Systems*, vol. 12, no. 3, pp. 1189-1204, 2018.
- [30] A. Jalal, J. T. Kim, and T.-S. Kim, "Human activity recognition using the labeled depth body parts information of depth silhouettes," in *Proceedings of the 6th international symposium on Sustainable Healthy Buildings*, 2012, pp. 1-8.
- [31] M. Muller and T. Roder, "Motion templates for automatic classification and retrieval of motion capture data," in *SIGGRAPH/Eurographics symposium on computer animation*, 2006, pp. 137-146.
- [32] X. Seidenari, C. Varano, Y. Berretti, C. Bimbo and Y. Pala, "Recognizing actions from depth cameras as weakly aligned multi-part bag-of-poses," in *Conference on Computer Vision and Pattern Recognition Workshops*, 2013, pp. 479-485.
- [33] B. Ni, G. Wang and P. Moulin, "RGBD-HuDaAct: A color-depth video database for human daily activity recognition," in *Conference on Computer Vision Workshops*, 2011, pp. 1147-1153.
- [34] A. Jalal, S. Kamal and D. Kim, "Depth Silhouettes Context: A new robust feature for human tracking and activity recognition based on embedded HMMs," in *International Conference on Ubiquitous Robots and Ambient Intelligence*, 2015, pp. 294-299.
- [35] A. Jalal, Y. Kim, S. Kamal, A. Farooq and D. Kim, "Human daily activity recognition with joints plus body features representation using Kinect sensor," in *International Conference on Informatics, electronics and vision*, 2015, pp. 1-6.
- [36] X. Yang and Y. Tian, "Eigenjoints-based action recognition using naive-bayes-nearest-neighbor," in *Conference on Computer vision and pattern recognition workshops*, 2012, pp. 14-19.
- [37] A. Jalal, S. Kamal and D. Kim, "Human depth sensors-based activity recognition using spatiotemporal features and hidden markov model for smart environments," *J. of computer networks and communications*, pp. 1-11, 2016.
- [38] L. Xia and J. Aggarwal, "Spatio-Temporal Depth Cuboid Similarity Feature for Activity Recognition Using Depth Camera," in *International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2834-2841.
- [39] F. Lv and R. Nevatia, "Recognition and segmentation of 3-D human action using HMM and multi-class adaboost," in *European Conference on Computer Vision*, 2006, pp. 359-372.
- [40] A. Jalal, S. Kamal and D. Kim, "Shape and motion features approach for activity tracking and recognition from Kinect video camera," in *International Conference on Advanced Information Networking and Applications Workshops*, 2015, pp. 445-450.
- [41] Shaharyar Kamal and Ahmad Jalal, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian J. of Science and Engineering*, vol. 41, no. 3, pp. 1043-1051, 2016.
- [42] C. Wang, Y. Wang and A. Yuille, "An Approach to Pose-Based Action Recognition," in *International Conference on Computer Vision and Pattern Recognition*, 2013, pp. 915-922.
- [43] M. Zanfir, M. Leordeanu and C. Sminchisescu, "The Moving Pose: An Efficient 3D Kinematics Descriptor for Low-Latency Action Recognition and Detection," in *International Conference on Computer Vision*, 2013, pp. 2752-2759.
- [44] V. B. Semwal, J. Singha, P. K. Sharma, A. Chauhan and B. Behera, "An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification," *Multimedia Tools and Applications*, vol. 76, no. 22, pp. 24457-24475, 2017.
- [45] V. B. Semwal, M. Raj, and G. C. Nandi, "Biometric gait identification based on a multilayer perceptron," *Robotics and Autonomous Systems*, vol. 65, pp. 65-75, 2015.
- [46] V. B. Semwal, K. Mondal, and G. C. Nandi, "Robust and accurate feature selection for humanoid push recovery and classification: deep learning approach," *Neural Computing and Applications*, vol. 28, no. 3, pp. 565-574, 2017.



Ahmad Jalal

A. Jalal received his M.S. degree in Computer Science from Kyungpook National University, Republic of Korea. He received his Ph.D. degree in the Department of Biomedical Engineering at Kyung Hee University, Republic of Korea. His research interest includes human computer interaction, image processing, and computer vision.



Shaharyar Kamal

S. Kamal received his M.S. degree in Computer Engineering from Mid Sweden University, Sweden. He obtained Ph.D. degree from the Department of Radio and Electronics Engineering, Kyung Hee University, Republic of Korea. His research interest includes 5G, IoT, Cyber Security and Image & Signal Processing.



# Contour Enhancement Algorithm for Improving Visual Perception of Deutan and Protan Dichromats

M. Madalena G. Ribeiro<sup>1,2</sup>, Abel J.P. Gomes<sup>3,4</sup> \*

<sup>1</sup> Polytechnic Institute of Castelo Branco, ESART, 6000-084 Castelo Branco (Portugal)

<sup>2</sup> Centro de Investigação em Património, Educação e Cultura, 6000-084 Castelo Branco (Portugal)

<sup>3</sup> Universidade da Beira Interior, 6200-001 Covilhã (Portugal)

<sup>4</sup> Instituto de Telecomunicações, Delegação da Covilhã, 6200-001 Covilhã (Portugal)

Received 3 May 2019 | Accepted 21 May 2019 | Published 27 May 2019

**unir**  
LA UNIVERSIDAD  
EN INTERNET

## ABSTRACT

A variety of recoloring methods has been proposed in the literature to remedy the problem of confusing red-green colors faced by dichromat people (as well by other color-blinded people). The common strategy to mitigate this problem is to remap colors to other colors. But, it is clear this does not guarantee neither the necessary contrast to distinguish the elements of an image, nor the naturalness of colors learnt from past experience of each individual. In other words, the individual's perceptual learning may not hold under color remapping. With this in mind, we introduce the first algorithm primarily focused on the enhancement of object contours in still images, instead of recoloring the pixels of the regions bounded by such contours. This is particularly adequate to increase contrast in images where we find adjacent regions that are color-indistinguishable from the dichromacy's point of view.

## KEYWORDS

Image Color Analysis, Color Vision Deficiency, Dichromacy, Image Naturalness, Contour Enhancement, CVD Adaptation.

DOI: 10.9781/ijimai.2019.05.003

## I. INTRODUCTION

THE visual perception of color involves the cone cells on retina. These photosensitive cells react to the light reflected by objects. There are three types of cone cells with different sensitivities: S-cones, M-cones, and L-cones, which are sensitive to small, medium, and large wavelengths of light, respectively. These three types of cones work well for an individual without color vision deficiency (CVD), and he/she is called a trichromat [1].

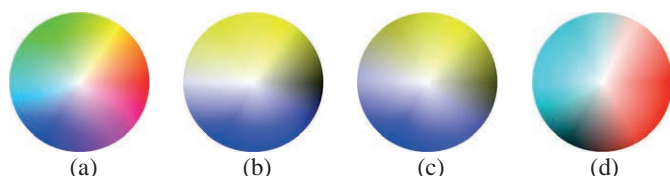


Fig 1. Color wheel when seen by (a) regular trichromat people; (b) protanope people, (c) deuteranope people and (d) tritanope people.

However, about 5% of world population have some sort a disturbance in color perception, which reduces their visible color spectrum [2]. This visual disturbance of CVD people may reduce their perception of the surrounding world, specifically when compared to trichromats (cf. Fig. 1). For example, a trichromat can distinguish two overlapping objects with distinct colors in an image, but a CVD person may not distinguish

them, i.e., two distinct colors are perceived as one.

### A. Color Vision Deficiencies

CVD mainly stems from the partial or total malfunctioning of a cone cell type, as shown in Fig. 2. In the case of partial malfunctioning of a single type of cones, the deficiency is called anomalous trichromacy, but if this malfunctioning is total, the impairment is called dichromacy; finally, in the case of total malfunctioning of two or three types of cone cells, the impairment is known as monochromacy [3] [4]. Besides, as illustrated in Fig. 2, the prefixes 'protano' (or protan), 'deuterano' (or deutan) and 'tritano' (or tritan) denote the handicap of L-, M- and S-cones, respectively.

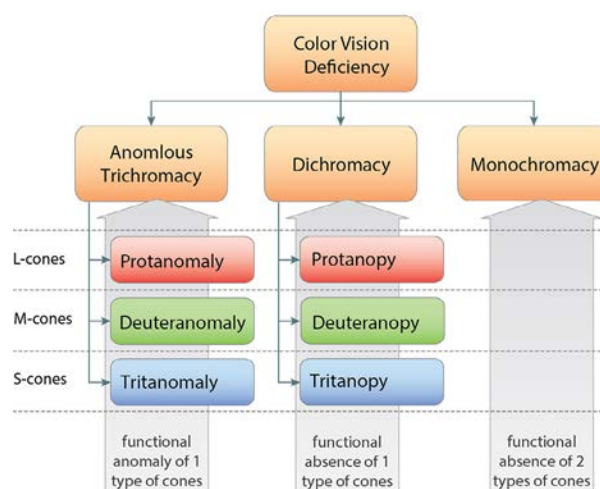


Fig 2. Degrees of severity of color vision deficiency and its subtypes.

\* Corresponding author.

E-mail addresses: mrribeiro@ipcb.pt (M. Madalena G. Ribeiro), agomes@di.ubi.pt (Abel J.P. Gomes).

In the present work, the research focus is on dichromacy, specifically deuteranopy and protanopy. As shown in Fig. 1, the deutan and protan dichromat individuals only see blues and yellows, while the tritan see bluish cyans and reds. More specifically, deutan/protan dichromats only see two different hues, considering the HSV (Hue, Saturation, and Value) color wheel depicted in Fig. 1: 60°-yellow and 240°-blue. That is, all the colors Fig. 1(b)-(c) seen by deutan and protan dichromats just result from varying saturation and brightness. Analogously, the tritan dichromats only see two hues, 0°-red and 180°-cyan, as the remaining colors observed in Fig. 1(d) are obtained from either 0°-red or 180°-cyan with varying values of saturation and brightness. These results were achieved using the simulation algorithm due to Viennot et al. [5]. Note that the dichromacy corresponds to anomalous trichromacy with a 100% of severity [6].

### B. Related Work

Currently, most color adaptation algorithms are based on color mapping, i.e., they map some colors onto other colors, as it the case of the following algorithms due to Martin et al. [7], Ichikawa et al. [8], Rasche et al. [9], Iaccarino et al. [10], Huang et al. [11], Kuhn et al. [12], Doliotis et al. [13], Wang et al. [14], Chen et al. [15], and Ribeiro and Gomes [16] [17]. However, in general, these algorithms tend to recoloring most pixels of an image, causing loss of naturalness of color as perceived by dichromat people [Ichikawa et al. 2003] [Wakita and Shimamura 2005] [Huang et al. 2007] [Kuhn et al. 2008] [Flatla et al. 2013], though the resulting contrast is more noticeable after all. But, this may have negative implications on the perceptual learning of CVD people, possibly causing more perceptual confusion in many cases. These naturalness- and contrast-related problems have led to the investigation of alternative color identification algorithms, called content-independent methods [18] [19], which overlay patterns (e.g., color names and meters) on colored visualization contents. But, clearly, these techniques have the downside of causing noise in the color perception, though they minimize ambiguities in color identification.

In this paper, we follow a distinct approach, which consists in enhancing the contours (not the interiors) of image regions featuring objects. As far as we know, this is the first contour-based adaptation algorithm for minimizing ambiguities if color perception of CVD people. As shown in paper, this allows for increasing the contrast without compromising the naturalness of imaging. In addition, this allows CVD people to see more objects than they see normally.

### C. Contributions

The idea of the adaptation algorithm proposed in this paper is to increase the contrast of neighbor pixels that are indistinguishable for dichromat people, but at the same time to maintain the naturalness of color as much as possible. Color contrast is obtained by increasing the perceptual difference between neighboring regions of a given image. For that purpose, the interior of each region remains unchanged, being the contrast increased by lightening or darkening its region contours. As far as we are aware, there is no similar contour-based algorithm in the literature.

### D. Organization Paper

The remainder of this paper is organized as follows. Section II describes our contour enhancement method. Section III presents the most relevant results about our algorithm with respect to the evaluation parameters: contrast and naturalness. Section IV describes our study of usability. Section V draws the most important conclusions about our work.

## II. CONTOUR ENHANCEMENT

The individuals with dichromacy see only two distinct hues; more specifically, blues and yellows for deutan and protan dichromats, and

reds and greenish blue for tritan dichromats, although with different values of saturation and brightness. For example, a deutan dichromat perceives a weakly saturated yellow as a moss green. The reduced chromatic range as perceived by dichromat people may lead to lack of discrimination between neighbor regions in an image, resulting in confusion about what is being seen in the image.

To mitigate this problem, we propose an approach that enhances contours between adjacent image regions, which are indistinguishable for dichromat individuals. The idea is to highlight the contours that separate contiguous regions represented with different colors, but seen as similar or identical by dichromat people.

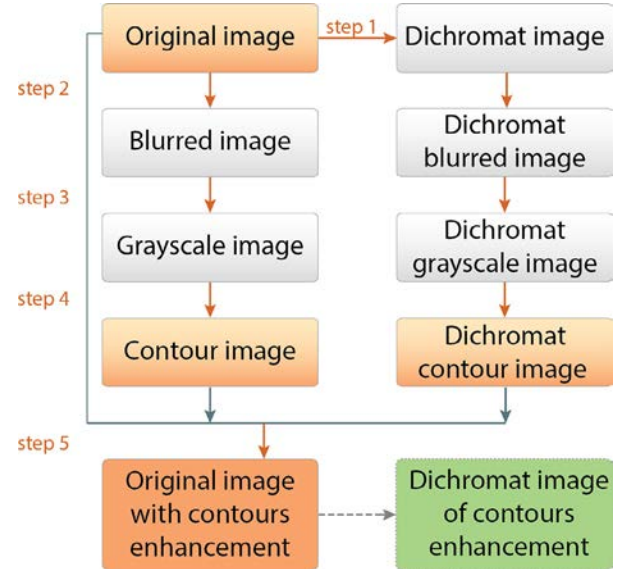


Fig 3. Diagram of the contour enhancement algorithm (CEA).

This contour enhancement procedure is illustrated in Figs. 3 and 4, and comprises the following steps:

1. Generate the image as seen by (deutan and protan) dichromat individuals from the original image.
2. Apply a Gaussian blur filter to both original and dichromat images.
3. Convert both original and dichromat images to grayscale.
4. Compute contours in both original and dichromat images.
5. Highlight those contours of the original image that are absent in the respective dichromat image.

### A. Dichromat Image Generation

Taking the original image as input data, the first step of our algorithm consists in generating the corresponding image as seen by a dichromat individual. This is accomplished using the simulation algorithm due to Viennot et al. [5].

Essentially, this algorithm comprises three steps: (i) the RGB LMS conversion [20] [21]; (ii) the dichromat simulation (e.g. deuteranope) in the LMS color space, as proposed by Viennot et al. [5]; (iii) the LMS- RGB conversion. Each step is associated to a specific matrix, so the entire process reduces to a product of three matrices, resulting in the following overall matrix:

$$M = \begin{bmatrix} 0.2928 & 0.7072 & 0 \\ 0.2928 & 0.7072 & 0 \\ 0.02234 & 0.02234 & 1 \end{bmatrix} \quad (1)$$

Summing up, we generate the image as seen by a dichromat individual by multiplying the matrix  $M$  above by the vector  $[R \ G \ B]^T$  associated to each pixel. This step may be adjusted to any sort of CVD using a distinct overall simulation matrix  $M$ .



Fig 4. The contour enhancement process applied to a hibiscus flower.

Interestingly, as shown in Figs. 4(a)-(b), a deutan dichromat person cannot see the hibiscus flower at all, because he/she cannot distinguish the green background from the red flower.

### B. Gaussian Blur Filter

Both images, the original image and dichromat image, are subject to a noise reducing procedure through a Gaussian blur filter (see [22] for

further details). This is a pre-processing step, and is performed using the Gaussian  $3 \times 3$  kernel matrix:

$$M = \begin{bmatrix} 0.077847 & 0.123317 & 0.077847 \\ 0.123317 & 0.7072 & 0.123317 \\ 0.077847 & 0.123317 & 0.077847 \end{bmatrix} \quad (2)$$

This matrix corresponds to a Gaussian kernel of width of 1. Recall that this sort of image-blurring filters is commonly used in image analysis and processing, particularly when they coupled with contour detectors, which are sensitive to noise.

### C. Grayscale Conversion

The conversion to a grayscale representation is performed by computing the luminance of each pixel [23], in conformity with the following expression:

$$I = 0.2989 R + 0.5866 G + 0.1145 B \quad (3)$$

where,  $R$ ,  $G$  and  $B$  are color components of the pixel. This conversion is justified by the fact that the human eye can perceive brightness changes better than color changes [22]. This fact is not only valid for trichromat, but also for dichromat people. Figs. 4(e) and (f) show the images that result from converting the images in Figs. 4(c) and (d) to grayscale, respectively.

### D. Contour Detection

The core of our method lies in the detection of contours of regions within the grayscale images of both original and dichromat images. Its leading idea is to identify contours (or part of them) in the grayscale original image that do not exist in grayscale dichromat image. This is illustrated in Figs. 4(e) and (f), where the contours of the flower are completely absent in the grayscale dichromat image (Fig. 4(f)). This is because a deutan dichromat individual cannot perceive the difference between the green background from the red flower at all.

Such contours are determined by applying the Sobel gradient masks in the  $x$  and  $y$  directions to each pixel [22]. As a result, we obtain two Sobelized images  $G_x$  and  $G_y$  in the  $x$  and  $y$  directions, respectively, which are then merged into a Sobelized image through the following expression:

$$S = \sqrt{G_x^2 + G_y^2} \quad (4)$$

Note that this process must be applied to both grayscale counterparts (Figs. 4(e) and (f)) of the original image and dichromat image. The resulting images are shown in Figs. 4(g) and (h).

### E. Contour Highlighting

The contour enhancement step in subtracting the contours of the grayscale dichromat image to those contours of the grayscale original image. In this way, we obtain the missing contours in the dichromat image, which are then highlighted with more or less luminance, depending on the current luminance of each pixel in the dichromat image, as illustrated in Figs. 4(i) and (j).

## III. RESULTS

To our best knowledge, CEA is the first contour-based color adaptation algorithm for CVD people. However, our testing was focused exclusively only on deutan and protan dichromat people, largely because their color perception is quite similar, in addition to the fact that deuteranomaly and protanomaly are the most common types of dichromacy we find in world male population.

### A. Setup

Testing was performed using a 64-bit Microsoft Windows laptop



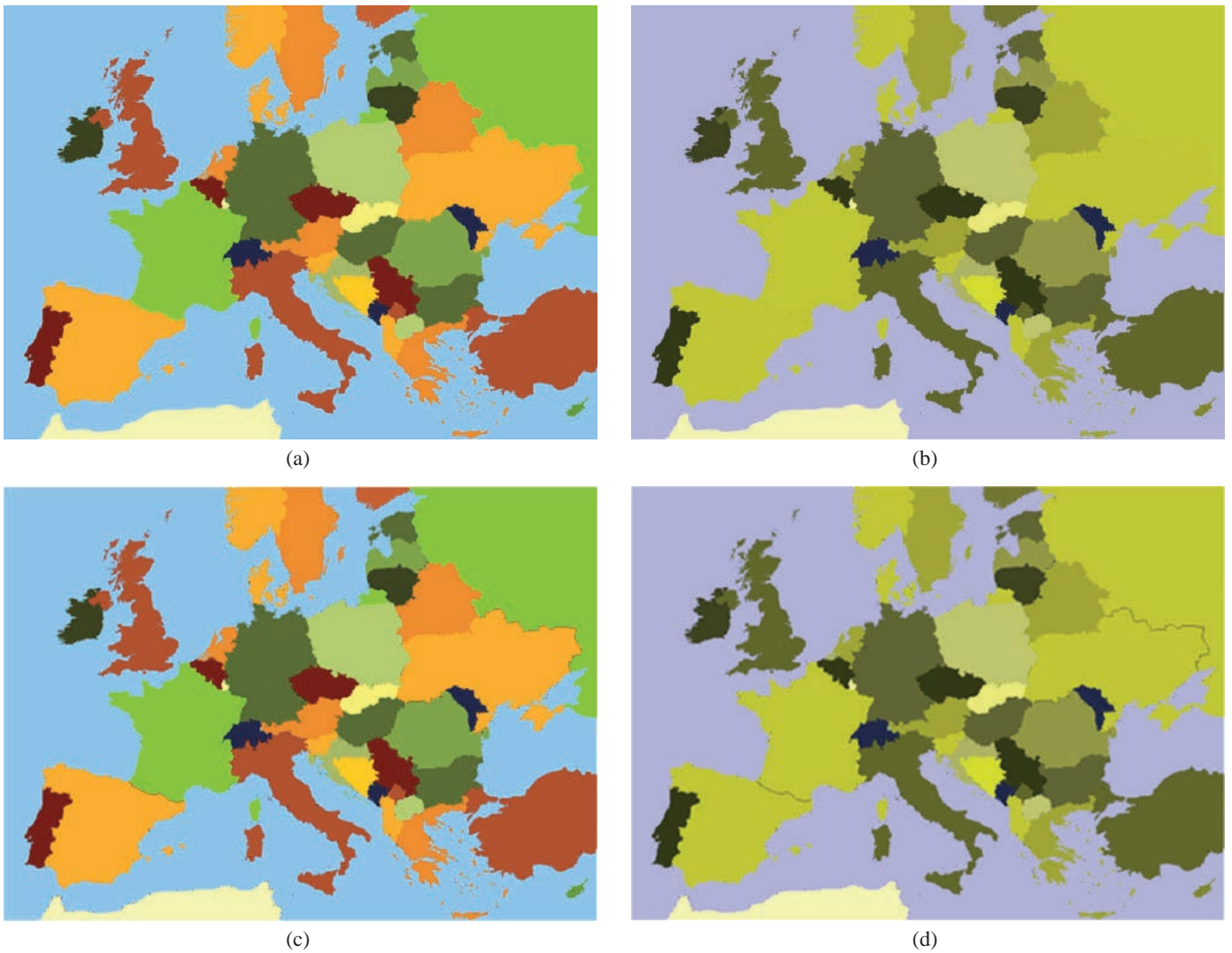


Fig 5. Europe map: (a) original image as seen by a trichromat individual; (b) original image as seen by a deutan deuteranope individual; (c) image as seen by a trichromat individual after adaptation through the CEA method; (d) image as seen by a deutan deuteranope individual after adaptation through the CEA method.

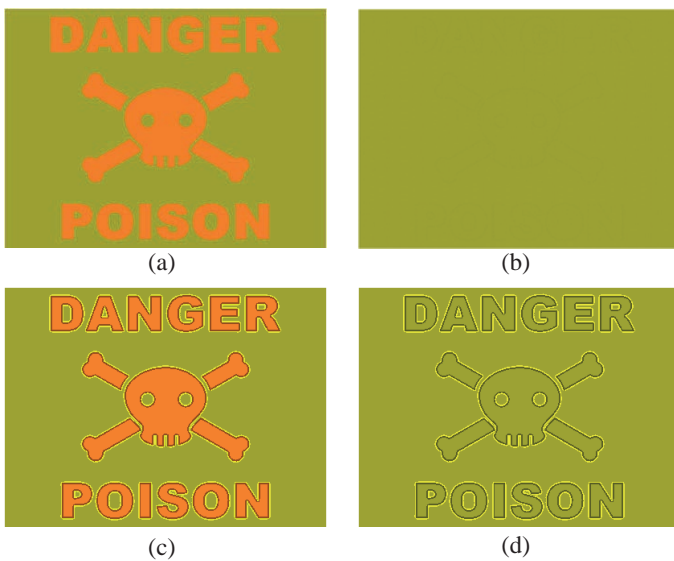


Fig 6. Nameplate of danger as seen by: (a) a trichromat individual; (b) a deutan dichromat individual; (c) a trichromat individual after CEA adaptation; (d) a deutan dichromat individual after CEA adaptation.

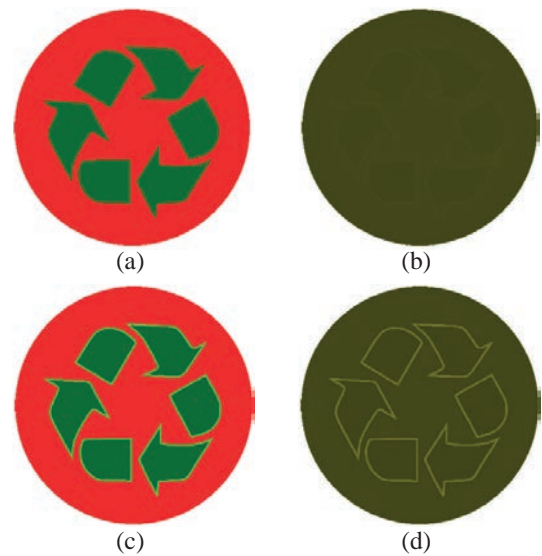


Fig 7. Recycling sign as seen by: (a) a trichromat individual; (b) a deutan dichromat person; (c) a trichromat person after CEA adaptation; (d) a deutan dichromat person after CEA adaptation.



equipped with an Intel Core i7-4750HQ CPU 2.0GHz, with 8GB RAM. CEA algorithm and its competitors (Ching-Sabudin method [24] and Iaccarino et al.'s method [10]) were coded in Javascript programming language for HTML5 web compliant browsers, including Firefox and Chrome. Note that we also implemented in Javascript the Vienot et al.'s algorithm [5] which simulates how deuteranope people see colors, and yet an algorithm to count regions in images.

Those two competitor algorithms (Ching-Sabudin method and Iaccarino et al.'s method) were chosen because they also apply to deutan and protan dichromat people. Besides, these algorithms have the further advantage of their codes are publicly available.

In methodological terms, as explained further below, we used three metrics to evaluate the efficiency of CEA algorithm: perceived region rate ( $\rho$ ), naturalness ( $v$ ), and contrast ( $C$ ).

### B. Perceived Region Rate

By definition, the perceived region rate ( $\rho$ ) is the ratio of the number of regions (or objects) seen by the CVD individual to the number of regions as seen by a trichromat individual. For example, the image depicted in Fig. 7(a) has 7 regions seen by a trichromat individual, but the same image as seen by a deutan dichromat individual has only one region; therefore, the value of  $\rho = 1/7$ . That is, there are 6 out of 7 regions that not seen by the deutan dichromat individual. But, after applying our CEA algorithm, every single deutan dichromat person could see those 6 unseen regions. This was accomplished without changing the color of pixels inside each region; only region contour pixels were changed where needed to avoid color confusion. This made it possible to distinguish between Spain and France in Fig. 5 along their common border.

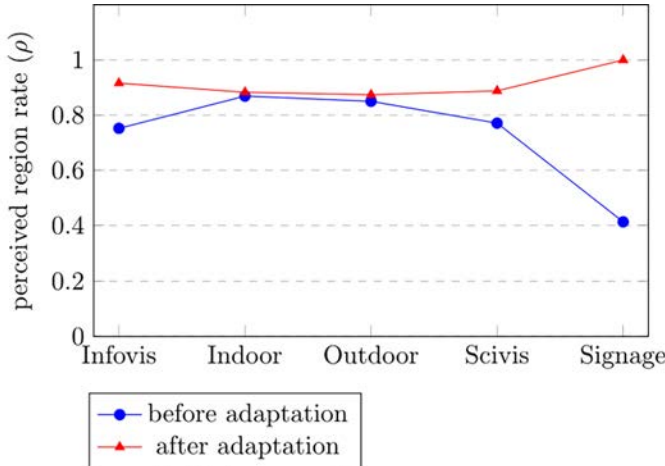


Fig 8. Perceived region rate for five categories of images before (in blue) and after (in red) adaptation through CEA.

In Fig. 8, we show the overall results for perceived region rate in respect to a dataset of 100 images divided into five categories, namely: infovis (information visualization), indoor, outdoor, scivis (scientific visualization), and signage. Examples of these images are shown in Figs. 9-10 further ahead. As observed from Fig. 8, the perceived region rate of our algorithm has significant gains for infovis-, scivis-, and signage-type images, when compared to the corresponding images as seen by deutan and protan dichromats before adaptation. That is, these CVD people see more image regions after adaptation than before adaptation, resulting in an increasing of their visual perception.

### C. Contrast

In the literature, there are various ways for contrast measuring of images. Our contrast metric  $C$  is based on Squared Laplacian [25], and is as follows:

$$C = \frac{1}{W \cdot H} \sum_{x=1}^W \sum_{y=1}^H G(x, y)^2 \quad (5)$$

being  $G(x, y)$  given by:

$$G(x, y) = \sum_{i=x-1}^{x+1} |I(x, y) - I(i, y)| + \sum_{j=y-1}^{y+1} |I(x, y) - I(x, j)| \quad (6)$$

where  $I(x, y)$  is the value of the intensity of the pixel at position  $(x, y)$ .  $W$  and  $H$  are the width and height of the image.

For the computation of the pixel intensity, we use the formula from Eq. 7, derived from by Poynton [23]:

$$I(x, y) = 0.299 \frac{R}{255} + 0.587 \frac{G}{255} + 0.114 \frac{B}{255} \quad (7)$$

After applying the simulation algorithm for deuteranopes described in [5], the mean average value of the contrast of the original dataset as seen by deuteranope people was  $\bar{C} = 0.0418$ . After applying our contour enhancement procedure, the average contrast increased to  $\bar{C} = 0.054$ , representing a gain of 28.5%. In respect to Iaccarino et al.'s method [10], the average contrast for the same dataset of 100 images was  $\bar{C} = 0.0417$ , so there was no gain in contrast. Finally, we obtained an average contrast of about 0.047 for Ching-Sabudin method [24], featuring an increase of 12.4% relative to the average contrast before the color adaptation process.

Notice that the contrast computation is achieved only after the deuteranope view simulation (which comes after the contour enhancement), since our goal is to improve the perception of images for deutan and protan dichromat people.

### D. Naturalness

According to Flatla et al. [26], the color naturalness of an image can be expressed as follows:

$$v = \frac{1}{W \cdot H} \sum_{i=1}^{W \cdot H} \Delta(P_i, P_i^*) \quad (8)$$

where  $W \times H$  denotes the image resolution,  $P_i$  is the color of the  $i$ -th pixel, and  $P_i^*$  is the color of this pixel after the color adaptation, while  $\Delta(P_i, P_i^*)$  denotes the color difference between  $P_i$  and  $P_i^*$  in conformity with the CIE76 color-difference formula expressed in CIE Lab space coordinates given by

$$\Delta = \sqrt{(L_i^* - L_i)^2 + (a_i^* - a_i)^2 + (b_i^* - b_i)^2} \quad (9)$$

where  $(L_i, a_i, b_i)$  and  $(L_i^*, a_i^*, b_i^*)$  represent the Lab colors of  $P_i$  and  $P_i^*$ , respectively. The smaller the value of  $v$ , more natural is the re-coloring procedure of each image.

Considering the reference dataset of 100 still images, we achieved an average value for the naturalness of about  $v = 2.0$  for deuteranope people. For the methods in comparison, Iaccarino et al. [10] and Ching and Sabudin [24], we got the naturalness scores of  $\bar{v} = 8.6$  and  $\bar{v} = 20.76$ , respectively. These values are expectable, since our method only changes the contour pixels, not interior pixels of regions; Iaccarino's et al.'s method changes colors to close colors, while Ching-Sabudin method changes colors to far away colors. Summing up, CEA method outperforms those two competitor algorithms in respect to color naturalness maintenance.

#### IV. SUBJECTIVE EVALUATION

To assess the CEA algorithm in the perceptual augmentation of CVD users, we carried a statistical study based on a questionnaire as described below.

##### A. The Universe of CVD People

The subjective evaluation involved 13 CVD male volunteers. Initially, they performed the D-15 Color Arrangement Test [27] for a more accurate characterization of the universe of CVD users. The following results were obtained: 2 people with strong protanomaly (ages 32 and 49), 1 person with moderate protanomaly (age 27), 5 people with strong deuteranomaly (ages 49, 51, 57 and 69), 4 people with moderate deuteranomaly (ages 17, 21, 26 and 46), and 1 person with deuteranopia (age 35). In spite of the fact that the CEA algorithm has been designed for deuteranope and protanope people, we also considered deuteranomalous and protanomalous people in our study because of their high degree of CVD severity, i.e., they see in a similar way to deutan and protan dichromats [28].

##### B. The Questionnaire

###### Methodology

Unlike most color adaption usability studies based on questionnaires [11] [12], we have not adopted the Law of Comparative Judgment (LCJ) of L.L.Thurstone [29], because this law only allows us to compare two alternatives or algorithms. In our study, we compare four alternatives: the original image without color adaptation (WCA), Iaccarino et al.'s method [10], Ching-Sabudin method [24], and our CEA method. Consequently, we decided to use descriptive statistics techniques [30] [31], particularly the following metrics: arithmetic mean ( $\bar{x}$ ), standard deviation ( $\sigma$ ), and coefficient of variation ( $v = \sigma / \bar{x}$ ), also called relative standard deviation. For each questionnaire image, those four alternatives are presented randomly to each CVD individual to not influence the choice of the respondents somehow.

###### Dataset of images

In our subjective evaluation study, we considered five categories of images (see Figs. 9 and 10): InfoVis, concerning visualization of information; Indoor, concerning indoor scenes, Outdoor, concerning outdoor scenes, SciVis, concerning scientific visualization; and Signage, concerning traffic and warning signs. More specifically, we selected six images by category, in a total of thirty images. Therefore, each CVD individual had to score thirty images times four alternatives in a total of thirty images. Therefore, each CVD individual had to score thirty images times four alternatives, in a total of 120 images; the scores were thus 1 (highest score), 2, 3, and 4 (lowest score).

It is worth noting that the dataset of images was selected in conformity with the principles of representativeness and diversity suggested by Shaffer and Zhang [31] to reduce the sampling error and, consequently, getting a significant statistical confidence interval.

###### Implementation

For the implementation of the questionnaire available at <http://cea.ipcb.pt/>, we used the Google Forms from the Google Docs application package to collect data from each CVD individual. Basically, the questionnaire consists of six web pages, one page per category of images. These images are disposed in the questionnaire in the same order as they are in Figs. 9 and 10.

In the questionnaire, each CVD volunteer expresses his/her satisfaction degree (or preference) by scoring the four alternatives of each image, regarding the discrimination of the contents (contrast) and naturalness. Scoring is performed using a (discrete) qualitative ordinal scale, from the more (highest score 1) to less (lowest score 4) preferred alternative. Note that this scoring scale is adequate up to five items [32] [33].

##### Validation

For the validation of the questionnaire, we benefited from the contribution of 2 statisticians and 2 CVD researchers. The statistics experts were important to adopt descriptive statistics techniques, as adequate for multiple options up to 5, as well as to design the questionnaire itself. In respect to CVD researchers, one of whom is also a dichromat individual, they played an important role in the formulation of the questions focused on color contrast and naturalness, as selected for the questionnaire.

##### C. Data Scoring and Collecting

Fig. 11 shows the raw quantitative results obtained from the questionnaire. These results express the CVD people's preferences relative to five image categories depicted in Figs. 9 and 10, namely: InfoVis, Indoor, Outdoor, SciVis, and Signage. Recalling that we have 6 images per category and a universe of 13 respondents, we see the data sample consists of 78 responses ( $= 6 \times 13$ ) per category; for example, considering the CEA method, the data sample for the SciVis category (see Fig. 11(d)), comprises 1 response with score 1, 16 responses with score 2, 22 responses with score 3, and 18 responses with score 4. Summing the number of responses for any other method, we always obtain 78 responses with the scores ranging in [1, 4]. Thus, the number of scores is equal to the number of methods under analysis.

##### D. Data Analysis

We based data analysis on two descriptive statistical tools: (i) box-and-whisker diagrams (see Fig. 12); (ii) coefficient of variation (see Table I). Such diagrams and statistical data were produced from raw data presented in Fig. 11. The box-and-whisker diagrams constitute a visual tool that helps us to observe the distribution of preferences of CVD respondents radially [34]. As its name says, a box-and-whisker diagram has one box and two whiskers (see Fig. 12). The box represents the consensus of preferences and includes at least 50% of the preferences of each method. More specifically, such box represents the preferences of the second and third quartiles, which are put apart through the median (cf. horizontal straight-line segment inside the box). The arithmetic mean shows off as a cross inside the box.

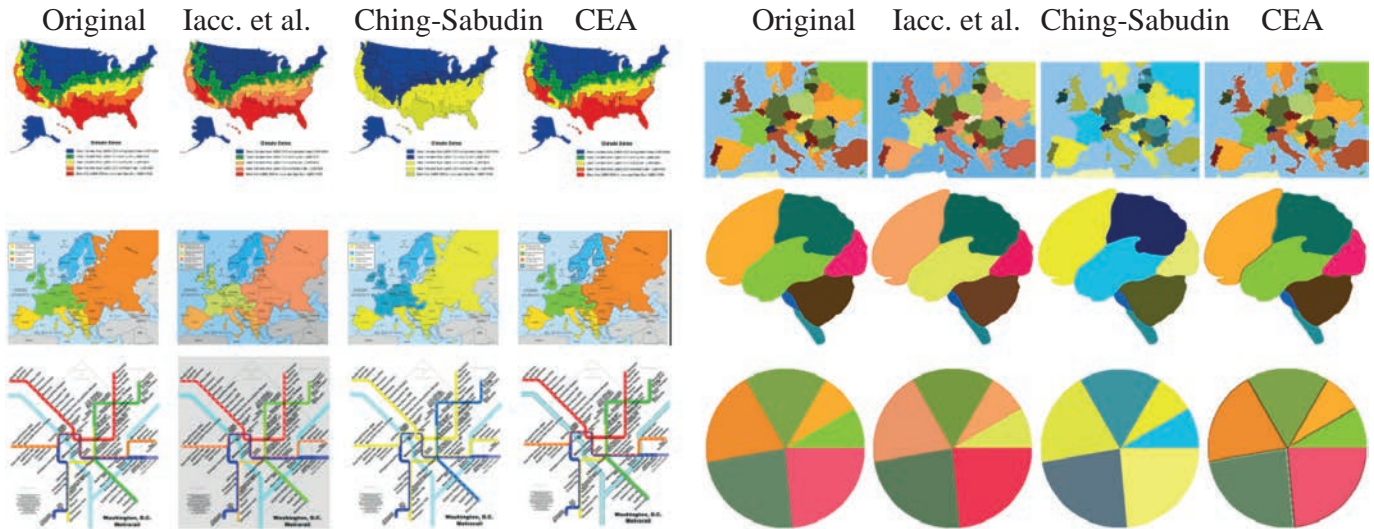
On the other hand, the coefficient of variation (CV, for brevity) measures the dispersion/concentration of the distribution of such preferences [30], as shown in Table I. Recall that the coefficient of variation  $v$  is given by the ratio  $\bar{x}/\sigma$ , where  $\bar{x}$  stands for the mean, and  $\sigma$  the standard deviation. Sometimes, the coefficient of variation is also called relative standard deviation.

Therefore, a brief glance at Figs. 11 and 12 and statistical data listed in Table I shows us the following:

- *Infovis*: When compared to other methods, CEA method is the one with the highest mean (2.67). Furthermore, it presents the lowest coefficient of variation (36%). So, in respect to infovis-type images, CEA method clearly performs better than any other testing method, including the WA approach.
- *Indoor*: Taking into consideration the diagram (b) in Fig. 12, we again see that both WA and CEA methods are clearly better than the other two methods, but their arithmetic means are similar, 3.04 and 2.94, respectively. Despite its higher dispersion (31%) face to the CEA method (26%), the WA method tends to count with higher scores (31 out of 78 preferences with score 4) than CEA method. Therefore, the WA method ranks first for the indoor-type images.
- *Outdoor*: In respect to outdoor-type images, the WA alternative reaches the higher mean (3.06) and the smaller dispersion (29%), so it ranks first for the outdoor-type images. Thus, the best color adaptation method for outdoor-type images seems to be not using any recoloring procedure at all.



## :: InfoVis ::



## :: Indoor ::



## :: Outdoor ::

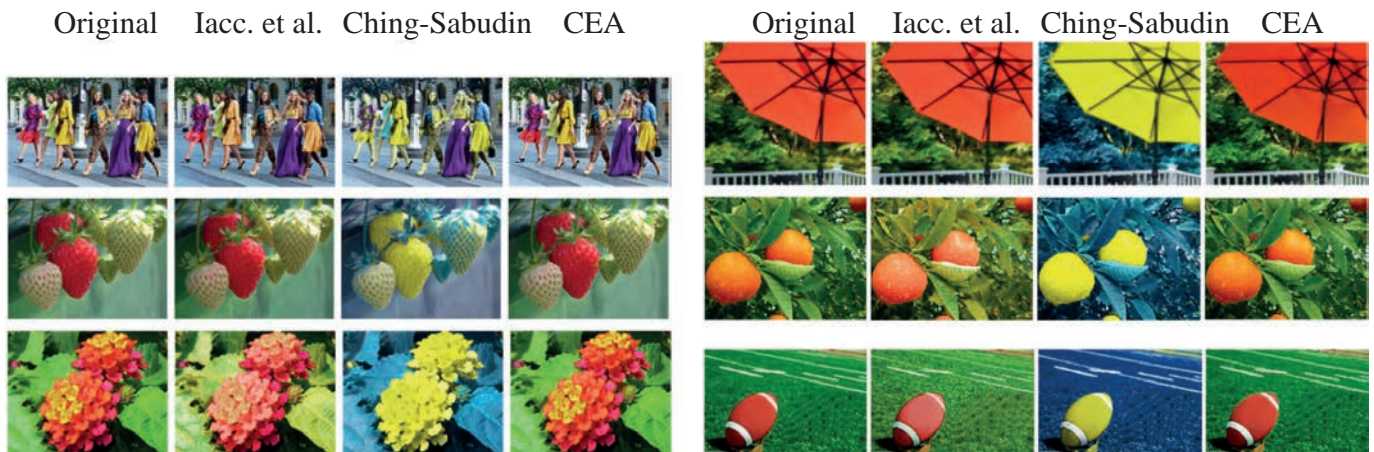


Fig 9. Thumbnails of infovis-, indoor-, and outdoor-type images used in the usability test.



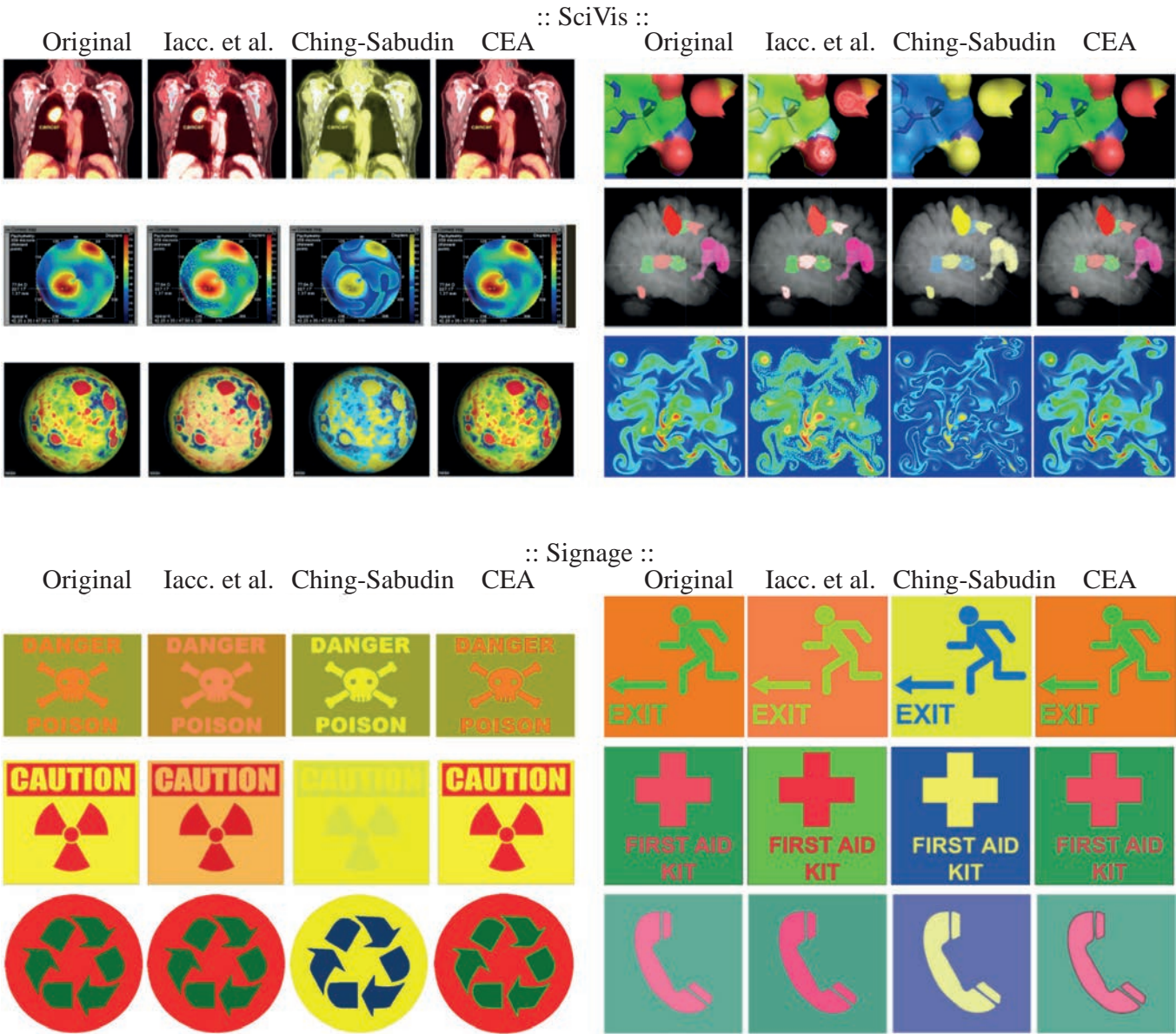


Fig 10. Thumbnails of scivis- and signage-type images used in the usability test.

	Without adaptation	Iaccarino et al.	Ching & Sabudin	Sobel based
1	17	14	36	11
2	23	28	7	20
3	19	22	6	31
4	19	14	29	16

(a)

	Without adaptation	Iaccarino et al.	Ching & Sabudin	Sobel based
	Indoor			
	6	32	38	2
	15	28	15	20
	26	7	7	37
	31	11	18	19

(b)

	Without adaptation	Iaccarino et al.	Ching & Sabudin	Sobel based
	OutDoor			
	2	12	57	7
	21	37	5	15
	25	18	0	35
	30	11	16	21

(c)

	Without adaptation	Iaccarino et al.	Ching & Sabudin	Sobel based
	SciVis			
	8	35	33	2
	17	19	20	21
	30	10	5	33
	23	14	20	22

(d)

	Without adaptation	Iaccarino et al.	Ching & Sabudin	Sobel based
	Signage			
	36	6	23	13
	15	32	2	29
	15	33	7	23
	12	7	46	13

(e)

Fig 11. Breakdown of the preference scores per category.



