

Tai-hoon Kim  
Hojjat Adeli  
Carlos Ramos  
Byeong-Ho Kang (Eds.)

Communications in Computer and Information Science

260

# Signal Processing, Image Processing and Pattern Recognition

International Conference, SIP 2011  
Held as Part of the Future Generation  
Information Technology Conference, FGIT 2011  
in Conjunction with GDC 2011  
Jeju Island, Korea, December 2011, Proceedings



Tai-hoon Kim Hojjat Adeli Carlos Ramos  
Byeong-Ho Kang (Eds.)

# Signal Processing, Image Processing and Pattern Recognition

International Conference, SIP 2011  
Held as Part of the Future Generation  
Information Technology Conference, FGIT 2011  
in Conjunction with GDC 2011  
Jeju Island, Korea, December 8-10, 2011  
Proceedings

Volume Editors

Tai-hoon Kim  
Hannam University, Daejeon, Korea  
E-mail: taihoonn@hannam.ac.kr

Hojjat Adeli  
The Ohio State University, Columbus, OH, USA  
E-mail: adeli.1@osu.edu

Carlos Ramos  
University of Porto, Portugal  
E-mail: csr@dei.isep.ipp.pt

Byeong-Ho Kang  
University of Tasmania, Hobart, TAS, Australia  
E-mail: byeong.kang@utas.edu.au

ISSN 1865-0929  
ISBN 978-3-642-27182-3  
DOI 10.1007/978-3-642-27183-0  
Springer Heidelberg Dordrecht London New York

e-ISSN 1865-0937  
e-ISBN 978-3-642-27183-0

Library of Congress Control Number: Applied for

CR Subject Classification (1998): I.4, I.2, I.5, H.3, C.2, J.3

© Springer-Verlag Berlin Heidelberg 2011

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

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

*Typesetting:* Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

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

# Foreword

Signal processing, image processing and pattern recognition are areas that attract many professionals from academia and industry for research and development. The goal of the SIP conference is to bring together researchers from academia and industry as well as practitioners to share ideas, problems and solutions relating to the multifaceted aspects of signal processing, image processing and pattern recognition.

We would like to express our gratitude to all of the authors of submitted papers and to all attendees for their contributions and participation.

We acknowledge the great effort of all the Chairs and the members of Advisory Boards and Program Committees of the above-listed event. Special thanks go to SERSC (Science and Engineering Research Support Society) for supporting this conference.

We are grateful in particular to the speakers who kindly accepted our invitation and, in this way, helped to meet the objectives of the conference.

December 2011

Chairs of SIP 2011

# Preface

We would like to welcome you to the proceedings of the 2011 International Conference on Signal Processing, Image Processing and Pattern Recognition (SIP 2011) — the partnering event of the Third International Mega-Conference on Future-Generation Information Technology (FGIT 2011) held during December 8–10, 2011, at Jeju Grand Hotel, Jeju Island, Korea.

SIP 2011 focused on various aspects of advances in signal processing, image processing and pattern recognition. It provided a chance for academic and industry professionals to discuss recent progress in the related areas. We expect that the conference and its publications will be a trigger for further related research and technology improvements in this important subject.

We would like to acknowledge the great effort of the SIP 2011 Chairs, Committees, Special Session Organizer, as well as all the organizations and individuals who supported the idea of publishing this volume of proceedings, including the SERSC and Springer.

We are grateful to the following keynote, plenary and tutorial speakers who kindly accepted our invitation: Hsiao-Hwa Chen (National Cheng Kung University, Taiwan), Hamid R. Arabnia (University of Georgia, USA), Sabah Mohammed (Lakehead University, Canada), Ruay-Shiung Chang (National Dong Hwa University, Taiwan), Lei Li (Hosei University, Japan), Tadashi Dohi (Hiroshima University, Japan), Carlos Ramos (Polytechnic of Porto, Portugal), Marcin Szczuka (The University of Warsaw, Poland), Gerald Schaefer (Loughborough University, UK), Jinan Fiaidhi (Lakehead University, Canada) and Peter L. Stanchev (Kettering University, USA), Shusaku Tsumoto (Shimane University, Japan), Jemal H. Abawajy (Deakin University, Australia).