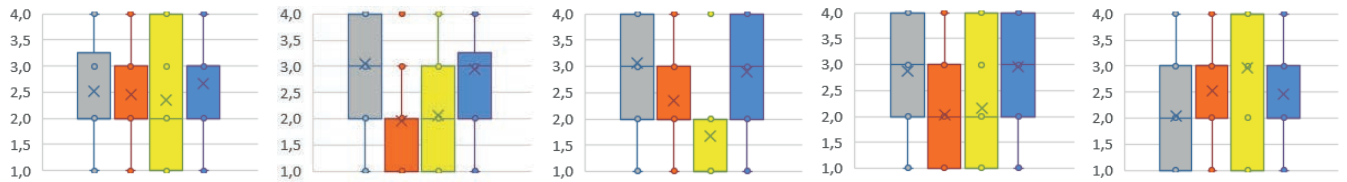


Fig 12. Box-and-whisker diagrams of the distribution of preferences per category: (a) infovis; (b) indoor; (c) outdoor; (d) scivis; and (e) signage. The mean is represented by a cross, while the median is represented by a horizontal line.

- *SciVis*: Regarding scivis-type images, the best adaptation method is the CEA method, since it has the higher mean (2.96) and the smaller dispersion (27%), i.e., it is more consensual than any other method.
- *Signage*: Ching-Sabudin method ranks first in this image category because it has the higher mean (2.97), despite its high coefficient of variation (45%). In fact, the coefficient of variation is not relevant in this case because most scores (46 out of 78) accumulate at bin 4. Note that the CEA method is also a good solution for signage-type images because it also allows for the identification of unseen image regions, as shown in Fig. 8.

Summing up, CEA method seems to be better than WA method in three categories of images, namely infovis-, scivis- and signage-type images. Regarding indoor- and outdoor-type images, the best option is to leave them as they are, i.e., the WA method, according which there is no need performing any computational recoloring procedure. These results agree with those shown in Fig. 8.

Recalling that it is too difficult to find CVD people who could collaborate in the kind of studies, most of published works in this area carried out the assessment without the opinion of this audience, or involving only a group with a small number of people. However, this constitutes a limitation in the results achieved and in the drawn conclusions, as with our study.

TABLE I. STATISTICAL RESULTS

category	metric	WA	Iaccarino <i>et al.</i>	Ching- Sabudin	CEA
InfoVis	$\bar{x}$	2.51	2.49	2.36	2.67
	$\sigma$	1.090	0.989	1.386	0.963
	$v$	43%	40%	59%	36%
Indoor	$\bar{x}$	3.04	1.96	2.06	2.94
	$\sigma$	0.946	1.038	1.231	0.779
	$v$	31%	53%	60%	26%
Outdoor	$\bar{x}$	3.06	2.36	1.68	2.90
	$\sigma$	0.873	0.911	1.211	0.906
	$v$	29%	39%	67%	31%
SciVis	$\bar{x}$	2.87	2.04	2.15	2.96
	$\sigma$	0.958	1.145	1.228	0.813
	$v$	33%	56%	57%	27%
Signage	$\bar{x}$	2.04	2.53	2.97	2.46
	$\sigma$	1.133	0.768	1.348	0.963
	$v$	56%	30%	45%	39%

Statistical metrics:

$\bar{x}$ : arithmetic mean;  $\sigma$ : standard deviation;  $v$ : coefficient of variation.

## V. CONCLUSIONS

In contrast to the state-of-the-art of the color adaptation methods, which use a pixel-wise recoloring procedure in an attempt of improving the color perception of deutan and protan dichromats, the focus of the CEA method is on the detection and highlighting of region contours where necessary, so the region interiors are left untouched. In other

words, the CEA method is contour-wise recoloring procedure. Thus, the CEA method is a disruptive technique when compared to the current state-of-the-art of color adaptation methods. This opens new perspectives for a new family of color adaptation methods based on image analysis and processing.

In fact, to discriminate between confusing neighbor regions of a given image, we have introduced a contour highlighting or enhancement algorithm that increases the image contrast, while keeping the naturalness of image color, since there are no color changes in the region interiors and most contours do not need to be highlighted. Consequently, deutan and protan dichromats can see more images regions than usual, i.e., their image perception increases to a rate close to trichromats' perception, but without disturbing their perceptual learning about the surrounding world.

## ACKNOWLEDGMENT

The authors would like to thank to Carla S. Pedro, Isabel Castanheira and Paulo Silveira for their support in statistical analysis of usability questionnaire, Marco Bernardo and Vasco Almeida for their criticism relative to color adaption techniques, as well as CVD participants for their help and time to answer the usability questionnaire.

This research has been partially supported by the Portuguese Research Council (Fundação para a Ciência e Tecnologia), under the FCT Project UID/EEA/50008/2019.

## REFERENCES

- [1] I. Paterson, *A Dictionary of Colour*. London: Thorogood Publishing Ltd., 2003.
- [2] H. K. Kolb, E. F. Fernandez, and R. N. Nelson, *WebVision: The Organization of the Retina and Visual System*. John Moran Eye Center, University of Utah, 2011.
- [3] J. Birch, *Diagnosis of Defective Colour Vision*, 2nd ed. Edinburgh: Elsevier Science, 2001.
- [4] B. J. Case, "Color blindness," Pearson Education, Tech. Rep., 2003.
- [5] F. Vienot, H. Brettel, and J. D. Mollon, "Digital video colourmaps for checking the legibility of displays by dichromats," *Color Research and Application*, vol. 24, no. 4, pp. 243–252, 1999.
- [6] S. Yang and Y. M. Ro, "MPEG-21 digital item adaptation for color vision deficiency," in *Proceedings of the 2003 International Workshop on Advanced Image Technology (WAIT'03)*, Nagasaki, Japan, January 21–22, 2003.
- [7] C. Martin, S. Rogers, and M. Kabrisky, "Digital production of color mach bands using a color human visual system model," *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol. 28, no. 1, pp. 127–130, 1998.
- [8] M. Ichikawa, K. Tanaka, S. Kondo, K. Hiroshima, K. Ichikawa, S. Tanabe, and K. Fukami, "Preliminary study on color modification for still images to realize barrier-free color vision," in *Proceedings of the 2004 IEEE International Conference on Systems, Man and Cybernetics*, vol. 1. IEEE Press, 2004, pp. 36–41.
- [9] K. Rasche, R. Geist, and J. Westall, "Detail preserving reproduction of color images for monochromats and dichromats," *Computer Graphics and Applications*, IEEE, vol. 25, no. 3, pp. 22–30, 2005.
- [10] G. Iaccarino, D. Malandrino, M. Del Percio, and V. Scarano, "Efficient edge-services for colorblind users," in *Proceedings of the 15th International*

- Conference on World Wide Web (WWW'06), Edinburgh, Scotland, May 23-26. ACM Press, 2006, pp. 919-920.
- [11] J.-B. Huang, Y.-C. Tseng, S.-I. Wu, and S.-J. Wang, "Information preserving color transformation for protanopia and deuteranopia," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 711-714, 2007.
- [12] G. Kuhn, M. Oliveira, and L. Fernandes, "An efficient naturalness preserving image-recoloring method for dichromats," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, pp. 1747-1754, 2008.
- [13] P. Doliotis, G. Tsekouras, C.-N. Anagnostopoulos, and V. Athitsos, "Intelligent modification of colors in digitized paintings for enhancing the visual perception of color-blind viewers," ser. *IFIP Advances in Information and Communication Technology*, I. Iliadis, M. Maglogiann, T. Tsoumakasis, V. Vlahavas, and B. Bramer, Eds. Springer Boston, 2009, vol. 296, pp. 293-301.
- [14] M. Wang, B. Liu, and X.-S. Hua, "Accessible image search for colorblindness," *ACM Transactions on Intelligent Systems and Technology*, vol. 1, no. 1, pp. 8:1-8:26, October 2010.
- [15] W. Chen, W. Chen, and H. Bao, "An efficient direct volume rendering approach for dichromats," *IEEE Transactions on Visualization and Computer Graphics*, vol. 17, no. 12, pp. 2144-2152, 2011.
- [16] M. M. Ribeiro and A. J. Gomes, "A skillet-based recoloring algorithm for dichromats," in *Proceedings of the 15th IEEE International Conference on e-Health Networking, Applications & Services (Healthcom'13)*, Lisbon, Portugal, October 9-12. IEEE Press, 2013, pp. 702-706.
- [17] "Creating accessibility to web contents for colorblind people," in *Proceedings of the 5th International Conference on Applied Human Factors and Ergonomics (AHFE'14)*, Krakow, Poland, July 19-23, P. Arezes and P. Carvalho, Eds., vol. 10. AHFE Press, 2014, pp. 646-656.
- [18] B. Sajadi, A. Majumder, M. M. Oliveira, R. G. Schneider, and R. Raskar, "Using patterns to encode color information for dichromats," *IEEE Transactions on Visualization and Computer Graphics*, vol. 19, no. 1, pp. 118-129, 2013.
- [19] D. R. Flatla, A. R. Andrade, R. D. Teviotdale, D. L. Knowles, and C. Stewart, "ColourID: Improving colour identification for people with impaired colour vision," in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI'15)*, Seoul, Republic of Korea, April 18-23. ACM Press, 2015, pp. 3543- 3552.
- [20] V. Smith and J. Pokorny, "Spectral sensitivity of the foveal cone photopigments between 400 and 500 nm," *Vision Research*, vol. 15, no. 2, pp. 161-171, 1975.
- [21] D. Travis, *Effective Color Displays: Theory and Practice*. London: Academic Press, 1991.
- [22] R. Gonzalez and R. Woods, *Digital Image Processing*, 3rd ed. Pearson, Prentice Hall, 2007.
- [23] C. Poynton, *Digital Video and HDTV: Algorithms and Interfaces*. The Morgan Kaufmann Series in Computer Graphics. Morgan Kaufmann Publishers, 2003.
- [24] S.-L. Ching and M. Sabudin, "Website image colour transformation for the colour blind," in *Proceedings of the 2nd International Conference on Computer Technology and Development (ICCTD'10)*, Cairo, Egypt, November 2-4. IEEE Press, 2010, pp. 255-259.
- [25] X. Xu, Y. Wang, J. Tang, X. Zhang, and X. Liu, "Robust automatic focus algorithm for low contrast images using a new contrast measure," *Sensors*, vol. 11, no. 9, pp. 8281-8294, 2011.
- [26] D. R. Flatla, K. Reinecke, C. Gutwin, and K. Z. Gajos, "SPRWeb: preserving subjective responses to website colour schemes through automatic recolouring," in *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'13)*, Paris, France, April 27 - May 02. ACM Press, 2013, pp. 2069-2078.
- [27] A. J. Vingrys and P. E. King-Smith, "A quantitative scoring technique for panel tests of color vision," *Investigative Ophthalmology & Visual Science*, vol. 29, no. 1, pp. 50-63, January 1988.
- [28] L. Sharpe, A. Stockman, H. Jagle, and J. Nathans, "Opsin genes, cone photopigments, color vision and color blindness," in *Color Vision*, K. Gegenfurtner and L. Sharpe, Eds. New York: Cambridge University Press, 1999.
- [29] L. Thurstone, "A law of comparative judgment," *Psychological Review*, vol. 34, no. 4, pp. 273-286, 1927.
- [30] J. W. Tukey, *Exploratory Data Analysis*. Addison-Wesley, 1977.
- [31] D. Shafer and Z. Zhang, *Introductory Statistics*. Saylor Foundation, 2012.
- [32] R. Taplin, "The statistical analysis of preference data," *Applied Statistics*, vol. 46, no. 4, pp. 493-512, 1979.
- [33] S. Abeyasekera, "Analysis approaches in participatory work involving ranks or scores," *Statistical Services Centre*, The University of Reading, and Natural Resources Institute, United Kingdom, Tech. Rep., August 2000.
- [34] K. Potter, "Methods for presenting statistical information: The box plot," in *Visualization of Large and Unstructured Data Sets*, ser. *Lecture Notes in Informatics*, H. Hagen, A. Kerren, and P. Dannenmann, Eds. Gesellschaft für Informatik, 2006, vol. S-4, pp. 97-106.



M. Madalena G. Ribeiro

M. Madalena G. Ribeiro reached her PhD in Computer Science at the University of Beira Interior, Covilhã, Portugal, in 2017, a Master's degree in Electrical Engineering and Computers at the Instituto Superior Técnico – University of Lisboa, Lisbon, Portugal, in 2001 and a degree in Mathematics (with Computer Graphics majors) at the University of Coimbra, Coimbra, Portugal, in 1994, having completed the final project as Erasmus student at the Fraunhofer Institute for Computer Graphics Research, in Darmstadt (Germany). She is professor at the School of Applied Arts at the Polytechnic Institute of Castelo Branco (Portugal) since 2005, after 10 years in the Superior School of Technologies (in the same Polytechnic). teaching subjects in the areas of Web design, UI and UX and Programming Languages. Prof. Madalena Ribeiro was collaborator in the Instituto de Telecomunicações (IT) for several years and currently she is an integrated member of the Centro de Investigação em Património, Educação e Cultura (CIPEC) since January'2018, being a founding member. Her expertise is focused in the Human-Computer Interaction (HCI) area. Interface design, usability, user experience, accessibility, color, image processing, web design and development and augmented reality are research interests.



Abel J. P. Gomes

Abel J. P. Gomes is an Associate Professor in Computer Graphics at the University of Beira Interior, Portugal. He obtained a PhD degree in geometric modeling at Brunel University, England, in 2000. He has over 100 publications, including journal and conference articles, and 1 book published by Springer-Verlag. He was Head of the Department of Computer Science and Engineering, University of Beira Interior, Portugal, and the leader of a research unit of Instituto de Telecomunicações, which is one of the biggest research centers in Portugal. He is also a licensed Professional Engineer and member of the IEEE, ACM, and Eurographics. His current research interests include color accessibility, computer graphics algorithms, molecular graphics, geometric computing, and implicit curves and surfaces.

# Handwritten Arabic Documents Segmentation into Text Lines using Seam Carving

M. Daldali, A. Souhar \*

LaRIT laboratory, Faculty of science, University Ibn Tofail, Kénitra (Morocco)

Received 22 December 2017 | Accepted 3 May 2018 | Published 8 June 2018



## ABSTRACT

Inspired from human perception and common text documents characteristics based on readability constraints, an Arabic text line segmentation approach is proposed using seam carving. Taking the gray scale of the image as input data, this technique offers better results at extracting handwritten text lines without the need for the binary representation of the document image. In addition to its fast processing time, its versatility permits to process a multitude of document types, especially documents presenting low text-to-background contrast such as degraded historical manuscripts or complex writing styles like cursive handwriting. Even if our focus in this paper was on Arabic text segmentation, this method is language independent. Tests on a public database of 123 handwritten Arabic documents showed a line detection rate of 97.5% for a matching score of 90%.

## KEYWORDS

Text Line Segmentation, Arabic Documents, Handwritten Character Recognition, Projection Profile, Seam Carving.

DOI: 10.9781/ijimai.2018.06.002

## I. INTRODUCTION

WITH the advent of digital means to share information, people are slowly abandoning paper as a medium and use digital devices and technologies instead. With the volume of processed information growing every day, businesses, organizations and public services adopt new digital technologies instead of paper documents. This transition brings a real need for scalable optical character recognition (OCR) systems capable of converting paper documents - handwritten or printed - into digital formats.

In addition to modern handwritten texts, there is a huge amount of historical documents in libraries and archives that have not been exploited yet. Systems capable of processing such manuscripts will be beneficial for building indexed databases and computer systems helping to preserve the cultural heritage that these documents represent, as well as easing the research process in historical libraries, which will be beneficial for researchers across disciplines.

OCR systems use many steps to process document images. In general, these steps can be summarized in four main steps, text blocks identifying, text lines segmentation, word segmentation, and finally character recognition. This is making the text lines segmentation a very important phase in the recognition process since the recognition rate depends strongly on it, as any error occurring at this stage will highly affect the recognition efficiency.

Most of the proposed solutions in the literature are based on Connected Components (CC) analysis. This type of approaches has shown some struggles dealing with low text-to-background contrast images such as historical manuscripts or damaged documents because of the binarization pre-processing as shown in [1], where binarization

may ensue critical information loss. An approach based on less filtered data would offer better results on such documents. The CC analysis may also be an overkill when applied to well-structured text lines, like in the case of printed documents.

Human readers can identify text lines without knowing exactly their content or the language which they are written in. This can be noticed when the document is presented to the reader from a considerable distance. As a result, we can safely assume that the text line segmentation is based on identifying the interline space where each line is defined by their neighbors as previously stated in the watershed approach [2], where the authors used CC analysis in order to define the locations of high and low information density, recognizing the text areas from the empty interlines. In a similar way, the seam carving is an image processing tool based on the notion of identifying high and low information areas on an image. Relying on dynamic programming, this can offer better computation efficiency and can show new opportunities for developing more complex and efficient approaches. Additionally, the seam carving offers an additional layer of versatility as only the gray scale of the document image is used, which can make an OCR system capable of processing many types of documents.

The next chapters will discuss the seam carving technique, its text line segmentation implementation, and practical tests results. Firstly, an overview on related works will be conducted in the second chapter. Followed by a presentation of the seam carving technique and its use for text line segmentation, then in the fourth chapter, practical test and accuracy results will be conducted. Finally, the robustness of this seam carving approach is discussed with practical illustrations leading to conclusions and perspectives for future works.

## II. RELATED WORKS

In general, there are two types of text lines segmentation systems: either by searching in a structural way for related physical units such as Connected Components, or through analyzing the documents in search for global image features.

\* Corresponding author.

E-mail addresses: mehdi.daldali@outlook.com (M. Daldali), houssouhar@gmail.com (A. Souhar).

In the article [2], a watershed approach was used to detect text components locations then trying to find the best arranged parts to form text lines, following a local vision or as more commonly known as “bottom-up” analysis. The segmentation process begins by identifying the multiple text components present on the image, then the use of the watershed technique to split the document image into well-defined areas containing each component. These areas are then analyzed in order to clustering them into well-aligned text lines. The same can be said for the other [3] approach where a multi-agent system was used to process each step of the line segmentation from de-noising to dealing with touching components. In case of [4], the problem was modeled into a decision making situation. Starting from a text component, the system tries to find the best next text component to add to a collection until a text line is finally defined using a Markov Decision Process (MDP). The work presented in [5] was also based on connected components analysis where text components are categorized into subsets according to their importance in defining potential text lines that are later analyzed using the Hough Transform. While the last works are based on text components detection, the approach used in [6] rely heavily on binarization of the document image in order to differentiate between text ink and blank spaces, these spaces are then used to define potential text blocks and text lines across the document.

In contrast, the methods using “top-down” analysis do not rely on rough estimations of physical components to detect text lines which require a binary representation of the document image. By using less filtered data avoiding binarization, these methods can extract more global features resulting in a better accuracy especially on handwritten documents using cursive styles like Arabic manuscripts or degraded documents showing a low text-to-background contrast as shown in [1] where the seam carving technique was used combined with a base line approximation technique in order to guide the seam carving process through the inter-line space. Unfortunately, the approach used was inefficient in case of informal handwritten documents where paragraphs and lines show a high amount of discord. [7] was a prior work to [1], and used the seam carving technique in a different context. The document image is first converted to binary representation, then a distance transform was used to measure distance separating the text ink (pixels), then by using this distance as energy, the seam carving technique was used to find the least cost seams passing through the text components. These approximations were then used to group the components into aligned text lines.

Interestingly, in [2] the authors used an hybrid approach to improve the system accuracy. Even if the analysis relies heavily on CC analysis, a projection profile analysis is used as an extension in order to help deciding which components are more likely to be from the same text line following a global vision of the document. This approach is based on watershed technique which relies on connected components analysis to create a similar image partitioning to the Voronoi diagram, where these partitions are then analyzed to merge them into text lines.

The practical tests showed that the watershed approach is efficiently accurate in most of the cases, but relying on CC analysis makes it highly inefficient in case of damaged or low-contrast documents. Additionally, the approach is hard to improve on, notably in case of runtime speedup using parallelization.

By using the seam carving technique, we were able to achieve better versatility which will offer new grounds to improve the segmentation accuracy and ability to build more complex systems.

### III. SEAM CARVING AND LINE SEGMENTATION

#### A. Seam Carving

The Seam Carving operator [8] was proposed first as a solution for resizing images. When re-targeting images, the classical way was to

either stretch the image by interpolating data, cropping the image, or by manual processing.

Seam Carving re-targets images automatically with the least distortion possible in order to preserve important areas dimensions, size ratios and details clarity in a content aware fashion. To achieve this goal, the method uses an energy map to calculate the least-cost seam to remove or duplicate. In more simple words, this seam travels from side to side through a path presenting the least of details on the image. The energy function used can be of several types such as gradient magnitude, visual saliency, and more.

#### B. Energy Map

In case of document images, the energy map must reflect accurately the position of pixels representing the text. One of the best energy maps is the gradient magnitude which offers good efficiency for low complexity, describing accurately the outline of text lines strokes. This map can be obtained by using a Sobel operator which is a matrix convolution of the original grayscale image by a weighted mask set adequately to get the local pixel gradient magnitude from the neighboring values.

After applying a Gaussian filter of a standard deviation  $\sigma$  for noise reduction purposes, the gradient magnitude is then computed using the simple energy function:

$$e(I) = \left| \frac{\partial}{\partial x} I \right| + \left| \frac{\partial}{\partial y} I \right| \quad (1)$$

with  $I \in \mathbb{R}^{n \times m}$  the original document image converted to gray scale having  $n$  rows and  $m$  columns. This is done by calculating the Sobel operator for the vertical and horizontal axis by matrix convolutions:

$$\frac{\partial}{\partial x} I = \begin{pmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{pmatrix} * I \quad (2)$$

$$\frac{\partial}{\partial y} I = \begin{pmatrix} +1 & +2 & +1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix} * I \quad (3)$$

After computing the energy map, a cost map is then computed using dynamic programming in order to deduce the least-cost path from one side of the image to the other.

#### C. Dynamic Programming

The cost map is generated from the energy map using a dynamic programming technique where each position on the cost map equivalent to the pixel position on the image has as value the sum of its own energy, and the least cost value of the 3 neighboring pixel positions from the last column processed:

$$C(I_{i,j}) = e(I_{i,j}) + \min_{k \in \{-1,0,1\}} (C(I_{i-1,j+k})) \quad (4)$$

Where  $I_{ij}$  designates the image value at the  $i^{\text{th}}$  row and the  $j^{\text{th}}$  column, with the origin at the upper left corner of the image, and  $C(I_{i,j})$  the cost value of the position  $(i, j)$ . To find the least-cost seam, the algorithm backtracks starting from the position with the minimum cost value on the last column of the cost map (Fig. 1).

The resulting seams have no constraints while path-finding as shown in Fig. 2. In the case of text lines, the seam calculation must be guided to not overlap potential text by staying in between the lines. For this purpose, medial seams are introduced to guide the seam carving process, using lines' axis approximations.



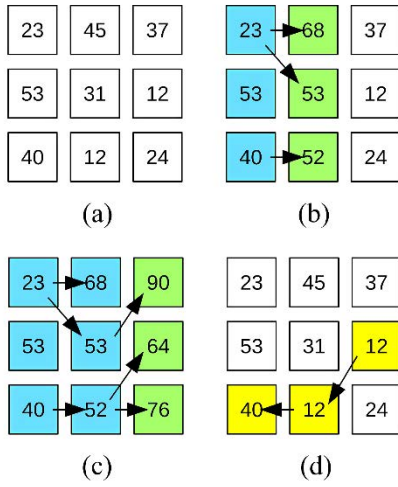


Fig. 1. Example of the dynamic programming used for Seam Carving: the original energy map (a), computing the second column costs (b), calculating the last column costs (c).

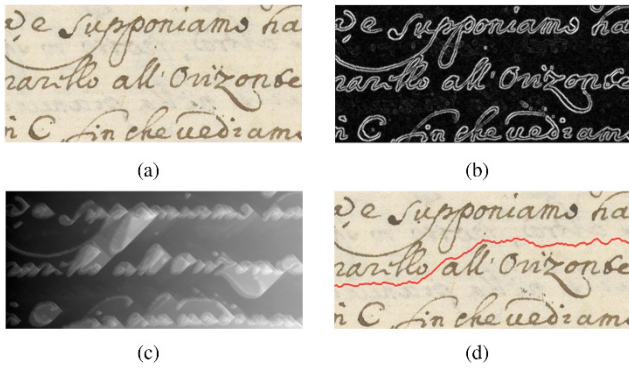


Fig. 2. Example of Seam Carving used on a manuscript image, where (a) is the original image, (b) the gradient magnitude of the image used as an energy map, (c) an intensity representation.

#### D. Medial Seams

The medial seams are approximations of the text lines axis computed by using a projection profile matching approach similar to the work presented in [9]. The image is split vertically into  $s$  slices of equal width  $w$ , then the energy map is used to approximate the axis positions on each slice. Calculating the projection profile is made by summing the values of the energy map for each row of the image slice, and finally extracting local maximum positions after smoothing the resulting data with a smoothing cubic spline filter (Fig. 3) using a smooth factor `smoothFactor`:

$$smooth = smoothFactor \times \sum_{i=1}^n PP(i) \quad (5)$$

Where  $n$  the number of rows present in the processed image,  $PP(i)$  is the projection profile value at the  $i^{th}$  row, and  $smooth$ , the constraint used while calculating the smoothed spline  $SP$ , this constraint is defined as:

$$smooth \geq \sum_{i=1}^n (PP(i) - SP(i))^2 \quad (6)$$

Ensuring that the resulting smoothed histogram is true to the original as possible as it can be. Thus the `smoothFactor` can be considered as

a maximal threshold limiting the difference between the original raw data and the smoothed representation.

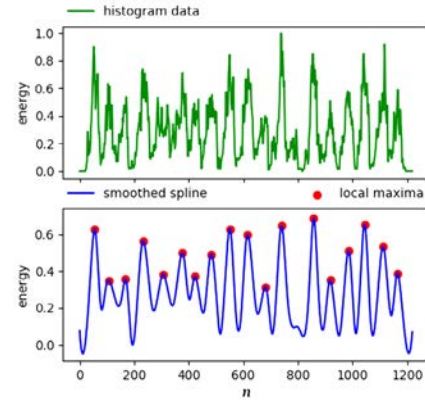


Fig. 3. The projection profiles of a document image slice, the smoothed cubic spline, and the local maximum positions found.

For each local maximums set of each slice, each local maximum 2D position is matched with the closest local maximum position from the adjacent slice using a distance matrix of size  $n1 \times n2$  computed by calculating the distances between each pair of local maximums from each pair of adjacent slices where  $n1$  and  $n2$  is the number of local maximums from each slice:

$$(d_{ij}) = \begin{pmatrix} d_{1,1} & d_{1,2} & \dots & d_{1,n_1} \\ d_{2,1} & d_{2,2} & \dots & d_{2,n_1} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n_2,1} & d_{n_2,2} & \dots & d_{n_2,n_1} \end{pmatrix} \in \mathbb{R}^{n_1 \times n_2} \quad (7)$$

$d_{ij}$  denotes the Euclidean distance between the position of the local maximum of index  $i$  from the first slice and the position of index  $j$  from the other adjacent slice. To match these local maximums, all positions pairs of distance  $d_{ij}$  which values are simultaneously a minimum column-wise and row-wise on the distance matrix and where:

$$\frac{w}{d_{ij}} \geq \sqrt{2}/2 \quad (8)$$

are considered a matched pair of local maxima, and thus, the pair is considered as a piece representing the position of a text line axis passing over the two slices. The last constraint (8) translates as a skew degree limit (Fig. 4) while matching local maximums, with the value  $\sqrt{2}/2$  equivalent to the cosine of a  $45^\circ$  angle. This value was inspired by the results of a neuroscience research in [10] answering questions about degraded text readability, showing a better reader response when the text skew angle is under  $45^\circ$  from the horizon in relation to the reader's eye.



Fig. 4. An example of the skew angle  $\alpha$  between the 2nd and 3rd positions starting from the right on a single Arabic text line.

For deducing the complete lines axis approximations, the solution on [1] proposed a connected components analysis approach (Fig. 5) used to fuse piece-wise approximations with some drastic decisions to avoid problems such as ignoring any seam not starting from the first slice. This makes the algorithm inefficient in most of handwritten documents where text lines present noticeable irregularities.

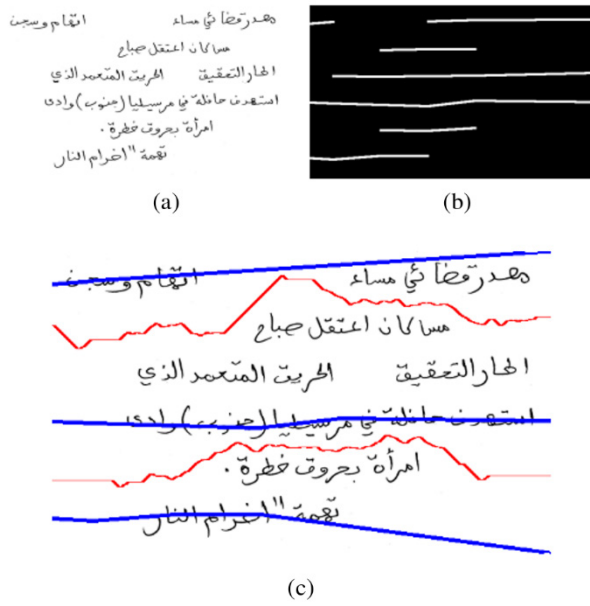


Fig. 5. Example of line extraction using the method on [1]: (a) the original image, (b) the piece-wise approximations detected represented as components, and (c) the resulting segmentation.

We propose a more numerical and geometrical approach, based on interline distances. After defining the matched maximums between all of the document image slices, the piece-wise approximations are assembled by matching extremities into full lines axis approximations using the algorithm on Fig.6, with an example shown on Fig.7.

In order to fuse discontinued lines and eliminate false positives in case of under-smoothing the projection profile histograms, all positions of each line are represented on a lines map, a sparse matrix of size  $s \times l$  where  $s$  is the number of the vertical slices and  $l$  the number of lines detected, while filling the unavailable data with a blank value.

The blanks are then filled using interline distance approximation and the minimum interline distance  $minh$  defined as two thirds of  $h$  the median of the distances between each vertically adjacent local maximums positions.

#### Algorithm

Input:

*Pieces*: list of detected pieces sorted by column

Output:

*Lines*: list of detected lines axis approximations

```

for each piece in Pieces do
    unmatched ← True
    for each line in Lines do
        if line[-1] = piece[0] do
            line ← Concatenate( line, piece )
            unmatched ← False
            break
        end if
    end for
    if unmatched = True do
        Lines ← Insert( piece, Lines )
    end if
end for

```

Fig. 6. Algorithm used for assembling the piece-wise approximations to generate full lines axis approximations. Where line[-1] represents the last element of the line.

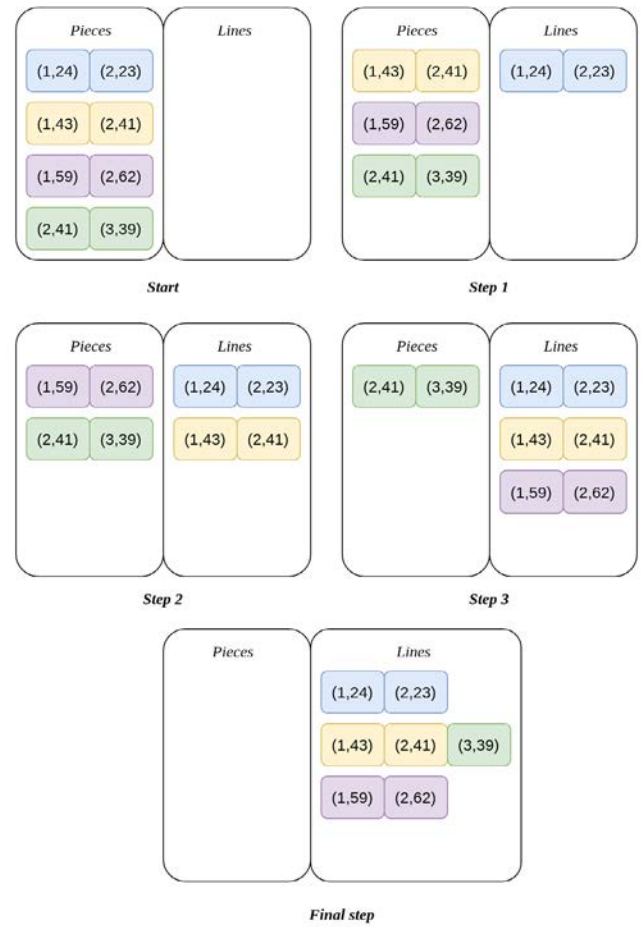


Fig. 7. An example showing the execution of the proposed piece-wise approximation matching algorithm. The values used represent the 2D coordinates of the local maximums positions detected by the projection profile method.

This distance  $minh$  is inspired from the natural spacing of handwritten text lines (also in printed documents). To ensure an acceptable level of readability, a minimum distance must be kept between each two neighboring baselines (the virtual lines on which the text components forming text lines are arranged). Knowing that almost in every handwritten text, there is a number of characters that extend either upward or downward, these extending characters tend to be roughly  $3/2$  times taller than normal height characters, which makes the minimal distance between each two adjacent baselines around  $2/3$  of the typical interline distance as shown in Fig. 8.

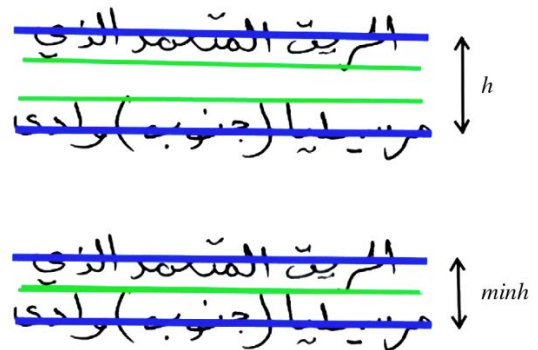


Fig. 8. The approximate minimum interline distance  $minh$  is equal to  $2/3$  of  $h$ , the typical distance between adjacent text lines' axis across the text block.

The interline distance between each consecutive line is calculated by measuring the mean distance between the position on the same column. If no consecutive positions exist, the distance is then calculated using the mean of the ordinate values of each line. Then comes the time to fill the missing values by using the interline distance between the filled line and the line above, without exceeding the value on the line under. A practical example and the algorithm are shown on Fig. 9 and Fig. 10.

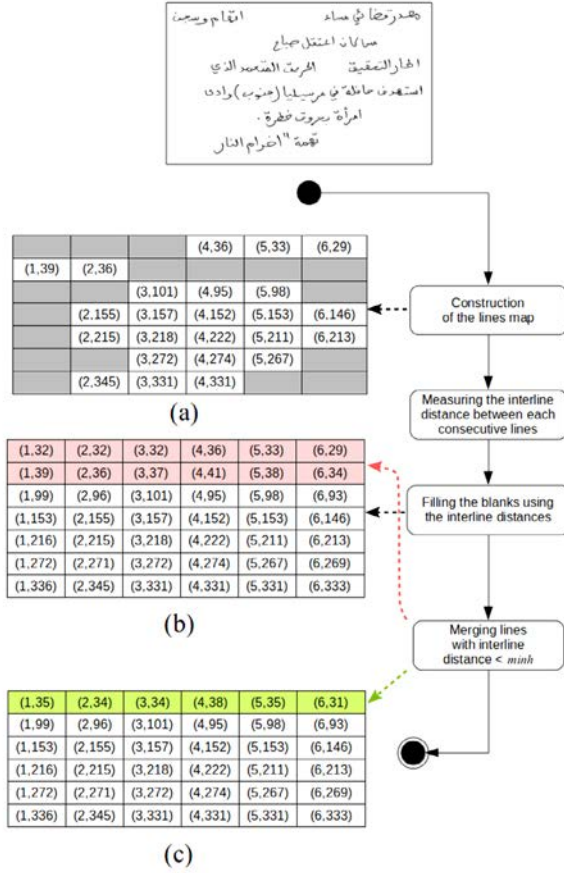


Fig. 9. Example showing the process used for post-processing the lines axis approximations data to generate the lines map: (a) the initial sparse matrix, (b) the filled lines map, (c) the final lines map generated from (b) by merging the lines with the interline distance less than  $minh$ .

#### Algorithm

Input:

*LinesMap*: map of detected lines' axis approximations

*IntDist*: interline distances

Output:

*LinesMap*: the completed map

```

for each line(i) in LinesMap do
  for each value(i,j) in line(i) do
    if value(i,j) = NULL do
      value(i,j) ← value(i-1,j)+IntDist[i]
      if value(i+1,j) ≠ NULL
        and value(i,j) ≥ value(i+1,j) do
          value(i,j) ← value(i+1,j)-1
        end if
      end if
    end if
  end for
end for
end for

```

Fig. 10. The algorithm used to fill the unavailable values on the lines map.

This new method showed better results than the one used in [1], and was able to process many documents exhibiting multiple irregularities, approximating the lines axis accurately. Fig. 11 shows the results when processing the same image used in Fig. 5.

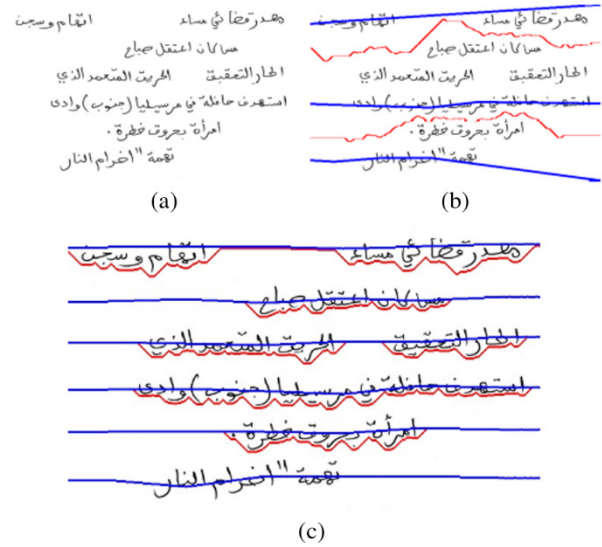


Fig.11. Comparison between method [1] and the method proposed using the same parameters: (a) the original image, (b) the resulting seams from method [1], and (c) the result using the proposed method.

#### IV. EXPERIMENTATION AND EVALUATION

To assess the accuracy of the proposed approach and to compare it to other methods, we faced a lack of test databases for gray-scale and color documents as it brings more challenges to define ground-truths in these cases.

Instead of developing a new database for gray-scale document images, we simply adapted the testing procedure proposed to the connected components oriented data base to suite our line extraction approach. This public dataset has 123 images of Arabic handwritten documents with a total of 1974 lines, available for download (Proximity Dataset, LMP Lab [11]). The method used to evaluate the performance of a system using this database is based on counting the number of matches between the extracted lines and the lines in the ground truth. And so, a criterion called Match Score is used to evaluate the number of detected zones according to the intersection of the sets of pixels on the results with those of the ground truth.

Unfortunately, this criterion is incompatible with the seam carving approach as the extracted lines are image stripes rather than connected components.

In order to circumvent the incompatibility issue, we used a new match score criterion MS2 in order to use the same approach of the harmonic mean F1-Score.

First, for each ground-truth of a document, we converted each line data into a set of pixels  $g_p$ , then, for each line zone extracted  $r_p$ , the new matching score criterion becomes:

$$MS_2(r_i, g_j) = 2 \times \frac{\text{linePrecision} \times \text{lineRecall}}{\text{linePrecision} + \text{lineRecall}} \quad (9)$$

$$\text{linePrecision} = \frac{TP_p}{TP_p + FP_p} \quad (10)$$







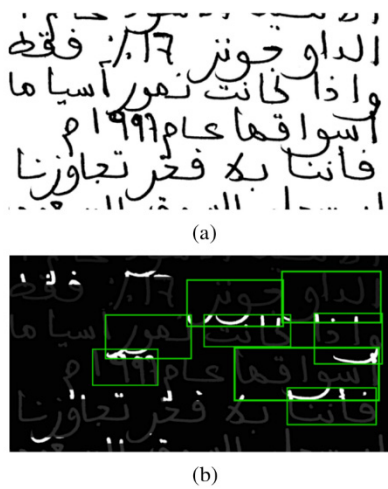


Fig. 13. Example of the overlapping components' ground-truth (convex envelopes in green) and the redundant pixels represented in white: (a) the original image, (b) the reconstructed image from the ground truth.

Even if the accuracy is sensible to parameters values, most of the inaccuracy is caused by undetected text lines due to their considerably short length ranging from single words to simple punctuation symbols.

Choosing different parameters manually for the 18 images, showed a non-negligible improvement of the detection rate. Table II compares the score of the proposed method with other methods tested on the same data set.

TABLE II. RESULTS OF DIFFERENT METHODS

METHODS	F1-Score	
	MS2 = 90%	MS2 = 95%
Souhar et al. [2]	Not Reported	93.3%
Boulid et al. [3]	97.4%	94.3%
Boulid et al. [4]	95.8%	90.5%
Kumar et al. [12]	95.6%	90.9%
Kumar et al. [13]	98.8%	Not Reported
Zhang et al. [14]	98.41%	Not Reported
<b>The proposed method</b>	<b>97.5%</b>	<b>93.4%</b>

Results of different methods obtained on the original data in (Proximity Dataset, LMP Lab [11]).

In case of the matching score threshold of 90%, the precision and recall were respectively 98.44% and 96.39%. This shows that the proposed approach only suffers from false negatives due to the slices width that can exceed text lines length in some cases, making the detection of such lines impossible during the projection profile maxima matching step.

## V. ROBUSTNESS

The text lines segmentation using seam carving use a global approach giving it more flexibility for adapting to various situations making it a language-independent method. It can tolerate different degrees of writing irregularities such as skew angles, irregular writing size, line discontinuities and low text-to-background contrast, frequent problems faced while processing handwritten document images and historical manuscripts (Fig. 14).

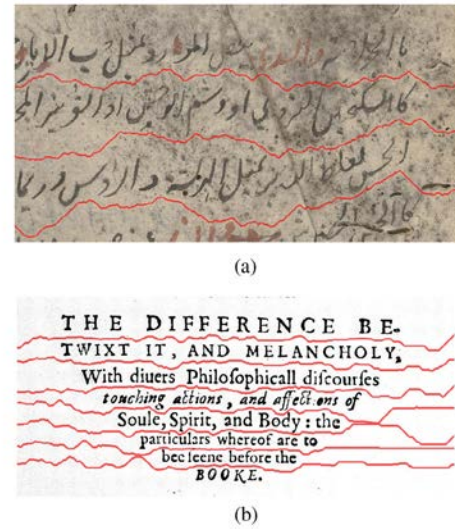


Fig. 14: Examples showing the seam carving text line segmentation on challenging documents: (a) damaged Arabic manuscript, and (b) a gray scale printed document with irregular font types and sizes.

In Fig. 15 many segmentation examples are shown. The method can be applied to languages written horizontally such as Arabic (right to left) or Latin (left to right), but also to exotic languages like Japanese and Chinese scrolls written vertically. This is simply achieved by using a transposed copy of the document image as input, and converting the horizontal seams to vertical seams afterwards like on Fig. 15. Another way is to calculate the cost map vertically instead of horizontally, in order to compute the vertical least-cost separating seams.



Fig. 15. Examples showing the adaptability of the seam carving text line segmentation, by processing documents of different languages, (a) Arabic manuscript, (b) Japanese scroll, and (c) Latin manuscript.

In case of binary document images, the Gaussian filter used for a noise reduction purpose before the gradient magnitude is computed, can affect the results. By increasing the Gaussian filter's kernel size it's possible to dilate the high energy areas, fusing the diacritics to the nearest text line and adding more clearance between the separating seam and the actual lines. Bigger kernel size, means less details are processed while computing the energy map.

To avoid this inconvenience, a small kernel size can be used only to reduce the image noise, then after computing the energy map, another Gaussian filter with a bigger kernel size can be applied on the energy map to dilate the high energy areas, an example is shown on Fig. 16 demonstrating the effect.

In the proposed implementation, the kernel size  $ksize$  of the Gaussian filter is dynamically calculated relatively to the standard deviation  $\sigma$  :

$$ksize = (\sigma \times 2) + 1 \quad (18)$$

ensuring that the resulting convolution matrix can sample the variations of the Gaussian bell.

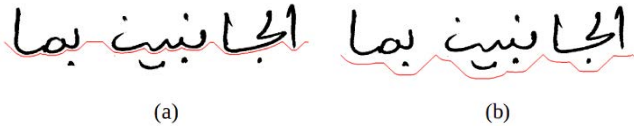


Fig. 16. The kernel size of the Gaussian filter used for a noise reduction purpose, can affect the result on document images with very high text-to-background contrast such as binary images: (a) use of a small kernel size can result in wrong segmentation of diacritics, this can be avoided by using a bigger kernel size (b).

## VI. CONCLUSION

The seam carving approach to text line segmentation is an agile method that can be applied to different styles of writing whether it be printed or handwritten, and in different languages, which makes it a language-independent approach, and able to process multilingual documents.

It must be mentioned that the detection rate of the seam carving segmentation proves its efficiency and adaptability to the different handwriting styles present in the benchmarking dataset used, showing promising results.

In future works, we intend to improve this method on several points, notably improving the runtime performance of the implementation by using parallelism and simplifying the algorithm parameters choice based on each document image features, adding additional robustness.

## ACKNOWLEDGMENT

We thank the Wellcome Library for providing a wide collection of old manuscript and digitized documents for free under creative commons license "Attribution 4.0 International (CC BY 4.0)" which grant researchers and enthusiasts the access to rare and interesting historical content, and we hope that other libraries take this as a model for sharing the knowledge and cultural heritage to the research community.

## REFERENCES

- [1] Arvanitopoulos, N., & Süsstrunk, S. (2014, September). Seam carving for text line extraction on color and grayscale historical manuscripts. In *Frontiers in Handwriting Recognition (ICFHR)*, 2014 14th International Conference on (pp. 726-731). IEEE.
- [2] Souhar, A., Boulid, Y., Ameer, E., & Ouagague, M. (2017). Segmentation of Arabic Handwritten Documents into Text Lines using Watershed Transform. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(6), 96-102.
- [3] Boulid, Y., Souhar, A., & Elkettani, M. Y. (2015, December). Arabic handwritten text line extraction using connected component analysis from a multi agent perspective. In *Intelligent Systems Design and Applications (ISDA)*, 2015 15th International Conference on (pp. 80-87). IEEE.
- [4] Boulid, Y., Souhar, A., & El Kettani, M. E. Y. (2016). Detection of

Text Lines of Handwritten Arabic Manuscripts using Markov Decision Processes. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(1), 31-36.

- [5] Louloudis, G., Gatos, B., Pratikakis, I., & Halatsis, C. (2008). Text line detection in handwritten documents. *Pattern Recognition*, 41(12), 3758-3772.
- [6] Nikolaou, N., Makridis, M., Gatos, B., Stamatoopoulos, N., & Papamarkos, N. (2010). Segmentation of historical machine-printed documents using Adaptive Run Length Smoothing and skeleton segmentation paths. *Image and Vision Computing*, 28(4), 590-604.
- [7] Saabni, R., & El-Sana, J. (2011, September). Language-independent text lines extraction using seam carving. In *Document Analysis and Recognition (ICDAR)*, 2011 International Conference on (pp. 563-568). IEEE.
- [8] Avidan, S., & Shamir, A. (2007, August). Seam carving for content-aware image resizing. In *ACM Transactions on graphics (TOG)* (Vol. 26, No. 3, p. 10). ACM.
- [9] Liwicki, M., Indermuhle, E., & Bunke, H. (2007, September). On-line handwritten text line detection using dynamic programming. In *Document Analysis and Recognition*, 2007. ICDAR 2007. Ninth International Conference on (Vol. 1, pp. 447-451). IEEE.
- [10] Cohen, L., Dehaene, S., Vinckier, F., Jobert, A., & Montavont, A. (2008). Reading normal and degraded words: contribution of the dorsal and ventral visual pathways. *Neuroimage*, 40(1), 353-366.
- [11] Handwritten Arabic Proximity Datasets. Language and Media Processing Laboratory. <https://lampsrv02.umiacs.umd.edu/projdb/project.php?id=65>
- [12] Kumar, J., Abd-Almageed, W., Kang, L., & Doermann, D. (2010, June). Handwritten Arabic text line segmentation using affinity propagation. In *Proceedings of the 9th IAPR International Workshop on Document Analysis Systems* (pp. 135-142). ACM.
- [13] Kumar, J., Kang, L., Doermann, D., & Abd-Almageed, W. (2011, September). Segmentation of handwritten textlines in presence of touching components. In *Document Analysis and Recognition (ICDAR)*, 2011 International Conference on (pp. 109-113). IEEE.
- [14] Zhang, X., & Tan, C. L. (2014, September). Text line segmentation for handwritten documents using constrained seam carving. In *Frontiers in Handwriting Recognition (ICFHR)*, 2014 14th International Conference on (pp. 98-103). IEEE.

### Mehdi Daldali



Mehdi Daldali received the License degree in mathematics and informatics in 2015 from University Ibn Tofaïl, Faculty of Science, Kénitra, Morocco. Currently, he is attending a Master's degree in Big Data and Cloud Computing at the same University, and expected to graduate in 2018. In 2017, he was certified as "Big Data Developer" and "Business Intelligence Analyst" by the IBM MEA Academy. His research interests include high performance computing and pattern recognition.

### Abdelghani Souhar



Abdelghani Souhar is a full Professor of computer science at the University of Ibn Tofaïl, Faculty of science Kénitra Morocco. He received the M.S. degree in applied Mathematics in 1992, PhD degree in computer science in 1997 from the University of Mohammed 5 in Rabat-Morocco. His habilitation thesis concerned Modeling complex systems as Arabic handwritten recognition and an intelligent system for generating mesh. His research interests include automatic processing of Arabic language, modeling complex systems, CAE/CAD, and Artificial Intelligence.

# IMCAD: Computer Aided System for Breast Masses Detection based on Immune Recognition

Leila Belkhodja\*, Djamila Hamdadou

Computer Science Laboratory of Oran (LIO) - University Oran1, Ahmed Benbella, Oran (Algeria)

Received 26 May 2018 | Accepted 8 November 2018 | Published 7 December 2018



## ABSTRACT

Computer Aided Detection (CAD) systems are very important tools which help radiologists as a second reader in detecting early breast cancer in an efficient way, specially on screening mammograms. One of the challenging problems is the detection of masses, which are powerful signs of cancer, because of their poor appearance on mammograms. This paper investigates an automatic CAD for detection of breast masses in screening mammograms based on fuzzy segmentation and a bio-inspired method for pattern recognition: Artificial Immune Recognition System. The proposed approach is applied to real clinical images from the full field digital mammographic database: Inbreast. In order to validate our proposition, we propose the Receiver Operating Characteristic Curve as an analyzer of our IMCAD classifier system, which achieves a good area under curve, with a sensitivity of 100% and a specificity of 95%. The recognition system based on artificial immunity has shown its efficiency on recognizing masses from a very restricted set of training regions.

## KEYWORDS

Breast Cancer, Computer Aided Detection, Mammogram, Immune Recognition, Fuzzy Segmentation, Feature Selection.

DOI: 10.9781/ijimai.2018.12.006

## I. INTRODUCTION

**W**HEN biological immune system fails to distinguish between what belongs and what does not belong to the body, the defense against the invaders weakens. Such invaders include microorganisms, parasites and cancer cells.

We call a cancer all diseases in which abnormal cells divide without control and can invade other tissues. The National Cancer Institute cites more than 100 kinds of cancer in the world [1].

Lung and bronchus, colorectum and breast cancer are the three most commonly diagnosed women cancers. Breast cancer alone is expected to account for 30% of all new cancer diagnosis in women [2]. It is the most frequently observed cancer among women in France, European Union and the United States and remains one of the leading cause of women cancer death.

Breast cancer starts in the breast tissue by an uncontrolled growth of cells in the mammary gland. These cells may remain in the breast or migrate into the body via the blood and lymph vessels. The majority of cancers start in the milk channels. If they remain in the channels the cancer is called in situ or non-invasive. However if the cells leave the wall of the channels, the term "invasive cancer" is used [3].

If detected at an early stage, cancer can be cured in 9 cases out of 10. Presently, there is no effective way to prevent this disease. However, to improve survival rates recently researches showed that screening mammograms helps finding precancerous lesions before they become cancerous by early detection. Thereby, the number of new cases can be reduced and deaths caused by this kind of cancer can be prevented.

Mammogram can be used for breast screening or diagnosing

abnormalities. The screening mammogram is an x-ray exam of the breasts and the most effective tool for early detection. It is used when women have no breast signs, in order to find breast cancer when it is too small to be felt by a woman or her doctor. This will greatly improve a woman's chance for successful treatment.

Masses and microcalcifications are two powerful cancer indicators that are commonly used in evaluating mammogram. Radiologists consider mass detection a more challenging problem than microcalcifications detection because of the poor image contrast of masses, not only for the large variation in size and shape in which masses can appear in a mammogram, but also because masses often exhibit poor image contrast due to breast density [4]. So, it is often difficult to separate normal and abnormal breast tissues. This engenders false positive cases that look like cancer.

Because of fatigued or inexperienced physicians during screening campaigns and the complex structure of the breast, radiology interpretation by visual perception can often miss true positive readings. To fix this problem, strategies, such as a second reading of screening mammograms, have been selectively used, which yield an increase in the cancer detection rate. This is a heavy major challenge for governments, medical organizations, and a difficult task to interpret screening mammograms in large numbers [5].

Computer aided systems for detection and diagnosis on mammograms are one of the automatic solutions that help the radiologist in detecting abnormalities in an efficient way as a second reader of digital mammograms.

To this end, we propose, in this paper, a methodology for computer-aided detection of breast masses on screening mammograms, which joins multidisciplinary axes such as medical domain, image processing and biological pattern recognition. For this, we focus on minimizing false positive findings and increasing true positive cases using all benefits of fuzzy processing and artificial immune recognition system.

\* Corresponding author.

E-mail address: belkhodja\_leila@yahoo.fr



In the following paragraphs, we first outline, in section II, some related work on computer detection for breast cancer and particularly masses symptoms, followed in section III by a state of the art on the artificial immune systems. Our contribution is described in section IV. Then we present the description of the proposed CAD system in section V. The adopted approach is given in detail in section VI. Finally, we present our experimental results and discuss them, respectively, in section VII and VIII.

## II. RELATED WORK

Generally, computer aided systems on breast medical images take two forms:

1. **Computer aided detection system (CAD)** which is able to identify the regions of Suspicion (ROS),
2. **Computer aided diagnosis system (CADx)** which can make a decision whether a ROS is benign or malignant.

We are going to focus on the first form. For more details of the second form the reader can refer to [6]

In the context of the CAD detection systems, the goal of the detection stage is to assist radiologists in locating abnormalities on asymptomatic women mammogram images especially during screening campaigns where a large numbers of mammograms must be analyzed.

In radiological routine the practice consists of applying visual perception by looking at a mammogram and then using cognition for interpreting what is seen. Prospective clinical studies have demonstrated an increase in breast cancer detection with CAD assistance [5].

CAD algorithms, which refers to pattern recognition software, must explore digital or digitized mammograms and search particular signs, which may be the first alarm of cancer. In this way many researchers have focused on particularly two markers: masses and calcifications. They take into the account single image or multiple images.

It is reported in [7] that the American College of Radiology (ACR) and the Breast Imaging Reporting and Data System (BI-RADS) define a mass as a three dimensional space occupying lesion which can be seen at least in two different projections (Cranio Caudal/Medio Lateral). It is characterized by its margins and shape.

Most CAD algorithms operate on a single image and are performed on oblique medio-lateral mammograms (MLO views) or cranio-caudal mammograms (CC views), on the right breast or the left breast, either to detect suspicious regions on a mammogram or to classify them as normal tissue or abnormal one. This engenders a large number of false positives which may be removed when classifying them. Reference [6] notes two types of algorithms: pixel based and region based detection methods.

The pixel based techniques work on features extracted from the local neighborhood of the pixel. In [8] for example, the authors proposed a method for lesion site selection using a morphological filtering enhancement combined with the stochastic model-based segmentation. Their results showed that with the proposed algorithm, the subtle masses could be segmented more accurately than those when the original image is used for extraction without enhancement. Another way to enhance masses is the adaptive thresholding. Reference [9] presented a dual stage adaptive thresholding method to identify the suspicious mass region. They used global histogram, to perform coarse level segmentation in order to locate abnormal regions, and local window thresholding method for each pixel to provide precise and fine segmentation results. Matsubara et al. [10] developed an adaptive threshold technique that uses histogram analysis to divide mammograms into three categories based on the density of the tissue ranging from fatty to dense. Masses were detected using multiple threshold values based on the category of the mammogram.

Contrary to pixel methods, region based detection ones use filtering techniques or segmentation to extract regions of interest and their features, which are later classified as suspicious or normal. [6] notes that a number of these methods are based on the idea of matched filtering where the image is filtered with a filter that is used as a model for a mass. For example, [11] [12] focused on circumscribed masses. The method in [11] uses modified median filtering to enhance mammogram images and template matching to detect the tumors. Herredsvela et al. [12] presented a method based on morphological hierarchical watersheds in the segmentation process. For circular and stellar masses [13] proposed a fuzzy pyramid linking method to detect tumors in mammogram image and classified detected regions to benign and malignant.

It is reported in [14] that because of many complex and changing characteristics of mass in mammogram images, with great difficulty in mass segmentation, region growing becomes a reliable method to accomplish it. Other segmentation techniques are cited in [15].

On the other hand, some methods use multiple images from right and left breasts or CC and MLO projections to search for asymmetries, which can be potential abnormalities. An example of such methods is developed in [16]. Multiple modalities, like mammography, ultrasonography and magnetic resonance imaging, can also be used [17]. More details for different imaging modalities are given in a review established by [18].

Note also that calcifications are the second very important marker of benign or malignant process after masses. They are tiny or big deposits of calcium. A number of different approaches have been applied for detection of calcifications. We cite as an example the work of [19].

Once regions of interest are extracted, some researchers, using CAD or CADx, focus on classifying them to normal and abnormal tissues or benign or malignant region respectively. The goal of this stage is to reduce false positive number for CAD and specify the nature of the mass for CADx. To deal with that, shape and texture features are used by a lot of researchers, for example [20].

To improve classification accuracy and stability of the system, various classification techniques have been used for classifying Regions of Interest (ROI) as normal or suspicious or as benign or malignant. Most of them use supervised methods. Up to present, popular approaches mainly include: artificial neural networks [21], swarm intelligence for neural network optimization [22], support vector machine [23], linear discriminant analysis [24], bayesian network [25], etc. In addition, deep learning, which is based on learning data representations and convolutional neural network, has shown its higher performance and has been effectively applied to breast mass detection and classification [26][27]. However, [28] mentioned that these methods work well on large data sets but exhibit certain limitations on small data sets. They propose a new exploratory method for the automatic detection of lesion based on gestalt psychology, which combines human cognitive characteristics and radiologist's knowledge. Most researches are applied on digitized mammograms from the MIAS [29] and the DDSM [30] databases, or on real clinical images from screening centers.

Recently, natural computing techniques have emerged in the artificial medical processing image domain and have proved their robustness to improve interpretation quality for radiologists. In this work, a particular attention is given to artificial immune recognition system (AIRS) whose details are outlined in the following section.

These kinds of systems (CAD/CADx) are generally evaluated using ROC plot (Receiver Operating Characteristic) and FROC plot (Free-response Receiver Operating Characteristic) [6]. They are standard methodologies for measurement of performance of detection and diagnosis algorithms in CAD systems. Raman et al. present more details about ROC in [31].



### III. ARTIFICIAL IMMUNE RECOGNITION SYSTEM: STATE OF THE ART

Also called immunological computation, AIS is a field of study dedicated to develop computational models based on biological immune mechanisms, which are used to solve hard computational problems.

The human immune system is a robust, complex, highly distributive learning system that is able, through adaptation, to distinguish between dangerous foreign antigens and the body own cells. It learns how to identify patterns and then uses memory cells to remember previously identified patterns. There are two types of defense mechanism: innate and acquired. Innate defense acts without taking into account the type of disease and is achieved by some specialized cells. The acquired response involves specialized cells called lymphocytes [32]. The immune system contains B lymphocytes originated from the bone marrow and T- lymphocytes originated from thymus. When a pathogen is identified, stimulated B cells, helped by T cells, use the mechanism defense to lock on it. Undergoing somatic and hypermutation cloning, these B cells produce antibodies and distribute them all over the body to prepare the next attack from the antigen which is destroyed by the T cells. Detailed information about the immune system can be found in [33].

Many researchers have been motivated by these self-defense biological concepts. Artificial Immune Recognition Systems (AIRS) have not emerged since a collection of other supervised and unsupervised artificial immune systems algorithms have been developed [34] [35] [36] [37] [38].

AIRS is a supervised learning approach inspired from the biologic immune system for pattern recognition proposed in the Masters work by Watkins [39] who published the first version AIRS1 in [40]. It was replaced after that by a new efficient version of the algorithm called AIRS2 proposed by Watkins and Timmis in [41]. The specified version of this algorithm is detailed in section VI.

Meng et al. demonstrate in [42] the reliability and accuracy of AIRS on benchmarking experiments. They find that AIRS consistently outperforms other algorithms and can be used for real world classification tasks.

In diagnosing disease, AIRS have been largely applied as a decision making tool in medicine for example for heart disease [43] and diabetes [44]. For breast cancer diagnosis, many developed AIRS based researches show very important accuracies on Wisconsin breast cancer dataset [45] [46]. AIRS was also applied by Katsis et al. [17] to detect early breast cancer using different examinations (i.e. mammography, ultrasonography and magnetic resonance imaging) with promising results.

The main factors handled by the artificial immune system are antigens, antibodies or B- memory cells and Artificial Recognition Ball (ARB). Antigens (AG) are a set of  $n$  training data AG with labelled instances  $C$ . Antibodies are feature vectors of potential solutions matching more to antigens. An ARB represents a number of identical B-Cells which are employed within a mechanism to reduce duplication and dictate survival within the population [41].

The idea of the technique is to prepare a set of real-valued vectors to classify patterns. The system generates a set of memory cells from training data. If these cells are insufficiently stimulated for a given input pattern, candidate memory cells are then generated to replace them by a process of cloning and mutation of cells for the most stimulated memory cell. To join the memory pool, clones compete based on stimulation and on the amount of resources used by each cell [47].

To fine tune the training process, the AIRS algorithm uses a set of configurable parameters outlined in the following:

- **Initialization Instances:** are randomly selected training antigens to initialize the memory cells pool.

- **Affinity Threshold (AT):** is the mean affinity value between all the antigens in the training set (1).

$$AT = \frac{\sum_{i=1}^n \sum_{j=i+1}^n affinity(ag_i, ag_j)}{\frac{n(n-1)}{2}} \quad (1)$$

Where:

- $n$  is the number of training antigens,
- affinity  $(ag_i, ag_j)$ =Euclidean distance  $(ag_i, ag_j)$ ,
- $ag_i$  and  $ag_j$  are the  $i^{th}$  and  $j^{th}$  training antigen.
- **Affinity Threshold Scalar (ATS):** It is a parameter used with AT for memory cell replacement in the training process. Its value is between 0 and 1.
- **Stimulation Threshold:** In the range of  $[0,1]$ , and generally around 0.9 this parameter regulates the ARB refinement process.
- **Clonal Rate:** It is an integer value, which determines the number of clones of the best memory cell or the number of clones and resources of each ARB in the refinement stage.
- **Mutation Rate:** It is used to determine the number of mutated clones that the best matching memory cell can create. The common value is 2.
- **Maximum Resources:** It limits the number of ARBs in the system. Usual values are between 150 and 300.

### IV. CONTRIBUTION

We present, in detail, our contribution named IMCAD, which is a computer aided masses detection for screening mammography that acts as a second reader, and have the goal of improving the detection performance.

The proposed system IMCAD offers to radiologist's community an important tool, during screening campaigns of breast cancer, to detect abnormalities and missed masses that can be fatal for women life.

The main purpose of IMCAD is not to provide a perfect decision because of the lack of information about asymptomatic patients, but much more attract the attention of radiologist on regions that can be the beginning of cancer.

In this study, our contribution consists in imitating exactly biological immune self-defense of human body by developing a full automatic system based on a powerful recognition classifier AIRS. IMCAD acts, on reduced data, as well as an adaptive natural immune system, which can learn via experience.

For this aim, we propose a methodology, over different research areas (medical image processing, pattern recognition, computer vision...) which processes, mammograms as input data and produces results decision as output.

### V. IMCAD SYSTEM DESCRIPTION

IMCAD is the proposed computer aided detection for breast masses based on a pattern recognition immune system which automatically identifies abnormal regions on screening mammograms. It is designed to provide a second opinion to aid rather than substituting the radiologist. Our CAD scheme is applied to a mammogram database and it is based on four sequential modules:

- **Subsystem1:** Preprocessing.

The goal of this module, which uses real mammogram as input data, is to minimize time and memory allocation and reduce noise on mammograms.

- **Subsystem2:** Fuzzy Segmentation.

The second module uses a fuzzy classifier to extract homogenous classes from mammogram. The classified images are then labelled by a recursive labelling method.

- **Subsystem3:** Characterization.

The third module of the CAD scheme converts results of the second module to quantitative information. In this stage, IMCAD computes a set of features for each region in the segmented mammogram.

- **Subsystem4:** Immune Recognition.

Extracted features are then injected in the last module based on immune learning and recognition to detect suspecting regions as being positive masses. Details are given in the following section.

Note that IMCAD contains an offline immune treatment which is detailed later. The overall methodology schema of the proposed method is illustrated in Fig. 1.

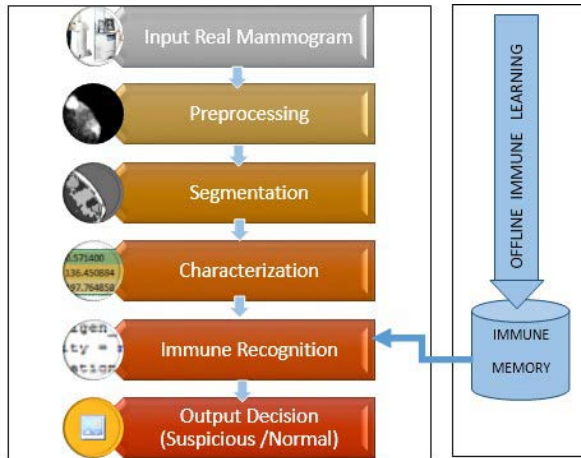


Fig. 1. The proposed system IMCAD: An overview.

## VI. IMCAD ADOPTED APPROACH

The decisional process, adopted by IMCAD, presented previously, is detailed on a flowchart in Fig. 2.

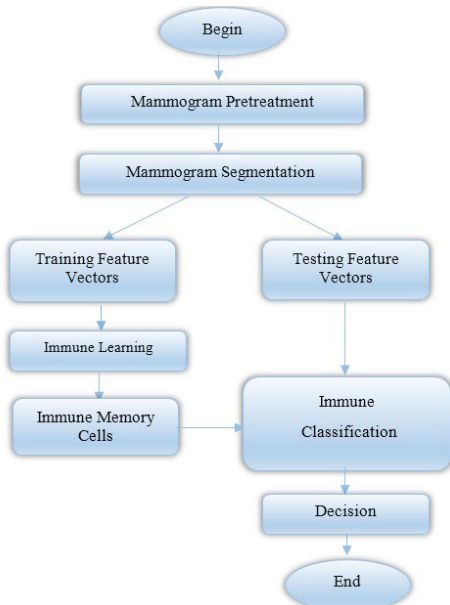


Fig. 2. Flowchart of the general approach adopted by IMCAD.

### A. About Data

IMCAD was conducted on the *Inbreast Database* which is created by the Breast Research Group from INESC Porto and acquired at the Breast Center in Centro Hospitalar of São João at Porto [48]. In opposition to usual digitized mammograms, *Inbreast* is built with full-field digital mammograms, with a wide variability of cases. The acquisition equipment was the MammoNovation Siemens FFDM (Full Field Digital Mammogram), with 14-bit in contrast resolution.

*Inbreast* has a total of 115 cases (screening, diagnostic and follow up). Images are in DICOM format (Digital Imaging and Communications in Medicine), with matrix size equal to 3328×4084 or 2560×3328 pixels, depending on the compression plate used in the acquisition (according to the breast size of the patient). This format gathers not only the image but also some related metadata.

The database contains examples of normal mammograms, mammograms with masses, mammograms with calcifications, architectural distortions, asymmetries, and images with multiple findings. An example of mammograms is given in Fig. 3.

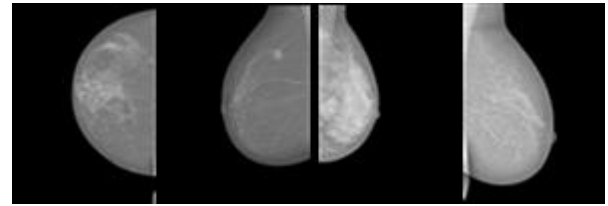


Fig. 3. Examples of mammograms from *InBreast* database.

### B. Mammogram Preprocessing

Because of mammogram's hardness interpretation, any CAD system needs a preprocessing and a preparation stage to improve image quality, remove noise and make more correct the image segmentation outcome.

Our IMCAD preprocessing algorithm, shown in Fig. 4, proceeds first by performing a Gauss pyramid reduction to mammograms because of high image sizes and slowness of running time. Then we apply iteratively a mean filter to reduced images.

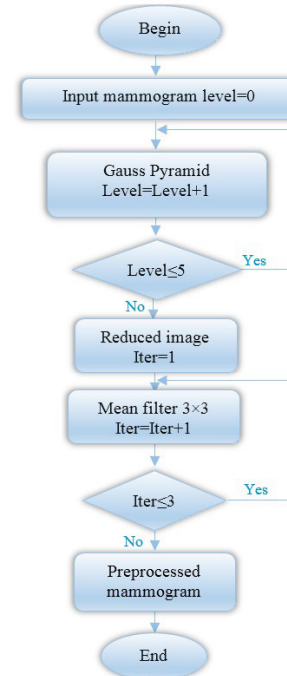


Fig. 4. Flowchart of the IMCAD preprocessing.

The Gaussian Pyramid is a multiple scale representation of the image. It allows the processing algorithm to work from the details up to the rough. To generate a pyramid, we iterate between two steps: smoothing and down-sampling. The smoothing operation removes high frequency components, which engenders fast changes that down-sampling would miss. The down-sampling reduces the image size by  $\frac{1}{2}$  at each level [49].

Note that images in level 0 are 3328x4084 or 2560x3328 pixels. The kernel used is cited below (2), with  $\alpha=0,375$ .

$$\begin{bmatrix} \frac{1}{4} - \frac{\alpha}{2} & \frac{1}{4} & \alpha \\ \frac{1}{4} & \alpha & \frac{1}{4} - \frac{\alpha}{2} \end{bmatrix} \quad (2)$$

After reduction to level 5 mammograms are then filtered using a median filter, three times in succession, to avoid later over-segmentation and reduce the number of very small classes and regions. The median filter is a nonlinear digital filtering technique, which is used to remove noise from images and to improve the results for later processing. The main idea is to run a square window  $3 \times 3$  through the image pixel by pixel replacing each entry with the median of neighboring pixels.

### C. Segmentation

Segmentation is the mid-level image processing which consists in partitioning an image into regions or objects and reducing them to a form suitable for high computer processing level (recognition). One of the most difficult tasks in digital image processing is automatic segmentation. The approach to apply usually depends on the context and type of image to be segmented.

Mammograms are delicate images to analyze even when masses are blur and hidden in dense tissue. The concept of fuzziness corresponds exactly to this problem, that is why we choose a partitioning method based on the fuzzy algorithms. Note that our goal is to locate only the anomaly in a computer aided detection. Therefore, a successful segmentation must preserve the whole real mass and avoid creating negative ones.

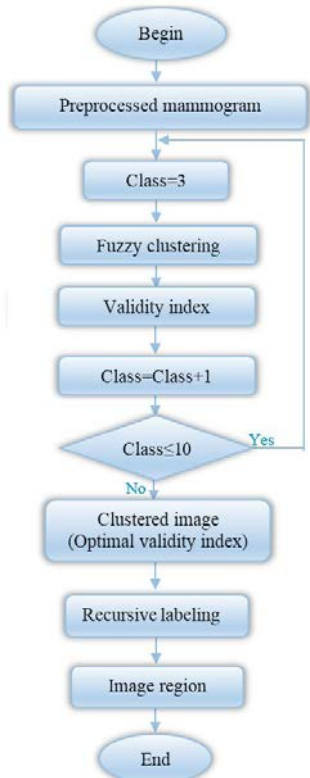


Fig. 5. Flowchart of the IMCAD fuzzy segmentation.

The proposed segmentation algorithm is carried out as follows:

- *Phase 1:* Loading reduced and filtered images from the preprocessed step.
- *Phase 2:* Automatic grayscale fuzzy image clustering.
- *Phase 3:* Iterative evaluation of the validity index for the clustering process.
- *Phase 4:* Recursive labeling of the clustered image.

A detailed description of the segmentation algorithm is illustrated in Fig. 5.

#### 1) Fuzzy Clustering

The fuzzy classification of the preprocessed mammogram is conducted by the Fuzzy C-means algorithm. It requires beforehand, the knowledge of the number of classes  $c$  and produces them in such a way that the objective function  $J_{fcm}$  (3) minimizes the total weighted mean-square error so:

$$J_{fcm}(\mu, x, c^{(k)}) = \sum_{i=1}^c \sum_{k=1}^n (\mu_{ik})^m \|x^{(k)} - c^{(i)}\|^2 \quad (3)$$

Where :

$C^{(i)}$  : is the center of the class  $i$ ,

$m$ : is a real number greater than 1 to control the fuzziness of cluster,

$X^{(k)}$  : is the  $k^{\text{th}}$  pixel in the preprocessed mammogram.

The optimization of this function is done iteratively. At each iteration the membership degrees  $\mu_{ik}$  of each pixel to the classes  $C$  and prototypes  $C^{(i)}$  of classes are updated respectively according to the following relations (4) and (5):

$$\mu_{ik} = \frac{\|x_k - c_i\|^{\frac{2}{1-m}}}{\sum_{j=1}^c \|x_k - c_j\|^{\frac{2}{1-m}}} \quad (4)$$

$$c_i = \frac{\sum_{k=1}^n \mu_{ik}^m x_k}{\sum_{k=1}^n \mu_{ik}^m} \quad (5)$$

We use a variant of the fuzzy algorithm proposed in [50], which is based on the original Fuzzy C-Means (FCM) [51]. The main steps of the algorithm are:

**Step 1:** Set the parameters

- $m$ : the fuzziness index, a real number greater than 1 to control the fuzziness of cluster,
- $c$ : the number of classes,
- $\varepsilon$ : convergence error = 0,001,
- $X$ : vector which contains all pixels of the preprocessed mammogram.

**Step 2:** Initialize the membership degree matrix  $\mu$  with random values.

**Step 3:** Standardize the initial weight over  $K$  (6).

$$\mu_{qk}^{normK} = \frac{\mu_{qk}}{\sum_{i=1}^K \mu_{qi}} \quad (6)$$

**Step 4:** Standardize cluster weights over  $Q$  (7).

$$\mu_{qk}^{normQ} = \frac{\mu_{qk}^{normK}}{\sum_{i=1}^Q (\mu_{ik}^{normK} - \min) / (\max - \min)} \quad (7)$$

**Step 5:** Compute new prototype centers according to relation (5).

**Step 6:** Compute new weights according to relation (6).

**Step 7:** Repeat steps 4, 5 and 6 until reaching the maximum number of iterations or satisfy the criterion (8).

$$\|J^{old} - J^{new}\| < \varepsilon \quad (8)$$

**Step 8:** Evaluate the K clustering by calculating the Xie-Beni validity index.

## 2) The Xie-Beni Validity Measure

Cluster validity measures are methods that evaluate clustering either by comparing the results of two different sets of cluster analysis to choose the best one or by determining the correct number of clusters in the data set.

Various indexes have been proposed in the literature. In this work we use the Xie-Beni measure XB [52] which is an index of fuzzy clustering and also applicable to crisp clustering. The XB index (9) focuses on two properties: compactness of the fuzzy partition and separation of clusters. A well-defined partition produces a small value of compactness and well separated centroids will give a high value of separation. Consequently, minimizing XB for  $c=2,3,\dots,cmax$  will determine the optimal partition of data.

$$XB = \frac{\sum_{i=1}^c \sum_{j=1}^n \mu_{ij}^m \|x_j - c_i\|^2}{n \cdot D_{min}} \quad (9)$$

Where

$C_i$ : Centroid of the cluster  $i$ ,

$X_j$ : Pixel  $j$ ,

$\mu_{ij}$ : The fuzzy membership of the pixel  $j$  belonging to class  $i$ ,

$D_{min}$ : indicates the minimum Euclidean distance between centroids.

## 3) Labeling of the Fuzzy Clustered Image

The purpose of this phase is to obtain the image region with connected components (Fig. 6.b). The main problem is that the fuzzy clustered image is class labeled (Fig. 6.a). A class may contain one or more regions semantically separated. To obtain these regions, we separate classes in different binary images, where we label each class 1 and the other pixels background 0. A region is then a connected set of 1 pixels.

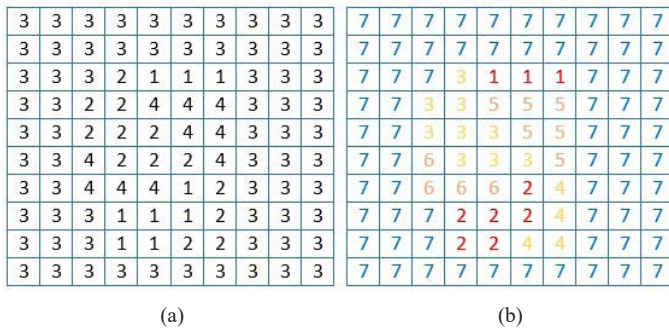


Fig. 6. (a) Example of fuzzy clustered image (b) and image region with connected components.

This means that from any pixel labeled 1 there is a path of 1's to any other pixel in its region. This path can be found by searching recursively all roots from the starting 1 pixel and its 8 neighbors until the destination 1 is reached. These steps are detailed in the recursive labeling algorithm 1.

## Algorithm 1: Fuzzy Clustered Recursive Labeling

**Input data:** Clustered image *Clus*

**Output data:** Labeled image *Lab*

*Procedure* Browse(*Lab*,*label*,*L*,*P*)

**BEGIN**<sub>Browse</sub>

Lab[*L*][*P*] ← *label*;

if (Lab[*L*][*P*-1] = -1) Browse(*Lab*,*L*,*P*-1);

if (Lab[*L*][*P*+1] = -1) Browse(*Lab*,*L*,*P*+1);

if (Lab[*L*-1][*P*] = -1) Browse(*Lab*,*L*-1,*P*);

if (Lab[*L*+1][*P*] = -1) Browse(*Lab*,*L*+1,*P*);

**END**<sub>Browse</sub>

**BEGIN**

Copy each class in *Clus* in a single image *Im*

**foreach** image *Im* **do**

Bin ← Binarize (*Im*);

Lab ← negate (bin); //Put all pixels with value 1 to -1

Label ← 0;

**for** i ← 0 to *N*

**for** j ← 0 to *M*

**if** Lab[*i*,*j*] = -1 **then**

label ← label + 1;

Browse(*Lab*,label,*i*,*j*);

**Endif**

**Endfor**

**Endfor**

**END.**

**return** *Lab*;

## D. Feature Selection

This process follows the output of a segmentation stage, which are generally pixel data. The purpose of this step is to extract attributes which give quantitative information able to distinguish between normal and abnormal regions.

It comes out, after a discussion with radiologists and a preliminary study, that the description of a mass is based on its intensity, size, density, shape, position, and edge characteristics. Radiologists define a critical size from which metastatic spread occurs. When the patient is treated before this size, she will not have metastases.

Given the wide variety of masses, it is extremely difficult to define a common set of attributes. As a result, we limited our research on circumscribed, spicules and poorly defined masses. The set of features we have calculated, for each region in the segmented image, includes only intensity and shape features to localize abnormal regions. IMCAD features are:

**Area (S):** Is the number of pixels inside the boundary of the region. We believe that this is an important feature because it corresponds to the size feature defined by the radiologists.

**Average gray level (AGL):** This feature (10) defines the average intensity of the region. It corresponds exactly to radiologists density attribute. Density is a measure used to describe mammogram's masses. A hyperdense region tends to be clearer than an hypodense region.

$$AGL = \frac{\sum_{k=1}^{k=S} Pixel_k}{S} \quad (10)$$

**Compactness (C):** This parameter can be used to detect compact and circumscribed regions (11).



$$C = \frac{4 \pi S}{P^2} \quad (11)$$

Where P represents the region perimeter (number of border pixels).

### E. Artificial Immune Masses Recognition

Recognition, based on Artificial Immune Recognition System, is the higher-level processing step in the proposed IMCAD system. It involves making sense to previous extracted features by performing cognitive functions normally associated with radiologist vision. Our challenge and purpose in using artificial immunity in detection of abnormalities in IMCAD is to simulate globally immune human defense and imitate the extraordinary powers of brain against danger.

AIRS proceeds in two steps: offline training and online classification. It tries to make memory cells, which are representative of the extracted training regions the model is exposed to, and are suitable for classifying unseen mammograms.

#### 1) Airs Training Phase:

The immune training process consists in building supervised classes from specific extracted features of normal and abnormal regions. AIRS learning [41][47], outlined in the Algorithm 2, turns on four stages: initialization, memory cell identification, competition of resources and refinement of memory cells.

##### 1. Initialization

The initialization step consists in:

- Normalizing all antigens (Input regions).
- Calculating the *affinity threshold* AT (1).
- Initializing  $Cells_{Memory}$  by choosing regions similar to antigens.
- Initializing ARB population to  $\emptyset$ .

##### 2. Memory cell identification ( $Cells_{Memory}$ ) and ARBs Generation ( $Cells_{clones}$ )

This process and the others described later are run for each antigenic input region one at a time, which makes AIRS a one-shot learning algorithm.

For each input region, this process consists of:

- Stimulating each initial memory cell B to the current antigen A (12).
- Selecting  $MC_{best}$ , the memory cell that more stimulates the antigen and adding it to ARB pool.
- Making ARB pool by cloning  $MC_{best}$   $Nclones$  (13) times and introducing diversification by mutating randomly each cloned cell.

$$Nclones = Stimulation(MC_{best}, ag) \times clonal_{rate} \times mutation_{rate} \quad (13)$$

##### 3. Competition of resources and development of a candidate memory cell

AIRS must maintain a population of memory cells for each class of antigen at the end of the algorithm. For this purpose the strongest ARBs ( $Cells_{clones}$ ), with important resources (14) must survive in this stage. This is performed via a resource allocation and competition mechanism which is used to control the size of the ARB pool.

$$Resources = Normalised\ stimulation * clonal_{rate} \quad (14)$$

Before allocation of resources, diversification is also introduced. Each ARB is cloned  $num\_clones$  times and then mutated. After that, resources are allocated to each ARB in the pool. The total of resources is computed and compared against  $max\_resources$ . ARBs with low resources are then removed from the pool. The stop condition for this

process occurs when the mean normalised stimulation exceeds the stimulation threshold

$$num\_clones = Stimulation * clonal_{rate} \quad (15)$$

Note that mutation and competition of resources allocation subroutines are referred to [47].

#### 4. Memory Cell Introduction

Once the competition stop condition is reached, the optimal ARB pool is selected. The ARB with maximum normalized stimulation value is designated to become the memory cell candidate:  $MC_{candidate}$ . This cell joins the memory cells ( $Cells_{Memory}$ ) if its stimulation value is better than  $MC_{best}$  stimulation, which is removed if their affinity is less than the product of the affinity threshold and the affinity threshold scalar.

#### Algorithm 2: Artificial immune Learning Of Abnormal Regions

**Input data:** Input regions, stimulation\_threshold, affinity\_threshold, mutation\_rate, clonal\_rate, max\_ressources,

**Output data:**  $Cells_{Memory}$

Initialization

**foreach**  $Input\_region_i \in$  Input regions **do**

Stimulate the  $Input\_region_i$  with each memory cell;

$MC_{best} \leftarrow$  most stimulated memory cell;

**if** (class( $MC_{best}$ )  $\neq$  class( $Input\_region_i$ )) **then**

$Cells_{Memory} \leftarrow Input\_region_i$ ;

Class( $Cells_{Memory}$ )  $\leftarrow$  Class( $Input\_region_i$ );

**else**

$Nclones \leftarrow Stim(MC_{best}) \times clonal\_rate \times mutation\_rate$ ;

$Cells_{clones} \leftarrow MC_{best}$ ;

**for** j to  $Nclones$  **do**

$Cells_{clones} \leftarrow$  Clone and Mutate  $MC_{best}$ ;

**end**

Stimulate ( $Cells_{clones}$ ,  $Input\_region_i$ );

**while** AverageNormalisedStimulation( $Cells_{clones}$ )  $\leq$  stim\_threshold

**do**

**foreach**  $Cell_j \in Cells_{clones}$  **do**

$Cells_{clones} \leftarrow$  Clone and Mutate  $Cell_j$ ;

**end**

Stimulate ( $Cells_{clones}$ ,  $Input\_region_i$ );

Resources allocation

ReducePoolToMaximumResources( $Cells_{clones}$ , max\_ressources);

**end**

$MC_{candidate} \leftarrow$  GetMostStimulated ( $Input\_region_i$ ,  $Cells_{clones}$ );

**if** Stim ( $MC_{candidate}$ )  $>$  Stim ( $MC_{best}$ ) **then**

$Cells_{Memory} \leftarrow MC_{candidate}$ ;

**if** Affinity( $MC_{candidate}$ ,  $MC_{best}$ )  $\leq$  affinity\_threshold **then**

DeleteCell( $MC_{best}$ ,  $Cells_{Memory}$ );

**end**

**end**

**end**

**end**

**return**  $Cells_{Memory}$ ;

#### 2) AIRS Testing Phase:

At the end of the training phase, all antigens are represented by a set of antibodies in the memory cell pool. A segmented mammogram is stimulated by all antibodies in order to classify the new antigenic regions. The criterion of classification is to attribute the class of the most stimulated antibody to each region in the image.

## VII. EXPERIMENTAL RESULTS

The objective of this work is to develop the automatic computer aided detection system IMCAD for screening mammograms, using artificial immunity, to help radiologists in preventing breast cancer early. IMCAD acts as a natural defense of the human immune system facing cancer. All experiment results were conducted on selected masses cases from Inbreast database.

All methods in IMCAD are performed using Embarcadero C++ Builder 2010 software, running on a laptop PC with a 2.50GHZ CPU(i5) and 4G RAM.

### 1) Nature of Data

Recall that the InBreast database includes a set of high quality mammogram images with different pathologies. IMCAD focus is on cases annotated as masses and normal, which are sorted and extracted manually. Original dicom mammograms are converted to BMP format. Corresponding annotated images are PNG files.

There are a total of 116 masses among 107 images (=1.1 masses per image). The average mass size is 479 mm<sup>2</sup> (with a standard deviation of 619 mm<sup>2</sup>), the smallest mass has 15 mm<sup>2</sup> and the biggest has an area of 3689 mm<sup>2</sup>.

The annotations were made by a specialist in the field, and validated by a second specialist, between April 2010 and December 2010. When there was a disagreement between the experts, the case was discussed until a consensus was obtained.

### 2) IMCAD Running

The proposed system proceeds in several steps.

#### 1. Preprocessing

This step prepares mammograms for high levels treatments. The aim objective is to improve segmentation in areas of interest. We applied a 5 level multiresolution with a gauss pyramid reduction (Fig. 7.b), followed by 3 times median filtering operation for the last level of the pyramid (Fig. 7.c). A case study with abnormal mass region is shown in Fig. 7.a.

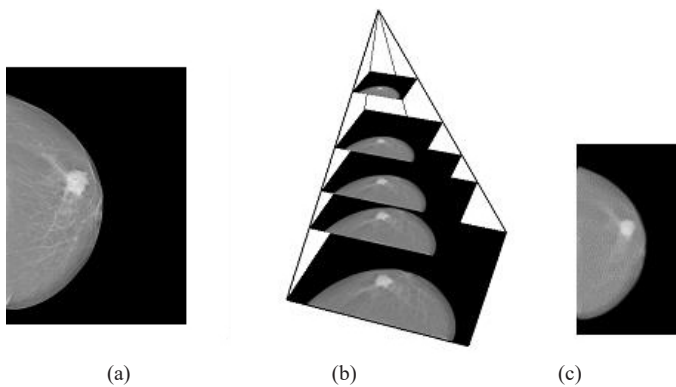


Fig. 7. Original mammogram with abnormal region (mass) (a) Five level Gauss Pyramid (b) Reduced filtered image (c).

#### 2. Fuzzy Segmentation

Automated FCM clustering, intensity based, gives a very interesting result for our IMCAD segmentation step. It is applied on the fifth level image in the gauss pyramid. Through our implementation, we set the parameter fuzzy index  $\mu = 2$  and we have varied the number of classes from 2 to 10. For each number the process stops when the optimal objective function is reached with  $\varepsilon = 0.001$

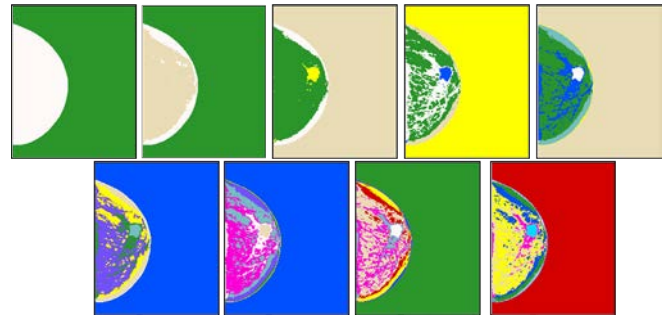


Fig. 8. FCM clustered images on original reduced mammograms (c=2..10).

Several tests were done with different settings of iteration number. First FCM works manually on original data considered as clean data (0% noise). Clustered images are shown in Fig. 8.

In order to reduce noise, avoid over-segmentation and minimize the number of regions to be processed, a median filter is applied to original reduced images. We can observe in Fig. 9 that results are much better and masses are more valued.

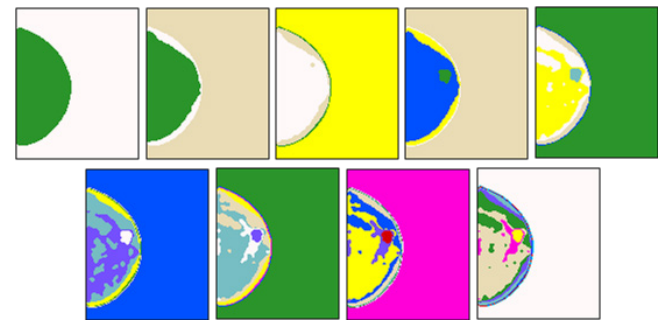


Fig. 9. FCM clustered images on reduced filtered mammograms (c=2..10).

To obtain connected regions, we submit clustered images to recursive labeling. An example is outlined in Fig. 10.

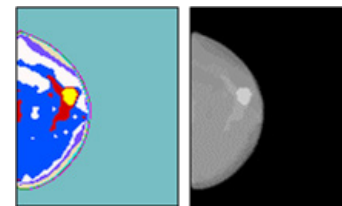


Fig. 10. Connected regions with recursive labeling on clustered mammogram (c=10).

With the objective to complete the IMCAD independence to parameters, we automated the FCM segmentation task by accurately identifying the optimal number of clusters from the Xie-Beni validity index. Results, shown on the graph in the Fig. 11, give 5 classes as optimal number with Xie-Beni=0.012375 over 50 iterations.



Fig. 11. Evolution of the Xie-Beni validity index over number of classes.

The number of iterations corresponds to minimum error between the objective function in the epoch  $t-1$  and the final epoch  $t$ . The best visual results (Fig. 9,  $c=5$ ) which conserve the whole abnormal region and avoid over-segmentation were obtained with the following parameters:  $\mu = 2$ , 70 iterations and  $\varepsilon \sim 0$  (Fig. 12).

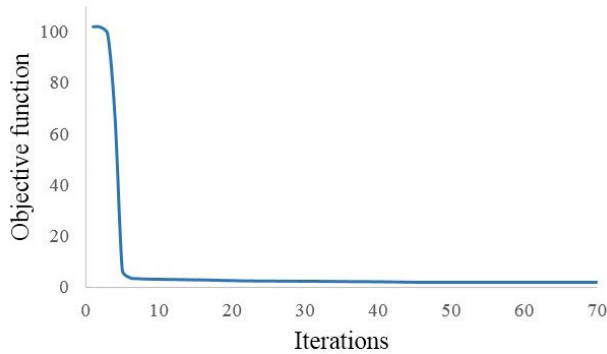


Fig. 12. Evolution of the objective function.