We would like to express our gratitude to all of the authors and reviewers of submitted papers and to all attendees, for their contributions and participation, and for believing in the need to continue this undertaking in the future.

This work was supported by the Korean Federation of Science and Technology Societies Grant funded by the Korean Government.

December 2011

Tai-hoon Kim  
Hojjat Adeli  
Carlos Ramos  
Byeong-Ho Kang

# Organization

## General Chair

Carlos Ramos ISEP-IPP, Portugal

## Program Co-chairs

Byeong-Ho Kang University of Tasmania, Australia  
Tai-hoon Kim GVSA and University of Tasmania, Australia

## Publicity Co-chairs

Junzhong Gu East China Normal University, China  
Hideo Kuroda FPT University, Vietnam  
Muhammad Khurram Khan King Saud University, Saudi Arabia  
Aboul Ella Hassanien Cairo University, Egypt

## Program Committee

Andrzej Dzielinski	Junzhong Gu	Roman Neruda
Andrzej Kasinski	Kenneth Barner	Rudolf Albrecht
Antonio Dourado	Kidiyo Kpalma	Ryszard Tadeusiewicz
Caroline Fossati	Kousuke Imamura	Salah Bourennane
Chng Eng Siong	Mei-Ling Shyu	Selim Balcisoy
Dimitris Iakovidis	Miroslaw Swiercz	Serhan Dagtas
Debnath Bhattacharyya	Makoto Fujimura	Shu-Ching Chen
Ernesto Exposito	Marie Babel	Tae-Sun Choi
Francesco Masulli	Mathieu Gineste	William I. Grosky
Gérard Medioni	Mototaka Suzuki	Xavier Maldague
Hideo Kuroda	N. Jaisankar	Xuejing Wu
Hong Kook Kim	Nadia Magnenat-Thalmann	Yi Fang
Janusz Kacprzyk	Nikos komodakis	Yi Lu Murphey
Jocelyn Chanussot	Nilesh Patel	Yue Lu
Joonki Paik	Paolo Remagnino	
Joseph Ronsin	Peter L. Stanchev	

## Special Session Organizer

D. Jude Hemanth

# Table of Contents

Image Content Detection Method Using Correlation Coefficient between Pixel Value Histograms . . . . .	1
<i>Kousuke Imamura, Hideo Kuroda, and Makoto Fujimura</i>	
Measuring Blockiness of Videos Using Edge Enhancement Filtering . . . . .	10
<i>Md. Mehedi Hasan, Kiok Ahn, and Oksam Chae</i>	
An Improved Joint Particle Filter Algorithm for Multi-target Tracking . . . . .	20
<i>Jin-Long Yang and Hong-Bing Ji</i>	
Enhanced Fuzzy-Based Models for ROI Extraction in Medical Images . . . . .	26
<i>Yasser El-Sonbaty, Sherin M. Youssef, and Karma M. Fathalla</i>	
A Variation of Local Directional Pattern and Its Application for Facial Expression Recognition . . . . .	36
<i>Tianwei Xu, Juxiang Zhou, and Yunqiong Wang</i>	
Application of Fourier Transform to Get an Original Satellite Image without Applying the SRM or 180 Degree Rotation . . . . .	48
<i>Subhasis Kesh and Srishty Chakravarty</i>	
Evaluating Inpainting Methods to the Satellite Images Clouds and Shadows Removing . . . . .	56
<i>Ana Carolina Siravenha, Danilo Sousa, Aline Bispo, and Evaldo Pelaes</i>	
Comparing Different High-Pass Filters to Improve the Accuracy of Classification of Satellite Imagery Obstructed by Clouds and Fog . . . . .	66
<i>Danilo Sousa, Ana Carolina Siravenha, and Evaldo Pelaes</i>	
A Fast Implementation of Semi-Markov Conditional Random Fields . . . . .	74
<i>La The Vinh, Sungyoung Lee, and Young-Koo Lee</i>	
Adapted Scan Based Lossless Image Compression . . . . .	82
<i>Tarek Ouni, Arij Lassoued, and Mohamed Abid</i>	
Localization of Passengers Inside Intelligent Vehicles by the Use of Ultra Wideband Radars . . . . .	92
<i>Philipp Galdia, Carsten Koch, and Anthimos Georgiadis</i>	