### 3. Immune learning

To train the artificial immune system, four abnormal mammograms are chosen from Inbreast Database (Fig. 13). Only masses and normal regions are featured and area, compactness and AGL are extracted to be presented to the system.

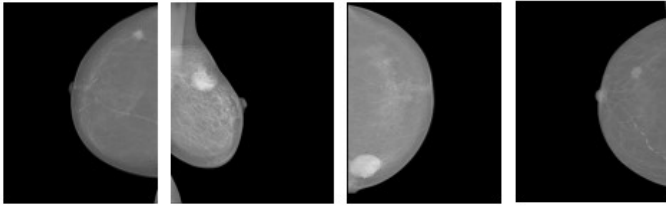


Fig. 13. Learning mammograms.

The set of training data is composed of 6 antigens which represents 4 masses regions and only 2 normal regions. Important masses and average masses form the class 1 and class 2 (Abnormal class) and the normal region forms the class 3 (Normal class). The affinity threshold obtained is 0,563. Values of training parameters are given in the table in Fig. 14.

During the ARB refinement process, the artificial recognition balls enter in competition of their resources. The average mean normalised stimulation value obtained is 0,97 ( $>$ stimulation threshold) for all iterations. The average value of departure  $\sim 0.45$

Note that the purpose of this step is to compute a set of memory cells which are used later to recognize regions of segmented mammograms. To obtain more Bcells, we choose to run the training process iteratively seven times.

The total number of the memory cells after the training process is 20. Antigens of the normal class (C3) are recognized by a set of 10 Bcells. We obtain 3 Bcells in the abnormal class (C1), and 7 Bcells in the abnormal class (C2).

Parameters	Values
Max Ressources	200
Mutation Rate	2
Clone Rate	10
Hypermutation Rate	2
Affinity Threshold Scalar	0,1
Stimulation Threshold	0,9

Fig. 14. Table of the immune training parameters.

### 4. Immune classification

After 7 iterations, the B memory cells generated from the learning process are used in the classification step (test) on a total of 342 regions of 32 mammograms (16 with masses and 12 normal). Very small regions were ignored.

We stimulate each region of the segmented mammogram test by all the memory cells. Then we affect each region in the class of the most stimulated Bcell if the stimulation exceeds a certain threshold  $T$ .

## VIII. IMCAD EVALUATION AND DISCUSSION

The proposed system IMCAD is a computer aided detection system for breast masses computerized to support radiologists in achieving their interpretation task in detecting abnormalities on screening mammograms.

In order to evaluate the effectiveness of our IMCAD system and to extract the best threshold  $T$  for classification of regions, we choose the Receiver Operating Characteristic curve and compare our results to radiologist's annotations in the Inbreast database.

The ROC curve is a graphical representation of the false positive rate (1-specificity) on the X axis and the true positive rate (sensitivity) on the Y axis, calculated for all possible thresholds  $T$ .

For all training and testing regions of all studied mammograms, we compute true positive and false positive rates (TPR, FPR). Results with different cut-off used are outlined on the ROC plot in Fig. 15.

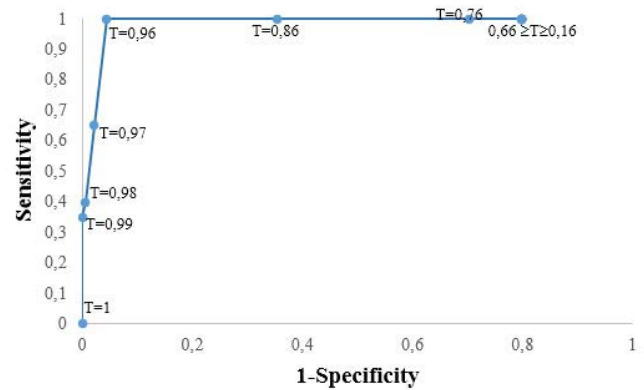


Fig. 15. IMCAD Receiver Operating Characteristic Curve.

The accuracy of the IMCAD test is measured by the Area Under Curve (AUC) which is equal to 0,78. The best cut-off, where sensitivity and specificity are close to 1, corresponds to  $T=0,96$  (Sensitivity=1, Specificity=0,956). Decisional results on some abnormal and normal mammograms are exposed, respectively, in Fig. 16 and Fig. 17.

Normal images are processed exactly in the same way as abnormal mammograms. Some results are displayed in Fig. 17.

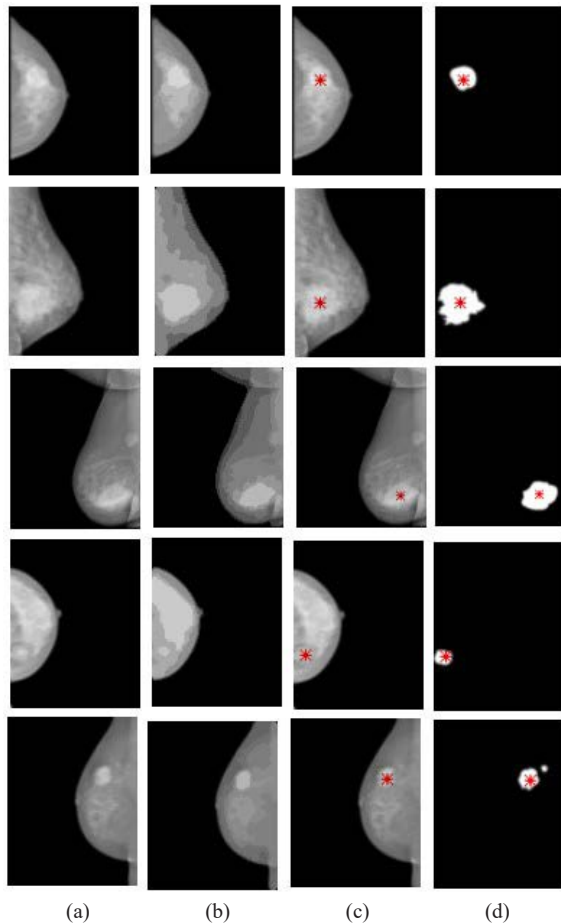


Fig. 16. Examples of IMCAD segmentation images (b) and detection results (c) on original images (a). Comparison by superposition on annotated masks of Inbreast Database (d).

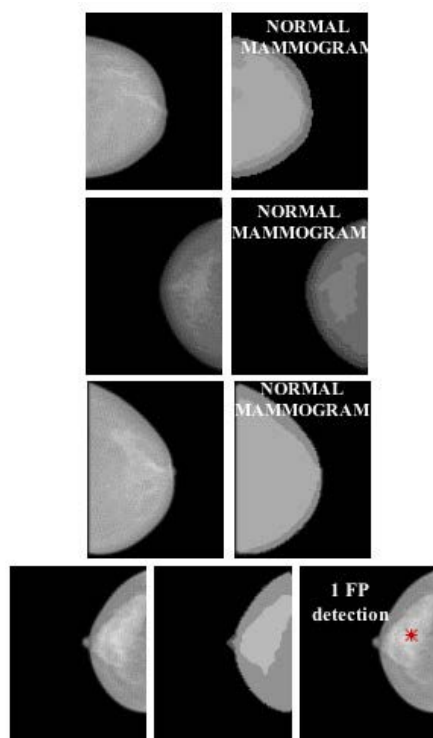


Fig. 17. IMCAD results on normal mammograms.

From these results, we can note that:

- Statistically, an area under curve of 0.5 represents a worthless test, while an area under curve of 1 represents a perfect test. IMCAD classifier has reached an AUC~0,8 which makes it a good system capable of identifying more true positives while minimizing the number of false positives.
- We obtain a very good sensitivity and a good specificity, which does not affect the objective of our IMCAD system.
- The detection of a mass is affected by the automatic computation of the number of classes  $C$ . In the data sample used, the only case, with 2 masses per mammogram (Example 5 in Fig. 16), one detected and one missed, is due to FCM segmentation. Indeed, the second mass was merged during the segmentation process (Example 5 in Fig. 16.b) and thus lost for the immune system (Fig. 16.d)
- Large original images require a high computation time and a powerful hardware. We solve this problem by applying a five level multiresolution Gauss pyramid that had no effect on defining ROIs. However, a very high level of reduction can make disappear some masses.
- We believe that the false positive rate reported, on some normal mammograms and normal regions in abnormal mammograms, by the IMCAD system, is not alarming since these suspicions could be removed by adding other complementary exams or performing other features to eliminate them.
- The artificial immune recognition system shows its efficiency on recognizing masses and normal regions from a very restraint set of training data. However, the choice of immune parameters is a very delicate task and it is done in an empirical way.
- Because abnormalities are often hidden in dense breast tissue, some learning mammograms were contrast adjusted.
- Unlike the deep learning which needs a large amount of data, our automatic IMCAD manages to classify a large number of regions from only 4 learned masses and 3 features (intensity, compactness and area).
- The proposed computer aided detection system based on artificial immunity works automatically from input mammogram to final decision. It takes into account the set of memory cells computed during the learning step. However, it has a significant time consuming (~2mn), when computing the Xie-Beni validity index and the optimal number of classes.
- The direct comparison of systems for detecting mammographic abnormalities is difficult because few studies have been reported on a common database and have not the same working conditions (for example [45][46] on Wisconsin database and [17] on multimodality images ). For this reason, we relied on the decision making model of H. Simon [53] which refers to the expert in the evaluation step and we compared our results with the annotations of Inbreast Database radiologists (Fig 16.d).

## IX. CONCLUSION AND FUTURE WORK

We presented in this work an automatic Computer Aided Detection system IMCAD which combines medical image processing, bio-inspired pattern recognition areas and others methods in computer vision.

The aim of this work is to support radiologists as a second reader of screening mammograms to search subtle lesions that might otherwise be missed visually, and thus a contribution to reduce the mortality rate caused by breast cancer.

The methodology presented in this paper takes advantages of several robust approaches. Reduced data by a multiresolution Gauss



pyramid allows the system to work easily by reducing the processing time and resource allocation. Automatic segmentation was performed by the fuzzy c-means approach with a recursive labeling of regions. One of the important features of FCM algorithm is the membership function and the belonging of an object to several classes, with different degrees. This is an important supportive tool for medical CAD systems. Artificial immune recognition system in a morphological feature space lets our CAD act like a natural human system facing danger. It was designed to differentiate masses from normal regions from only three features.

It can be concluded that IMCAD succeeded widely in automatic detection of abnormal regions. Indeed, AIRS achieved a good ROC curve with AUC of 0,78, sensitivity of 100% and specificity of 95%, on the studied images.

Studies are in progress to increase specificity, treat ill-defined masses and minimize time processing. In our future works, we plan to extend our system to a computer aided diagnosis by analyzing the ROI extracted from the actual system and specifying its degree of malignancy and benignity.

#### ACKNOWLEDGMENT

The authors would like to thank Professor Jaime S. Cardoso and the Breast Research Group from the INESC Porto (Universidade do Porto) for providing the INBREAST database and for useful discussions.

#### REFERENCES

- [1] National Cancer Institute (2015, Feb). What Is Cancer? Available: <http://www.cancer.gov>
- [2] Rebecca L. Siegel, Kimberly D. Miller, Ahmedin Jemal (2017). Cancer statistics. CA: A Cancer Journal for Clinicians, 65(1), pp. 7–30, Wiley Publishing.
- [3] Ahmad Taher Azar, Sundarapandian Vaidyanathan (2014). Computational Intelligence Applications in Modeling and Control. Springer.
- [4] Cheng HD, Cai X, Chen X, Hu L, Lou X (2003). Computer-aided detection and classification of microcalcifications in mammograms: a survey. Pattern Recognition, 36(12), pp. 2967–91.
- [5] Ronald A Castellino (2005). Computer aided detection (CAD): an overview. Cancer Imaging, 5(1), pp. 17–19.
- [6] Mehul P. Sampat, Mia K. Markey, and Alan C. Bovik (2005). Computer-Aided Detection and Diagnosis in Mammography, Handbook of image and video processing, 2(1), pp. 1195-1217.
- [7] Debra M. Ikeda (2010). Breast imaging: the requisites. Mosby Elsevier, 2nd edition.
- [8] H. Li, Y. Wang, K. J. Liu, Lo SC, Freedman MT. (2001). Computerized radiographic mass detection—part I: Lesion site selection by morphological enhancement and contextual segmentation. IEEE Transactions on Medical Imaging, 20(4), pp. 289–301.
- [9] J. Anitha, J. Dinesh Peter, S. Immanuel Alex Pandian (2017). A dual stage adaptive thresholding (DuSAT) for automatic mass detection in mammograms. Computer methods and programs in biomedicine. 138©, pp. 93–104.
- [10] T. Matsubara, H. Fujita, T. Endo, et al. (1996). Development of mass detection algorithm based on adaptive thresholding technique in digital mammograms. K. Doi, M. L. Giger et al. eds. pp. 391–396.
- [11] S. M. Lai, X. Li, and W. F. Bischof (1989). On techniques for detecting circumscribed masses in mammograms. IEEE Transaction on Medical Imaging. 8(4), pp. 377–386.
- [12] Jostein Herredsvella, Thor Ole Gulsrud, Kjersti Engan (2005). Detection of circumscribed masses in mammograms using morphological segmentation. SPIE Proceedings Volume 5747, Medical Imaging 2005: Image Processing.
- [13] Brzakovic, D., Luo, X.M., Brzakovic, P. (1990). An Approach to Automated Detection of Tumors in Mammograms. IEEE Transaction on Medical Imaging. 9(3), pp. 232–241.
- [14] Lixin Song, Yanan Lv, Bin Yang, Yuhong Wang (2013). Segmentation of breast masses using adaptive region growing. 8th International Forum on Strategic Technology (IFOST), Ulaanbaatar, Mongolia, 28 June-1 July 2013.
- [15] Marcomini K.D., Schiabel H. (2015). Investigating automatic techniques in segmentation accuracy of masses in digital mammography images. In: Jaffray D. (eds) World Congress on Medical Physics and Biomedical Engineering, June 7-12, 2015, Toronto, Canada. IFMBE Proceedings, vol. 51. Springer, Cham.
- [16] Frédéric J. P. Richard, Laurent D. Cohen (2003). A new Image registration technique with free boundary constraints: application to mammography. Computer Vision and Image Understanding, Vol. 89, pp. 166-196.
- [17] C.D. Katsis, I. Gkogkou, C.A. Papadopoulos, Y. Goletsis, P.V. Boufounou (2013). Using Artificial Immune Recognition Systems in Order to Detect Early Breast Cancer, I.J. Intelligent Systems and Applications, vol.2, pp. 34-40.
- [18] Bushra Mughal, Muhammad Sharif, Automated Detection of Breast Tumor in Different Imaging Modalities: A Review (2017). Current Medical Imaging Reviews, 13(2), pp. 121-139.
- [19] Zobia Suhail, Mansoor Sarwar, and Kashif Murtaza (2015). Automatic detection of abnormalities in mammograms, BMC Medical Imaging, Elsevier, 2015.
- [20] G. M. te Brake, N. Karssemeijer, and J. H. Hendriks (2000). An automatic method to discriminate malignant masses from normal tissue in digital mammograms. Physics Med. Biol. 45(10), pp. 2843–2857.
- [21] M. M. Mehdy, P. Y. Ng, E. F. Shair, N. I. Md Saleh, and C. Gomes (2017). Artificial Neural Networks in Image Processing for Early Detection of Breast Cancer. Computational and Mathematical Methods in Medicine, vol. 2017, Article ID 2610628, 15 pages, 2017. <https://doi.org/10.1155/2017/2610628>.
- [22] Dheeba J., Albert Singh N., Tamil Selvi S. (2014). Computer-aided detection of breast cancer on mammograms: a swarm intelligence optimized wavelet neural network approach. Journal of Biomedical Informatics 49, pp. 45–52.
- [23] Mohamed Meselhy Eltoukhy, Ibrahima Faye (2013). An Adaptive Threshold Method for Mass Detection in Mammographic Images. IEEE international conference on Signal and image processing applications (ICSIPA), pp.374-378.
- [24] Petrick N, Chan HP, Wei D, Sahiner B, Helvie MA, Adler DD (1996). Automated detection of breast masses on mammograms using adaptive contrast enhancement and texture classification, Med Phys., 23(10), pp. 1685-1696.
- [25] B. Zheng, Y.H. Chang, X.H. Wang, W.F. Good, D. Gur (1999). Application of a Bayesian belief network in a computer-assisted diagnosis scheme for mass detection. SPIE Conference on Image Processing, 3661 (2), pp. 1553–1561.
- [26] D Ribli, A Horváth, Z Unger, P Pollner, I Csabai (2018). Detecting and classifying lesions in mammograms with Deep Learning. Scientific reports 8 (1), 4165.
- [27] R. Agarwal, O. Diaz, X. Lladó, R. Martí (2018). Mass detection in mammograms using pre-trained deep learning models. Proc. SPIE 10718, 14th International Workshop on Breast Imaging (IWBI 2018).
- [28] Hongyu Wang, Jun Feng, Qirong Bu, Feihong Liu, Min Zhang, Yu Ren, and Yi Lv (2018). Breast Mass Detection in Digital Mammogram Based on Gestalt Psychology. Journal of Healthcare Engineering. Volume 2018, Article ID 4015613, 13 pages.
- [29] Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I., Boggis, C., Ricketts, I., et al. (2015). Mammographic Image Analysis Society (MIAS) database v1.21 [Dataset]. <https://www.repository.cam.ac.uk/handle/1810/250394>.
- [30] Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore and W. Philip Kegelmeyer (2001). The Digital Database for Screening Mammography. Proceedings of the Fifth International Workshop on Digital Mammography, M.J. Yaffe, ed., 212-218, Medical Physics Publishing.
- [31] Valliappan Raman, Putra Sumari, H.H. Then, and Saleh Ali K. Al-Omari (2011). Review on Mammogram Mass Detection by Machine Learning Techniques. International Journal of Computer and Electrical Engineering, Vol. 3, pp. 873-879.
- [32] Dipankar Dasgupta, Luis Fernando Niño (2009). Immunological computation: Theory and applications, CRC Press Taylor & Francis Group.

- [33] Anthony Brabazon, Michael O'Neill, Seán McGarraghy (2015), Natural Computing Algorithms, Springer.
- [34] Carter, J. H. (2000). The immune systems as a model for pattern recognition and classification. Journal of the American Medical Informatics Association, 7(1), pp. 28-4.
- [35] Jon Timmis, Mark Neal, and John Hunt (2000). An Artificial Immune System for Data Analysis. Biosystems, 55(1-3), pp. 143-150.
- [36] Jon Timmis and Mark J. Neal (2000). A Resource Limited Artificial Immune System for Data Analysis Research and Development in Intelligent Systems XVII, pp. 19-32.
- [37] Jon Timmis and Mark Neal (2000). Investigating the evolution and stability of a resource limited artificial immune system. Special Workshop on Artificial Immune Systems, Genetic and Evolutionary Computation, Conference (GECCO) 2000, Las Vegas, Nevada, U.S.A., pp. 40-41.
- [38] Leandro N. de Castro and Fernando J. Von Zuben (2000). The Clonal Selection Algorithm with Engineering Applications. GECCO 2000, Workshop on Artificial Immune Systems and Their Applications, Las Vegas, USA, pp. 36-37.
- [39] A. B. Watkins (2001). AIRS: A resource limited artificial immune classifiers. PhD Thesis, Mississippi State University.
- [40] A. B. Watkins and L. C. Boggess (2002). A Resource Limited Artificial Immune Classifier. In Part of the 2002 IEEE World Congress on Computational Intelligence held in Honolulu, 2002.
- [41] A. Watkins and J. Timmis (2002), Artificial Immune Recognition System (AIRS): Revisions and Refinements. 1st International Conference on Artificial Immune Systems (ICARIS2002).
- [42] Meng L, van der Putten P, Wang H (2005). A comprehensive benchmark of the artificial immune recognition system (AIRS). Advanced Data Mining and Applications, Springer Berlin/Heidelberg, vol.3584, pp. 575-582.
- [43] Polat K, Güneş S, Tosun S (2006) Diagnosis of heart disease using artificial immune recognition system and fuzzy weighted preprocessing. Pattern Recogn 39, pp. 2186-2193.
- [44] Chikh MA, Saidi M, Settouti N (2012). Diagnosis of diabetes diseases using an artificial immune recognition system2 (AIRS2) with fuzzy k-nearest neighbor. Journal of Medical Systems, 36(5), pp. 2721-2729.
- [45] Saybani, M.R., Wah, T.Y., Aghabozorgi, S.R. et al. (2016). Diagnosing breast cancer with an improved artificial immune recognition system. Soft computing, 20(10), pp. 4069-4084.
- [46] B.V.Kavethaa, Dr. P. Mohankumarb (2016). An artificial immune system based investigation of the breast cancer classification. International Journal of Pharmacy & Technology, 8(4), pp. 23097-23107.
- [47] Jason Brownlee (2012). Clever algorithms, Nature-Inspired Programming Recipes, ISBN:1446785068 9781446785065.
- [48] Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012). INbreast: Toward a Full-field Digital Mammographic Database. Acad. Radiol., 19(2), pp. 236-248.
- [49] Burt and Adelson (1983). The Laplacian Pyramid as a Compact Image Code. IEEE Transactions on Communications, COM-31(4), pp. 532-540.
- [50] Liyan Zhang (2001). Comparison of fuzzy c-means algorithm and new fuzzy clustering and fuzzy merging algorithm, Professional Paper, University of Nevada, May 2001.
- [51] Liyan Zhang (May 2001). Comparison of Fuzzy c-means algorithm and New Fuzzy Clustering and Fuzzy Merging Algorithm. Professional Paper, Computer Science Department, University of Nevada, Reno.
- [52] Xuanli Lisa Xie and Gerardo Beni (1991). A validity measure for fuzzy clustering. IEEE Trans. Pattern Anal. Mach. Intell., 13(8), pp. 841-847.
- [53] Herbert Siomon (1977). The New Science of Management Decision. Prentice Hall, 1977.



Leila Belkhodja

Leila Belkhodja is an Assistant Professor at the national institute of industrial security and maintenance in the University of Oran2. She received her engineering degree in computer science (2002) and her Master in electronic filed (2006) from the University of Sciences and Technology Mohammed Boudiaf, Algeria. She is preparing her PhD thesis within the Computer Science Department in the University of Oran1. In research field, she works on medical image processing, Computer Aided systems, bio-inspired algorithms, and artificial recognition.



Djamila Hamdadou

D. Hamdadou received her Engineering degree in Computer Science and her Master of Science degree from the Computer Science Institute in 1993 and 2000, respectively. She also obtained her Doctorate in 2008. She received her PHD in 2012 from the Computer Science Department. Currently, responsible of the Research Team "Spatio- Temporal Modeling and Artificial Vision: from the Sensor to the Decision" of the Computer Science Laboratory of Oran LIO. Her research interests include Business Intelligence, Artificial Intelligence, Spatio Temporal Modeling and Image Processing.

# Sentiment Analysis on IMDb Movie Reviews Using Hybrid Feature Extraction Method

H. M. Keerthi Kumar<sup>1</sup>, B. S. Harish<sup>2</sup>, H. K. Darshan<sup>3</sup> \*

<sup>1</sup> JSSRF, JSS TI Campus, Mysuru, Karnataka (India)

<sup>2</sup> Department of Information Science and Engineering, JSS Science and Technology University, Mysuru, Karnataka (India)

<sup>3</sup> Department of Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka (India)

Received 12 June 2018 | Accepted 17 November 2018 | Published 7 December 2018



## ABSTRACT

Social Networking sites have become popular and common places for sharing wide range of emotions through short texts. These emotions include happiness, sadness, anxiety, fear, etc. Analyzing short texts helps in identifying the sentiment expressed by the crowd. Sentiment Analysis on IMDb movie reviews identifies the overall sentiment or opinion expressed by a reviewer towards a movie. Many researchers are working on pruning the sentiment analysis model that clearly identifies and distinguishes between a positive review and a negative review. In the proposed work, we show that the use of Hybrid features obtained by concatenating Machine Learning features (TF, TF-IDF) with Lexicon features (Positive-Negative word count, Connotation) gives better results both in terms of accuracy and complexity when tested against classifiers like SVM, Naïve Bayes, KNN and Maximum Entropy. The proposed model clearly differentiates between a positive review and negative review. Since understanding the context of the reviews plays an important role in classification, using hybrid features helps in capturing the context of the movie reviews and hence increases the accuracy of classification.

## KEYWORDS

Classification, Hybrid Features, Short Text, Sentiment Analysis.

DOI: 10.9781/ijimai.2018.12.005

## I. INTRODUCTION

**S**OCIAL media has become an integral part of human living in recent days. People want to share each and every happening of their life on social media. Nowadays, social media is used for showcasing one's pride or esteem by posting photos, text, video clips, etc. The text plays a vital aspect in information shared, where users share their opinions on trending topics, politics, movie reviews, etc. These opinions which people share on social networking sites are generally known as Short Texts (ST) because of its length [1]. ST have gained its importance over traditional blogging because of their simplicity and effectiveness in influencing the crowd. They are even used by search engines in the form of queries. Apart from their popularity, ST has certain challenges like identification of sarcasm, sentiment, use of slang words, etc. Therefore it becomes important to understand short texts and derive meaningful insights from them, which is generally known as Sentiment Analysis (SA) [2].

SA played an important role in the US Presidential Elections 2016 [3]. People shared their likes and dislikes regarding a particular political party on micro-blogs such as Twitter and Facebook. Those blogs were analyzed and candidates pruned their tweets based on these analyses. Thus, SA helped them to increase their popularity and followers. SA is widely used by most of the companies because of its capacity to

analyze a large number of documents at once, which manually would take more time. In the business sector, companies use SA to derive new strategies based on the customer feedback [4].

Reviews are short texts that generally express an opinion about movies or products. These reviews play a vital role in the success of movie or sales of the products [5]. People generally look into blogs, review sites like IMDb to know about movie cast, crew, review and ratings. Hence it is not only the Word of Mouth that brings the audience to the theatres; reviews also play a prominent role in this regard. In other words, SA on movie reviews makes the task of Opinion Summarization [6] easier by extracting the sentiment expressed by the reviewer.

The task of SA on movie reviews mainly include – Preprocessing [7], Feature Extraction followed by Selection [8], Classification [9] and finally the analysis of results. Preprocessing involves removal of stop words, abbreviating short forms, replacing slangs, etc. which are scrutinize for the task of classification. Feature Extraction involves identifying the features that represent the documents in the vector space. Many feature extraction [10] methods that exist will extract the features from the reviews, mainly by statistical based and lexicon based approaches. In statistical feature extraction methods [11], the words present in the review are used as features by calculating various weighing measures like Term Frequency (TF), Inverse Document Frequency (IDF) and Term Frequency-Inverse Document Frequency (TF-IDF) [31]. In Lexicon [12] based feature extraction methods, textual features are extracted by deriving the patterns among the words, deriving from Parts of Speech of the words tagger, using Lexicon Dictionaries, etc. Lexicon based methods generally capture the semantics of the text by considering the ordering of text in the

\* Corresponding author.

E-mail addresses: hmkeerthikumar@gmail.com (H. M. Keerthi Kumar), bsharish@jssstuniv.in (B. S. Harish), bharadwajdarshan@gmail.com (H. K. Darshan).

review. Hybrid approaches [13] involving both Statistical and Lexicon based feature extraction methods will increase the overall accuracy of the model. Once the extraction of features is done, relevant features are identified using feature selection methods [14] which eliminate the features that do not contribute towards effective classification. Classification involves identifying the polarity of the review and classifying it as either positive or negative sentiment.

This paper proposes a Hybrid method, where the features are extracted by using both statistical and lexicon methods. In addition, we apply various feature selection methods such as Chi-Square, Correlation, Information Gain and Regularized Locality Preserving Indexing (RLPI) [15] for the features extracted by statistical methods. This maps the higher dimension input space to the lower dimension input space. The Lexicon based feature extraction method extract features based on the Lexicon dictionaries. Features from both methods are combined to form a new feature set which is of lower dimension when compared to the initial dimension of the input space. The new features set is classified using various classifiers such as Support Vector Machines (SVM), Naïve Bayes (NB), K- Nearest Neighbor (KNN) and Maximum Entropy (ME) classifiers on IMDb movie review dataset.

The contents of the paper are divided into five sections. Section II presents an overview of the literature survey of previous works on sentiment analysis. Section III presents the methodology. Section IV shows experimental results and the paper is concluded in section V.

## II. LITERATURE SURVEY

The process of Sentiment Analysis involves the construction of the input vector space from the existing document vector space. Mainly there are two approaches to carry out vector space mapping. The machine learning based or statistical based feature extraction methods are widely used because extraction of features is done by applying statistical measures directly. Earlier works on sentiment classification using machine learning approaches were carried by Pang et al. in 2002 [16]. Sentiment analysis was performed on IMDb movie reviews using n-gram approaches and Bag of Words (BOW) as features. The model was trained using different classifiers like Naïve Bayes (NB), Maximum Entropy (ME), and Support Vector Machines (SVM). The unigram features outperformed when compared to other features [16]. Similar work was done by Tripathy et al. [5], where TF, TF-IDF was used for the conversion of the text file to a numerical vector. Experimentation was done with n-gram approaches and its combination are tried to get the best results.

Apart from the word features which are considered for the classification task, special symbols which are present with words -known as emoticons (☺, ☹, ☺, ...) can also be used as features. Neetu et al. [17] used these special features along with the word features. The use of an ensemble classifier which classifies based on the results obtained by different classifiers like NB, ME and SVM is the major highlight of the work. Many researchers have worked on extracting features based on the parts of speech tagger. Geetika et al. [18] used unigram model to extract adjective as a feature which in turn describes the positivity or negativity of the sentence.

Identifying the semantics or the meaning of the text by a machine learning algorithm is a challenging task. Lexicon features are used in this regard to extract the opinions expressed in the text. Sarcasm detection is one of the major advantages of choosing lexicon features. Anukarsh et al. [19] focused on the slangs and emojis which were present in the text to detect sarcasm. Use of slang and emoji dictionaries during preprocessing increased the efficiency of sarcasm detection. Capturing the sentiment orientation of the text towards a topic helps in identifying the overall polarity of the text. Taboda et al., in [12], used

dictionaries to calculate the Semantic Orientation (SO) and termed it as Semantic Orientation CALculator (SO-CAL). Various factors such as Parts of Speech (Adjectives, Nouns, Verbs and Adverbs), Intensifiers (Somewhat, Very, Extraordinary etc.), Negations, etc., were considered to calculate sentiment orientation. Results showed that the Lexicon based sentiment analysis gives better results and can be applied to wide domains. Similarly in [32], Dehkharghani developed lexicon for sentiment analysis.

Melville et al. [20], worked on extracting features using lexicon methods. Positive and negative word counts that are present in the text were used as the background lexicon knowledge and then the probability that a document belongs to a particular class was calculated. Use of pooling multinomial classifiers which incorporate both training examples and the background knowledge is the major contribution. Kolchyna et al., in [21], used both machine learning and lexicon approaches to perform sentiment analysis on Twitter data. Special lexicon features such as N-grams, Lexicon sentiment, Elongated words number, Emoticons, Punctuations, etc., were used. Use of these features increased the overall accuracy of the model. The hybrid method combines the features generated by both machine learning approach and lexicon approach. Use of a hybrid approach reduces the complexity of the overall model by retaining only the important features and thus increases time efficiency. The main advantage of using the lexicon features is that it captures the meaning or the semantics expressed in the reviews thereby contributing to the effective classification. The experimental results showed that the review classification was more accurate because of the use of semantics of the review as a feature and is comparable with the human review classification.

The polarity of a review depends on the intensity of each word present in the review and the context used by the reviewers to express their opinion. Therefore, identifying the features that extract the intensity of words based on context that inclines the polarity either towards positive or negative polarity is a challenging task. The proposed work captures the polarity of a word and determines how important the word is for the classification task. The capturing phase is done through the features generated using Hybrid Feature Extraction Method (HFEM). The HFEM combines the reduced Machine learning features with the Lexicon features to increase the performance of the model.

The major contribution of this paper includes:

- Identifying the lexicon features such as Positive word count, Negative word count, Positive Connotation count, Negative Connotation count, which helps to identify the semantics of the reviews.
- Use of RLPI feature selection method to reduce high dimensional features.
- Comparison of the classification accuracy and F-measure of Machine learning feature, lexicon feature and HFEM using different supervised learning algorithms.

## III. METHODOLOGY

In the proposed model, sentiment analysis is employed on IMDb Movie Reviews. The input for the proposed model is the set of reviews whose polarity needs to be determined. The output corresponds to reviews with polarity assigned to each of them. The task of sentiment analysis is carried out in the following phases: preprocessing the dataset, feature Extraction (Both Statistical and Lexicon approach), feature selection and finally classification using hybrid features. Fig. 1 gives the overall workflow of the proposed model.



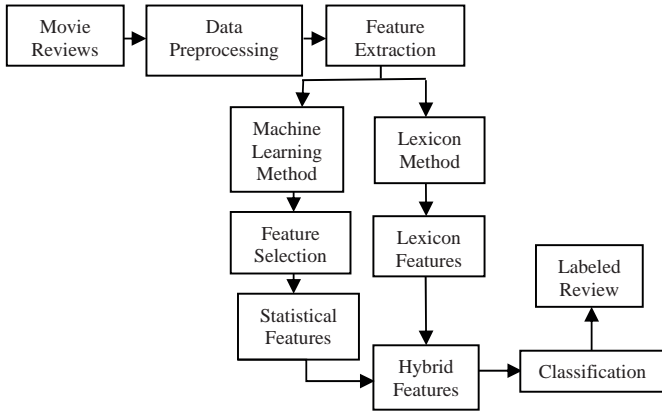


Fig. 1. Proposed model for sentiment classification.

### A. Preprocessing

The reviews which need to be analyzed consist of words, numbers, and special symbols as its constituents. Consider the following review, “The great and underrated Marion Davies shows her stuff in this late (1928) silent comedy that also showcases the wonderful William Haines. A must for any serious film buff or for anyone interested in the still-maligned Marion Davies!”. The review contains the year 1928 and punctuations like ‘.’, ‘!’, ‘,’ that does not influence on sentiment analysis because of its neutral polarity [33]. Hence the numbers and punctuations are removed. Many words such as ‘a’, ‘an’, ‘the’, ‘should’, etc., which are commonly known as stopwords are also eliminated. Many words that are present in the reviews will not be in their root forms. For example, words like ‘studying’, ‘studied’ belong to same root word ‘study’. This process is known as Lemmatization [22] where the ineffectual endings of the words are removed by bringing to the root form with the help of vocabulary. Further, the preprocessed dataset is used for feature extraction in the next phase.

### B. Feature Extraction

Feature Extraction identifies the features that have a positive effect towards classification. In this work, feature extraction is carried in two different parallel stages namely- Machine learning based feature extraction and Lexicon based features extraction.

Machine learning based feature extraction method is used to extract the features using popularly known technique Bag of Words, wherein the column corresponds to words and row corresponds to value of weighing measures such as Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF).

Lexicon based feature extraction method which is used in the proposed work extracts 4 different features from the review. They are as follows: Positive word Count (PC), Negative word Count (NC), Positive Connotation Count (PCC) and Negative Connotation Count (NCC).

Positive and negative words present in the reviews are identified by using Positive word dictionary and Negative word dictionary respectively. Connotation refers to the abstract meaning of the word depending on the context. For example, “Desire” is a positive connotation which is absent in the positive dictionary and “Avoid” is a negative connotation which is absent in the negative dictionary. Hence, Positive and Negative connotation lexicon is used in addition to the regular positive and negative words dictionary to identify PCC and NCC.

Combination of features extracted through different feature extraction methods will increase the overall performance of the model. Combining machine learning based features with the features extracted by using a Positive-Negative lexicon and Positive-Negative connotation lexicon is the key idea of the paper. This combination of

features helps in identifying the overall polarity of the review more accurately. Use of Hybrid Feature Extraction Method (HFEM) is the major contribution of this paper.

### C. Feature Selection

Features extracted in the previous phase may contain hundreds to thousands of features for a small set of reviews. Handling large number of features increases the burden on the classification algorithms. Feature selection methods are generally applied to reduce the dimension of the feature space by selecting only the important features. Reduction in the feature dimension should not affect the classification accuracy. Importance of a feature is calculated by various statistical methods. Various conventional feature selection methods such as Chi-Square [23], Correlation [24], and Information Gain [25] are used in literature to perform sentiment analysis. Many feature selection methods which are used in other domains such as medical image processing, document indexing and clustering etc., can also be used to perform sentiment analysis effectively. RLPI [15] is one such unconventional feature selection method which is generally used for document indexing and representation. RLPI reduces the dimension of the features by performing Eigen vector decomposition on feature space and then selects top Eigen vectors to represent the features. Thus, RLPI helps in handling large number of features which can be further reduced to smaller dimension feature space. Use of RLPI features along with Lexicon features for testing and training the learning algorithms is the major contribution of this paper.

In the proposed approach, we apply feature selection method to the feature set generated by an statistical approach because of its larger dimension. On the other hand, we extracted 4 features using the lexicon approach. The selected feature set from statistical approach is combined with the Lexicon features to form a matrix which is the input for the classification algorithm.

### D. Classification

Classification is the process of assigning labels to the reviews whose label is unknown. In the proposed work, supervised learning algorithms such as Naïve Bayes [17], Maximum Entropy [16], Support Vector Machines (SVM) [13] and K- Nearest Neighbor (KNN) [26] [30] are used. Naïve Bayes and Maximum Entropy work on the principles of probability and hence they are known as probabilistic classifiers. Naïve Bayes works on the principle of independence of features and calculates the probability of a review belonging to particular class using Bayes theorem. Maximum Entropy classifies the review by calculating the conditional probability. Maximum Entropy does not assume independence of features. SVM is independent of the number of features in the feature space. SVM uses a hyper plane to separate the samples of two classes. KNN classifies the review with the unknown label by comparing it with the reviews present in the test data. The comparison is done by applying various similarity measures and identifies the most similar review (nearest neighbor). Further, it assigns the label of nearest neighbor to the unlabeled review.

## IV. EXPERIMENTAL SETUP

### A. Dataset

IMDb is the most commonly used website for getting information about a movie throughout the world. Because of its popularity and due to the presence of large number of reviews related to a particular movie, IMDb Movie Review Dataset [27] is used in the proposed work. It is one of the standard benchmark datasets used for Sentiment Analysis on Movie reviews. The dataset contains 25,000 positive and negative reviews each. However, due to the limitations of computational resources, we have randomly chosen 5000 reviews for experimentation.

## B. Lexicons

In the proposed work, the following lexicons are used: 1. Opinion Lexicon created by Hu et al., [28] which is used as Positive-Negative Lexicon. It consists of around 6800 words including positive and negative words. 2. Connotation Lexicon created by Feng et al., [29] which contains positive and negative connotations.

## C. Experimentation

The experimentation is carried on a machine running on Ubuntu 16.04 operating system with R Studio version 3.4.2 environment. For experimentation, 5-fold validation technique is used i.e., 5000 reviews were randomly selected from the dataset and were again split into 5 batches containing 1000 reviews in each batch. The final result shown in Table I is the average of results obtained in the 5 batches. Initially, 13,346 features are extracted using machine learning feature extraction method. Because of its larger dimension, feature selection methods such as IG, Correlation, Chi-Square were applied. The reduced dimension of the input space was varied from 10% to 60% of the initial number of features. Feature count is varied from 1000 to 8000 features and the best results were obtained for the feature counts 2000, 5000 and 8000. In case of RLPI feature selection method, the reduced dimension of the input space was varied from 50 to 150 because lesser number of features are sufficient to represent feature in terms of Eigen vectors and hence feature count varies around 1% of the initial feature count. Better results were obtained for fewer feature counts like 50, 100 and 150. Four lexicon features namely Positive word count, Negative word count, Positive Connotation count and Negative Connotation count are used along with the reduced machine learning features. The hybrid features are used to train the model using learning algorithms such as Naïve Bayes, SVM, Maximum Entropy and KNN.

## D. Result Analysis and Discussion

This section presents the analysis of results obtained using various classifiers with different feature selection methods (FSM). Machine learning features described in this section consists of features generated using different weighing schemes like TF, TF-IDF. When these features are merged with the Lexicon features mentioned previously, it generates the Hybrid features. Fig. 2-4 show the comparison of accuracy obtained using Machine Learning features, Lexicon features and Hybrid features on classifiers like SVM, NB, ME and KNN during 5 batches of experimentation. The accuracies shown in the figures are the highest accuracies in that particular batch using different FSM on different classifiers.

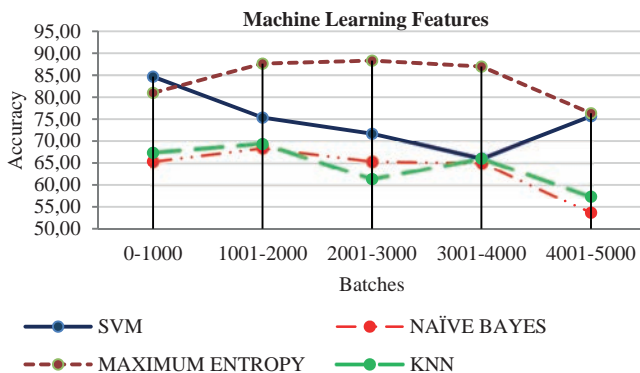


Fig. 2. Comparison of accuracy obtained using Machine Learning features for 5 batches.

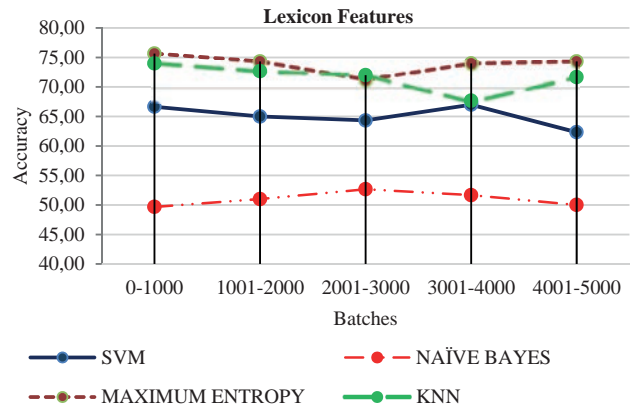


Fig. 3. Comparison of accuracy obtained using Lexicon features for 5 batches.

When compared to various feature selection methods, RLPI gives the best accuracy of 74.66% with less number of features using Maximum Entropy classifier (see Table I). This is because RLPI selects the discriminating features which are smaller in number but are highly sufficient to represent the feature space. This can be considered as the best result since the complexity of the input data is reduced to 0.4% of the original dimension. However, accuracy is the tradeoff when we consider lesser complex input data which contains less number of features. Maximum Entropy is the best performing classifier with the highest accuracy of 83.93% when correlation is used as a feature selection method, because the features selected by correlation are highly correlated with the class and have effective contribution towards classification. Hybrid features outperforms machine learning features and gives the best result in terms of both accuracy and F-measure irrespective of the feature selection method and classifier used. Further, addition to lexicon features along with machine learning features i.e., use of hybrid features increases the accuracy using SVM classifier with Chi-Square feature selection method.

The results summarized in Table I can be analyzed by considering two parameters namely complexity of the input data and highest accuracy achieved. Percentage of improvement in number of features, accuracy and F-measure is presented in Table I. The percentage change corresponds to the percentage increase or decrease in number of features, accuracy and F-measure of Hybrid features when compared to the Machine Learning Features.

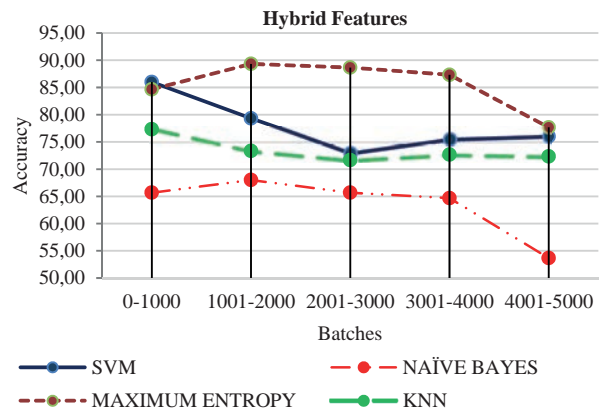


Fig. 4. Comparison of accuracy obtained using Hybrid features for 5 batches.

It is evident from the Table I that the highest percentage of decrease in the input space is about 37.5% and the highest increase in percentage of accuracy is 36.38% and that of F-measure is 78.19%. The results obtained by using Hybrid Feature Extraction Method are promising both in terms of accuracy, F-measure and complexity.

TABLE I. RESULTS USING FSM WITH CLASSIFIERS USING HYBRID FEATURES

FS Method	Classifier	Hybrid Features			% of improvement in no. of features	% of improvement in accuracy	% of improvement in F-measure
		Number of features	Accuracy	F-Measure			
IG	SVM	5000	75.467	0.752	-37.5	32.554	59.660
	Naïve Bayes	8000	54.733	0.608	60	2.62	10.144
	Maximum Entropy	<b>8000</b>	<b>78.333</b>	<b>0.780</b>	<b>0</b>	<b>7.600</b>	<b>8.18</b>
	KNN	5000	72.267	0.723	150	24.454	17.84
Correlation	SVM	5000	76.600	0.764	150	34.859	75.632
	Naïve Bayes	2000	60.667	0.573	0	-0.108	-0.174
	Maximum Entropy	<b>5000</b>	<b>83.933</b>	<b>0.837</b>	<b>-37.5</b>	<b>1.94</b>	<b>2.32</b>
	KNN	2000	72.000	0.721	0	26.021	15.733
Chi Square	SVM	5000	75.467	0.752	-37.5	36.386	78.199
	Naïve Bayes	8000	54.733	0.608	0	0	0
	Maximum Entropy	<b>8000</b>	<b>78.333</b>	<b>0.780</b>	<b>0</b>	<b>7.600</b>	<b>8.18</b>
	KNN	5000	72.267	0.723	150	24.454	21.717
RLPI	SVM	100	73.600	0.734	0	1.939	2.370
	Naïve Bayes	50	63.400	0.567	0	0	-0.176
	Maximum Entropy	<b>50</b>	<b>74.667</b>	<b>0.739</b>	<b>0</b>	<b>2.564</b>	<b>2.354</b>
	KNN	150	71.933	0.720	200	24.022	22.033

## V. CONCLUSION

Sentiment Analysis on short informal text is a challenging task. Due to the limited number of characters, huge dimensional features and sparseness, which increases complication. In this paper, Hybrid Feature Extraction Method (HFEM) is used to extract features from machine learning and lexicon based feature extraction methods. Initially, machine learning features are in high dimensional in nature. The feature selection methods such as Information Gain, Correlation, Chi Square and RLPI are applied on the machine learning features to reduce high dimensional features. On the other hand, lexicon based features such as Positive word Count (PC), Negative word Count (NC), Positive Connotation Count (PCC) and Negative Connotation Count (NCC) are extracted. Combining machine learning features with the lexicon features captures the orientation of words and thus identifies the context of the review. To demonstrate the effectiveness of the proposed work, we used four different classifiers such as SVM, KNN, Maximum Entropy and Naïve Bayes on IMDb movie review dataset. The Maximum Entropy with correlation shows the best results in terms of both accuracy and F-measure when compared to other classifiers. Use of Hybrid Feature Extraction Method (HFEM) makes the model more efficient in terms of accurate classification by adding the advantages of individual feature extraction method. HFEM improves the space complexity by reducing the input space to minimal number of features that are sufficient to represent the review content. Thus, results obtained are highly promising both in terms of space complexity and classification accuracy. In future work, we will include more lexicon features to the feature subset and thereby expect to increase the classification accuracy.

## ACKNOWLEDGMENT

H M Keerthi Kumar has been financially supported by UGC under Rajiv Gandhi National Fellowship (RGNF) Letter no: F1-17.1/2016-17/ RGNF-2015-17-SC-KAR-6370/(SA-III Website), JSSRF (University of Mysore), Karnataka, India.

## REFERENCES

- [1] C. D. Santos and M. Gatti. "Deep convolutional neural networks for sentiment analysis of short texts." In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers, 2014, pp. 69-78.
- [2] A. Ortigosa, J. M. Martín, and R. M. Carro. "Sentiment analysis in Facebook and its application to e-learning." Computers in Human Behavior Vol. 31, pp.527-541. 2014.
- [3] R. Ahmad, A. Pervaiz, P. Mannan, and F. Zaffar. "Aspect Based Sentiment Analysis for Large Documents with Applications to US Presidential Elections 2016." Social Technical and Social Inclusion Issues (SIGSI), 2017, pp. 13.
- [4] K. Xu, S. S. Liao, J. Li, and Y. Song. "Mining comparative opinions from customer reviews for Competitive Intelligence." Decision support systems, Vol. 50, no. 4, pp.743-754. 2011.
- [5] A. Tripathy, A. Agrawal, and S.K. Rath. "Classification of sentiment reviews using n-gram machine learning approach." Expert Systems with Applications, Vol. 57, pp. 117-126. 2016.
- [6] M. E. Moussa, E. H. Mohamed, and M. H. Haggag. "A survey on Opinion Summarization Techniques for Social Media." Future Computing and Informatics Journal (2018). In press.
- [7] I. Hemalatha, G. P. S. Varma, and A. Govardhan. "Preprocessing the informal text for efficient sentiment analysis." International Journal of Emerging Trends & Technology in Computer Science (IJETTCS) 1, no. 2: pp.58-61. 2012.
- [8] A. S. Manek, P. D. Shenoy, M. C. Mohan, and K. R. Venugopal. "Aspect term extraction for sentiment analysis in large movie reviews using Gini Index feature selection method and SVM classifier." World Wide Web Vol. 20, no. 2, pp.135-154. 2017.
- [9] A. Kennedy and D. Inkpen. "Sentiment classification of movie reviews using contextual valence shifters." Computational intelligence, Vol. 22, no. 2, pp.110-125. 2006.
- [10] M. Z. Asghar, A. Khan, S. Ahmad, and F. M. Kundi. "A review of feature extraction in sentiment analysis." Journal of Basic and Applied Scientific Research, Vol. 4, no. 3, pp.181-186. 2012.
- [11] A. Sharma and S. Dey. "A comparative study of feature selection and machine learning techniques for sentiment analysis." In Proceedings of the 2012 ACM research in applied computation symposium, pp. 1-7. ACM, 2012.
- [12] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, and M. Stede. "Lexicon-

based methods for sentiment analysis.” Computational linguistics, Vol. 37, no. 2, pp.267-307. 2011.

- [13] A. Mudinas, D. Zhang, and M. Levene. “Combining lexicon and learning based approaches for concept-level sentiment analysis.” In Proceedings of the first international workshop on issues of sentiment discovery and opinion mining, pp. 5. ACM, 2012.
- [14] L. Zheng, H. Wang, and S. Gao. “Sentimental feature selection for sentiment analysis of Chinese online reviews.” International journal of machine learning and cybernetics, Vol. 9, no. 1, pp.75-84. 2018.
- [15] D. Cai, X. He, W. V. Zhang, and J. Han. “Regularized locality preserving indexing via spectral regression.” In Proceedings of the sixteenth ACM conference on Conference on information and knowledge management, pp. 741-750, ACM, 2007.
- [16] B. Pang, L. Lee, and S. Vaithyanathan. “Thumbs up?: sentiment classification using machine learning techniques.” In Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10, Association for Computational Linguistics, pp. 79-86. 2002.
- [17] M. S. Mubarak, Adiwijaya, and M. D. Aldhi. “Aspect-based sentiment analysis to review products using Naïve Bayes.” In AIP Conference Proceedings, vol. 1867, AIP Publishing, no. 1, pp 1-8.2017.
- [18] G. Gautam, and D. Yadav. “Sentiment analysis of twitter data using machine learning approaches and semantic analysis.” In Contemporary computing (IC3), 2014 seventh international conference on, pp. 437-442. IEEE, 2014.
- [19] A. G. Prasad, S. Sanjana, S. M. Bhat, and B. S. Harish. “Sentiment analysis for sarcasm detection on streaming short text data.” In Knowledge Engineering and Applications (ICKEA), 2017, 2nd International Conference on, pp. 1-5. IEEE, 2017.
- [20] P. Melville, W. Gryc, and R. D. Lawrence. “Sentiment analysis of blogs by combining lexical knowledge with text classification.” In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 1275-1284. 2009.
- [21] O. Kolchyna, T. T. P. Souza, P. Treleaven, and T. Aste. “Twitter sentiment analysis: Lexicon method, machine learning method and their combination.” arXiv preprint arXiv:1507.00955. 2015.
- [22] Y. Bao, C. Quan, L. Wang, and F. Ren. “The role of pre-processing in twitter sentiment analysis.” In International Conference on Intelligent Computing, pp. 615-624. Springer, 2014. Cham.
- [23] J. Brooke, M. Tofiloski, and M. Taboada. “Cross-linguistic sentiment analysis: From English to Spanish.” In Proceedings of the international conference RANLP-2009, pp. 50-54. 2009.
- [24] L. Deng, Y. Hu, J. P. Y. Cheung, and K. D. K. Luk. “A Data-Driven Decision Support System for Scoliosis Prognosis.” IEEE Access 5, pp. 7874-7884. 2017.
- [25] F. K. Ahmad. “Comparative Analysis of Feature Extraction Techniques for Event Detection from News Channels’ Facebook Page.” Journal of Telecommunication, Electronic and Computer Engineering (JTEC) Vol. 9, no. 1-2 , pp.13-17. 2017.
- [26] H. M. Kumar, B. S. Harish, S. V. Kumar, and V. N. Aradhya. “Classification of sentiments in short-text: an approach using mSMTP measure”. In Proceedings of the 2nd International Conference on Machine Learning and Soft Computing, pp. 145-150. ACM. 2018.
- [27] A. L. Maas, R. E. Daly, P. T. Pham, D. Huang, A. Y. Ng, and C. Potts. “Learning word vectors for sentiment analysis.” In Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies-volume 1, Association for Computational Linguistics, pp. 142-150. 2011.
- [28] M Hu, and B. Liu. “Mining and summarizing customer reviews.” In Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining, ACM, pp. 168-177. 2004.
- [29] S. Feng, J. S. Kang, P. Kuznetsova, and Y. Choi. “Connotation lexicon: A dash of sentiment beneath the surface meaning.” In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics, Vol. 1, pp. 1774-1784. 2013.
- [30] K. S. Srujan, S. S. Nikhil, H. Raghav Rao, K. Karthik, B. S. Harish, and H. M. Kumar. “Classification of Amazon Book Reviews Based on Sentiment Analysis.” In Information Systems Design and Intelligent Applications, pp. 401-411. Springer, Singapore, 2018.
- [31] M. B. Revanasiddappa, B. S. Harish. A New Feature Selection Method based on Intuitionistic Fuzzy Entropy to Categorize Text Documents, International Journal of Interactive Multimedia and Artificial Intelligence,

(2018), <http://dx.doi.org/10.9781/ijimai.2018.04.002>

- [32] R. Dehkharghani. Building Phrase Polarity Lexicons for Sentiment Analysis, International Journal of Interactive Multimedia and Artificial Intelligence, (2018), <http://dx.doi.org/10.9781/ijimai.2018.10.004>

- [33] H. M. Kumar and B. S. Harish. “Classification of Short Text Using Various Preprocessing Techniques: An Empirical Evaluation.” In Recent Findings in Intelligent Computing Techniques pp. 19-30. Springer, Singapore, 2018.



H. M. Keerthi Kumar

He received B.E in Information Science and Engineering and M.Tech in Software Engineering from Visvesvaraya Technological University, India. He is currently pursuing Ph.D degree in Computer Science from University of Mysore, India. His area of research includes Data Mining, Pattern Recognition and Machine Learning.



B. S. Harish

He obtained his B.E in Electronics and Communication (2002), M.Tech in Networking and Internet Engineering (2004) from Visvesvaraya Technological University, Belagavi, Karnataka, India. He completed his Ph.D. in Computer Science (2011); thesis entitled “Classification of Large Text Data” from University of Mysore. He is presently working as an Associate Professor in the Department of

Information Science & Engineering, JSS Science & Technology University, Mysuru. He was invited as a Visiting Researcher to DIBRIS - Department of Informatics, Bio Engineering, Robotics and System Engineering, University of Genova, Italy from May-July 2016. He delivered various technical talks in National and International Conferences. He has invited as a resource person to deliver various technical talks on Data Mining, Image Processing, Pattern Recognition, and Soft Computing. He is also serving and served as a reviewer for National, International Conferences and Journals. He has published more than 50 International reputed peer reviewed journals and conferences proceedings. He successfully executed AICTE-RPS project which was sanctioned by AICTE, Government of India. He also served as a secretary, CSI Mysore chapter. He is a Member of IEEE (93068688), Life Member of CSI (09872), Life Member of Institute of Engineers and Life Member of ISTE. His area of interest includes Machine Learning, Text Mining and Computational Intelligence.



H. K. Darshan

He is currently pursuing B.E in Information Science and Engineering, Sri Jayachamarajendra College of Engineering, Mysuru, Karnataka, INDIA, His area of interest includes Pattern Recognition, Machine Learning, and Text Mining.



# GIFT: Gesture-Based Interaction by Fingers Tracking, an Interaction Technique for Virtual Environment

Muhammad Raees\*, Sehat Ullah

Department of Computer Science and IT, University of Malakand, KPK (Pakistan)

Received 28 October 2018 | Accepted 31 December 2018 | Published 25 January 2018



## ABSTRACT

Three Dimensional (3D) interaction is the plausible human interaction inside a Virtual Environment (VE). The rise of the Virtual Reality (VR) applications in various domains demands for a feasible 3D interface. Ensuring immersivity in a virtual space, this paper presents an interaction technique where manipulation is performed by the perceptive gestures of the two dominant fingers; thumb and index. The two fingertip-thimbles made of paper are used to trace states and positions of the fingers by an ordinary camera. Based on the positions of the fingers, the basic interaction tasks; selection, scaling, rotation, translation and navigation are performed by intuitive gestures of the fingers. Without keeping a gestural database, the features-free detection of the fingers guarantees speedier interactions. Moreover, the system is user-independent and depends neither on the size nor on the color of the users' hand. With a case-study project; Interactions by the Gestures of Fingers (IGF) the technique is implemented for evaluation. The IGF application traces gestures of the fingers using the libraries of OpenCV at the back-end. At the front-end, the objects of the VE are rendered accordingly using the Open Graphics Library; OpenGL. The system is assessed in a moderate lighting condition by a group of 15 users. Furthermore, usability of the technique is investigated in games. Outcomes of the evaluations revealed that the approach is suitable for VR applications both in terms of cost and accuracy.

## KEYWORDS

Gesture-based Interactions, Virtual Environments, Human Computer Interfaces, Image Processing.

DOI: 10.9781/ijimai.2019.01.002

## I. INTRODUCTION

**V**IRTUAL Reality is emerging in various domains of robotics [1-3], computer engineering [4,5], physical sciences, health-related issues [6-8], natural sciences [9] and industrial academic [10]. To ensure the feelings of being there while interacting with a 3D virtual space, an interface with a high degree of immersivity is required. Interaction via keyboard and mouse lags far behind to properly engross VR users [11]. The gestures of hands and/or fingers are used as the courier of feelings and thoughts in daily life. As gestures have meaning in real-world, hand gestures can be used for distinct interactions. Hence, with the gesture-based interactions, a VR interface can be made natural and perceptual [12].

Different interaction techniques have been proposed for acoustic mechanical and magnetic devices to make human-computer dialogue realistic. The complex setup of the devices make them a rare choice for interactions [13]. However, due to the recent developments in image processing, RGB and depth sensor cameras are becoming the prominent alternatives [14, 15]. Various state-of-the-art techniques have been proposed for HCI, based on RGB camera [16-18] and depth sensors [19-22]. Besides hand tracking, vision-based systems have been proposed for face tracking [23-25], action tracking [26-28] and gait tracking [29, 30]. Although, the outcomes of the contemporary research works favour gesture-based interfaces for interactions [31, 32], such systems are fairly more error-prone [33].

However, the errors and complexities may be minimized if simple

gestures are used for interactions [34].

This research work aims to introduce a low-cost 3D interaction technique based on the positions of the fingers in the dynamic image streams. With the simple and perceptive gestures of the fingers, a 3D object is selected and/or manipulated. At the detection of the finger-tips, a Virtual Hand (VH), indicating the user's position in the VE, is activated. The VH moves freely along x, y and z-axis with the movements of the hand. Navigation and panning are performed by the feasible movements of the hand. As bare-handed gestures are critical [35] and the recognition accuracy of gestures can be improved with coloured-markers [35], therefore the two contrast coloured markers are used for the robust and reliable detection of the fingers. Instead of extracting features from an entire hand posture [36], interactions are performed by the positions of the two fingers, hence fewer computation is ensured.

Using the libraries of OpenCV and OpenGL, the technique is implemented in a project; IGF. Each time, the dynamic positions of the fingers in a scanned image frame are traced using an ordinary camera. By the relative 2D positions of the fingers, a 3D interaction is performed inside the designed VE. In a normal lighting environment, the system is evaluated for a total of 570 interactions by 15 participants. The satisfactory accuracy rate (94.3%) approves suitability of the technique in VR applications.

The paper is organized into 8 main sections. Section II is about the previous related research. The technique with its algorithm and mathematical details is covered in Section III. Details about the interactions are covered in Section IV. Implementation with evaluation details is explained in Section V. A comparative analysis is performed in Section VI whereas applicability in games is investigated in Section VII. The last section of the paper is about conclusion and future work.

\* Corresponding author.

E-mail address: visitrais@yahoo.com

## II. LITERATURE REVIEW

With the privilege of interaction, a VR-user maintains the belief of being there and acknowledges a VE as real as it gets. To achieve a high degree of naturalness, the interface of a VR system should be simple and consistent [37]. The conventional way of interaction with keyboard and mouse are insufficient to properly engross users [11]. As human gestures are flexible and natural [38], therefore gesture/posture based interactions are suitable for VR applications [15]. Various gesture-based techniques have been proposed in the literature of VR for intuitive interactions. The magnetic tracker's based system [39] provides 6-DOF for selection and DeSelection necessitates the use of both hands simultaneously. The haptic feedback interface for virtual objects manipulation of [40] needs the cumbersome setup of wires. The system is useable only inside a restricted workspace. The approach of different gestures for different commands is followed in the system; IVEAS (*Immersive Virtual Environment Authoring System*) designed by Lee et al. [41]. The technique is applicable for the basic interactions, however, the use of CyberGlove and Polhemus Fastrak for the detection of the finger joints and the required setup adds in discomfort. The tangible 3D system designed by Kim et al. [42] for the manipulation of 3D objects supports only five static gestures of hand. Reifinger et al. [43] presented their infrared based tracking system where infrared markers are used to track both static and dynamic hand gestures. The system is applicable in a wide lighting range but needs an array of six IR (Infrared) cameras besides a high-cost Head Mounted Display (HMD). Valuable research has been investigated about the use of Fiducial markers in VR and AR systems [44]. Gestures of the hand are traced by the FingARtips system of Buchmann et al. [45] for the manipulation of a virtual object. The unique Fiducial markers are used to distinguish different fingers. The use of full body tracking device; *Microsoft Kinect* has also been investigated for 3D interactions [46, 47]. Although Kinect is applicable for tracking body-gestures [48], it cannot differentiate individual fingers [49]. The technique of Jin et al. [50] detects the fingers movements with the help of the multi-sensors Leap Motion Controller (LMC). The system supports faster navigation with a satisfactory accuracy rate if used with HMD. However, due to the limited workspace of LMC [51], the accuracy of the system falls if gestures are posed outside a short range. Summarizing the challenges of gesture-based interactions, Benko [52] pointed out the key issue of false tracking [41]. This research work is an attempt to minimize false tracking by recognizing gestures of the two fingers with the help of an ordinary camera.

## III. GIFT: THE PROPOSED TECHNIQUE

This research work intends to present a gesture-based direct interaction [11] technique for interactions in a VE. All the interactions are performed on the basis of the dynamic positions of the two fingers.

Instead of following the computationally costlier methods of gesture recognition, the two finger-caps of colour green (for index) and pink (for thumb) are used. Lest a same-colour background object is detected, first a Region Of Interest (*ROI\_Img*) is extracted from dynamically scanned Frame-Image (*Fr\_Img*) on the basis of skin color. The *ROI\_Img* is then thresholded for green color to trace tip of index finger. The *Hue*, *Saturation* and *Values* (HSV) color space are used for the detection and tracking of the finger-caps. The *VH* representing the user's position in the digital world is activated at the detection of the finger-caps. From the 2D position of the fingertip-thimbles, the Central Point (*CP*) and the central areas for Index and Thumb ( $CA_{index}$  and  $CA_{Thumb}$ ) are computed from the first five dynamic images. On the basis of the *CP*,  $CA_{index}$  and the  $CA_{Thumb}$ , the *Selection-Manipulation* (SM) zone is defined. The *SM* is a limited area around a user's hand within which gestures for *selection*, *rotation*, *translation* and *scaling* can be posed with ease, see Fig. 1(a).

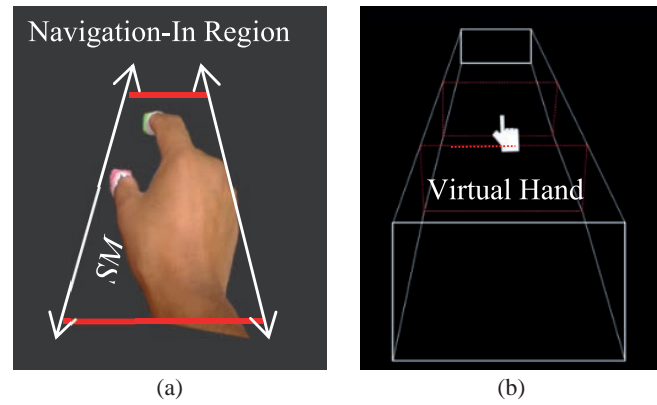


Fig. 1. (a) SM zone with Navigation regions (b) Perspective viewport of the VE.

The system works in three phases; *Exploration Phase* (EP), *Translation Phase* (TP) and *Manipulation Phase* (MP). Initially, the system starts with the EP and supports navigation and panning to explore a synthetic world. Translation and manipulation (scaling and rotation) are performed along an arbitrary axis in the TP and MP respectively. Switching between any two phases is activated by posing a distinct gesture. The three intuitive gestures of the two fingers; *Open-pinch*, *Closed-pinch* and *Cross-sign* are used to switch the system from one phase to another. Open-pinch is the posture where the two fingers are gently apart, see Fig. 2(a). With the posture of closed-pinch, tip of index touches the tip of index finger, see Fig. 2(b). Rotating the fingers along the look-at vector makes the cross-sign gesture, see Fig. 2(c).

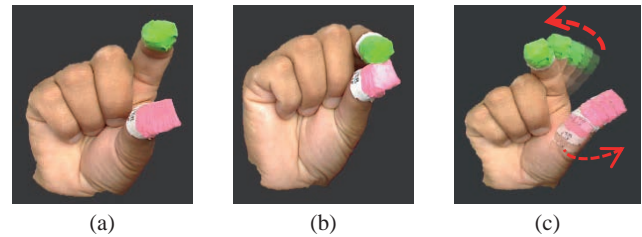


Fig. 2. The postures of (a) Open-pinch (b) Closed-pinch and (c) Cross-sign.

The open-pinch and closed-pinch gestures are used to activate MP and TP respectively. The cross-gesture is to switch the system back to the default EP. The object to be manipulated is inscribed inside two concentric cubes, as shown in Fig. 3. The outer cube inscribing the whole of an object specifies the Outer Aura (OA). The inner cube marking the central portion of an object represents the Inner Aura (IA). To translate an object, the IA selection of the object needs to be performed using the closed-pinch gesture. Similarly to activate manipulation of the object, the object should be selected for the OA using the open-pinch gesture. In order to visualize the switching between different phases, the appropriate cube around an object is highlighted.

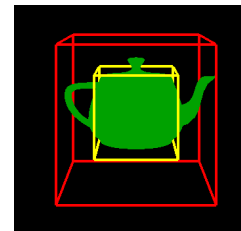


Fig. 3. Object inscribed inside two concentric cubes.

To explore a synthetic world, the system starts with the EP and supports only navigation and panning. Navigation and panning are performed outside the SM zone. Dynamic Areas (DAs) of the fingers are

compared with CAs for going inside into the VE (forward navigation) or for reverse navigation (backward navigation). Within the SM zone, translation and manipulation are performed based on the attained state of the system (TP or MP).

#### A. The Architecture of the Proposed Technique

Besides rendering a virtual scene at the front-end, the system extracts the *ROI\_Img* based at the back-end. The *ROI\_Img* is converted to HSV and is then thresholded for a broader range of green colour. At the detection of the thimble of the index finger, the *ROI\_Img* is split horizontally at the center of the cap of the index finger to get the Down-Section Image (*DS\_Img*), see Fig. 4(b).

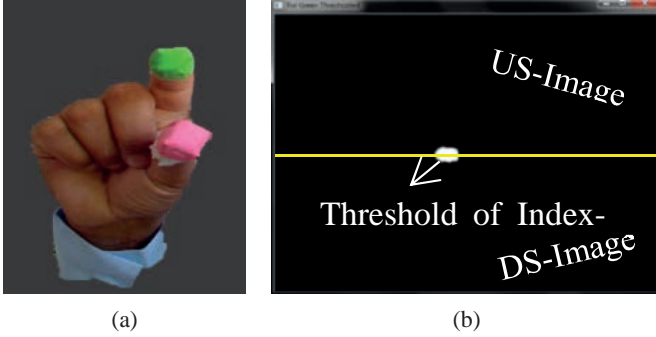


Fig. 4. (a) A hand pose in the *Fr\_Img* and (b) partition of the image into up and down sections with threshold of Index.

The *DS\_Img* is thresholded for pink color to find out the thumb-tip in the image. When both the fingers are traced, coordinates mapping between the fingers in image frame and the VH in the VE is initiated. As long as both the caps are visible, the VH can move freely with the movements of the fingers to access a 3D object. Navigation is performed by tracing a change in areas of the finger-caps while panning is performed on the basis of dynamic positions of the fingers.

If an object is selected with the open-pinch posture (OA selection), the MP is activated. Similarly, bringing the VH over an object and posing a closed-pinch (IA selection), the object is selected for translation (TP). User is informed about either of the states by highlighting the respective cube inscribing the object. Schematic of the entire system is shown in Fig. 5.

#### B. Image Segmentation

Image segmentation refers to the extraction of a set of pixels [53]. At the outset, *ROI\_Img* is extracted to avoid the chances of false detection if similar colors are present in the background. The *ROI\_Img* is supposed to be the most probable section of *FR\_Img* containing both of the fingers. As the *YCbCr* space is best to differentiate between skin and non-skin colors [54], therefore the *YCbCr* model is followed for skin color segmentation.

$$Fr\_Img\_YCbCr \begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 0.2(\frac{219}{255}) & 0.7(\frac{219}{255}) & 0.08(\frac{219}{255}) \\ -(\frac{0.212}{1.8})(\frac{224}{255}) & (\frac{224}{255}) - (\frac{0.7}{1.8})(\frac{224}{255}) & 0.5(\frac{224}{255}) \\ 0.5(\frac{224}{255}) & -(\frac{0.7}{1.5})(\frac{224}{255}) & -0.6(\frac{224}{255}) \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

After getting the binary image of the *FR\_Img*, see Fig. 6, the *ROI\_Img* with rows 'm' and columns 'n' is extracted from the *FR\_Img* using our designed algorithm [55] as,

$$ROI\_Img(m, n) = \left( \begin{array}{c} \bigcup_{r=D_m}^{Fr\_Img.Row(0)} (Fr\_Img) \\ \bigcup_{c=L_m}^{Fr\_Img.Column(R_m)} (Fr\_Img) \end{array} \right) \quad (2)$$

where  $L_m$ ,  $R_m$  and  $D_m$  represents *Left-most*, *Right-most* and *Down-most* skin pixels.

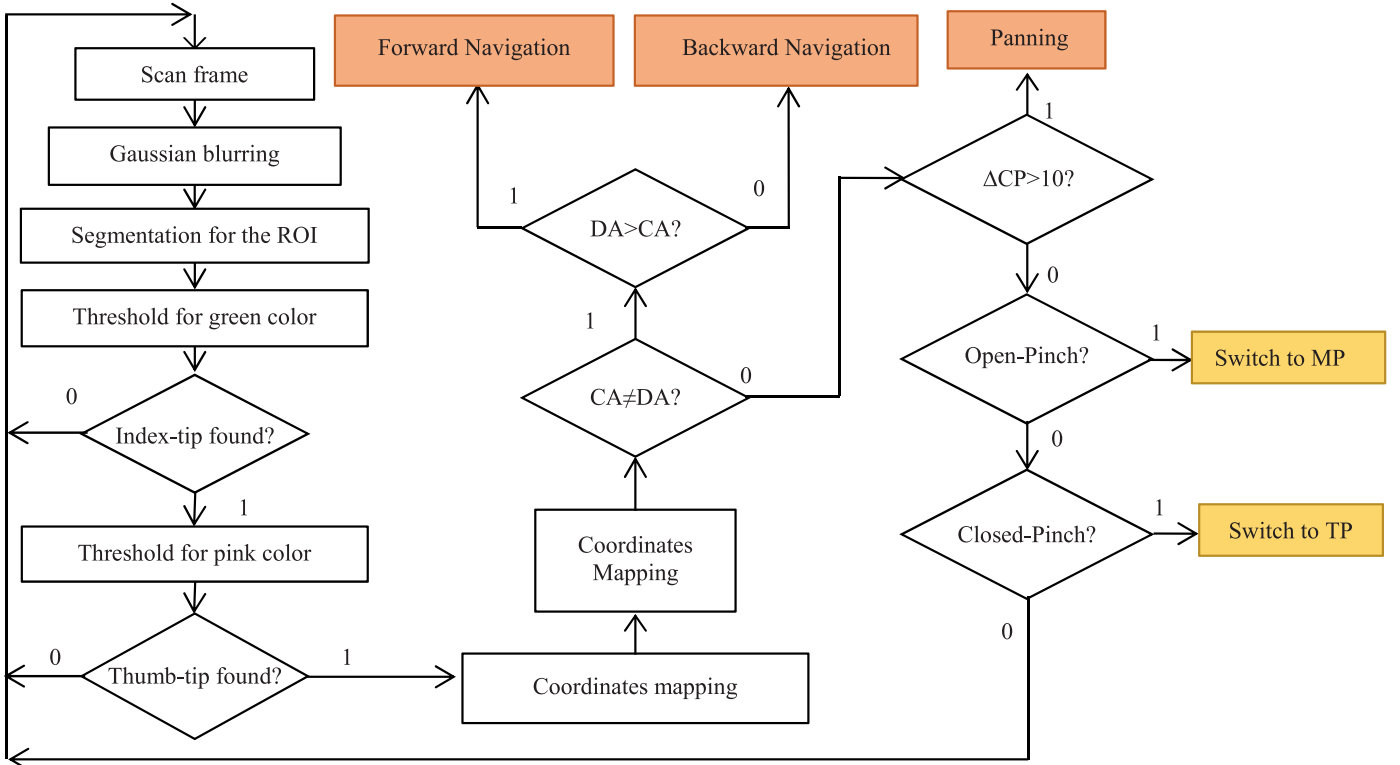


Fig. 5. Schematic of the proposed system.

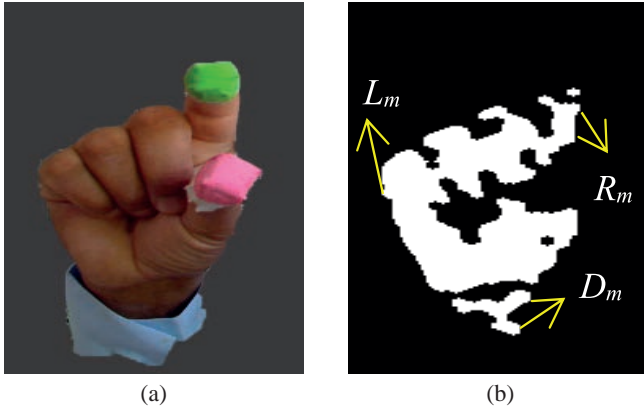


Fig. 6. The  $Fr\_Img$  in (a) RGB and (b) after conversion into YCbCr.

The segmented  $ROI\_Img$  is then thresholded for green and pink colors using the HSV color space.

$$ROI\_Img = \begin{cases} 1 & \text{if } 47 \leq ROI\_Img.H(x,y) \leq 94 \\ 1 & \text{if } 100 \leq ROI\_Img.S(x,y) \leq 187 \\ 1 & \text{if } 102 \leq ROI\_Img.V(x,y) \leq 255 \\ 0, & \text{Otherwise} \end{cases} \quad (3)$$

### C. Coordinates Mapping

A challenging task in the research was to harmonize the pixels representing the finger-caps with the coordinates of the VH. The coordinate system of OpenCV is different from the coordinate system of OpenGL. In OpenCV, a frame image starts with O(0,0) at top left. The origin O(0,0) of the OpenGL lies at the centre of a rendered frame, see Fig. 7.

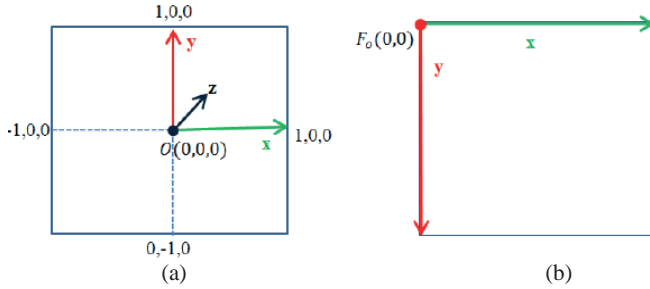


Fig. 7. The coordinates of (a) OpenGL and (b) OpenCV.

To harmonize the dissimilar coordinate systems, we devised our own mapping function,  $f$  [55]. The image frame is virtually split into four regions  $R_1$  to  $R_4$  as shown in Fig. 8. Mapping is made by the corresponding function taking 'x' and 'y' of a pixel of  $R_n$  as independent variables. The OpenCV pointer variable; CP is used to locate and move the virtual hand based on the position of both the fingers.

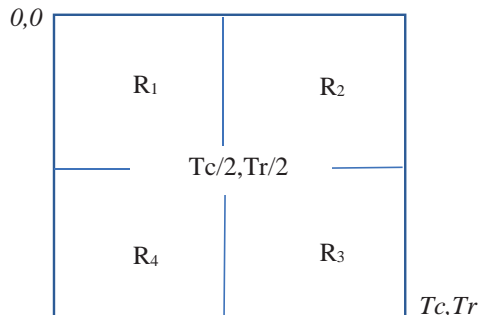


Fig. 8. The virtual division of an image frame.

If  $CP_i \in \mathbb{R}^2$  represents the CP in the initial frames and  $CP_d \in \mathbb{R}^2$  represents the dynamic position of the CP in any scanned frames, then the virtual hand's position;  $VHP: VHP \in \mathbb{R}^3$  is calculated as,

$$CP.x = (Index\_Pos.x + Thumb\_Pos.x)/2 \quad (4)$$

$$CP.y = (Index\_Pos.y + Thumb\_Pos.y)/2 \quad (5)$$

$$\Delta Px = (CPd.x - CPi.x) \quad (6)$$

$$\Delta Py = (CPd.y - CPi.y) \quad (7)$$

$$w(x; y) = ((CPx/Tc)k, (Cpy/Tr)k) \quad (8)$$

$$VHP(x, y) = w(CPf, CPd) \quad (9)$$

Where  $k$  is the speed constant; greater the value of  $k$  speedier will be the movements of the VH. The value of 'Tc' and 'Tr' represents the total number of columns and rows respectively. We kept a moderate value of  $k$  in the IGF project. Furthermore, to ignore the unintentional movements of the fingers the mapping is performed only when the value of  $\Delta Px$  or  $\Delta Py$  is greater than five pixels. The mapping process is shown in Fig. 9.

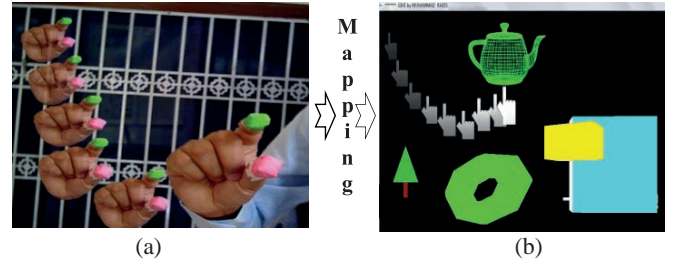


Fig. 9. The mapping between (a) the fingers move in image frame to (b) virtual hand in VE.

## IV. INTERACTIONS SUPPORT

All the basic 3D interactions are performed by the intuitive gestures of the fingers. The 2D positions of the fingers traced in the initial frames are supposed to be in a moderate range from the camera. Hence, the zone; SM is set as soon as the fingers' tips are recognized by the system.

### A. Navigation

Navigation is to explore a VE. Mostly, a user needs to navigate to a proper location before performing selection or manipulation. In the proposed technique, forward or backward movement of the hand outside the SM zone performs navigation. As plausible, the forward navigation is carried out by the forward movement of the hand (see Fig. 10). Similarly, backward navigation is performed by the backward movement of hand outside the SM zone.

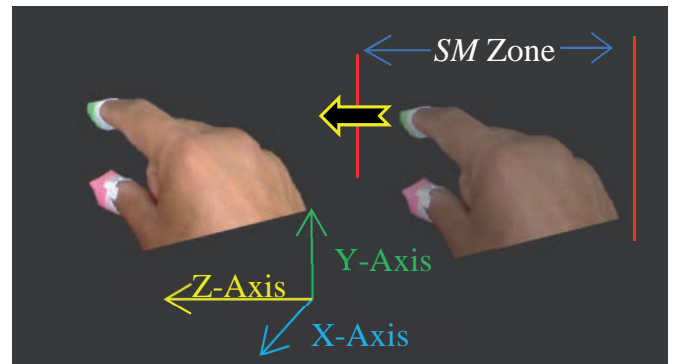


Fig. 10. The forward movement of hand for forward-navigation.



To deduce the forward or backward hand movements in a scanned 2D image, the DAs (Dynamic Areas) of the fingers-caps are calculated and are compared with the CAs at runtime. By crossing the forward limit of the SM in the z-axis, DAs increase, see Fig. 11(b).

$$DA_x = (y)(CA_x) \quad (10)$$

where,  $x = \{\text{index, thumb}\}$  and  $y > 1$ .

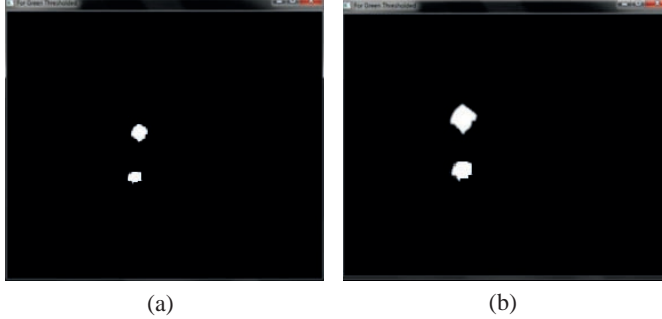


Fig. 11. (a) The CAs in initial image frame and (b) the DAs after forward hand movement.

To avoid the possibility of unintended increase/decrease in the DAs, the Fingers Average Dynamic (FAD) positions along the x-axis (FAD.x) and y-axis (FAD.y) are also checked against CP.x and CP.y, as clear from the following pseudo-code.

```

if (DAIndex > CAIndex AND DAThumb > CAThumb) AND
    (FAD.x ≤ CP.x + 10 AND FAD.x ≥ CP.x - 10)
    AND (FAD.y ≤ CP.y + 10 AND FAD.y ≥ CP.y - 10)
    Forward Navigation

if (DAIndex < CAIndex AND DAThumb < CAThumb) AND
    (FAD.x ≤ CP.x + 10 AND FAD.x ≥ CP.x - 10)
    AND (FAD.y ≤ CP.y + 10 AND FAD.y ≥ CP.y - 10)
    Backward Navigation

```

## B. Selection DeSelection

Selection is choosing an object for interaction. Selection for scaling and rotation is made by bringing the VH over the object and posing an open-pinch gesture. With this, the OA selection is performed whereas the outer cube inscribing the object is highlighted. Similarly, an object is selected for translation by posing a closed-pinch to perform the IA selection. By posing a cross gesture by the two fingers deselects the object. The system recognizes the gesture when the fingers exceed a specified threshold along x-axis from the horizontal central positions of the fingers. If *ID* and *TD* represent Index-Dynamic and Thumb-Dynamic positions then, pseudo code for the DeSelection is given as,

```

If (ID.x > CP.x + 30 AND TD.x + 30 < CP.x)
    AND (ABS(ID.y - CP.y) > 10 AND (ABS(TD.y - CP.y) > 10)
    AND ABS(DAIndex - CAIndex) < 20
    AND ABS(DAThumb - CAThumb) < 20
    DeSelect

```

Once an object is deselected, the system switches back to the default EP.

## C. Translation

Translation is changing the position of a virtual object along x, y and/or z-axis in a VE. Like gripping an object in the real world, the closed-pinch gesture is used to select an object for translation. To perform translation of an object, the system needs to be in the TP. The transition from the EP to the TP is made by hovering the VH over an

object and posing the pinch-gesture. Besides switching from the EP to TP, the IA selection of the object is performed. After the selection, translation is performed in the VE by the movements of the hand.

The following condition is checked to ensure the closed-pinch posture.

```

if (ABS(DAIndex - CAIndex) < 20
    AND (ABS(DAThumb - CAThumb) < 20
    AND (ABS(ID.x - TD.x) < 10)
    AND (ABS(ID.y - TD.y) < 10)
    Select for Translation

```

Unless the fingers' tips are made apart, the IA selection of the object remains intact. During translations, the dynamic coordinates of the VH are assigned to the selected object to translate the object by the free movement of hand accordingly. After the IA selection of an object *Obj*, the 3D coordinates of the *Obj* are obtained as,

$$\begin{pmatrix} Obj.x \\ Obj.y \\ Obj.z \end{pmatrix} = \begin{pmatrix} VHP.x \\ VHP.y \\ VHP.z \end{pmatrix} \quad (11)$$

## D. Rotation

Within the SM zone, a selected object is rotated using the ChordBall technique [56]. As perceivable, moving the VH along the x-axis with the open-pinch rotates the object along the y-axis. Similarly, movement of the VH along y-axis rotates the object along the x-axis. The angle  $\theta$  of rotation is calculated using the distance ( $d$ ) of the fingers from the central position (*CP*). Greater the distance, larger will be the angle of rotation.

$$d = \sqrt{(CP.x - FAD.x)^2 + (CP.y - FAD.y)^2} \quad (12)$$

$$\theta = d / (Tc/2) \quad (13)$$

The following algorithm is followed to perform rotation about an axis,

```

if (ABS(FAD.y - CP.y) < 10)
    AND (ABS(DAIndex - CAIndex) < 20)
    AND (ABS(FAD.x - CP.x) > 20)
    Rotation about y-axis

if (ABS(FAD.x - CP.x) < 10)
    AND (ABS(DAIndex - CAIndex) < 20)
    AND (ABS(DAThumb - CAThumb) < 20)
    AND (ABS(FAD.y - CP.y) > 20)
    Rotation about x-axis

if (ABS(FAD.x - CP.x) > 20)
    AND (ABS(DAIndex - CAIndex) > 20)
    AND (ABS(DAThumb - CAThumb) < 20)
    AND (ABS(FAD.y - CP.y) > 20)
    Rotation about z-axis

```

## E. Scaling

After the OA selecting with the open-pinch, an object can be scaled up or down. With the activation of the MP, the initial distance between the two fingers is traced. In the MP, a selected object is scaled or scale-down by comparing the dynamic distances with the initial distance between the fingers. A gentle increase in the distance between the thumb and index fingers scales up the object and vice versa. If the gesture is parallel to the x-axis then up or downscaling along the x-axis is performed. Scaling or downscaling along the y-axis is performed if the gesture is parallel to the y-axis as shown in Fig. 12(a) and Fig. 12(b).

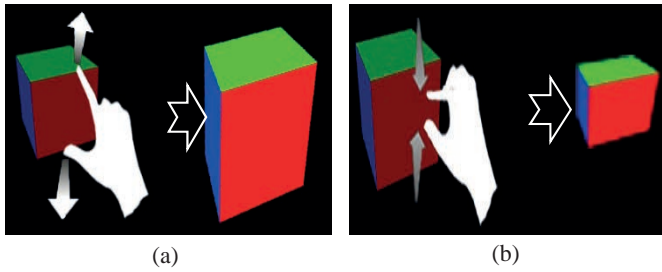


Fig. 12. (a) Gesture for scaling along y-axis and (b) Gesture for downscaling along y-axis.

Scaling or scale-down about the z-axis is enacted by increasing or decreasing the diagonal distance respectively, see Fig. 13.

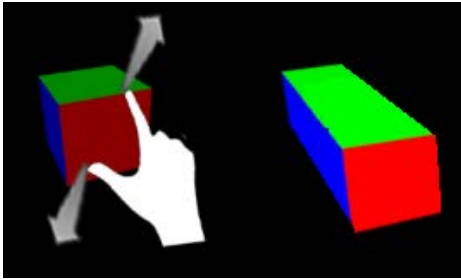


Fig. 13. The scaling along the z-axis.

## V. IMPLEMENTATION AND EVALUATION OF THE TECHNIQUE

To systematically evaluate the proposed approach, the technique is implemented in the case-study application; IGF (Interaction by the Gestures of Fingers). The two open libraries; OpenGL and OpenCV are used for the front-end rendering and for the back-end image processing respectively. A Corei3 laptop with 2.30 GHz processor and 4GB RAM was used for the implementation and evaluation. The built-in camera of the laptop (resolution 640x480) was used to capture the image streams at runtime. Fifteen participants, all male, of ages between 25 and 40 (mean=31, SD=6) performed the six predefined tasks. Each participant performed two trials of the tasks. Before starting the evaluation session, participants of the evaluation were introduced to the system and pre-trials were performed by the users.

In the IGF application, the VH moves freely in the environment where the coordinates of virtual camera change with the fingers movement accordingly. The z-axis of the VH is intentionally kept constant so that to be visible for interaction everywhere in the VE. Activation of the entire system is conditioned to the visibility of the fingers' caps. The users are constantly informed by the text "DETECTED" displayed in the upper centre part of the scene. Similarly, a user is notified both with text and a beep-signal (audio) when an interaction is activated, see Fig. 14.

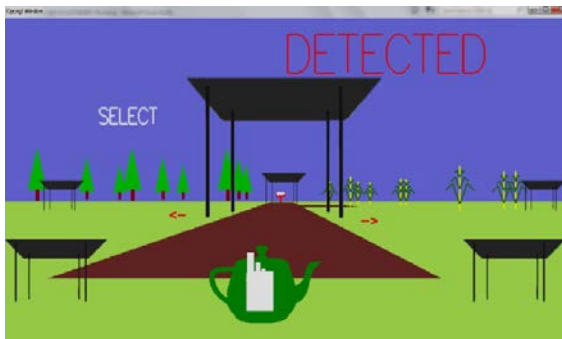


Fig. 14. The virtual scene for evaluation of the technique.

## A. Testing Environment and the Experimental Tasks

The 3D environment designed for the evaluation of the technique contains six tables; two on either side and two in the middle. The VH made of cubes represents the user's position in the environment. For easy noticing, endpoint of the scene is marked by a board with the text "Stop". To enhance immersivity, the scene is designed with different 3D objects at different positions. By hitting the *Enter* key, the system restarts with a new trial/task.

Participants were asked to perform the following six tasks in the designed 3D environment. The tasks cover the basic interaction operations, i.e. *Selection*, *DeSelection*, *Translation*, *Scaling*, *Rotation* and *Navigation*. A 3D object (Teapot) is rendered in the mid of the scene for selection and manipulation.

**Task-1:** Select the object, translate it to the far left table and then deselect it.

**Task-2:** Select the object, translate it to the far right table and then deselect it.

**Task-4:** Navigate to the endpoint marked by the stop-board and then navigate back to the starting point.

**Task-5:** Select the object, scale it up/down along x-axis/y-axis and then deselect it.

**Task-3:** Select the object, scale it up/down along the z-axis and then deselect it.

**Task-6:** Select the object, rotate it along the x-axis, y-axis and z-axis and then deselect it.

While performing these tasks, selection and deslection are evaluated five times, rotation three times while translation, navigation and scaling two times each in a single trial. Missed or false detection of the system, after posing the specified gestures, were counted as errors. With this setup, overall accuracy rate for all the interaction tasks, as shown in Table I, is 94.3%. Comparatively more errors occurred during translation and navigation which were due to quicker move of hand or both of the fingers at the same time. In such cases, the camera misses tips of the fingers. This implies that the system performance can be improved with a high-efficiency camera.