Spectral Density Analysis: Theta Wave as Mental Stress Indicator . . . . .	103
<i>Saidatul Ardeenawatie Awang, Paulraj Murugesu Pandiyan, Sazali Yaacob, Yusnita Mohd Ali, Fadzly Ramidi, and Fauziah Mat</i>	
Leaf Image Analysis towards Plant Identification . . . . .	113
<i>Debnath Bhattacharyya, Tai-hoon Kim, and Gang-soo Lee</i>	
Watermarking Using Multiresolution Cosine Transformation: A Review . . . . .	126
<i>Debnath Bhattacharyya, Tai-hoon Kim, and Gang-soo Lee</i>	
Use of Artificial Neural Network in Bengali Character Recognition . . . . .	140
<i>Debnath Bhattacharyya, Tai-hoon Kim, and Gang-soo Lee</i>	
New Algorithm for Skewing Detection of Handwritten Bangla Words . . . . .	153
<i>Rajib Ghosh, Debnath Bhattacharyya, Tai-hoon Kim, and Gang-soo Lee</i>	
An Extended Set of Haar-like Features for Bird Detection Based on AdaBoost . . . . .	160
<i>Chih-Cheng Huang, Chun-Yi Tsai, and Horng-Chang Yang</i>	
A Non-blind Digital Image Watermarking Method Based on the Dyadic Wavelet Transform and Interval Arithmetic . . . . .	170
<i>Teruya Minamoto and Ryuji Ohura</i>	
Song Classification: Classical and Non-classical Discrimination Using MFCC Co-occurrence Based Features . . . . .	179
<i>Arijit Ghosal, Rudrasis Chakraborty, Bibhas Chandra Dhara, and Sanjoy Kumar Saha</i>	
Gabor Based Gender Classification with Classifier Independent Feature Selection . . . . .	186
<i>Aun Irtaza, M. Arfan Jaffar, and Tae-Sun Choi</i>	
Data Hiding in Images Using Some Efficient Steganography Techniques . . . . .	195
<i>Chandreyee Maiti, Debanjana Baksi, Ipsita Zamider, Pinky Gorai, and Dakshina Ranjan Kisku</i>	
Context Based Speech Analysis of Bengali Language as a Part of TTS Conversion . . . . .	204
<i>Nabanita Mukherjee, Imon Mukherjee, Debnath Bhattacharyya, and Tai-hoon Kim</i>	
Image Searching with Eigenfaces and Facial Characteristics . . . . .	215
<i>Ayesha Kurukulasooriya and Anuja T. Dharmarathne</i>	

Design of DT-CNN for Imputing Data at Unobserved Location of Geostatistics Image Dataset . . . . .	225
<i>Sathit Prasomphan, Hisashi Aomori, and Mamoru Tanaka</i>	
Enhanced Edge Localization and Gradient Directional Masking for Moving Object Detection . . . . .	234
<i>Pranab K. Dhar, Mohammad I. Khan, D.M.H. Hasan, and Jong-Myon Kim</i>	
Automatic Detection of Face and Facial Landmarks for Face Recognition . . . . .	244
<i>Hajra Momin and Jules-Raymond Tapamo</i>	
A Tool for Ranking and Enhancing Aesthetic Quality of Paintings . . . . .	254
<i>W.A.P. Wickramasinghe, Anuja T. Dharmaratne, and N.D. Kodikara</i>	
Aging Progression of Elderly People Using Image Morphing . . . . .	261
<i>L.L. Gayani Kumari and Anuja T. Dharmaratne</i>	
Off-line Signature Verification Based on Combination of Modified Direction and Microstructure Features . . . . .	270
<i>Danfeng Yang, Yuzhu Qin, Zhimin Huang, and Yue Lu</i>	
Heart Sound Feature Reduction Approach for Improving the Heart Valve Diseases Identification . . . . .	280
<i>Mostafa A. Salama, Aboul Ella Hassanien, Aly A. Fahmy, and Tai-hoon Kim</i>	
Comparison of Different Ontology-Based Query Expansion Algorithms for Effective Image Retrieval . . . . .	291
<i>C.H.C. Leung and Yuanxi Li</i>	
Fast Reconstruction Technique for Medical Images Using Graphics Processing Unit . . . . .	300
<i>Mohammad Nazmul Haque, Mohammad Shorif Uddin, M. Abdullah-Al-Wadud, and Yoojin Chung</i>	
Adaptive Image Zooming Based on Bilinear Interpolation and VQ Approximation . . . . .	310
<i>Yu-Chen Hu, Wu-Lin Chen, and Jun-Rong Zeng</i>	
An Enhanced Fuzzy C-Means Clustering (ECFMC) Algorithm for Spot Segmentation . . . . .	320
<i>A. Sri Nagesh, G.P. Saradhi Varma, A. Govardhan, and B. Raveendra Babu</i>	
On Intuitionistic Fuzzy T-ideals in TM-Algebra . . . . .	328
<i>Megalai Kandasamy and Tamilarasi Angamuthu</i>	