TABLE I. STATISTICS OF THE INTERACTION TASKS

Interaction	Correct	Total	%age
Selection	146	150	97.3
DeSelection	57	60	95
Translation	55	60	91.6
Scaling	86	90	95.5
Rotation	56	60	93.3
Navigation	140	150	93.3
<b>Total</b>	<b>540</b>	<b>570</b>	<b>94.3</b>

## B. Learning Effect

The learning effect was assessed from the errors occurrence rate. The paired two-sample T-test was used to analyze differences in means of the two trails. With the null hypothesis ( $H_0$ ), we assumed that the mean difference ( $\mu_d$ ) is 0. The hypothesis was rejected as there was a significant difference between the outcomes of Trail-1 (M=92.8, SD=1.7) and Trail-2 (M=96, SD=1.29) conditions; (t(5)=-3.23, p=0.0091). The graph shown in Fig. 15 and Fig. 16 indicate the vivid decrease in errors.

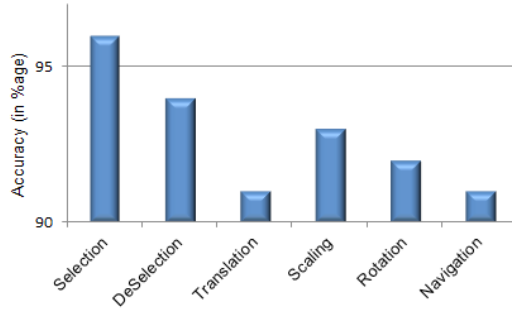


Fig. 15. Accuracy of the interactions performed in trial-1.

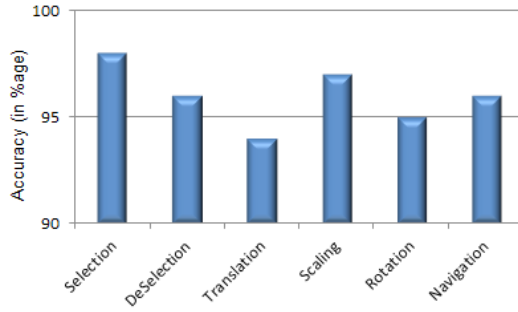


Fig. 16. Accuracy of the interactions performed in trial-2.

### C. Subjective Analysis

To analyze the response of the participants, a questionnaire was presented to the users at the end of the evaluation session. The questionnaire was to measure the three factors; *Ease of use*, *Fatigue* and *Suitability* of the technique in VEs. The post-assessment questionnaire is shown in Table II.

TABLE II. THE POST-ASSESSMENT QUESTIONNAIRE ABOUT THE FOUR FACTORS

Question No.	Statement	Your response (tick one)?				
		Strongly agree	Agree	Indifferent	Disagree	Strongly disagree
1	As a whole, the system was easy to use.					
2	No fatigue, stress or strain was felt during or after performing the interaction tasks					
3	The technique is suitable for VR/AR applications.					

The percentage of users' response to the three factors is shown in Fig. 17.

## VI. COMPARATIVE STUDY

Following the interaction standards presented by various researchers [57-59], the proposed approach is systematically compared with the recent state-of-the-art interaction techniques. Based on the available information (acquired from the literature), a score ('1') is assigned to a technique if a particular interaction rule ( $R_i$ ) is followed. The final score is calculated as,

$$Score = \sum_{i=1}^N R_i / N \quad (14)$$

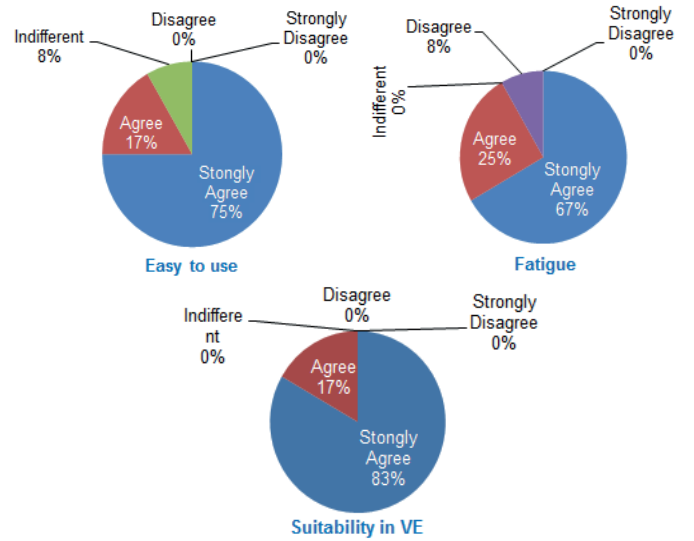


Fig. 17. Response of the participants (in %) about the three factors.

Where  $N$  is the total number of the standard rules of 3D interactions (see Table III).

TABLE III. THE BASIC FIVE STANDARDS OF INTERACTIONS

Standards	Description
$R_1$	Interactive support
$R_2$	Reliable sensory feedback
$R_3$	Simple and inexpensive
$R_4$	Useable while sitting
$R_5$	Wireless connection

A total of eight state-of-the-art interaction techniques (including the proposed approach) were selected for comparative analysis. Details of the techniques with the *possible challenge(s)* are presented in Table IV. After extracting information from the research works about the five standards, the final score was computed, see Table V.

TABLE V. EVALUATION OF THE TECHNIQUES BASED ON THE FIVE RULES

Technique No.	$R_1$	$R_2$	$R_3$	$R_4$	$R_5$	Score
1	1	1	0	1	0	0.6
2	1	1	0	1	0	0.6
3	1	1	0	1	0	0.6
4	1	1	0	1	0	0.6
5	1	1	0	1	0	0.6
6	1	1	1	1	0	0.8
7	1	1	0	1	0	0.6
8	1	1	1	1	1	1

## VII. APPLICABILITY IN GAMES

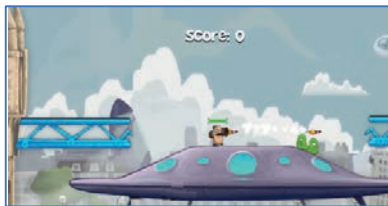
The game technology is shifting from 2D graphics to realistic 3D VR. Recent research works have proved that gesture-based games are suitable for games [67] [68]. The free hand and/or fingers gestures have promising potential not only in education, training and medical application [69-71] but also in the contemporary VR games [67]. It has been proved that the use of gestures improves overall engagement of the gamers [72] [73]. To make VR games interesting and intuitive, a number of gesture-based techniques have been proposed. The technique proposed by [67] and [68] are based on Leap motion and Nintendo Wii controllers. However such systems suffer from the limited work space of the controllers and training of gestures [68].

The proposed technique is applicable in the games where the gamer

TABLE IV. THE DETAILS OF THE CONTEMPORARY INTERACTION TECHNIQUES

Tech. No.	Author(s)	Year	Tracking device used	Title	Possible challenge(s)
1	Linn, Andreas [60]	2017	HTC Vive with Tobii Eye-tracker	Gaze Teleportation in Virtual Reality	Highly Sensitive to calibration errors
2	Thammatthip Piumsomboon [61]	2017	HMD with eye-tracker	Exploring Natural Eye-Gaze-Based Interaction for Immersive Virtual Reality	Support only selection
3	Mohamed Khamis et al. [62]	2018	HTC Vive eye-tracker	VRPursuits: Interaction in Virtual Reality using Smooth Pursuit Eye Movements	Selection while walking is erroneous and time-consuming
4	Stellmach et al. [63]	2012	Tobii T60 eye-tracker	Designing Gaze-based User Interfaces for Steering in Virtual Environments	Continuous gazing at destination point may lead to asthenopia
5	Wen-jun Hou et al. [64]	2018	HTC Vive	User Defined Eye Movement-Based Interaction for Virtual Reality	Multiple tasks by the same gestures may lead to fuzzy input
6	G. Prabhakar and P. Biswas [65]	2018	Eye-gaze trackers	Eye Gaze Controlled Projected Display in Automotive and Military Aviation Environments	Supports only pointing and selection
7	Qi Sun et al. [66]	2018	HMD with SMI gaze tracker	Towards Virtual Reality In_nite Walking: Dynamic Saccadic Redirection	Suffers from tracking latency and redirection during blinking.
8	M. Raees and S. Ullah	2018	Ordinary camera	The proposed technique	Lighting is required to trace the finger-tips

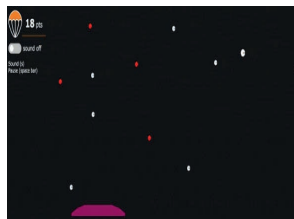
needs to select and/or controls actions of an avatar/character in the game environment. Unlike other gesture-based systems [74], the proposed technique neither keeps feature sets [75-77] nor needs searching and classification [78-80], hence ensure speed in playing games. To properly assess applicability of the proposed technique in the contemporary VR games we selected the three well-liked games; see Fig. 18.



(a)



(b)



(c)

Fig. 18. The scenes of (a) Alien invasion (b) Play breads puzzle and (c) Skyfall games.

Skyfall is a simple game created using the physics engine; planck.js [81]. There are three types of balls in the game that fall from the top of the screen in a random order. White and green balls have worth +10 points while red balls are of -10 points. Players need to earn more

points by moving a paddle to catch the good balls (white and green balls). However, at the same time a gamer should avoid bad balls (red balls). The GIFT technique is applicable to control the paddle by the movements of the fingers.

Similarly, the Beads Puzzle [82] is a ball-shooter game where a gamer has to fire a ball at a time. Direction of the cannon is controlled by the mouse movements. A fire is made by the mouse down event. The game can be played with the proposed technique more easily than with a mouse.

Unity's Alien Invasion [83] is another popular game where a player has to defend a city from aliens. The player needs to control a character that is able to collect bombs and fire. With the proposed technique, a gamer can more efficiently control actions of the character. One suitable configuration of the game with the technique is 1) to control movements of the character with the dynamic positions of the fingers. 2) To use the *Pinch-gesture* for collecting the bombs and 3) to use the *Cross-gesture* for firing the bullets. An additional research is required to assess applicability of the technique in HMD based VR games.

## VIII. CONCLUSION AND FUTURE WORK

To cope with the pace of VR developments in various fields, simple and natural interfaces are in need. With this contribution, we proposed a novel gesture-based interaction technique which needs no expensive device other than an ordinary camera and pieces of paper. Simple and intuitive gestures are used to ensure naturalism while performing the basic interaction; navigation, selection, scaling, rotation and translation. Experimental results show that the proposed approach has reliable recognition and accuracy rates. The system neither needs training of images nor use any feature extraction, hence guarantying fast preprocessing. The proposed system is feasible and well suitable



in a wide spectrum of HCI including virtual prototyping, 3D gaming, robotics, industrial architecture and simulation. The work also covers the smooth integration of image processing and VE and can be used in the designing of a realistic VR application.

We believe that the technique is suitable for the basic 3D interactions, however the technique suffers from illumination variation [15]. Moreover, the technique is not applicable for the two-player shooter games. An additional research is required to reduce dependency of the system on the lighting condition and to make the technique suitable for the collaborative VE. In future, we are determined to enhance the technique for the emerging augmented and mixed VR setups.

## REFERENCES

- [1] Uddin, M. Taufeeq, and M. A. Uddiny, "Human activity recognition from wearable sensors using extremely randomized trees." In International Conference on Electrical Engineering and Information Communication Technology (ICEEICT), pp. 1-6, (2015).
- [2] J. Ahmad, N. Sarif, J. T. Kim, and T.S. Kim, "Human activity recognition via recognized body parts of human depth silhouettes for residents monitoring services at smart home." *Indoor and built environment*, 22, no. 1, pp. 271-279, (2013).
- [3] Zhan, Yi, and T. Kuroda, "Wearable sensor-based human activity recognition from environmental background sounds," *Journal of Ambient Intelligence and Humanized Computing* 5, no. 1, pp. 77-89, (2014).
- [4] J. Ahmad, "Security architecture for third generation (3g) using gmhs cellular network," In International Conference on Emerging Technologies, pp. 74-79, (2007).
- [5] J. Ahmad, and M. A. Zeb, "Security enhancement for e-learning portal," *International Journal of Computer Science and Network Security*, 8, no. 3, pp. 41-45, (2008).
- [6] J. Ahmad and Y. A. Rasheed, "Collaboration achievement along with performance maintenance in video streaming," In Proceedings of the IEEE Conference on Interactive Computer Aided Learning, Villach, Austria, vol. 2628, p. 18, (2007).
- [7] J. Ahmad, and A. Shahzad, "Multiple facial feature detection using vertex-modeling structure," In Proceedings of the IEEE Conference on Interactive Computer Aided Learning, Villach, Austria, vol. 2628, (2007).
- [8] J. Ahmad, S. Kim, and B. J. Yun, "Assembled algorithm in the real-time H. 263 codec for advanced performance." In Proceedings of 7th International Workshop on Enterprise networking and Computing in Healthcare Industry, pp. 295-298, (2005).
- [9] J. Ahmad, and S. Kim, "Advanced performance achievement using multi-algorithmic approach of video transcoder for low bit rate wireless communication," *ICGST International Journal on Graphics, Vision and Image Processing* 5, no. 9, pp. 27-32, (2005).
- [10] J. Ahmad and S. Kim, "Algorithmic Implementation and Efficiency Maintenance of Real-Time Environment using Low-Bitrate Wireless Communication," In Proc. of the 4th IEEE Workshop on Software Technologies for Future Embedded and Ubiquitous Systems, and the Second International Workshop on Collaborative Computing, Integration, and Assurance (SEUS-WCCIA'06), pp. 81-88, (2006).
- [11] S. S. Rautaray, "Real Time Hand Gesture Recognition System for Dynamic Applications," *Int. J. UbiComp*, vol. 3, no. 1, pp. 21-31, (2012).
- [12] Sahane, M. S., H. D. Salve, N. D. Dhawade, and S. A. Bajpai, "Visual Interpretation Of Hand Gestures For Human Computer Interaction," *environments (VEs)*, 2, p.53, (2014).
- [13] R. Itkarkar and A. Nandy, "A Study of Vision Based Hand Gesture Recognition for Human Machine Interaction," vol. 1, no. 12, pp. 48-52, (2014).
- [14] Pirker, Johanna, M. Pojer, A. Holzinger and C. Gütl, "Gesture-based interactions in video games with the leap motion controller," In International Conference on Human-Computer Interaction, pp. 620-633, (2017).
- [15] Kim, Kwangtaek, J. Kim, J. Choi, J. Kim and S. Lee, "Depth camera-based 3D hand gesture controls with immersive tactile feedback for natural mid-air gesture interactions," *Sensors* 15, no. 1, pp. 1022-1046, (2015).
- [16] Koller, Dieter, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan, "Real-time vision-based camera tracking for augmented reality applications," In Proceedings of the ACM symposium on Virtual reality software and technology, pp. 87-94, (1997).
- [17] J. Ahmad, Y. Kim, and D. Kim, "Ridge body parts features for human pose estimation and recognition from RGB-D video data," In International Conference on Computing, Communication and Networking Technologies (ICCCNT), pp. 1-6, (2014).
- [18] J. Ahmad, J. T. Kim, and T-S Kim, "Development of a life logging system via depth imaging-based human activity recognition for smart homes," In Proceedings of the International Symposium on Sustainable Healthy Buildings, Seoul, Korea, vol. 19, (2012).
- [19] J. Ahmad, M. Z. Uddin, and T-S. Kim, "Depth video-based human activity recognition system using translation and scaling invariant features for life logging at smart home," *IEEE Transactions on Consumer Electronics* 58, no. 3, pp. 863-871, (2012).
- [20] Kamal, Shaharyar, J. Ahmad and D. Kim, "Depth images-based human detection, tracking and activity recognition using spatiotemporal features and modified HMM," *J. Electr. Eng. Technol* 11, no. 3, pp. 1921-1926, (2016).
- [21] J. Ahmad, S. Lee, J. T. Kim and T-S. Kim, "Human activity recognition via the features of labeled depth body parts," In International Conference on Smart Homes and Health Telematics, pp. 246-249, (2012).
- [22] J. Ahmad, and S. Kamal, "Real-time life logging via a depth silhouette-based human activity recognition system for smart home services," In International Conference on Advanced Video and Signal Based Surveillance (AVSS), pp. 74-80, (2014).
- [23] Huang, Qiao, J. Yang, and Y. Qiao, "Person re-identification across multi-camera system based on local descriptors," In Sixth International Conference on Distributed Smart Cameras (ICDSC), pp. 1-6, (2012).
- [24] J. Ahmad and S. Kim, "Global security using human face understanding under vision ubiquitous architecture system," *World academy of science, engineering, and technology* 13, pp. 7-11, (2006).
- [25] Farooq, Faisal, J. Ahmed and L. Zheng, "Facial expression recognition using hybrid features and self-organizing maps," In International Conference on Multimedia and Expo (ICME), pp. 409-414, (2017).
- [26] Farooq, Adnan, J. Ahmad and S. Kamal, "Dense RGB-D map-based human tracking and activity recognition using skin joints features and self-organizing map," *KSII Transactions on Internet and Information Systems (TIIS)* 9, no. 5, pp.1856-1869, (2015).
- [27] Kamal, Shaharyar, and J. Ahmad, "A hybrid feature extraction approach for human detection, tracking and activity recognition using depth sensors," *Arabian Journal for science and engineering* 41, no. 3, pp.1043-1051, (2016).
- [28] J. Ahmad, S. Kamal and D. Kim, "A Depth Video-based Human Detection and Activity Recognition using Multi-features and Embedded Hidden Markov Models for Health Care Monitoring Systems," *International Journal of Interactive Multimedia & Artificial Intelligence* 4, no. 4, (2017).
- [29] Yoshimoto, Hiromasa, N. Date and S. Yonemoto, "Vision-based real-time motion capture system using multiple cameras," In International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp. 247-251, (2003).
- [30] J. Ahmad, S. Kamal and D-S. Kim, "Detecting Complex 3D Human Motions with Body Model Low-Rank Representation for Real-Time Smart Activity Monitoring System," *KSII Transactions on Internet & Information Systems* 12, no. 3, (2018).
- [31] A. Shamaie and A. Sutherland, "Accurate recognition of large number of hand gestures," *Proc Iran. Conf. Mach. Vis. Image Process. Univ. Technol. Tehran*, (2003).
- [32] X. Yan and N. Aimaiti, "Gesture-based interaction and implication for the future," vol. 11, no. 1, p. 25, (2011).
- [33] K. Katsuragawa, A. Kamal, and E. Lank, "Effect of Motion-Gesture Recognizer Error Pattern on User Workload and Behavior," *Proc. 22nd Int. Conf. Intell. User Interfaces - IUI '17*, pp. 439-449, (2017).
- [34] P. H. An. "The challenge of hand gesture interaction in the Virtual Reality Environment: evaluation of in-air hand gesture using the Leap Motion Controller," <http://urn.fi/URN:NBN:fi:amk-2018091215089>, (Accessed on 4 Dec. 2018).
- [35] Yuan, Xiaobu, and J. Lu. "A Vision-Based Approach of Bare-Hand Interface Design in Virtual Assembly," In *Information Technology For Balanced Manufacturing Systems*, pp. 281-290, (2006).
- [36] P. Xu, "A Real-time Hand Gesture Recognition and Human-Computer

- Interaction System,” pp. 1–8, (2017).
- [37] C. Grönegress, “Designing Intuitive Interfaces for Virtual Environments,” *Simulation*, no. August, pp. 1–128, (2001).
- [38] K. S. Hale and K. M. Stanney, *Handbook of Virtual Environments: Design, Implementation, and Applications*. (2015).
- [39] Kiyokawa, Kiyoshi, H. Takemura, Y. Katayama, H. Iwasa, and N. Yokoya. “Vlego: A simple two-handed modeling environment based on toy blocks,” *Proc. Of ACM VRST*, 96, (1996).
- [40] J. Hua and H. Qin, “Haptic sculpting of volumetric implicit functions,” *Proc. - Pacific Conf. Comput. Graph. Appl.*, pp. 254–264, (2001).
- [41] Lee, Chan-Su, J. D. Choi, K. M. Oh, and C.J. Park. “Hand Interface for Immersive Virtual Environment Authoring System,” *In Proc. of the International Conference on Virtual Systems and MultiMedia*, pp. 361–366, (1999).
- [42] H. Kim, G. Albuquerque, S. Havemann, and D. D. W. Fellner, “Tangible 3D: Hand Gesture Interaction for Immersive 3D Modeling,” *Proc. 11th Eurographics Conf. Virtual Environ.*, pp. 191–199, (2005).
- [43] S. Reifinger, F. Wallhoff, M. Ablassmeier, T. Poitschke, and G. Rigoll, “Static and dynamic hand-gesture recognition for augmented reality applications,” *Human-Computer Interact. HCI Intell. Multimodal Interact. Environ. HCI 2007. Lect. Notes Comput. Sci.*, vol. 4552, pp. 728–737, (2007).
- [44] Rabbi, Ihsan, S. Ullah, and D. Khan. “Automatic generation of layered marker for long range augmented reality applications,” *Kuwait Journal of Science* 44, no. 3, (2017).
- [45] Buchmann, Volkert, S. Violich, M. Billinghurst, and A. Cockburn. “FingARtips: gesture based direct manipulation in Augmented Reality,” *In Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pp. 212–221, (2004).
- [46] Song, Peng, W. B. Goh, W. Hutama, C. W. Fu, and X. Liu. “A handle bar metaphor for virtual object manipulation with mid-air interaction,” *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 1297–1306, (2012).
- [47] S. Oprisescu and E. Barth, “3D hand gesture recognition using the hough transform,” *Adv. Electr. Comput. Eng.*, vol. 13, no. 3, pp. 71–76, (2013).
- [48] F. Weichert, D. Bachmann, B. Rudak, and D. Fisseler, “Analysis of the accuracy and robustness of the Leap Motion Controller,” *Sensors (Switzerland)*, vol. 13, no. 5, pp. 6380–6393, (2013).
- [49] M. Nabiyouni, B. Laha, and D. A. Bowman, “Poster: Designing effective travel techniques with bare-hand interaction,” *IEEE Symp. 3D User Interfaces 2014, 3DUI 2014 - Proc.*, no. March 2015, pp. 139–140, 2014.
- [50] H. Jin, Q. Chen, Z. Chen, Y. Hu, and J. Zhang, “Multi-LeapMotion sensor based demonstration for robotic refine tabletop object manipulation task,” *CAAI Trans. Intell. Technol.*, vol. 1, no. 1, pp. 104–113, (2016).
- [51] Benko, Hrvoje. “Beyond flat surface computing: challenges of depth-aware and curved interfaces,” *In Proceedings of the 17th ACM international conference on Multimedia*, pp. 935–944, (2009).
- [52] A. V. Devadoss, R. Malaishamy, and M. Subramaniam, “Performance improvement using an automation system for recognition of multiple parametric features based on human footprint,” *Kuwait Journal of Science* 42, no. 1 (2015).
- [53] Phung, S. Lam, A. Bouzerdoun, and D. Chai “A novel skin color model in ycbcr color space and its application to human face detection,” *In Proceedings of International Conference on Image Processing*, vol. 1, pp. I-I, (2002).
- [54] M. Raees, S. Ullah, S. U. Rahman, and I. Rabbi “Image based recognition of Pakistan sign language,” *Journal of Engineering Research* , 4, no. 1 (2016).
- [55] M. Raees, S. Ullah, and S. U. Rahman. “VEN-3DVE: vision based egocentric navigation for 3D virtual environments,” *International Journal on Interactive Design and Manufacturing (IJIDeM)*, pp. 1–11, (2018).
- [56] M. Raees, S. Ullah, and S. U. Rahman. “CHORDBALL: A ROTATION TECHNIQUE FOR 3D VIRTUAL ENVIRONMENTS,” *Pakistan Journal of Science*, 69, no. 1, p. 85, (2017).
- [57] Alqahtani, A. Saeed, L. F. Daghestani, and L. F. Ibrahim. “Environments and System Types of Virtual Reality Technology in STEM: a Survey,” *International Journal of Advanced Computer Science and Applications (IJACSA)* 8, no. 6 (2017).
- [58] Zielasko, Daniel, S. Horn, S. Freitag, B. Weyers, and T. W. Kuhlen. “Evaluation of hands-free HMD-based navigation techniques for immersive data analysis.” *In 3D User Interfaces (3DUI)*, 2016 IEEE Symposium on, pp. 113–119, (2016).
- [59] Tapu, Ruxandra, B. Mocanu, and E. Tapu. “A survey on wearable devices used to assist the visual impaired user navigation in outdoor environments.” *In Electronics and Telecommunications (ISETC)*, 2014 11th International Symposium on, pp. 1–4, (2014).
- [60] Linn, Andreas. “Gaze Teleportation in Virtual Reality,” (2017).
- [61] Piumsomboon, Thammathip, G. Lee, R. W. Lindeman, and M. Billinghurst. “Exploring natural eye-gaze-based interaction for immersive virtual reality,” *In Symposium on 3D User Interfaces (3DUI)*, pp. 36–39, (2017).
- [62] Khamis, Mohamed, C. Oechsner, F. Alt, and A. Bulling. “VRpursuits: interaction in virtual reality using smooth pursuit eye movements,” *In Proceedings of the International Conference on Advanced Visual Interfaces (AVI’18)*. ACM, New York, NY, USA, vol. 7. (2018).
- [63] Stellmach, Sophie, and R. Dachsel. “Designing gaze-based user interfaces for steering in virtual environments,” *In Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 131–138, (2012).
- [64] Hou, Wen-jun, K. Chen, H. Li, and H. Zhou. “User defined eye movement-based interaction for virtual reality,” *In International Conference on Cross-Cultural Design*, pp. 18–30, (2018).
- [65] Prabhakar, Gowdham, and P. Biswas. “Eye Gaze Controlled Projected Display in Automotive and Military Aviation Environments.” *Multimodal Technologies and Interaction*, 2, no. 1 (2018).
- [66] Sun, Qi, A. Patney, L.Y. Wei, O. Shapira, J. Lu, P. Asente, S. Zhu, M. McGuire, D. Luebke, and A. Kaufman “Towards virtual reality infinite walking: dynamic saccadic redirection,” *ACM Transactions on Graphics (TOG)* 37, no. 4, p.67, (2018).
- [67] Pirker, Johanna, M. Pojer, A. Holzinger and C. Gütl, “Gesture-based interactions in video games with the leap motion controller,” *In International Conference on Human-Computer Interaction*, pp. 620–633, (2017).
- [68] Kratz, Louis, M. Smith and F. J. Lee, “Wiizards: 3D gesture recognition for game play input,” *In Proceedings of the 2007 conference on Future Play*, pp. 209–212, (2007).
- [69] Khademi, Maryam, H. M. Hondori, A. McKenzie, L. Dodakian, C. V. Lopes and S. C. Cramer, “Free-hand interaction with leap motion controller for stroke rehabilitation,” *In CHI’14 Extended Abstracts on Human Factors in Computing Systems*, pp. 1663–1668, (2014).
- [70] Bizzotto, Nicola, A. Costanzo, L. Bizzotto, D. Regis, A. Sandri and B. Magnan, “Leap motion gesture control with OsiriX in the operating room to control imaging: first experiences during live surgery,” *Surgical innovation*, 1, no. 2, pp. 655–656, (2014).
- [71] Potter, L. Ellen, J. Araullo and L. Carter, “The leap motion controller: a view on sign language,” *In Proceedings of the 25th Australian computer-human interaction conference: augmentation, application, innovation, collaboration*, pp. 175–178, (2013).
- [72] Birk, Max, and R. L. Mandryk, “Control your game-self: effects of controller type on enjoyment, motivation, and personality in game,” *In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp. 685–694, (2013).
- [73] Shafer, M. Daniel, P. Corey and L. Popova, “Spatial presence and perceived reality as predictors of motion-based video game enjoyment,” *Presence: Teleoperators and Virtual Environments*, 20, no. 6, pp. 591–619, (2011).
- [74] J. Ahmad, S. Kamal and D. Kim, “Shape and motion features approach for activity tracking and recognition from kinect video camera,” *In International Conference on Advanced Information Networking and Applications Workshops (WAINA)*, pp. 445–450, (2015).
- [75] Piyathilaka, Lasitha, and S. Kodagoda, “Gaussian mixture based HMM for human daily activity recognition using 3D skeleton features,” *In 8th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pp. 567–572, (2013).
- [76] J. Ahmad, Y.H. Kim, Y.J. Kim, S. Kamal, and D. Kim, “Robust human activity recognition from depth video using spatiotemporal multi-fused features,” *Pattern recognition* 61, pp. 295–308, (2017).
- [77] J. Ahmad and Y. Kim, “Dense depth maps-based human pose tracking and recognition in dynamic scenes using ridge data,” *In International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pp. 119–124, (2014).
- [78] J. Ahmad, S. Kamal and D. Kim, “A depth video sensor-based life-

logging human activity recognition system for elderly care in smart indoor environments,” *Sensors* 14, no. 7, pp. 11735-11759, (2014).

- [79] J. Ahmad, S. Kamal and D. Kim, “Individual detection-tracking-recognition using depth activity images,” In 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 450-455, (2015).
- [80] Wu, Haitao, W. Pan, X. Xiong and S. Xu, “Human activity recognition based on the combined svm&hmm.” In International Conference on Information and Automation (ICIA), pp. 219-224, (2014).
- [81] <https://www.github.com/victordibia/skyfall> (Accessed on 15 Nov. 2018)
- [82] <http://www.freeaddictinggames.com/game/beads-puzzle/> (Accessed on 15 Nov. 2018)
- [83] 2d platformer - asset store. <https://www.assetstore.unity3d.com/en/#!/content/11228> (Accessed on 15 Nov. 2018)



Muhammad Raees

Muhammad Raees is currently serving the department of Higher Education, KPK, Pakistan, as an Assistant Professor of Computer Science. He did M.Sc.(Computer Science) from Peshawar University, Pakistan in 2003. He completed M.Phil. with distinction in Virtual Reality Programming from University of Malakand in 2015. He is pursuing Ph.D. (Virtual Reality & Intelligent Systems) from University of Malakand. His research interests include Virtual Reality Programming, Computer Vision and Machine Learning.



Sehat Ullah

Sehat Ullah has been serving the Department of Computer Science and IT, University of Malakand Pakistan, since 2001. Currently he is Associate Professor of Computer Science and HOD of the Virtual Reality Department. He has Ph.D. degree in the field of VR and 3D Interactions from University of Évry, France. He has supervised more than ten Ph.D. scholars and has about 30 research publications to his credit.

# Forecasting the Behavior of Gas Furnace Multivariate Time Series Using Ridge Polynomial Based Neural Network Models

Waddah Waheeb\*, Rozaida Ghazali

Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, Johor (Malaysia)



Received 13 October 2018 | Accepted 22 April 2019 | Published 26 April 2019

## ABSTRACT

In this paper, a new application of ridge polynomial based neural network models in multivariate time series forecasting is presented. The existing ridge polynomial based neural network models can be grouped into two groups. Group A consists of models that use only autoregressive inputs, whereas Group B consists of models that use autoregressive and moving-average (i.e., error feedback) inputs. The well-known Box-Jenkins gas furnace multivariate time series was used in the forecasting comparison between the two groups. Simulation results show that the models in Group B achieve significant forecasting performance as compared to the models in Group A. Therefore, the Box-Jenkins gas furnace data can be modeled better using neural networks when error feedback is used.

## KEYWORDS

Error Feedback, Nonlinear Autoregressive Moving-Average Model, Multivariate Time Series, Recurrent Network.

DOI: 10.9781/ijimai.2019.04.004

## I. INTRODUCTION

**T**IME series data are everywhere. Examples of time series data are the daily weather temperature, monthly sales, and annual milk production.

Time series can be categorized into univariate and multivariate time series. Univariate time series can be obtained by recording a single phenomenon over time, for example, recording air pollution concentration. Multivariate time series, on the other hand, can be obtained by recording more than one phenomenon over time, for example, recording air pollution concentration and some weather information.

Multivariate time series can be found in many real problems such as in business and engineering domains. The importance of having multivariate time series is to take advantage of the available additional information from related time series in order to enhance the forecasting accuracy for each series individually.

Various models have been used for multivariate time series forecasting such as vector autoregression models, adaptive network-based fuzzy inference system, belief rule based system, neural networks and other hybrid models [1]-[4].

Neural networks (NNs) have been applied extensively in time series forecasting due to some distinguishing features as reported in [5]. These reported features are the universal function approximator property in NNs that allow them to approximate any continuous function with an arbitrary degree of accuracy. Furthermore, NNs are a nonlinear data-driven method with few a priori assumptions about underlying models. In addition, their generalization ability is good.

Higher order neural network (HONN) is a type of NNs that utilizes high order terms (e.g., multiplicative units) besides the commonly used summing units. HONNs are simple in their architecture and have fewer trainable parameters to deliver the input-output mappings as compared to multilayered NNs.

An example of a HONN is the feedforward ridge polynomial neural network (RPNN) [6]. RPNN has good mapping capabilities and it utilizes univariate polynomials that are easy to handle, unlike other HONNs which suffer an explosion of free parameters due to the use of multivariate polynomials [6]. Furthermore, the RPNN provides a more efficient and regular architecture compared to the ordinary HONNs while maintaining their fast learning property [6].

Based on the structure of the RPNN model, there are three recurrent neural networks, namely the dynamic ridge polynomial neural network (DRPNN) [7]-[8], the ridge polynomial neural network with error feedback (RPNN-EF) [9], and the ridge polynomial neural network with error and output feedbacks (RPNN-EOF) [10]. The DRPNN has a feedback based on its network output as an additional input to the network, whereas the feedback in the RPNN-EF is its network error. The RPNN-EOF has both feedbacks.

The feedforward RPNN and the three recurrent networks can be grouped into two groups. The first group, which is referred in this paper as Group A, contains the RPNN and DRPNN that use only autoregressive inputs (i.e., lagged variables of one or more time series), whereas Group B contains the RPNN-EF and RPNN-EOF that use autoregressive and moving-average (i.e., error feedback) inputs.

Based on studies about ridge polynomial based neural network models retrieved in this paper, the four existing ridge polynomial based neural network models have not yet been applied in multivariate time series forecasting.

\* Corresponding author.

E-mail address: waddah.waheeb@gmail.com



The well-known multivariate time series called Box-Jenkins gas furnace time series was used in this paper [3]. This time series has been frequently used for the assessment of new identification and modeling techniques. A combustion process of a methane-air mixture was used to record this series. The gas flow rate was kept constant, but the methane rate was randomly changed. The resulting carbon dioxide concentration ( $\text{CO}_2$ ) in the output gases was measured.

The main objective of this paper is to investigate and compare the forecasting efficiency of neural network models in the two groups in forecasting this well-known multivariate time series. To the best of our knowledge, this is the first study that has attempted to achieve this objective. This is important to investigate the difference in the forecasting ability between the two groups. Also, a comparison with some models reported in the literature with this time series was reported.

We organized our paper as follows. The neural network models used in this paper are described in Section II. Section III discusses the methodology used to conduct the experiments. The findings are presented and discussed in Section IV, and in Section V the conclusion and future works are given.

## II. RELATED WORKS

Neural networks provide a promising alternative tool for forecasters [11]. Numerous applications have been reported on using different types of neural networks to model time series. A basic example to show how a NN is used to learn the nonstationary time series is shown in Fig. 1. Simply, during training, the past time series (i.e., lagged variables) are fed to the neural network. Network output (i.e., forecast) based on the given inputs is produced then compared with the desired output to calculate the error. This error is used to update the network parameters until a specified goal is attained.

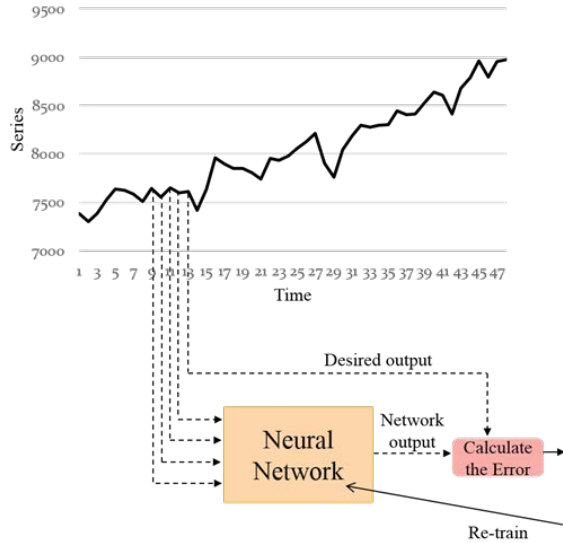


Fig. 1. Learning time series with a neural network.

The RPNN is a type of neural network that has only one layer of adjustable weights, and it uses the pi-sigma neural network (PSNN) as basic building blocks [6]. According to [6], the RPNN utilizes a constructive learning algorithm which allows it to automatically determine the necessary network order. For a given problem, the RPNN starts with a small order of the PSNN block(s), and then an extra block of higher orders is added during the learning process. The learning process stops when the desired level of approximation error is attained, or when the maximum number of epochs or the appropriate order of the network has been reached [6]-[9].

According to [6], the RPNN has the ability to uniformly approximate any continuous function on a compact set in a multidimensional input space with arbitrary degree of accuracy. Moreover, the RPNN utilizes univariate polynomials which are easy to handle, thus it avoids the explosion of the number of trainable parameters as the number of inputs increases, which happens with some neural networks [6]. Fig. 2 shows an example of the RPNN. It can be seen that the weights linking the hidden and the output units are fixed to unity. The inputs to the RPNN are *only* the lagged variables from the time series.

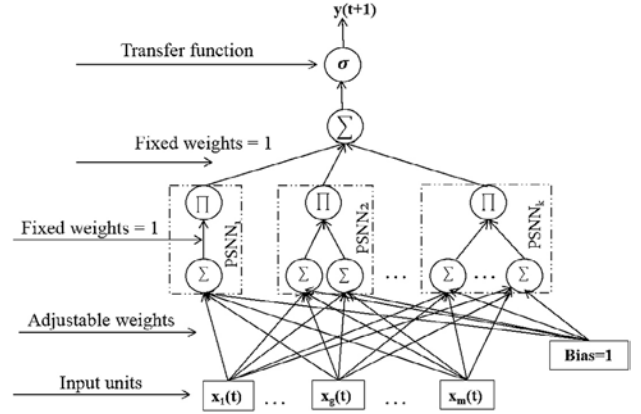


Fig. 2. Architecture of the ridge polynomial neural network [6].

Like any feedforward network, the output of the RPNN is only a function of the current input; in other words, the node equations are memoryless. Therefore, the dynamic RPNN which is a recurrent version of the RPNN was proposed [7]. The DRPNN uses its network output value as an additional input to the input layer. Therefore, it is provided with memory that helps to retain information to be used later to affect the processing of future inputs [7]. Fig. 3 shows an example of the DRPNN.

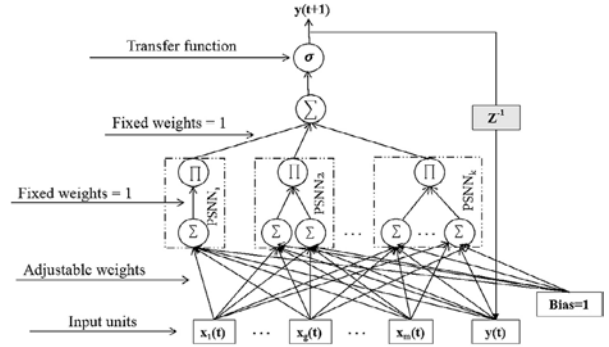


Fig. 3. Architecture of the dynamic ridge polynomial neural network [7].

Due to the existence of the recurrent feedback, learning instability problem could occur in the DRPNN. Therefore, the learning rate was controlled by a Lyapunov function [8]. This solution improves the forecasting accuracy of the DRPNN and reduces its training time. In this paper, we used the DRPNN that uses Lyapunov function and referred to it as DRPNN.

The inputs of the DRPNN is given as follows:

$$z_g(t) = \begin{cases} x_g(t) & 1 \leq g \leq m \\ y(t) & g = m+1 \end{cases} \quad (1)$$

where  $x_g(t)$  is the lagged variables from the time series,  $m$  is the number of past lagged values from the time series, and  $y(t)$  is network output at time  $t$ .

Both the RPNN and DRPNN use only autoregressive inputs (i.e., lagged variables of one or more time series). However, an alternative to autoregressive modeling has been suggested by feeding back network error to the input layer of a neural network [9]-[10], [12]-[13] in order to model the nonlinear moving-average processes more directly and parsimoniously [13].

The RPNN-EF [9] is a network based on the RPNN and is incorporated with network error as shown in Fig. 4. As shown in Fig. 4, the error term is calculated by subtracting the target from network output (i.e., forecast).

The RPNN-EF was used to forecast univariate time series [9]. It was found that the RPNN-EF showed better forecasting performance with respect to the RPNN and DRPNN using four univariate time series. Furthermore, the forecasting performance of the RPNN-EF was found better than some existing techniques with some benchmark time series.

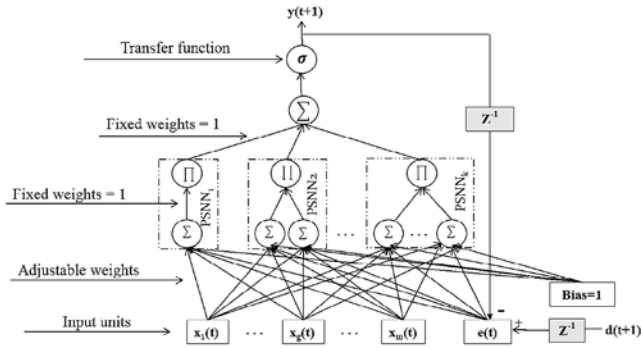


Fig. 4. Architecture of the ridge polynomial neural network with error feedback [9].

The inputs to the RPNN-EF are given as follows:

$$z_g(t) = \begin{cases} x_g(t) & 1 \leq g \leq m \\ e(t) = d(t) - y(t) & g = m + 1 \end{cases} \quad (2)$$

where  $x_g(t)$  is the lagged variables from the time series,  $m$  is the number of past lagged values from the time series,  $e(t)$  is the error,  $y(t)$  is the network output, and  $d(t)$  is the desired output.

The RPNN-EOF [10] is a network based on the DRPNN and is incorporated with network error as shown in Fig. 5. The inputs for the RPNN-EOF is given by the following equation:

$$z_g(t) = \begin{cases} x_g(t) & 1 \leq g \leq m \\ e(t) = d(t) - y(t) & g = m + 1 \\ y(t) & g = m + 2 \end{cases} \quad (3)$$

where  $x_g(t)$  is the lagged variables from the time series,  $m$  is the number of past lagged values from the time series,  $e(t)$  is the error,  $y(t)$  is network output, and  $d(t)$  is the desired output.

The RPNN-EOF was evaluated using the Mackey-Glass differential delay equation time series [10]. Based on the findings in [10], the RPNN-EOF produced smaller error as compared to some state-of-the-art models.

The ridge polynomial based neural network models have been utilized in different applications as tabulated in Table I. The feedforward RPNN was used in various applications, whereas the three recurrent networks were used only for time series forecasting. With regard to time series forecasting, all the studies in Table I were applied to forecast only univariate time series [7]-[10], [14], [19]. Based on [7], [9], and [14], the recurrent RPNN models are better than the feedforward RPNN model for univariate time series forecasting. Moreover, the RPNN-EF is better than those without error feedback (i.e., RPNN and DRPNN) on average [9].

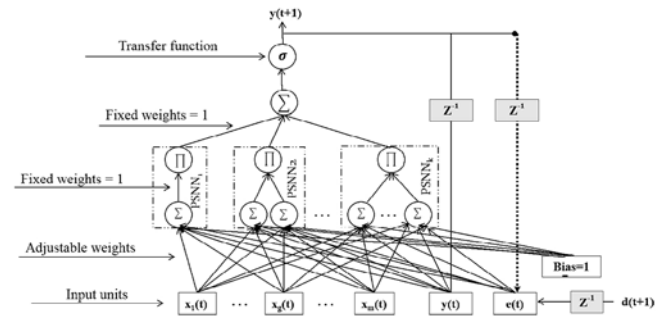


Fig. 5. Architecture of the ridge polynomial neural network with error-output feedbacks [10].

Based on the aforementioned, there is a scope to do further research for applying ridge polynomial based neural network models for multivariate time series forecasting. Therefore, the main objective of this paper is to apply and compare the forecasting ability of two groups of neural networks in multivariate time series forecasting: neural network models that use only autoregressive inputs (i.e., RPNN and DRPNN), simply referred to as Group A; and neural network models that use autoregressive and moving-average (i.e., RPNN-EF and RPNN-EOF) inputs, simply referred to as Group B.

### III. METHODOLOGY

#### A. Box-Jenkins Gas Furnace Data

It is one of the well-known and frequently used benchmark problems [3] that has been frequently used for the assessment of new identification and modeling techniques. This time series consists of 296 records and was downloaded from the following website <http://www.stat.purdue.edu/~chong/stat520/bjr-data/gas-furnace>. A combustion process of a methane-air mixture was used to record this series. The gas flow rate was kept constant, but the methane rate was randomly changed. The resulting carbon dioxide concentration ( $\text{CO}_2$ ) in the output gases was measured. The sampling interval is 9 seconds.

There are different fitting models used by researchers. In this paper, the following fitting model was chosen to compare our findings with [1] and [2]:

$$y(t) = F(v(t-4), y(t-1)) \quad (4)$$

where  $v(t)$  is the methane gas flow into the furnace, and  $y(t)$  is the  $\text{CO}_2$  concentration in the outlet gas.

Based on (4), there are 292 input/output data pairs. By following [1] and [2], these pairs were partitioned in 200 and 92 pairs for training and out-of-sample sets, respectively. Fig. 6 and 7 show the training and out-of-sample values of the  $\text{CO}_2$  concentration  $y(t)$ .

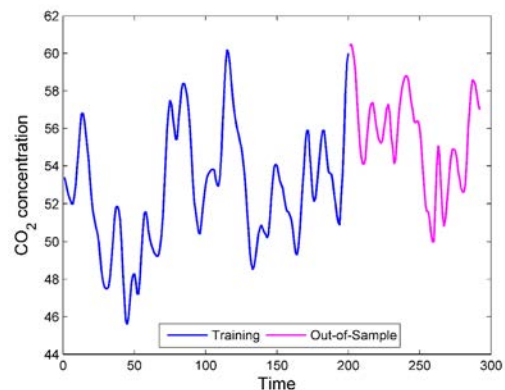


Fig. 6.  $\text{CO}_2$  concentration time series for the Box-Jenkins gas furnace data.

TABLE I. STUDIES APPLIED THE RIDGE POLYNOMIAL BASED NEURAL NETWORK MODELS

Study	Model	Inputs	Application
[6]	RPNN	Autoregressive	Multivariate function approximation, realization of multivariate polynomial, and data classification
[15]			Image compression
[16]			OFDM (Orthogonal Frequency Division Multiplexing) systems signals as a nonlinear compensator
[17]			Character recognition
[18]			Modeling nonlinear chemical kinetics
[20]			Microwave characterization of dielectric materials
[21]			Data classification
[19]			Univariate time series forecasting
[7]-[8], [14]	Dynamic RPNN	Autoregressive	
[9]	RPNN with error feedback	Autoregressive & Moving-average	
[10]	RPNN with error-output feedbacks	Autoregressive & Moving-average	

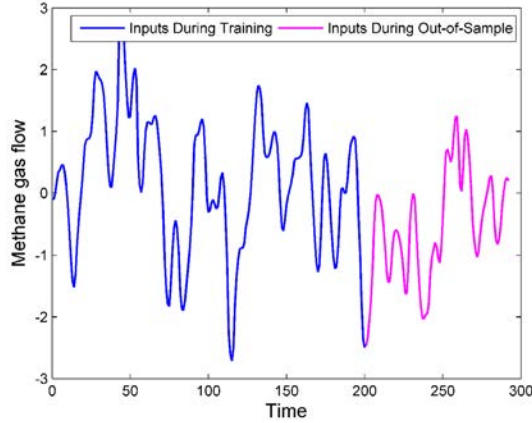


Fig. 7. Methane gas flow time series for the Box-Jenkins gas furnace data.

The time series were scaled to the range [0.2 - 0.8] by using the minimum and maximum normalization method which is given by:

$$\hat{y} = (\max_2 - \min_2) * \left( \frac{y - \min_1}{\max_1 - \min_1} \right) + \min_2 \quad (5)$$

where  $\min_1$  and  $\max_1$  are the respective minimum and maximum values of all observations,  $\min_2$  and  $\max_2$  refer to the desired minimum and maximum of the new scaled series,  $y$  refers to the original value, and  $\hat{y}$  is the normalized version of  $y$ . The  $\min_1$  and  $\max_1$  for the methane gas flow into the furnace are equal to -2.716 and 2.834, respectively. Similarly, the  $\min_1$  and  $\max_1$  for the CO<sub>2</sub> concentration in the outlet gas are equal to 45.6 and 60.5, respectively.

### B. Network Topology and Parameters' Settings

Network structure and the values used to train the four models are shown in Table II. These settings are based on previous works using these models for time series forecasting [8]-[10].

### C. Performance Metric

Root mean squared error (RMSE) metric was used in this paper. The equation for the RMSE metric is given by:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (6)$$

where  $N$ ,  $y$  and  $\hat{y}$  represent the number of test pairs, actual output and network output, respectively.

TABLE II. NETWORK TOPOLOGY AND TRAINING FOR THE RIDGE POLYNOMIAL BASED NEURAL NETWORK MODELS

Setting	Value
Initial weights	[-0.5,0.5]
Network order	Incrementally grown from 1 to 5.
Stopping criteria	Minimum Squared Error = 0.000001 or after accomplishing the 5th order network learning or reaching the maximum number of epochs = 3000. Extra criterion for the DRPNN, RPNN-EF and RPNN-EOF: network becomes unstable.
Learning rate ( $n$ )	[0.01-1]
Decreasing factors for $n$	0.8
Momentum	[0.4-0.8]
Threshold of successive PSNN addition ( $r$ )	[0.00001-0.1]
Decreasing factors for $r$	[0.05, 0.2]

### D. Model Selection

As mentioned before, the data were partitioned into two sets only; training and out-of-sample sets as used in [1] and [2]. In order to evaluate the forecasting performance of the four models, the adjustable parameters after finishing the training of each network's order were selected for generalization purpose (i.e., out-of-sample forecasting) [7]. Then, the training is continued with increasing network's order, and this continues until the stopping criteria are met [7]. Network structure with the lowest RMSE on the out-of-sample set is considered the best model.

### E. Wilcoxon Sign-rank Test

In order to know if there is any significance in the forecasting performance, the Wilcoxon sign-rank test [22] was used. This test is a nonparametric and distribution free test, thus it is statistically safer and more robust than parametric tests [23]. The Wilcoxon signed-rank test was conducted using IBM SPSS software.

### F. Comparison with Other Forecasting Models

In this paper, we compared the forecasting performance of the four models from the two groups with other forecasting models reported in [1] and [2]. In addition, we used the well-known multilayer feed-forward network model (MLP) and extreme learning machines (ELM) [24] models in the comparison.

The function *nnetar* in the R package 'forecast' [25] was used to build the MLP model. The number of hidden neurons are selected between 1 and 25. For that, 40 simulations for each hidden neuron size were conducted, yielding 1000 total simulations. The data is scaled by subtracting the column means and dividing by their respective standard deviations

because it gives better forecasting performance based on our experiments.

To build the ELM model, we used the function *elm* in the R package ‘nnfor’ [26]. We used a multiplication of 10 to 100 for the number of hidden nodes as well as the available option to be determined automatically by the algorithm. Ridge regression with cross validation was used as an estimation type for output layer weights because it gives better forecasting accuracy as compared to the other three types based on our experiments. Direct input-output connections to model strictly linear effects was used. We ran 1000 simulations using the ELM model.

#### IV. RESULTS AND DISCUSSION

This section presents one-step forecasts comparison between two neural network models that use only autoregressive inputs (i.e., Group A: RPNN and DRPNN) and two neural network models that use autoregressive and moving-average inputs (i.e., Group B: RPNN-EF and RPNN-EOF) in forecasting the Box-Jenkins gas furnace time series.

After finishing the training of each network's order with 225 different settings, the final values for the trainable parameters were selected for generalization purpose.

Among 1,125 simulations, the average RMSE results of best 20, 50 and 100 simulations for each model are shown in Table III. In addition, the results of the Wilcoxon sign-rank test are shown in Table IV.

TABLE III. AVERAGE OF BEST TOP SIMULATION RESULTS COMPARISON BETWEEN NEURAL NETWORK MODELS USING THE RMSE (BASED ON THE NORMALIZED DATA)

No. Simulations	Group A		Group B	
	RPNN	DRPNN	RPNN-EF	RPNN-EOF
20	0.0214	0.0207	<b>0.0159</b>	0.0165
50	0.0231	0.0215	<b>0.0173</b>	0.0176
100	0.0245	0.0225	<b>0.0182</b>	<b>0.0182</b>

Best results are in **boldface**.

TABLE IV. RESULTS OF THE WILCOXON SIGN RANK TEST

No. Simulations	20	50	100
RPNN-EF vs. RPNN	-3.920 ( $p < 0.000089$ )	-6.154 ( $p < 0.000000$ )	-8.682 ( $p < 0.000000$ )
RPNN-EOF vs. DRPNN	-3.920 ( $p < 0.000089$ )	-6.154 ( $p < 0.000000$ )	-8.682 ( $p < 0.000000$ )
DRPNN vs. RPNN	-3.920 ( $p < 0.000089$ )	-6.154 ( $p < 0.000000$ )	-8.682 ( $p < 0.000000$ )
RPNN-EOF vs. RPNN	-3.920 ( $p < 0.000089$ )	-6.154 ( $p < 0.000000$ )	-8.682 ( $p < 0.000000$ )

As shown in Table III and Table IV, the models in Group B produce significantly better forecasting performance as compared to the models in Group A in all cases. In other words, the RPNN-EF is significantly better than the RPNN, and the RPNN-EOF is significantly better than the DRPNN. Based on that, it can be concluded that incorporating error feedback to neural network models makes the models more suitable to deal with this time series. This is because the models in Group B have autoregressive and moving-average inputs, unlike the models in Group A that have only autoregressive inputs. Since an approach to identify the component of a non-linear model similar to the Box-Jenkins identification approach, which is used for linear model identification, is still not available, it is difficult to prove that the time series possess a non-linear moving-average component. However, it was found that neural networks with error feedback are more suitable than neural networks without error feedback in modelling univariate time series that possess a moving-average component [12]- [13].

We observe from Table III and Table IV that all recurrent models achieve significant results as compared to the feedforward RPNN that produced the highest forecasting error. Moreover, the RPNN-EF has the best average forecasting performance.

The best simulation results for the four models in the two groups are shown in Table V. In addition, the best results obtained by the MLP and ELM models are shown in Table V. For fair comparison with the models reported in [1] and [2], all the results in Table V are after the de-normalization (i.e., returned to the original data scale). Even though the models in Table V have different structures as compared to the four models, it is widely accepted to compare the final results with those reported since all the models use the same model mentioned in (4), the training and out-of-sample sets are identical, and the reported results are in the original scale. It is good to note that the best model using the MLP is with 14 hidden neurons while the best model for the ELM used 100 hidden neurons.

TABLE V. COMPARISON OF THE PERFORMANCE OF VARIOUS EXISTING MODELS (BASED ON THE DE-NORMALIZED DATA)

Model	RMSE
MLP	0.9098
ELM	0.6736
RPNN	0.4920
DRPNN	0.4804
Belief rule based (BRB) system with the number of 5 * 5 referential values [1]	0.4616 <sup>a</sup>
Adaptive network-based fuzzy inference system (ANFIS) [2]	0.4053 <sup>a</sup>
Pseudo-Gaussian basis function network (PG-BF) [2]	0.3962 <sup>a</sup>
RPNN-EF	0.3652
RPNN-EOF	0.3617
Hybrid neural fuzzy inference system (HyFIS) [2]	0.3074 <sup>a</sup>
Generalized fuzzy neural network (G-FNN) [2]	0.2728 <sup>a</sup>

<sup>a</sup> The original work reported the results in MSE

It can be seen from Table V that among the neural network models used in this paper, the RPNN-EOF produces the best simulation. Similar to the results shown in Table III, the models in Group B produce the better forecasting performance as compared to the models in Group A. The recurrent models also produce better results as compared to the feedforward model.

Moreover, as tabulated in Table V, the RPNN-EF and RPNN-EOF outperform hybrid models such as the MLP, ELM, BRB, ANFIS and PG-BF models. However, the hybrid HyFIS and G-FNN models outperform the RPNN-EF and RPNN-EOF. Therefore, using more than one error feedback in the RPNN-EF and RPNN-EOF could help to improve their forecasting performance.

Regarding the number of final parameters (i.e., weights and biases), the minimum number of parameters is with the RPNN where it equals 18 parameters. The maximum number of parameters is 60 with the DRPNN. The RPNN-EF and RPNN-EOF have 24 and 30 parameters, respectively.

The learning curve for these best models are shown in Fig. 8 - Fig. 11. It can be seen from these figures that the MSE per epoch decreases monotonically. The fastest training model based on these figures is the RPNN-EF with less than 70 epochs.



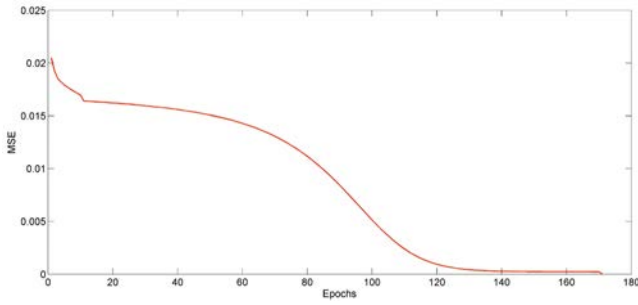


Fig. 8. Learning curve for the RPNN.

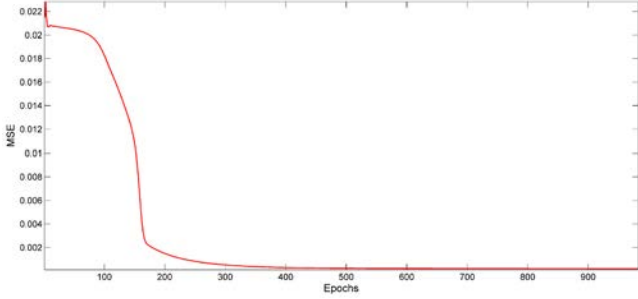


Fig. 9. Learning curve for the DRPNN.

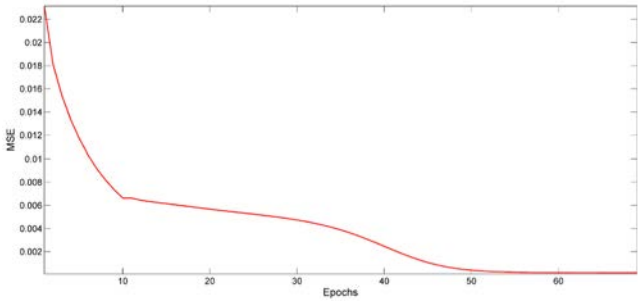


Fig. 10. Learning curve for the RPNN-EF.

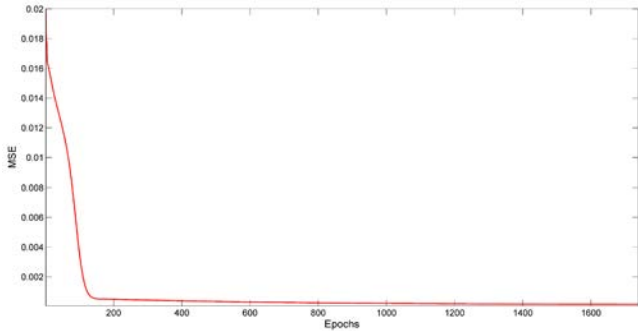


Fig. 11. Learning curve for the RPNN-EOF.

Fig. 12 and Fig. 13 show the best out-of-sample forecasting result using the four models. The idea behind dividing the results into two figures is to show the difference in the forecasting ability between the model with and without error feedback. Therefore, the RPNN is drawn with the RPNN-EF in Fig. 12, whereas the DRPNN is drawn with the RPNN-EOF in Fig. 13. It is clear that all models are able to produce forecasts close to the targets in most cases. However, the RPNN-EF and RPNN-EOF have better ability to follow the peaks and troughs as compared to the RPNN and DRPNN, respectively.

The best simulation performance among the four models is achieved using the RPNN-EOF model as shown in Table V. This model is as follows:

$$Eq1 = 0.004471 * x_1 + 0.201585 * x_2 - 0.33254 * e(t) - 0.325995 * y(t) - 0.088927 \quad (7)$$

$$Eq2 = (-0.410376 * x_1 - 0.100439 * x_2 - 0.223905 * e(t) + 0.127541 * y(t) + 0.411272) * (-0.279539 * x_1 + 0.167540 * x_2 - 0.299313 * e(t) - 0.399306 * y(t) + 0.072087) \quad (8)$$

$$Eq3 = (-0.641134 * x_1 - 0.585069 * x_2 - 0.078077 * e(t) - 0.458598 * y(t) - 0.941347) * (-0.788508 * x_1 - 0.510193 * x_2 - 0.308327 * e(t) - 0.171573 * y(t) - 0.891403) * (-0.720051 * x_1 + 0.952238 * x_2 + 0.868305 * e(t) - 0.064836 * y(t) - 0.021427) \quad (9)$$

$$y(t+1) = \text{Sigmoid}(Eq1 + Eq2 + Eq3) \quad (10)$$

where  $x_1$  is the methane gas flow into the furnace,  $x_2$  is the  $\text{CO}_2$  concentration in the outlet gas,  $e(t)$  is the past network error, and  $y(t)$  is the past network output. The inputs should be normalized, as explained in Eq. (5). The  $e(t)$  and  $y(t)$  to be used with the first out-of-sample sample are -0.001954 and 0.781819, respectively. It is worth noting that the numbers (i.e., weights) shown in Eq. (6) – Eq. (9) are rounded to 6 decimal places.

The main contribution of this paper is that it shows how good the forecasting performance is when error feedback is incorporated into the structure of the neural networks as compared to their counterparts without error feedback when forecasting the Box-Jenkins gas furnace data. This better forecasting performance can be attributed to the direct modelling of the non-linear moving-average component via error feedback.

The limitation of the study is that we conducted only one-step forecasts comparison. However, to the best of our knowledge, neural network models available in the literature that use error feedback have not yet tested for recursive multi-step forecasting. That is because the error cannot be observed until the real value for the current time is available. However, the unobserved errors can be replaced with zeros as in the autoregressive and moving-average (ARMA) model, or other solutions can be further investigated for recursive multi-step forecasting. Another limitation is the comparison with the MLP and ELM models. There are no options to select different ranges for the learning parameters such as learning rate with both functions used. Therefore, 1000 simulations were conducted to find best performance for both models.

## V. CONCLUSION

In this paper, one feedforward and three recurrent neural networks based on the ridge polynomial neural network were used to forecast the well-known Box-Jenkins gas furnace multivariate time series. These four models are grouped into two groups. Group A consists of models that use only autoregressive inputs, whereas Group B consists of models that use autoregressive and moving-average (i.e., error feedback) inputs. Overall, the following points can be concluded from the obtained results:

- The models in Group B achieve better and more significant forecasting performance as compared to the models in Group A in forecasting this time series.
- The recurrent models are better and more significant than the feedforward RPNN model to forecast this time series.
- The significance in the forecasting performance for Group B models reveals that the time series could possess moving-average component. Therefore, the used error feedback with the models helps to better modeling the series.

In future, more multivariate time series could be explored with the models in the two groups. Incorporating more loops into the recurrent

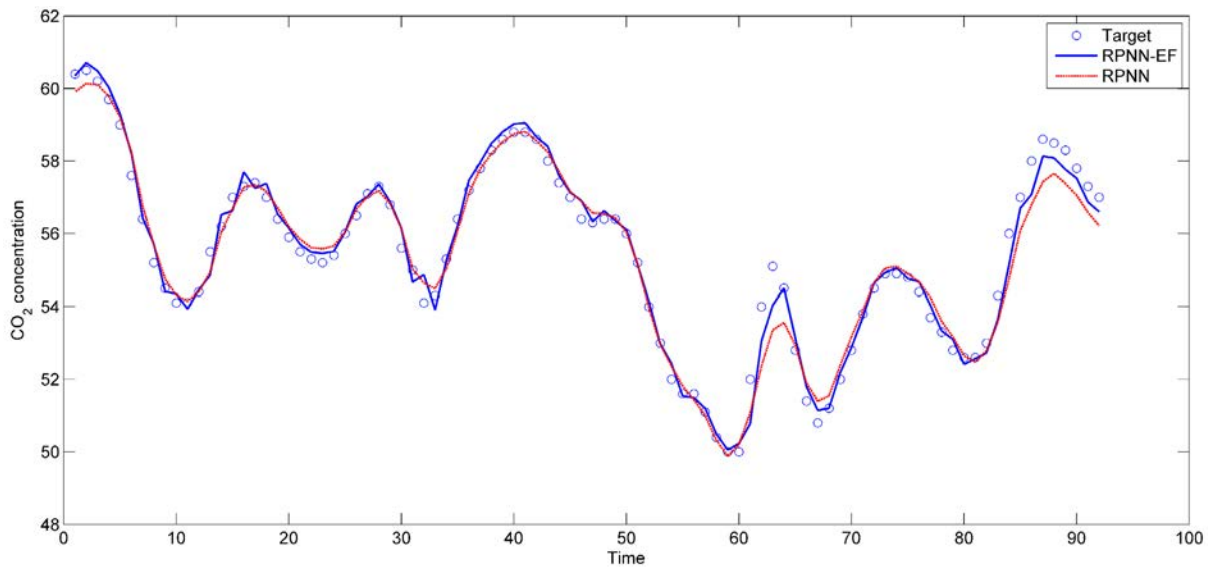


Fig. 12. The best out-of-sample forecasting using the RPNN and RPNN-EF.

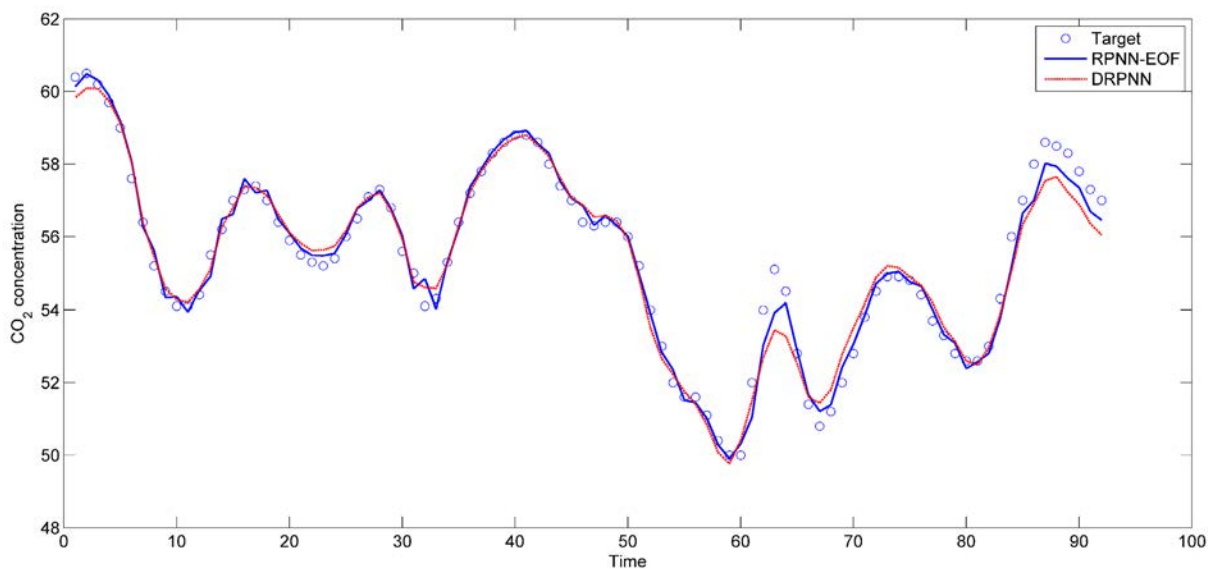


Fig. 13. The best out-of-sample forecasting using the DRPNN and RPNN-EOF.

models are also a subject of the future work. Moreover, recursive multi-step forecasting will be further explored with univariate and multivariate time series.

#### ACKNOWLEDGMENT

The authors would like to thank Universiti Tun Hussein Onn Malaysia and the Office for Research, Innovation, Commercialization and Consultancy Management (ORICC) for funding this research under the Postgraduate Research Grant (GPPS), VOT# U612.

#### REFERENCES

- [1] Y.W. Chen, J.B. Yang, D.L. Xu and S.L. Yang, "On the inference and approximation properties of belief rule based systems," *Information Sciences*, vol. 234, pp. 121-135, 2013.
- [2] Y. Gao and M. J. Er, "NARMAX time series model prediction: feedforward and recurrent fuzzy neural network approaches," *Fuzzy sets and systems*, vol. 150, no. 2, pp. 331-350, 2005.
- [3] G. E. Box, G. M. Jenkins, G. C. Reinsel and G. M. Ljung, *Time series analysis: forecasting and control*. New Jersey: John Wiley & Sons, 2015.
- [4] K. Chakraborty, K. Mehrotra, C. K. Mohan and S. Ranka, "Forecasting the behavior of multivariate time series using neural networks," *Neural networks*, vol. 5, no. 6, pp. 961-970, 1992.
- [5] G. Zhang, B. E. Patuwo and M. Y. Hu., "Forecasting with artificial neural networks: The state of the art," *International journal of forecasting*, vol. 14, no. 1, pp. 35-62, 1998.
- [6] Y. Shin and J. Ghosh, "Ridge polynomial networks," *IEEE Transactions on neural networks*, vol. 6, no. 3, pp. 610-622, 1995.
- [7] R. Ghazali, A. J. Hussain, N. M. Nawi, and B. Mohamad, "Non-stationary and stationary prediction of financial time series using dynamic ridge polynomial neural network," *Neurocomputing*, vol. 72, no. 10-12, pp.2359-2367, 2009.
- [8] W. Waheeb, R. Ghazali, and A. J. Hussain, "Dynamic ridge polynomial neural network with Lyapunov function for time series forecasting," *Applied Intelligence*, vol. 48, no. 7, pp.1721-1738, 2018.
- [9] W. Waheeb, R. Ghazali, and T. Herawan, "Ridge polynomial neural network with error feedback for time series forecasting," *PloS one*, 11(12), p.e0167248, 2016.
- [10] W. Waheeb and R. Ghazali, "Multi-step Time Series Forecasting Using Ridge Polynomial Neural Network with Error-Output Feedbacks," in *International Conference on Soft Computing in Data Science*, Kuala

Lumpur, 2016, pp. 48-58.

- [11] G. P. Zhang, "Neural Networks for Time-Series Forecasting," in *Handbook of Natural Computing*, Berlin, Heidelberg, Springer, 2012, pp. 461-477.
- [12] J. T. Connor, R. D. Martin and L. E. Atlas, "Recurrent neural networks and robust time series prediction," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 240-254, 1994.
- [13] A. N. Burgess and A. N. Refenes, "Modelling non-linear moving average processes using neural networks with error feedback: An application to implied volatility forecasting," *Signal Processing*, vol. 74, no. 1, pp. 89-99, 1999.
- [14] R. Ghazali, A. J. Hussain, and P. Liatsis, "Dynamic Ridge Polynomial Neural Network: Forecasting the univariate non-stationary and stationary trading signals," *Expert Systems with Applications*, vol. 38, no. 4, pp.3765-3776, 2011.
- [15] P. Liatsis and A. J. Hussain, "Nonlinear 1D DPCM image prediction using polynomial neural networks," in *Applications of Artificial Neural Networks in Image Processing IV*, San Jose, 1999, pp. 58-69.
- [16] S. Tertois, A. L. Glaunec and G. Vaucher, "Compensating the non-linear distortions of an OFDM signal with neural networks," in *the 9th International Workshop on Systems, Signals and Image Processing*, Manchester, 2002, pp. 484-488.
- [17] C. Voutriaridis, Y. S. Boutalis and B. G. Mertzios, "Ridge polynomial networks in pattern recognition," in *Video/Image Processing and Multimedia Communications*, Zagreb, 2003.
- [18] N. Shenvi, J. M. Geremia and H. Rabitz, "Efficient chemical kinetic modeling through neural network maps," *The Journal of chemical physics*, vol. 120, no. 21, pp. 9942-9951, 2004.
- [19] R. Ghazali, A. J. Hussain, P. Liatsis, and H. Tawfik, "The application of ridge polynomial neural network to multi-step ahead financial time series prediction," *Neural Computing and Applications*, vol. 17, no. 3, pp.311-323, 2008.
- [20] T. Hacib, Y. L. Bihan, M.-K. Smaïl, M. R. Mekideche, O. Meyer and L. Pichon, "Microwave characterization using ridge polynomial neural networks and least-square support vector machines," *IEEE Transactions on Magnetics*, vol. 47, no. 5, pp. 990-993, 2011.
- [21] N. K. S. Behera and H. S. Behera, "Firefly based ridge polynomial neural network for classification," in *Advanced Communication Control and Computing Technologies (ICACCCT)*, Ramanathapuram, 2014, pp. 1110-1113.
- [22] F. Wilcoxon, "Individual comparisons by ranking methods," *Biometrics bulletin*, vol. 1, no. 6, pp. 80-83, 1945.
- [23] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *Journal of Machine learning research*, 2006, pp. 1-30.
- [24] Huang, G. B., Zhu, Q. Y., & Siew, C. K. (2006). Extreme learning machine: theory and applications. *Neurocomputing*, 70(1-3), 489-501.
- [25] R. Hyndman, G. Athanasopoulos, C. Bergmeir, G. Caceres, L. Chhay, M. O'Hara-Wild, F. Petropoulos, S. Razbash, E. Wang, F. Yasmien, forecast: Forecasting functions for time series and linear models. R package version 8.5, 2019, <http://pkg.robjhyndman.com/forecast>.
- [26] N. Kourentzes, nnfor: Time Series Forecasting with Neural Networks. R package version 0.9.6, 2019, <https://cran.r-project.org/web/packages/nnfor/index.html>.



Waddah Waheeb

Waddah Waheeb received his MSc (Soft Computing) from Universiti Tun Hussein Onn Malaysia in 2015. He is currently pursuing his PhD degree at the same university. His scientific work has been published in peer-reviewed journals and conferences. He has won numerous awards for his research such as the Best Paper Award at the 2nd International Conference on Soft Computing in Data

Science and at the 3rd International Conference of Reliable Information and Communication Technology. Moreover, his ensemble method used in the worldwide M4 forecasting competition is listed among the 17 most accurate methods as compared to benchmark methods, ranked 15th among 50 submissions around the world on average, and ranked 6th among 50 submissions around the world with the weekly time series.



Rozaida Ghazali

Rozaida Ghazali is currently a Professor at the Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia (UTHM). She graduated with a Ph.D. degree in Higher Order Neural Networks from the School of Computing and Mathematical Sciences at Liverpool John Moores University, United Kingdom in 2007. Earlier, in 2003 she completed her M.Sc. degree in Computer Science from Universiti Teknologi Malaysia (UTM). She received her B.Sc. (Hons) degree in Computer Science from Universiti Sains Malaysia (USM) in 1997. In 2001, Rozaida joined the academic staff in UTHM. Her research area includes neural networks, swarm intelligence, optimization, data mining, and time series prediction. She has successfully supervised a number of PhD and master students and published more than 100 articles in various international journals and conference proceedings. She acts as a reviewer for various journals and conferences, and as an editor in a few Springer conference proceedings. She has also served as a conference chair, and as a technical committee for numerous international conferences.

# Marketing Intelligence and Big Data: Digital Marketing Techniques on their Way to Becoming Social Engineering Techniques in Marketing

Prof. Dr. habil. Jan Lies

FOM University of Applied Science, Dortmund (Germany)

Received 28 April 2019 | Accepted 6 May 2019 | Published 9 May 2019



## ABSTRACT

This contribution reviews the vast scope of digital application areas, which shape the *digital marketing landscape* and coin the present term “*marketing intelligence*” from a marketing *technique* point of view. Additionally, marketing intelligence as *social engineering techniques* are described. The review ranges from digital IT- and big data marketing until marketing 5.0 as digitalized trust marketing. The multiplicity of applications and interdependencies of the digital and social techniques reviewed should show that big data and marketing intelligence have already become a *marketing reality*. It becomes clear that marketing is witnessing a *methodological, technical and cultural paradigm shift* that augments and amplifies *traditional outbound marketing* with *inbound marketing*.

## KEYWORDS

Marketing Intelligence, Digital Marketing, Big Data Marketing, Social Engineering, Marketing 5.0.

DOI: 10.9781/ijimai.2019.05.002

## I. INTRODUCTION

**M**ARKETING intelligence is certainly not a new term by any means. Nevertheless, it is currently quite common in both theory and practice. The following descriptions of techniques and social engineering methods of marketing intelligence have three goals:

1. The explanations are, without claiming to be overly exhaustive, intended to point out the vast scope of digital application areas, which shape the digital marketing landscape and coin the present term “marketing intelligence” from a marketing technique point of view.
2. Additionally, marketing intelligence as social engineering techniques are described. The multiplicity of applications and interdependencies of the digital techniques reviewed should show that big data and marketing intelligence have already become a marketing reality. What happens in marketing practice concerning marketing Intelligence? It will be shown that digital marketing undergoes a shift from digital technologies to social engineering practices and vice versa. Thus, snapshots of the digital marketing evolution will be taken. They will show that marketing intelligence has already developed beyond its “embryonic status.” [1]
3. Particularly marketing techniques like “search engine marketing” and “content marketing” are currently shaping IT Marketing as well as online and social media marketing. They also have an effect *beyond* the online world. Their impact on marketing practice reveals a new paradigm of marketing, known as Marketing 4.0, which means an integration of value-based marketing and digitalization.

By utilizing and reflecting on literature regarding marketing techniques

and methods such as the applied research method of this contribution, it becomes clear that we are witnessing a *paradigm shift* that augments and amplifies traditional outbound marketing with inbound marketing.

## II. MARKETING INTELLIGENCE AS DIGITAL MARKETING TECHNIQUES

### A. Introduction of Marketing Intelligence from a Digital Point of View

This section introduces marketing from a digital-technological point view, following the popular digitalization processes of corporations. This review thus also describes the current landscape of digital marketing.

Digitalization is supposed to be “unavoidable” [2], i.e. it becomes the foundation of applied marketing. With regards to digital marketing in combination with web-based services, marketers are experiencing an epoch “(...) of technology explosion and easy access to it by consumers (...)” [3]

#### 1) Marketing Intelligence as IT-marketing and Big Data Marketing

The introductory part of this paper concentrates on the term “marketing intelligence” in order to trace the path from digital data techniques to social engineering.

The term “intelligence” is anything but new. As early as the 1960s, Kelley formulated the significance of marketing intelligence for management in the age of the information revolution. Even then, an increase in data volumes had been diagnosed [4]: Market information can be compared with that of military intelligence services. Top management staff must set up information channels at all data collection points. He proposes to create a novel position of “vice president of intelligence service” and emphasizes the importance of the

\* Corresponding author.

E-mail address: jan.lies@jan-lies.de



data for economic research, marketing research, market information, and management information. This information must then be verified and validated, i.e., checked and evaluated, for its significance and relevance for a respective company's divisions.

Today – almost 60 years later – marketing 4.0 is the term used to discuss digitalization in marketing, particularly in IT marketing. This term will be the topic of the second section of this contribution (see III.9), where marketing intelligence as social engineering will be discussed, as marketing 4.0 doesn't refer to a specific digital marketing technique, but expresses a social application of it.

Basically “the term marketing intelligence refers to developing insights obtained from data for use in marketing decision-making. Data mining techniques can help to accomplish such a goal by extracting or detecting patterns or forecasting customer behavior from large databases.” [5] Traditionally data research relied on market surveys to do consumer research. With the analysis tools of big data, key factors for marketing decisions can be automatically monitored, e.g. by mining social media data. [5]

The implementation-oriented beginning of IT marketing could already be witnessed in the 1990s. IT Marketing refers to the analysis, planning, and implementation of marketing with the help of information technology (IT) and with that the current increase of web-based applications. In the 1990s, many companies began to transfer customer information to electronic databases. At the end of the 1990s, the CRM (customer relationship management) debate arose, often with the goal of integrating sprawling data collection and, from today's perspective, marking a starting point for digitalization in marketing. However, this had already begun before with desktop publishing and digitalized media production at the end of the 1980s and continued with the introduction of call center technologies. [6] Email marketing from around the 1990s continued this development even further. [7]

At about the same time, business and *marketing intelligence* was being discussed. Jenster/Solberg Soilen describe the traditional paths of intelligence research. This originally began with military information in the 1960s in the US and today shapes a primarily technical-analytical data perspective. [8]

The status quo of marketing intelligence research is still considered to be in its infancy and occasionally even described as “embryonic”. [9] The implementation of marketing intelligence systems in practice will depend on whether and how quickly it is possible to find a way out of the complexity of economic statistics, IT and technocracy and to transfer it into simple applications, so that *pragmatics* [10], i.e., pragmatic analysis methods, find their way into the breadth of marketing practice. Here, applications are making rapid progress in the form of cloud, i.e., web-based, marketing software.

The following overview of marketing techniques and methodical marketing influences will show that marketing intelligence has long been an implicit office standard, irrespective of specific big data solutions, systems or projects. Marketing intelligence is already an *implicit* standard, because the term *intelligence* is often not even mentioned in science and practice, but nevertheless impacts marketing practice as social engineering.

## 2) Marketing Intelligence as Search Engine Marketing

Search engine marketing and the content marketing based thereon will show in the following that the marketing paradigm, at least in online marketing, is currently determined by an outside-in approach. The term “digital marketing” has developed with the spread of digital communication instruments, channels, and processes [11] and uses digital customer data to secure and implement marketing decisions.

“(…) Search Engine Optimization or SEO is essentially tweaking your website so that it comes up naturally or organically in search results

on Google, Yahoo, Bing or any other search engine.” [3] In recent years, what search engines discern as *good* content has become more and more similar to what a human internet user discerns as good content. In the past – until around 2010 – it was sufficient to place keyword-optimized texts with a keyword density of five percent on a simple website and then buy masses of backlinks on the web. Since the end of 2010, Google has been using various algorithm updates to conduct content quality offensives, and has implemented sanctions including harsh penalties for misdemeanors by listing websites with lower or no hits at all.

The basic technologies for recognizing good search results are *artificial intelligence* (software that can perceive, justify, act and adapt), *machine learning* (generic term for all processes that enable machines to generate knowledge from experience) and *deep learning* (working with artificial neural networks to achieve particularly efficient learning successes). [12] Hence, here we experience a popular example for the convergence of digital and social marketing techniques.

## 3) Marketing Intelligence as Social Media Marketing and (Micro) Targeting Tools

There are many more digital marketing techniques which use big data and lead to new interaction of marketing and customers: Social media describe a variety of online channels and platforms. They are able to facilitate collaborative creation and dissemination of information and provide the basis for social engineering techniques (see chapter III). They include forums and message boards, review and opinion sites, social networks, blogging, microblogging, bookmarking, and media sharing. [13]

From the marketing intelligence point of view, social media platforms are *self-seeding* data sources, as the users provide data by posts, share, likes, comments, photos etc.: “Every social media platform provides its own analytical data to help businesses make decisions about advertising spending and to help them choose their particular advertising tactics.” [14] Social media metrics include the opportunities to identify audiences, emerging business relevant topics, brand perceptions, user attitudes and behavior, market segments etc. [13]

*Targeting technologies*, in collaboration with *mobile technologies*, enable marketers to identify and address specific customer communities and/or customers personally. This procedure is known as “mass customization” which refers to group-related or even individualized mass production: the individualized car, the personally branded chocolate or the personally configured sports shoe. That means the consequence of *micro targeting* is *micro marketing*. [15] In the specified application of *geo marketing*, micro marketing involves small segments, down to the personal level (segment of one).

As early as 1990, the growth of micro marketing was recognised, resulting from a fragmented society and increasingly saturated markets: “Consumers do not form a nationwide mass, but are a product of their region or neighborhood.” [16] With *mobile and sensor technology* in particular, advertising posters can directly recognize and address the passing public, for example in shopping malls, in front of shop windows, in cinemas or at bus stops. Micro marketing is primarily implemented through *mobile marketing* and through *location-based services*. [17] McKinsey sees the micro-marketing strategy as a development with strong potential. [18] Hence, micro marketing is on its way to changing from a digital marketing technique to social engineering.

## 4) Marketing Intelligence as Real-Time Marketing

Big data is also associated with analysis, evaluation and (re-)action in real time, which is one of the core capabilities of digital marketing techniques. The term *timely*, in this case, represents the implementation within milliseconds thereby enabling marketing techniques: Real-time marketing is based on digitalized marketing processes. *Streaming analytics* is an identification of analysis methods for data streams in

real time. [19] Its goal is to process real time data so quickly that the analyses are also available in real time, i.e. *live*. Streaming analytics systems must meet high demands with regards to the throughput speed of the data, in order to guarantee fast evaluations. Streaming analytics record the routines of credit card use in real time for example: When customers proceed to the checkout to pay at a supermarket, they slide their card through the card reader or place their order in the online shopping basket.

These live-features are new for traditional marketing as *outbound marketing*, as the technical possibilities of web solutions have increased the expectations of marketing: “Real-time Marketing depicts the real-time response to customer demand, which is technically representable with real-time advertising, real-time bidding (online auction prices for advertising) and other applications.” [20]

Current applications of real-time marketing also include, for example *social listening*, i.e. the recording, analysis, and evaluation of posts, shares, and likes in social media to capture the mood and the opinion of the media as a form of real-time marketing research. In addition, *advertising and ad placement* proceed live: real-time bidding similar to an auctioneering principle for determining the price of advertising space as well as recommendation marketing and dynamic pricing are examples of real-time marketing as a digital marketing technique. The consequences of real time abilities are to be shown later, when marketing intelligence is discussed as social engineering (see chapter III.2).

### 5) Marketing Intelligence as Recommendation Marketing and Dynamic Pricing

Product recommendations, which are determined on the basis of the data of user behavior of online shop visitors, are already used particularly frequently. *Recommender systems* are intended to reduce the complexity of the choice process. They generate personalized predictions about product liking, by filtering the past behavior and preference statements of consumers. [21] The social engineering technique, where trust is exchanged for reputation, became a digital version known as *digital recommendation marketing*, so that digital techniques and social measures are able to become two sides of the same marketing coin.

Another technique with which to use digital customer data are *automated price adjustments*. “Dynamic price adjustment is common practice on the web. Amazon is considered a pioneer in that regard. The mail-order giant changes its prices countless times a day. Users accept this as long as the offer and service are right.” [22] The approach of multiple price changes per day is also known in the offline world: for example, in petrol stations. In the course of cross-channel marketing, i.e., the product and price policy integrated across channels, retailers will increasingly adapt their prices to developments on the internet in the real world as well. In consumer electronic stores, for example, digital price tags are increasingly being placed on shelves to serve this development. Hence, marketing intelligence impacts marketing decisions here and they in turn impact customer decisions. Therefore, we experience a snapshot of an evolving digitized-social system that impacts itself here.

### 6) Marketing Intelligence as Mobile and Proximity Marketing

*Mobile devices*, especially the smartphone, accompanied by the *semantic web* (web 3.0) enable web-based services to interact by voice recognition with web users and made marketing *mobile*. The mobility itself is not new: Listening to the radio has also migrated out of the home, as consumers are used to listening to radio in their cars. [23] What is new in the digitized marketing era is content-specific, individualized, visual and/or conversational and geo-located marketing, and thus customized to one-marketing, as smartphones are usually equipped with GPS (global positioning system). Digital mobile

devices enable marketers to provide digital *location based services*. [24] This development marks a paradigmatic technological shift in marketing, sometimes characterized as a “*mobile revolution*”. [25] Marketing began to *individually* accompany the customers almost everywhere: in the car, at work, in the restaurant, into the fitness club etc. Thus, *mobile (social media) marketing* is the digital platform for influencer marketing (see chapter III.2) and a source of big data.

*Proximity technologies* can automatically analyze significant touchpoints as parts of the customer’s journey, which are both social engineering techniques (see chapter III.5). *Nearfield communication* (NFC) is an important service and at the same time, an important data source for the requirements of proximity marketing. NFC is a further development of RFID (radio-frequency identification). This is a radio standard for wireless data transmission, which is the communication between two elements e.g., mobile phone and cash register, located close to each other. This technology handles the interaction between the customer’s smartphone and the point of sale. [24] Accordingly, proximity technologies frame the backbone of mobile touchpoints within the customer journey (see chapter III.5).

### 7) Marketing Intelligence as Semantic Marketing

“The arrival of the Semantic Web represents a revolution for the form of access and storage of information.” [26] From a marketing strategic point of view, a crucial change from web 2.0 to web 3.0 is the recognition of “*meaning*”. “The benefit of adding semantics consists of bridging nomenclature and terminological inconsistencies to include underlying meanings in a unified manner.” [26] Also, voice recognition technologies of the smartphone or in the car depend on the ability of semantic technologies.

A central popular function of the semantic marketing became *chatbots*, a blending of words derived from “chat” for conversation and “bot” for robots. Chatbots as voice robots interact with people using spoken or written text. In the messenger service, they enable direct offers in the context of a chat or a message. These are virtual assistants as computer-based, virtual conversation partners or more precisely: intelligent voice-controlled applications, e.g. search or purchase services by voice. They belong to the so-called *semantic web*, i.e., “Web 2.0 + artificial intelligence”. [27] Until now, chatbots have mostly been using stored language rules or FAQs (frequently asked questions) in the form of databases that reacted to certain rules, such as keywords. Questions were categorized, and corresponding answers were prioritized. In the meantime, there are also chatbots who learn independently via machine learning. They have artificial intelligence and can learn semantics in the respective context of conversations and are thus a part of marketing intelligence. This development is based on four pillars: artificial intelligence, speech recognition technology, messaging services, and virtual assistants. [28] Chatbots and open API are two integral parts of conversational commerce [29]: *Open API* (application programming interface) are open programming interfaces which allow external providers to integrate content and services into the messaging app. Data can thus be exchanged and processed barrier-free without the user having to switch to another application or install additional apps. The internet, thus, becomes an essential tool for finding out what is happening at the moment, what the competition is doing, what customers are demanding, or even discovering technological trends, innovations and expert opinions. [30]

### 8) Marketing Intelligence as Predictive Marketing

*Predictive intelligence* is a method for machine learning, which is already known from image and face identification as well as speech recognition and language translation and is now finding its way into digital marketing. For example, insurance companies calculate potential predictors for driving safety by determining a person’s age, gender and driving experience, e.g. derived from the *geo data* of drivers, which

foreshadows the rising meaning of mobile technologies (“mobile first”). From this, the insurance company then determines the probable risk of accident and the amount of motor vehicle insurance. Predictive pricing will probably become more and more popular especially in the context of customized mass services.

The “predictive” approach initially is nothing new, not even in marketing, if one thinks of the fields of action of marketing and trend research, which always include prognostic tasks. What’s is new however, is the fact that today there are standard technologies and applications on the market which can also be used by companies without data scientists, market researchers or software programming. [31]

### 9) Marketing Intelligence as Touchpoint Marketing and Marketing Automation

Customers leave digital data: from shares, likes, and posts, for example, to a branded product, from the search behavior in the shop, to payment transactions or cancellations and mobile movement in shopping malls and/or the point of sales. The analysis of customer retention times, search paths, contact points, queues at checkouts also make proximity marketing a field of action for marketing intelligence to derive assortments, shelf campaigns, check out openings or newly arranged product presentations from. For example, the longer a customer stays in a certain zone of a store or the customer heading for goods islands, info counters or departments, etc. can be interpreted as “exploration” of the goods on offer presented there. Proximity can support every touchpoint, from the first perception of a product in the shopping street, to evaluation, purchase, repurchase, and recommendation. Digitalization of marketing processes implicates the digitalization of touchpoints, i.e. interfaces between customers and products or services, which frame customer decisions. [32]

The opportunity of touchpoints within the customer journey leads to the integrated digital *marketing automation*. [33] The core value of automation is the repeatability of digital marketing measures in which people do not have to intervene. “What does automation mean to *marketing people*? (...) New methods of automation will make marketing information available much faster than before and provide data that was not available before.” [34] What sounds like a rather recent statement is a quote from the 1960s. It shows that marketing has always included an attempt at standardization. This foreshadows for example the term *programmable advertising*. It describes the data-supported trading of online advertising spaces. Campaigns can be booked fully automatically and can thus be adapted in real time to the situation of (mobile) users. Campaigns and messages can be adapted in real time to the needs and communication channels of consumers. [35]

### B. Summarizing Discussion

Looking back upon the selected digital marketing techniques described above (see summarizing Table I.), it is ultimately the *search engines* and, in this case, big data in the form of data on the search behavior of internet users, that are driving the current significance of content marketing and inbound marketing, as the methodical marketing discussion below will show.