Mitigating Congestion and Improving the Performance of Wireless Sensor Networks . . . . .	336
<i>S. Raj Barath, C. Kezi Selva Vijila, and A. Jaya Prakash</i>	
Robust Key Points Matching by Ordinal Measure . . . . .	346
<i>S. Lakshmi and V. Sankaranarayanan</i>	
Performance Enhanced Hybrid Kohonen-Hopfield Neural Network for Abnormal Brain Image Classification . . . . .	356
<i>D. Jude Hemanth, C. Kezi Selva Vijila, A. Immanuel Selvakumar, and J. Anitha</i>	
Middleware for Physical and Logical Context Awareness . . . . .	366
<i>Junzhong Gu</i>	
The Use of Biorthogonal Wavelet, 2D Polynomial and Quadtree to Compress Color Images . . . . .	379
<i>Loay E. Goerge and Bushra A. Sultan</i>	
A Robust Method for Head Orientation Estimation Using Histogram of Oriented Gradients . . . . .	391
<i>Dinh Tuan Tran and Joo-Ho Lee</i>	
Fusion of Gait and Facial Feature Using PCA . . . . .	401
<i>Sanjeev Sharma, Ritu Tiwari, Anupam Shukla, and Vikas Singh</i>	
Shockwave Velocity Estimation from Laser Induced Breakdown Images . . . . .	410
<i>Jaemyoung Lee</i>	
A Temporal Item-Based Collaborative Filtering Approach . . . . .	414
<i>Lei Ren, Junzhong Gu, and Weiwei Xia</i>	
A New Fusion Algorithm for Dim Target Detection Based on Dual-Wave Infrared Images . . . . .	422
<i>Jin Liu, Shao-Hua Wang, and Hong-Bing Ji</i>	
Fractal Analysis and the Effect of Aging on the Heart Rate and Breathing Frequency Relationship . . . . .	430
<i>Wilson Bucaoto, Han Jong Kim, and Artem Lenskiy</i>	
Relational Features for Texture Classification . . . . .	438
<i>Wan Nural Jawahir Hj Wan Yussof and Hans Burkhardt</i>	
<b>Author Index . . . . .</b>	<b>449</b>

# Image Content Detection Method Using Correlation Coefficient between Pixel Value Histograms

Kousuke Imamura<sup>1</sup>, Hideo Kuroda<sup>2</sup>, and Makoto Fujimura<sup>3</sup>

<sup>1</sup> Institute of Science and Engineering, Kanazawa University  
Kakuma-machi, Kanazawa, Ishikawa 920-1192 Japan  
imamura@ec.t.kanazawa-u.ac.jp

<sup>2</sup> Department of Computing Fundamentals, FPT University  
Cau Giay Hanoi, Vietnam

<sup>3</sup> Graduate School of Engineering, Nagasaki University  
1-14 Bunkyo-machi, Nagasaki 852-8521 Japan

**Abstract.** An extraction method for searching for unauthorized copies of an image on the Internet is required in image search to make use of digital watermarks. In this paper, we propose an efficient two-stage image search method for extraction of illegal copies of a target image. The first stage is a pre-search, which searches for candidate of illegal images by some simple method. The next stage is the main search, which extracts embedded copyright information from the candidate and decides whether the image is an illegal copy of the target. In addition, we propose a simple image search method which uses the correlation coefficient between pixel value histograms of images as a pre-search method. The proposed pre-search method is useful because the proposed method is possible to combine with the current extraction technologies of embedded information. The performance of the proposed pre-search method is evaluated through computer simulations.