The open question is, in how far these digital techniques become *marketing practice*. To answer this question this paper focuses in the following on currently utilized marketing practices. This resorts to the digital marketing techniques as *applied social engineering*.

## III. MARKETING INTELLIGENCE AS SOCIAL ENGINEERING

### A. Introduction of Marketing Intelligence from a Social Engineering Point of View

“Technology is changing the context of and practice of marketing.”

[37] Currently marketing practices are experiencing the shift from “big data to big impact.” [38] Applied marketing practices can be interpreted as *social engineering*. Social engineering is frequently discussed and often associated with negative, often politically motivated interventions impacting individual behavior. In contrary, sometimes social engineering is something positive, e.g. if it is seen as a kind of art: “Social engineering is the art of getting users to compromise information systems.” [39] Therefore, marketing intelligence becomes “artificial intelligence”, not just in a digital, but also in a social sense. A variety of definitions of social engineering exist and certain elements are agreed upon. One of these agreed aspects is that social engineering means a set of applied methods for social impact and thus, social change. [40] This means that social engineering are practices applied to influence people. Hence, marketing, PR-management, advertising etc. are examples of social engineering.

In the following, methodical shifts in marketing are identified. For this purpose the described marketing techniques above are used as applied social engineering. It will be discussed that this also leads to methodological shifts in marketing. Above all, search engine marketing technology emphasizes the current importance of *inbound marketing* as a method of market cultivation. This is in contrast to *outbound marketing*, which has long made the *product* the most important *p* in the marketing mix. However, from a current perspective, *customer needs* should stand in the center of this view as a methodical paradigm, if one regards customer management and customer relationship management as a central field of action in marketing.

### 1) Marketing Intelligence as Content Marketing

*Content marketing* is teeming with marketing measures to generate and provide content. As the expression “social media” emphasizes, these media are “social”. Of course, every used media is “social” as they are social phenomenon. But in contrast to TV, radio or the press just the social media are *explicitly* named “social”. Classic media and social media obey *distinctive logics*. Mass media provide a flow of topics by organizing linear programs to gain attention. [41] The key factor for the success of social media is not just transmitting information, but conversation and, thus, interaction. “Social media is not only a place to market the products and services of a company, but also a place to interact with the customers”. [42] Hence, the logic of classic media obeys an *inside-out* approach by transmitting information to gain attraction in contrast to the logic of social media, which concept is configured “*outside-in*” on the basis of conversation in order to gain *interaction*.

If we trace the evolution of social media, we currently experience a transformation process within (online) marketing across media. Traditionally marketing had been placing ads to gain attraction. This manner of “*paid media*” is also to be found in social media. But from the user’s point of view, social media becomes attractive because of its interesting, entertaining and sharable content. This content is the basis for so called “viral processes”. As soon as these social interaction-processes become vivid, these media are used to call “*earned media*” as far as a specific brand, product or corporation is subject of the viral processes. The ownership of content is the corporation and/or the *recipient*. Thus, viral effects within social media are known as “earned media” as they are distributed at no direct cost to the corporation. [43] “Sharing online content is an integral part of modern life.” [44] To achieve viral effects, studies suggest providing content which activates and evokes emotions because such content is likely to be shared. [44] Thus, we experience a *transformation process of marketing*. [45] Today social media is recognized as a *hybrid element* of the promotion mix: In a traditional sense, it enables companies to talk to their customers. In a *nontraditional* sense, it enables customers to talk directly to one another. [46]

While classical marketing instruments draw the attention of consumers to the *product* directly, content marketing focuses more on



TABLE I. POPULAR DIGITAL MARKETING TECHNIQUES

Marketing intelligence	Idea
Big data or data driven marketing	Big Data or data driven marketing is the systematic alignment of marketing measures with a digitally recordable target, process, and data result, as well as the corresponding analysis results based on <i>digital data</i> .
Search engine marketing	<ul style="list-style-type: none"> <li>Placing contributions in the search results of <i>search engines</i> such as Google as high up as possible.</li> <li>Since Google has a market share of over 90 percent, search engine marketing is currently actually google marketing.</li> </ul>
Social media marketing and personalization	<ul style="list-style-type: none"> <li>Social Media as big data sources.</li> <li>Social media data as a <i>digital farm</i> seeded by the users themselves.</li> <li><i>Personalization</i> is the adaptation of information, services or products to the defined or presumed needs of a person. The adaptation can take place on the basis of the person's profile, the current situation of the person, but also through active "personalization" by the user. [36]</li> </ul>
Mass customization and (micro-) targeting	<ul style="list-style-type: none"> <li><i>Targeting</i> means the formation of target groups through (market) segmentation. On the basis of the evaluation of mass data, it deviates from conventional target group formation.</li> <li><i>Micro-targeting</i> as big data targeting or data-based marketing refers to the evaluation of large amounts of data, the definition of special target groups and, in marketing, the (quasi-) individualized approach to each individual customer with personalized advertising content.</li> </ul>
Recommendation marketing	<ul style="list-style-type: none"> <li>Simple <i>recommendation systems</i> already work when a webshop operator simply stores a number of products as recommendations when visitors search for certain other products in his shop.</li> <li>Marketing intelligence refers to recommendations that are based on the analysis of mass data.</li> </ul>
Dynamic pricing	<ul style="list-style-type: none"> <li>Marketing intelligence is also partly applied to price by <i>optimizing the price</i> on the basis of mass data, especially in online trading.</li> <li>This turns marketing intelligence into <i>pricing intelligence</i>. It refers to automated price optimization in real time, so-called repricing.</li> </ul>
Mobile and proximity marketing with location based services and nearfield communications	<ul style="list-style-type: none"> <li>Proximity communication refers to the part of <i>mobile</i> corporate communication or product communication at the location of a company, e.g., in retail outlets, shopping malls or stadiums.</li> <li><i>Location based services (LBS)</i> are applications of mobile commerce that generate added value for the user through localization. Here, a combination of mobile image management and marketing with call-to-action, i.e., appeals for action such as the redemption of discounts, takes place.</li> <li><i>Nearfield communication (NFC)</i> enables services as cashless payment.</li> </ul>
Semantic marketing	<ul style="list-style-type: none"> <li><i>Voice recognition</i> systems enable digital conversation.</li> <li><i>Web 3.0</i> as semantic web enables real time conversation.</li> </ul>
Predictive marketing	<ul style="list-style-type: none"> <li>Predictive marketing depicts the application of analytics to calculate <i>probabilities of occurrence</i> of events and the application of marketing measures from them.</li> <li><i>Predictive analytics</i> is a method used to identify recurring patterns in data and then use sophisticated algorithms to predict their future development, e.g. with predictive prices.</li> </ul>
Touchpoint marketing	<ul style="list-style-type: none"> <li>Collecting, analyzing and optimizing the data customers leave at any <i>touchpoint</i>: web searches, online purchases, geo data, payments, posts etc.</li> </ul>
Marketing automation	<ul style="list-style-type: none"> <li>Marketing automation currently refers primarily to the software-supported networking of defined marketing channels with the goal of planning and implementing marketing as well as public relations-management and sales activities in a cost-efficient and time-optimized manner.</li> <li>One application of marketing automation is <i>programmatic advertising</i>. The term describes the data-supported trading of online advertising spaces.</li> </ul>

*publishing by sharing* media content. From a methodological point of view, the term *content* today thus stands less for information, but for "edutainment" i.e., appealing, playful and entertaining communication to trigger viral effects, especially in social media. With regard to implementation, this refers to the preparation of former advertising messages in the form of texts, images or videos, which are intended to increase awareness regarding the special interest of website visitors. The classic advertising marketing communication ("appeal to buy") is currently considered *unsuitable* here.

## 2) Marketing Intelligence as Inbound Marketing and Soft Selling

Studies reveal the effect "from social to sale" or with other words from "hard to soft selling": "The clear messages from our study are that social media marketing matters and that managers should embrace it to communicate and nurture relationships with customers." [47] Marketing intelligence emphasizes the trend of *inbound marketing*. It is closely linked to content marketing, as social media marketing does not tolerate *direct appeals to buy*. "Blogs as a tool for digital marketing have successfully created an impact for increasing sales revenue, especially for products where customers can read reviews and write

comments about personal experiences." [3]

The current prominence of content marketing is a result of (mobile) marketing intelligence. Instead of keyword management for search engine optimization, content is important nowadays. Companies that engage in search engine marketing are dependent on content to ensure that search engines display their sites as high up in the hit lists as possible. Hence, online marketing as search engine marketing triggers a new technique of social engineering: *digital inbound marketing*. Instead of interrupting advertising, it is based on interactivity and commitment, i.e., *pull marketing*, as it relies on the initiative of interested parties. Here the currently popular role of *influencer marketing* within social media marketing should be mentioned. For marketing, this means considering four fields of action: content provision, search engine optimization, social media marketing and brand-oriented marketing communication. [48]

The *mobile and customized technology* driven opportunities are also a precondition for the rise of inbound marketing, as many customer decisions are made whilst out and about. "Mobile advertising seems to have raised considerable interest as mobile technology has advanced." [49] Mobile micro marketing and recommender marketing offer the ability to react in *real-time*. This real-time competence is reinforced



by the *expectations* of social media users, who expect timely (re-) actions from companies. Here we experience the *interdependency* of the technological and shifts in applied marketing communications becoming not just a social engineering technique, but having a *cultural-sociological impact*: Social Media shapes the generation Z and thus its customer behavior (“generation facebook”). [50] Mobile phones and social media are two examples of big data farms steadily seeded by the users. Taking the meaning of content and/or social media marketing as a given, some argue that *traditional marketing* is no longer a viable option as it works with *push messages* in line with traditional transaction marketing. [48]

### 3) Marketing Intelligence as Creative Marketing

The meaning of content marketing leads to the impetus for *creative marketing* driven by marketing intelligence, as well as the consequences for automation and evaluation.

Events such as the Marketing Convention 2016 in Munich titled “Man or Machine: Marketing between Creation and Automation” [51] addressed the problem of the potential for standardization and automation, juxtaposed with the individualization and service potential of digitalization: on the one hand, the potentials are emphasized from a marketing perspective with cost savings, through standardization by automation. On the other hand, there is new potential for success through creativity, new products, and services, such as the aforementioned fraud warning for bank customers, based on the evaluation of transaction anomalies.

Creativity can be described as the creation of new and useful ideas. From a results perspective, creativity can mean new products and services, so that creativity leads to innovation and at the same time, can just *be* creativity. Creativity is thus defined both as a *characteristic* and a *process*, but also in terms of the *environment, culture or results*. [52] From the perspective of marketing intelligence, it can be used wherever people come together with their mobile phones and thus generate data for creative potential. Accordingly, marketing intelligence is addressed as *collective intelligence* for solving creative problems. [53]

As mentioned above, the digital technique “targeting” stands for the digital search with the aim of addressing customers, with the targeted control of mobile and situation-dependent identification of context-dependent target groups. It enables *location-dependent communication*: Advertising for the same brand in the respective context of the location, for example, sports advertising in stadiums or taxi advertising after busses have been delayed and the delay has been digitally recorded online by the local transport companies. Another opportunity for creative communication is provided by *weather-dependent communication*: the department store’s umbrella advertising for globetrotters in the shopping mile, the wellness hotel’s discount campaign for North Sea vacationers whose holiday season is cold and wet. This (creative) potential is supported above all by *mobile devices* (i.e., such as notebooks, smartphone/-watch, tablets) and/or other specific media (such as connected car, social media, displays) as mentioned above, when mobile marketing was introduced.

Mobile marketing intelligence is also understood as the *art of creating attractive customer experiences* at the right time: a new marketing capability with digital data for digital communication channels. [54] Creative marketing combines creativity with innovation [52] by also including the *crowd* as creative potential in value creation. This foreshadows the opportunities of *innovation* by creativity.

### 4) Marketing Intelligence as Innovation Marketing

Marketing intelligence is considered to have great potential in the development of new products and services and is labeled as *open source intelligence*. [55] Accordingly, many terms are entwined around digital innovations and product developments, which are referred to as *data-*

*driven innovation, user-driven innovation, event-driven innovation or social r&d* (social research and development). With the combination of more competence regarding technological data analysis and interaction orientation, a study of around 150 US companies by Trainor et al. showed that both relationship orientation and marketing intelligence competence have a positive influence on new product development. [56] Of course, these innovation approaches don’t necessarily need big data or intelligence to operate. But the more data is available to support innovation, the more useful insights may be identified, e.g. with the evaluation of data as portals for product improvement, data of crowdsourcing for the development of new ideas by the community etc.

### 5) Marketing Intelligence as Customer Journey Marketing

The customer journey describes the *journey* of a potential customer through various contact points (so-called “touch points”) with a product or service, a brand or a company, from inspiration and the realization of needs, to the procurement of information, search and final action. It could be seen as an enlarged customer decision process.

The social engineering technique arising from the mainly digital shaped touchpoint management is frequently named as the *customer journey*. The social relevance of the customer journey becomes very obvious when it differentiates between brand owned, partner owned, customer owned and social/external owned touchpoints. [54] It is intensively dealt with, especially in online and mobile marketing, but also includes the offline world.

The customer journey in online marketing is preferable to digital marketing because it is particularly easy to measure the success of established contact points with tracking technologies, by investigating the click and search behavior of internet users. But the “dichotomy of online–offline” today largely is outdated. Instead, the mapping of touchpoints is yet to be examined in the light of consumer decision-making, i.e. the integration of relevant channels is crucial to understand customer shopping behavior. [57]

### 6) Marketing Intelligence as Conversational Marketing

The touchpoint *voice* will develop into a central data source in marketing. Some experts are talking about a *voice first revolution* - i.e., in addition to “mobile first” *give way for voice services*. [58] This development leads to conversational commerce, i.e., to conversation-based commerce, which leads particularly to the extension of web 3.0. The term *conversational commerce* is partly attributed to Messina, who already used this term in 2015 and declared 2016 as the year of conversational commerce. [59] [60] However, the term “commerce” does not go far enough because service marketing, in particular, is benefiting from this development.

“Instead of tormenting themselves on their PC at home with ordering and payment processes, customers will soon be booking flights, ordering food, ordering concert tickets or buying shoes via voice control on their smartphone. What’s more, they will also use their messaging app to process claims, make use of services and receive real-time advice: Conversational Commerce, customer interaction with the help of artificial intelligence.” [29] In essence, conversational marketing is a language-based development of dialogue marketing in the sense of long-term customer interaction. [61]

### 7) Marketing Intelligence as Customized Lifecycle Marketing

The *life cycle model* is often found in marketing: Companies, products, customers - they all go through lifecycles, derived from the biological processes of existence. Subject to the regarded stage of the cycle, corresponding marketing action is to be applied. In contrast to the product lifecycle, the customer lifecycle emphasizes the *individual* customer approach. While the product life cycle, which is often contained in conventional marketing textbooks, follows standardized mass marketing, the customer life cycle emphasizes individual,

personalized customer interaction.

Big data supports the understanding of customers' life cycle and behavior. "Customers voluntarily generate a huge amount of data daily by detailing their interest and preference about products or services to the public through various channels." [62] Profiling for customers becomes important for business to make sure that the whole CRM life cycle (sales, marketing, and customer service) is offering personalized and customized services, so that each customer will have a different experience according to their needs and interest. [62]

Lifecycle marketing is also known as a field of action in CRM. Its goal is to build long-term customer loyalty to the company and its employees by considering the various specific life situations and phases of the customers. [63] "CRM represents (...) in relation to the customer markets a radical departure from the classical marketing concept, which in the sense of effective sales management has a very strong instrumental function regarding the dominance of product and communication." [63]

### 8) Marketing Intelligence as Performance Marketing

The creative potential of big data mentioned above methodically stands in contrast to the approach of *performance marketing*: With the current significance of content marketing, one could assume that soft goals such as web authority, i.e., the resonance through linking, commenting, etc. of the community, would become the focus of marketing intelligence. At the same time however, the ability to *measure the performance* of marketing rises. For example, digital campaigns with multivariate analysis methods and experimental research designs can be used to determine which dimensions make them particularly successful. For example, in terms of customer value along the life cycle of Search Engine Marketing at Google. [64]

(Digital) key performance indicators are gaining new attention, as marketing is coming under pressure to succeed. [65] "Marketing practitioners are under increasing pressure to demonstrate their contribution to firm performance." [66] Studies reveal that some companies emphasize short-term financial metrics at the expense of measuring long-term impacts. [67] "Performance Marketing in its purest form is purely success-oriented. Successful campaign modules (texts, keywords, tools, advertising media) are accelerated and expanded. Less successful ones are optimized and eliminated if the defined goals are still missed." [68] As creativity or content belong to the idea of inbound marketing or "soft selling" a conceptual gap of

measuring success arises between Marketing 1.0 and Marketing 4.0.

### 9) Marketing Intelligence as Marketing 4.0

Following the development of industry 4.0 with the networking of man and machine, a marketing 4.0 debate is taking place on the performance and cost-effective availability of (mobile) internet. Looking back, this debate shows a development that focuses on human-centric development through digitalization: [69]

- **Marketing 1.0:** The origin as a prototype puts the core competence of Marketing on a product and its distribution. Marketing activities are geared towards this, so that the market is at the center.
- **Marketing 2.0:** The focus shifts to the consumer. Companies further differentiate from each other as consumers become more self-confident (from the 1970s).
- **Marketing 3.0:** The focus is on people. They are determined by values that depend on their environment. Customer management instead of market-oriented corporate management is prevalent, as human centricity characterizes Marketing (from the 1980s onwards).
- **Marketing 4.0:** The focus here is on digitalization and thus the convergence of technologies, without losing sight of the previous stage. This means an online-offline integration (from approx. 2010 onwards).

The evolution of marketing will continue. The next version, Marketing 5.0, already is in discussion.

- **Marketing 5.0:** Expectations regarding further developments include current popular discussions like block chain and platform marketing as a part of the blockchain economy. [70] Analogue to the backbone of bitcoins block chain will provide marketing services which depend on the *trust* provided by *closed IT systems*, which are worthy e.g. for guaranteeing selected target media of programmatic advertising or to avoid fake accounts in social media. So, marketing 5.0 will probably be the epoch of *digital trust marketing*.

### B. Summarizing Discussion

Whether or not the concept and *terminology* of marketing *intelligence*, as smart data marketing on an IT basis, will hold its own or prevail is currently rather doubtful due to its technical-analytical past. It will presumably be absorbed in a multitude of possible

TABLE II. MARKETING: POPULAR SOCIAL ENGINEERING TECHNIQUES

Marketing intelligence	Idea
Content marketing	Content marketing refers to the informative, consultative and/or entertaining provision of company information with the goal of making them deal with information in the first place.
Inbound marketing/ Influencer marketing/social media marketing	<ul style="list-style-type: none"> <li>• Pull marketing as inbound marketing (outside-in thinking).</li> <li>• The content especially of social media users impact the rules of marketing wording: the shift from hard to soft selling.</li> </ul>
Creative marketing	Marketing intelligence is addressed as collective intelligence for generating creative solutions.
Innovation marketing	Open source intelligence as the source of the development of new products and services.
Customer journey marketing	Marketing impacting the journey of a potential customer through various contact points (so-called touch points) with a product or service, a brand or a company.
Conversational marketing	The current voice first revolution as an addition to the new mobile paradigm.
Customized lifecycle marketing	<ul style="list-style-type: none"> <li>• Product life cycle follows standardized mass marketing.</li> <li>• Customer life cycle emphasizes individual, personalized customer interaction.</li> </ul>
Performance marketing	Performance marketing refers above all to digital marketing measures and processes relating to online and mobile marketing measures, which can be particularly well measured and thus optimized.
Marketing 4.0 & 5.0	<ul style="list-style-type: none"> <li>• Value based marketing (3.0) and digital marketing (4.0).</li> <li>• Trust marketing (5.0) on the basis of closed systems like block chains.</li> </ul>

applications of marketing as social engineering techniques (see above/ see summarizing Table II).

Regarding the shift of marketing methodology is to state: If the prototype marketing with marketing 1.0 consists of sales-related mass-market orientation, then marketing is originally characterized by the methodological paradigm of *inside-out*. [71] The methodical approach is based on product solutions that open up markets. This is another reason why the product mix is regarded as the “heart of marketing” [72] In fact, marketing goes back to consumer goods marketing and thus to mass markets. The classic 4P marketing mix also shows that the original marketing did not focus on the customer. [73] Interestingly enough from today’s point of view, neither the customer nor the salesperson is originally considered in the conventional marketing mix, but instead places the *product* at its center. This *inside-out* approach was later criticized as being too narrow, and Hunt’s Nature of Marketing added a micro-macro-approach [74] or stakeholders beyond markets respectively [75]. Marketing 3.0 focuses on the values of customers also *beyond* markets, hence, they converge with “soft selling” approaches, which are more subtle and indirect than the traditional marketing sales. [76] “Soft selling” became the paradigm to gain earned media.

#### IV. CONCLUSION: MARKETING INTELLIGENCE AS A SHIFT FROM DIGITAL MARKETING TECHNIQUES TO SOCIAL ENGINEERING

Looking back and summarizing the evolution of digital marketing, we arrive at a multiplicity of “marketing-as...”-landscape. Table I to III retrace how marketing intelligence already has started to impact marketing practice with a big scope of social engineering techniques (see Table III).

TABLE III. THE LANDSCAPE OF DIGITAL MARKETING TECHNIQUES AND DIGITAL MARKETING AS SOCIAL ENGINEERING

Marketing as digital techniques	Marketing as social engineering
<ul style="list-style-type: none"> <li>• Big data or data driven marketing</li> <li>• Search engine marketing</li> <li>• Social media marketing and personalization</li> <li>• Mass customization and (micro-) targeting</li> <li>• Recommender marketing</li> <li>• Dynamic pricing</li> <li>• Mobile and proximity marketing with location based services and nearfield communications</li> <li>• Semantic marketing</li> <li>• Predictive marketing</li> <li>• Touchpoint marketing and marketing automation</li> </ul>	<ul style="list-style-type: none"> <li>• Content marketing</li> <li>• Inbound marketing/ influencer marketing/social media marketing</li> <li>• Creative marketing</li> <li>• Innovation marketing</li> <li>• Customer journey marketing</li> <li>• Conversational marketing</li> <li>• Customized lifecycle marketing</li> <li>• Performance marketing</li> <li>• Marketing 4.0 &amp; 5.0</li> </ul>

Content marketing, for instance, today wouldn’t be so successful without the capability of search engine on the basis of big data-analysis. Of course, marketing had been benefiting from other techniques also before the beginning of digitalization, as the discussion of marketing automation in the 1960th shows (see above, chapter II.9). Nevertheless, this contribution focuses on the current shifts driven by *digitalization*.

Marketing intelligence has long since led to a paradigm shift in marketing, with the described digital techniques shifting to the social engineering techniques – even for companies that are *not actively* involved in big data initiatives. If companies aim to be listed high up on search engine search results, they need interesting content from the point of view of internet users. This *outside-in* methodology is the top of the paradigm shift in marketing methodology. Thus, we are

currently experiencing another methodical shift. “When people say, ‘There is a method to this’, it is usually supposed to mean ‘systematic approach.’” [77] The term *method* thus refers to formative paradigms and principles as a cross-instrument approach to achieving specific goals. Methodology characterizes the totality of methods. The term *method* (from Greek *methódos* = way of proceeding) describes the way of thinking and recognizing. [77]

- **Methodological paradigm shift:** Originally, marketing was solely sales-, product- or market-oriented, i.e., Marketing 1.0 and 2.0 followed the inside-out perspective. In inbound marketing, which from the company’s point of view is considered a passive approach, the initiative comes from the customer who is actively establishing contact with a company. [61] With Content Marketing, on the other hand, the *inside-out* perspective has been supplemented by the *outside-in* perspective. This type of marketing is actually *PR management*, in that appeals to buy are inadmissible and instead entertainment-centric, humorous, target-group-oriented content is demanded by the stakeholders. [77] It is currently unclear whether and how intensely this paradigm shift will shape offline activities as well. One could now jump to the conclusion that customer needs and the value orientation of *marketing 3.0* have won (see Table IV). – However, this would be an incorrect assumption, as so-called *performance marketing* is also being introduced at the same time within marketing 4.0: This refers to data-supported and automated digital marketing as a particularly efficient form of marketing. With marketing automation as a marketing technique, it can be assumed that many services will fall victim to the superficial measurement of marketing efficiency. At the same time, however, big data marketing will also create new *creative potential*. The integration of customers into value creation is not new if one thinks of furniture giant Ikea, which leaves the final assembly of their products to their customers. But the multitude and bandwidth of interactive marketing leads to cooperative added value: the evaluation of orders and customer wishes for the development of individual order options for otherwise standardized products and/or services up to the so-called customer as co-designer of the fast food menu or sneaker.

TABLE IV. METHODOLOGICAL CORNERSTONES OF OUTBOUND AND INBOUND MARKETING [78]

Key Points	Traditional Marketing	Inbound Marketing
Basis	Interruption	Organic Interaction
Goal	Interaction as new customer acquisition through short-term sales increases	Interaction as requests from new interested parties for the development of long-term customer relationships
Method	Push Marketing as Outbound Marketing (inside-out thinking)	Pull Marketing as Inbound Marketing (outside-in thinking)
Target	Mass markets, groups	Fans, individual
Location	At home	At home and mobile
Time	Planning and approval processes make marketing slow	This real-time competence is reinforced by the expectations of social media users, who expect timely (re-)actions from companies.
Instruments	Conventional advertising	Blogs, viral communication, etc.
Sales approach	Hard	Soft
Marketing X.0	Marketing 1.0 & 2.0	Marketing 3.0 & 4.0

- **Technical paradigm shift:** Digital marketing based on big data



is more than just a multitude of digital marketing tools. Real-time marketing with its emphasis on reaction speed and digital marketing processes – from search to purchase – can be implemented automatically along the customer life cycle. Thus, process-related marketing automation takes place: Automated, cross-channel campaign management in real time and/or along a customer's life cycle by addressing the customer through cloud marketing via e-mail, WhatsApp, banner advertising on the Web or smart TV advertising along the customer journey. It is even possible to use *micro marketing* with market segmentation at the *individual* level. This can be realized in *real time* as *proximity marketing* of the retail branch, and thus describes the part of the mobile, corporate or product communication in or at the location of a company, for example in the retail trade, at the shopping mall or the football stadium. This also includes, for example, price communication, in other words, the evaluation of the search behavior for price setting depending on the online search profile of the passing customer.

- **Cultural paradigm shift:** Some experts pretend that with the availability of powerful data analysis technology, simplified (marketing) decisions are also available. In fact, *data dissemination*, the dissemination and availability of *smart data*, is likely to be a critical point for the success of marketing intelligence applications. The basis and consequence of this is a *corporate culture* which first turns to these technical innovations. Intelligence dissemination must, however, be organized. Contact management between data managers in marketing and those who can and should use their data – for example in sales, distribution or purchasing – is only one source of potential success. [79] Success in marketing intelligence is therefore also dependent on *internal company social networking*. In this respect, big data marketing requires a data-related knowledge culture in many companies.

## REFERENCES

- [1] A. Amado, P. Cortez, P. Rita, and S. Moro. (2018), "Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis", *European Research on Management and Business Economics*, Volume 24, Issue 1, January–April 2018, pp. 1-7, doi: doi.org/10.1016/j.iedeen.2017.06.002
- [2] Y. Durmaz, I.H. Efendioglu (2016), "Travel from Traditional Marketing to Digital Marketing", *Global Journal of Management and Business Research: E-Marketing*, Volume 16 Issue 2, pp. 35-40.
- [3] M. Bala, D. Verma (2018), "A Critical Review of Digital Marketing", *International Journal of Management, IT & Engineering*, Vol. 8 Issue 10, October 2018, pp. 321–339.
- [4] W.T. Kelley (1965), "Marketing Intelligence for the Top-Management", *Journal of Marketing*, Volume 29, October, pp. 19-24.
- [5] S. Fan, R.Y.K. Lau, and J.L. Zhao (2015), "Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix", *Big Data Research*, 2, pp. 28–32, doi: dx.doi.org/10.1016/j.bdr.2015.02.006.
- [6] P.R. Prabhaker, M.J. Sheehan, and J.I. Coppett (1997), "The power of technology in business selling: call centers", *Journal of Business & Industrial Marketing*, Vol. 12 Issue: 3/4, pp. 222-235, doi: doi.org/10.1108/08858629710188054.
- [7] M. Hartemo (2016), "Email Marketing in the era of the empowered consumer", *Journal of Research in Interactive Marketing*, Vol. 10, Issue 3, pp. 212-230, doi: doi.org/10.1108/JRIM-06-2015-0040.
- [8] P.V. Jenster, K. Solberg Soilen (2009), *Market Intelligence: Building Strategic Insight*, Gylling, Copenhagen Business School Press, Narayana Press.
- [9] A. Amado, P. Cortez, P. Rita, and S. Moro. (2018), "Research trends on Big Data in Marketing: A text mining and topic modeling based literature analysis", *European Research on Management and Business Economics*, Volume 24, Issue 1, January–April 2018, pp. 1-7, doi: doi.org/10.1016/j.iedeen.2017.06.002
- [10] C. A. Brea (2012), "Pragmalytics: Practical Approaches to Marketing Analytics in the Digital Age", Bloomington, iUniverse.
- [11] P.K. Kannan, H.A. Li (2016), "Digital Marketing: A framework, review and research agenda", *International Journal of Research in Marketing*, 34, 2017, pp. 22–45, doi: 10.1016/j.ijresmar.2016.11.006.
- [12] Y. Yuniarthe, (2017), "Application of Artificial Intelligence (AI) in Search Engine Optimization (SEO)", *International Conference on Soft Computing, Intelligent System and Information Technology*, pp. 96-101.
- [13] U. Ruhi (2012), "Social Media Analytics as a Business Intelligence Practice: Current Landscape & Future Prospects", *Journal of Internet Social Networking & Virtual Communities*, pp. 1-12, doi: 10.5171/2012.920553.
- [14] S. Lou (2017), "Applying Data Analytics to Social Media Advertising: A Twitter Advertising Campaign Case Study", *Journal of Advertising Education*, Vol 21, Issue 1, doi: doi.org/10.1177/109804821702100106.
- [15] E. Sivadasa, R. Grewalb, and J. Kellarisc (1998), "The Internet as a Micro Marketing Tool: Targeting Consumers through Preferences Revealed in Music Newsgroup Usage", *Journal of Business Research*, Volume 41, Issue 3, March 1998, pp. 179-186, doi: doi.org/10.1016/S0148-2963(97)00060-X.
- [16] S.L. Hapoieniu (1990), "The Rise of MicroMarketing", *Journal of Business Strategy*, November/December, pp. 37-42.
- [17] L.M. Hilty, B. Oertel, M. Wölk, and K. Pärli (2012), "Lokalisiert und identifiziert: wie Ortungstechnologien unser Leben verändern", Zürich, vdf. Accessed: May 2, 2019. [Online]. Available: [https://files.ifi.uzh.ch/hilty/t/Literature\\_by\\_RQs/RQ%20227/2012\\_Hilty\\_Oertel\\_Woelk\\_Paerli\\_Lokalisiert\\_und\\_identifiziert\\_BUCH.pdf](https://files.ifi.uzh.ch/hilty/t/Literature_by_RQs/RQ%20227/2012_Hilty_Oertel_Woelk_Paerli_Lokalisiert_und_identifiziert_BUCH.pdf)
- [18] M. Goyal, M.Q. Hancock, and H. Hatami (2012), "Selling into Micromarkets", *Harvard Business Review*, July–August 2012, Accessed: May 2, 2019. [Online]. Available: <https://hbr.org/2012/07/selling-into-micromarkets>.
- [19] C. Lanquillon, H. Mallow (2015) "Advanced Analytics mit Big Data", in: J. Dorschel (ed.), *Praxishandbuch Big Data: Wirtschaft – Recht – Technik*, Wiesbaden, SpringerGabler, pp. 55–89.
- [20] J. Lies (2017), "Die Digitalisierung der Kommunikation im Mittelstand – Auswirkungen von Marketing 4.0", Wiesbaden, SpringerGabler.
- [21] T. Hennig-Thurau, A. Marchand, P. Marx (2012), "Can Automated Group Recommender Systems Help Consumers Make Better Choices?", *Journal of Marketing*, 89, Volume 76, September 2012, pp. 89-109, doi: DOI: 10.1509/jm.10.0537.
- [22] A. Kalka, R. Krämer (2016), "Dynamic Pricing: Verspielt Amazon das Vertrauen seiner Kunden?", *Absatzwirtschaft online*, Accessed: May 2, 2019. [Online]. Available: <http://www.absatzwirtschaft.de/dynamic-pricing-verspielt-amazon-das-vertrauen-seiner-kunden-75271/>
- [23] S. Banerjee, R.R. Dholakia (2008), "Mobile Advertising: Does Location Based Advertising Work?", *International Journal of Mobile Marketing*, December 2008, Vol. 3, No. 2, pp. 68-74.
- [24] G. Arcese, G. Campagna, S. Flammini, and O. Martucci (2014), "Near Field Communication: Technology and Market Trends", *Technologies* 2014, 2(3), pp. 143-163, doi:10.3390/technologies2030143.
- [25] Y.B.B. Öztaş (2015), "The Increasing Importance of Mobile Marketing in the Light of the Improvement of Mobile Phones, Confronted Problems Encountered in Practice, Solution Offers and Expectations", *World Conference on Technology, Innovation and Entrepreneurship, Procedia - Social and Behavioral Sciences*, 195 (2015), pp. 1066-1073, doi: 10.1016/j.sbspro.2015.06.150.
- [26] A.G. Crespo, R. Colomo-Palacios, J.M. Gómez-Berbis, F. Paniagua-Martin (2010), "Customer Relationship Management in Social and Semantic Web Environments", *International Journal of Customer Relationship Marketing and Management*, pp. 1-10, doi: 10.4018/jcrrm.2010040101.
- [27] I. Sjurts (2011) (ed.), "Web 3.0", *Gabler Lexikon Medienwirtschaft*, Wiesbaden, Gabler.
- [28] S. Tuzovic, S. Paluch (2018), "Conversational Commerce – A new area for Service Business Development," in: M. Bruhn, K. Hadwich (eds.): *Service Business Development: Strategien – Innovationen – Geschäftsmodelle; Forum Dienstleistungsmanagement*, Band 1, Wiesbaden, SpringerGabler, S. 81-100.
- [29] K. Dörner, B. Hosseini (2016), "Conversational Commerce entwickelt sich zum Zukunftstrend – dank Messenger-Boom und Bots", McKinsey & Company, *Absatzwirtschaft*, Accessed: May 2, 2019. [Online]. Available: <http://www.absatzwirtschaft.de/conversational-commerce-entwickelt-sich-zum-zukunftstrend-dank-messenger-boom-und-bots-89619/>
- [30] F.J. Garrigos-Simon, R.L. Alcamí, and T.B. Ribera (2012), "Social networks and Web 3.0: their impact on the management and marketing of



- organizations", *Management Decision*, Vol. 50 Issue: 10, pp.1880-1890, doi: doi.org/10.1108/00251741211279657.
- [31] O. Artun, D. Levin (2015), "*Predictive Marketing: Easy Ways Every Marketer Can Use Customer Analytics*", Hoboken, Wiley.
- [32] A. Fölstad, K. Kvale (2018), "Customer journeys: a systematic literature review", *Journal of Service Theory and Practice*, Vol. 28, Issue 2, pp.196-227, doi: doi.org/10.1108/JSTP-11-2014-0261.
- [33] J. Wolny, N. Charoensuksai (2014), "Mapping customer journeys in multichannel decision-making", *Journal of Direct, Data and Digital Marketing Practice*, Vol. 15, No. 4, pp. 317-326, doi: 10.1057/dddmp.2014.24.
- [34] G.W. Head (1960), "What does automation mean to the Marketing man?", *European Journal of Marketing*, 24(4), pp. 35-37.
- [35] J.C. González, F. Mochón (2016), "Operating an Advertising Programmatic Buying Platform: A Case Study", *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 3, No. 6, pp. 6-15, doi: 10.9781/ijimai.2016.361
- [36] A. Klahold (2009), "*Empfehlungssysteme: Recommender- Systems - Grundlagen, Konzepte und Lösungen*", Wiesbaden, SpringerVieweg.
- [37] M. Bala, D. Verma (2018), "A Critical Review of Digital Marketing", *International Journal of Management, IT & Engineering*, Vol. 8 Issue 10, October 2018, pp. 321-339.
- [38] H. Chen, R.H.L.Chiang, and J.M.R. Storey (2012), "Business intelligence and analytics: from big data to big impact", *Mis Quarterly*, Vol. 36 No. 4, December 2012, pp. 1165-1188.
- [39] K. Krombholz, H. Hobel, M. Huber, and E. Weippl (2014), "Advanced Social Engineering Attack", *Journal of Information Security and Applications*, 22, pp. 113-122, doi: 10.1016/j.jisa.2014.09.005.
- [40] A.-M. Kennedy (2012), "Macro-social marketing and social engineering: a systems approach", *Journal of social marketing*, Vol. 2, Issue 1, pp.37-51, doi: doi.org/10.1108/20426761211203247
- [41] J. van Dijk, T. Poell (2013), "Understanding Social Media Logic", *Media and Communication*, Volume 1, Issue 1, pp. 2-14, doi: 10.12924/mac2013.01010002.
- [42] S. Edosomwan, S. Prakasan, D. Kouame, J. Watson, and T. Seymour (2011), "The history of social media and its impact on business", *Journal of Applied Management and Entrepreneurship*, No. 16.3 (2011), pp. 79-91.
- [43] T.L. Tuten, M.R. Salomon (2017), "*Social Media Marketing*", London, Sage.
- [44] J. Berger, K.I. Milkman (2011), "What Makes online Content Viral?", *Journal of Marketing Research*, pp. 1-14, doi: 10.1509/jmr.10.0353.
- [45] M. Kilgour, S. Sasser, and R. Larke (2015), "The social media transformation process: curating content into strategy", *Corporate Communications An International Journal*, 20(3), pp. 326-343, doi: 10.1108/CCIJ-07-2014-0046.
- [46] G. Mangold, D.J. Faulds (2009), "Social media: The new hybrid element of the promotion mix", *Business Horizons*, Volume 52, Issue 4, July–August 2009, pp. 357-365, doi: https://doi.org/10.1016/j.bushor.2009.03.002.
- [47] A. Kumar, R. Bezawada, R. Rishika, R. Janakiraman, and P.K. Kannan (2016), "From Social to Sale: The Effects of Firm-Generated Content in Social Media on Customer Behavior", *Journal of Marketing*, Vol. 80 (January 2016), pp. 7-25, doi: dx.doi.org/10.1509/jm.14.0249.
- [48] A. Opreana, S. Vinerean (2015), "A New Development in Online Marketing: Introducing Digital Inbound Marketing", *Expert Journal of Marketing*, Volume 3, Issue 1, pp. 29-34.
- [49] E.-E. Sunny, O.J. Anael (2016), "Mobile Marketing in a Digital Age: Application, Challenges & Opportunities", *British Journal of Economics, Management & Trade*, 11(1), pp. 1-13, doi: 10.9734/BJEMT/2016/19925
- [50] Á. Nagy, A. Kölcsey (2017), "Generation Alpha: Marketing or Science?", *Acta Technologica Dubnicae*, volume 7, 2017, issue 1, pp. 107-115, doi: 10.1515/atd-2017-0007.
- [51] F. v. Lewinski, (2016), "Lasst die Daten frei – neue Spielräume für Kreativität und Effizienz", *w&w special, Marketing convention 2016*, Accessed: Feb. 15, 2019. [Online]. Available: www.wuv.de/specials/w\_v\_Marketing\_convention\_2016/.
- [52] I. Fillis, R. Rentschler (2006), "*Creative Marketing: An Extended Metaphor for Marketing in a New Age*", New York, Palgrave Macmillan.
- [53] R.L. Flores, S. Negny, J.P. Belaud, and J.-M. Le Lann (2015), "Collective intelligence to solve creative problems in conceptual design phase", World Conference: TRIZ FUTURE, TF 2011-2014, *Procedia Engineering*, 131 (2015), pp. 850-860.
- [54] K.N. Lemon, P.C. Verhoef (2016), "Understanding Customer Experience Throughout the Customer Journey", *Journal of Marketing*, AMA/MSI Special Issue, Vol. 80 (November 2016), pp. 69-96, doi: 10.1509/jm.15.0420.
- [55] C.S. Fleisher (2008), "Using open source data in developing competitive and Marketing Intelligence", *European Journal of Marketing*, Vol. 42 Issue: 7/8, pp. 852-866, doi: doi.org/10.1108/03090560810877196.
- [56] K.J. Trainor, M.T. Krush, R. Agnihotri (2013), "Effects of relational proclivity and Marketing Intelligence on new product development", *Marketing Intelligence & Planning*, Vol. 31, Issue 7, pp.788-806, doi: doi.org/10.1108/MIP-02-2013-0028.
- [57] J. Wolny, N. Charoensuksai (2014), "Mapping customer journeys in multichannel decision-making", *Journal of Direct, Data and Digital Marketing Practice*, Vol. 15, No. 4, pp. 317-326, doi: 10.1057/dddmp.2014.24.
- [58] S. Tuzovic, S. Paluch (2018), "Conversational Commerce – A new area for Service Business Development," in: M. Bruhn, K. Hadwich (eds.): *Service Business Development: Strategien – Innovationen – Geschäftsmodelle; Forum Dienstleistungsmanagement*, Band 1, Wiesbaden, SpringerGabler, pp. 81-100.
- [59] C. Messina (2015), "Conversational commerce", *Medium*, 16. Januar 2015, Accessed: Feb. 15, 2019. [Online]. Available: https://medium.com/chris-messina/conversational-commerce-92e0bccfc3ff.
- [60] C. Messina, C. (2016), "20016 will be the year of conversational commerce", *Medium*, 19. Januar 2016, Accessed: Feb. 15, 2019. [Online]. Available: https://medium.com/chris-messina/2016-will-be-the-year-of-conversational-commerce-1586e85e3991.
- [61] H. Holland (2014) (ed.), "*Digitales Dialogmarketing: Grundlagen, Strategien, Instrumente*", Wiesbaden, SpringerGabler.
- [62] M. Anshari, M. N. Almunawar, S. A. Lim, and A. Al-Mudimigh (2018): "Customer relationship management and big data enabled: Personalization & customization of services", *Applied Computing and Informatics*, doi: doi.org/10.1016/j.aci.2018.05.004.
- [63] F. Lasogga (2014), "Grundlagen und Handlungsfelder für exzellentes Customer Relationship Management", in: Kliewe, T./ Kesting, T. (eds.): *Moderne Konzepte des organisationalen Marketing*, Wiesbaden, SpringerGabler, pp. 239-254.
- [64] T.Y. Chan, C. Wu, Y. Xie (2011), "Measuring the Lifetime Value of Customers Acquired from Google Search Advertising", *Marketing Science*, Vol. 30, No. 5, September–October 2011, pp. 837-850.
- [65] D.M. Hanssens, K.H. Pauwels (2016), "Demonstrating the value of Marketing", *Journal of Marketing*, Vol 80(6), pp. 173-190, doi: doi.org/10.1509/jm.15.0417.
- [66] D. O'Sullivan, A.V. Abela (2007), "Marketing Performance Measurement Ability and Firm Performance", *Journal of Marketing*, Vol. 71 (April 2007), pp. 79-93, doi: 10.1509/jmk.71.2.79.
- [67] J. Järvinen (2016), "*The Use of Digital Analytics for Measuring and Optimizing Digital Marketing Performance*", Jyväskylä University School of Business and Economics, University Library of Jyväskylä. Accessed: Feb. 15, 2019. [Online]. Available: https://jyx.jyu.fi/bitstream/handle/123456789/51512/978-951-39-6777-2\_vaitos21102016.pdf?sequence=1
- [68] E. Lammenett, (2015), "*Praxiswissen Online-Marketing, Affiliate- und E-Mail-Marketing, SuchmaschinenMarketing, OnlineWerbung, Social Media, Online-PR*", Wiesbaden, SpringerGabler.
- [69] P. Kotler, H. Kartajaya, and I. Setiawan (2017), "*Marketing 4.0: Moving from Traditional to Digital*", Hoboken, Wiley.
- [70] F.T. Lorne, S. Daram, R. Frantz, N. Kumar, A. Mohammed, and A. Muley (2018), "Blockchain Economics and Marketing", *Journal of Computer and Communications*, 2018, 6, pp. 107-117.
- [71] E. Constantinides (2006), "The Marketing Mix Revisited: Towards the 21st Century Marketing", *Journal of Marketing Management*, 2006, 22, pp. 407-438.
- [72] E. Schwedler (2000), "*Kompaktwissen Marketing: Grundlagen für Studium und Praxis*", Wiesbaden, SpringerGabler.
- [73] C.L. Goi (2009), "A Review of Marketing Mix: 4Ps or More?", *International Journal of Marketing Studies*, Vol. 1, No, pp. 2-15, doi: DOI:10.5539/ijms.v1n1p2.

- [74] S.D. Hunt D. (1976): The Nature and Scope of Marketing, *Journal of Marketing*, Vol. 40, No. 3 (July, 1976), pp. 17-28.
- [75] P. Kotler, S. Levy (1969), "Broadening the Concept of Marketing", *Journal of Marketing*, Vol. 38, pp. 10-15.
- [76] S. Okazaki, B. Mueller, and C.R. Taylor (2010), "Measuring Soft-Sell Versus Hard-Sell Advertising Appeals", *Journal of Advertising*, vol. 39, no. 2, summer 2010, pp. 5–20, DOI 10.2753/JOA0091-3367390201.
- [77] J. Lies (2019), "Marketing 4.0 als "Old School" des PR-Managements", in: M. Stumpf (ed.), *Digitalisierung und Kommunikation Konsequenzen der digitalen Transformation für die Wirtschaftskommunikation*, Europäische Kulturen in der Wirtschaftskommunikation, Band 31, Wiesbaden, SpringerVS, pp. 231-254.
- [78] In further development by A. Opreana, S. Vinerean (2015), "A New Development in Online Marketing: Introducing Digital Inbound Marketing", *Expert Journal of Marketing*, Volume 3, Issue 1, pp. 29-34.
- [79] E. Maltz, A.K. Kohli (1996), "Market Intelligence Dissemination across Functional Boundaries", *Journal of Marketing Research*, Vol. 33, No. 1, February 1996, pp. 47-61.



Prof. Dr. habil. Jan Lies

Prof. Dr. Jan Lies has a doctorate and postdoctoral habilitation in economics. Since 2013 he is professor for business administrations, especially corporate communications and marketing at FOM university of applied science in Dortmund, Germany. His research involves evolutionary and behavioral economics as well as digital marketing, pr-management and change communications. Digital

marketing is one of his research areas which demonstrates the heavy impact of evolutionary processes on corporate success.