**Keywords:** Digital watermark, Image search, Correlation coefficient, Pixel value histogram.

## 1 Introduction

Recently, the generation and transcription of digital content including images became sufficiently easy that it can be done by individuals, due to the high performance and high speed of current personal computers. Moreover, a large amount of content including images are sent and received by individuals, and thus content from/to servers is frequently downloaded and uploaded over the Internet.

The above-described situation, although convenient, also leads to copyright infringement due to unauthorized copying of content by third persons becoming a social problem. Methods to find unauthorized copying from among a large amount of content on the Internet are required. Digital watermark is one such method. The use of watermarks actually consists of two methods: one to embed copyright information in content, and a second to extract particular content from among a larger amount of

content. The latter, a method for extracting particular content from among the larger amount of content on the Internet, is similar to existing image search methods.

In general, the object being searched by image search method is in a database. In a general database, multiple feature vectors [1,2] that show the characteristics of particular images are registered. In an instance of an image search, the key word describing a characteristic of the content desired is input, and images for which this key word is included in the feature vector of the database are output as an extraction result.

However, in the case of a digital watermark, it is the Internet and not the database that is searched. Therefore, a key word describing a characteristic of the image cannot be used. Moreover, it is necessary not only to detect similar images but also to extract the embedded copyright information that proves ownership.

Because the above requirement is difficult, methods using key words have been proposed [3,4]. In these methods, a feature vector of content is registered with content in a watermark certificate center for proving ownership. Then, the feature vector of Internet content is extracted, and the feature vector registered at the watermark certificate center is retrieved. If the extracted feature vector matches the feature vector of the particular registered image using the latter feature vector as a key word, an illegal copy has been identified. This method is a reverse search procedure for searching for particular content, not from among content on the Internet, but from that registered at the center.

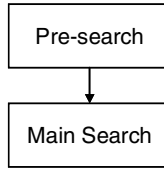
However, the procedure of the above-described methods is high complexity. Actually, though much research [5-7] on digital watermarks has already been done, only methods for the embedding and extraction of copyright information into/from content are discussed, where as methods for effectively searching for illegal copies from among the large amount of content on the Internet has not been sufficiently considered.

In this paper, to solve the above problem, we propose an efficient two-stage image search method for extraction of illegal copies of a target image. The first stage is a pre-search using some simple method, which searches for illegally copied images as an input candidate for next stage. The next stage is the main search, which extracts embedded copyright information from the candidate and decides whether it is an illegal copy of target image. In addition, we propose a simple image search method which uses the correlation coefficient between pixel value histograms of images as a pre-search method. It is the proposed image search method using this particular proposed pre-search algorithm that is the focus of this paper. We evaluate the extraction performance of this image search method for JPEG compressed images, which use the most common image compression format.

The remainder of this paper is organized as follows. Section 2 presents the proposed two-stage image search method. Section 3 describes the image search method using the correlation coefficient between pixel value histograms. Section 4 describes an image search for JPEG compression images. Section 5 presents the pre-search algorithm for the proposed image search method. Simulation and results are shown in Section 6, and concluding remarks are given in Section 7.

## 2 Proposed Two-Stage Image Search Method

Fig. 1 shows the proposed two-stage image search method. In the pre-search stage, image content from the Internet is input, and it is investigated whether this content is similar to owned content using a simple method. If the content is not similar, the search process is terminated, and the next image content is investigated. If the content is similar, the image content is output to the main search stage as a candidate of illegal copying. In the main search stage, the embedded copyright information in the image is extracted by a predefined method. When the extracted copyright information indicates the target image, a complaint is sent to the user by a predefined procedure. It is possible to search for illegal copies efficiently because the pre-search stage is simple.



**Fig. 1.** Two-stage image search method for digital watermarking

## 3 Image Search Method Using the Correlation Coefficient between Pixel Value Histograms

It is necessary to extract candidate illegal copies by a simple method in order to extract efficiently from among the large number of images on the Internet. In the process, avoiding missing an illegal copy is more important than avoiding extracting some many images. In addition, robustness to small changes in pixel values is required to avoid omissions because the pixel values of illegal copies may include various noises used to attack the watermarking.

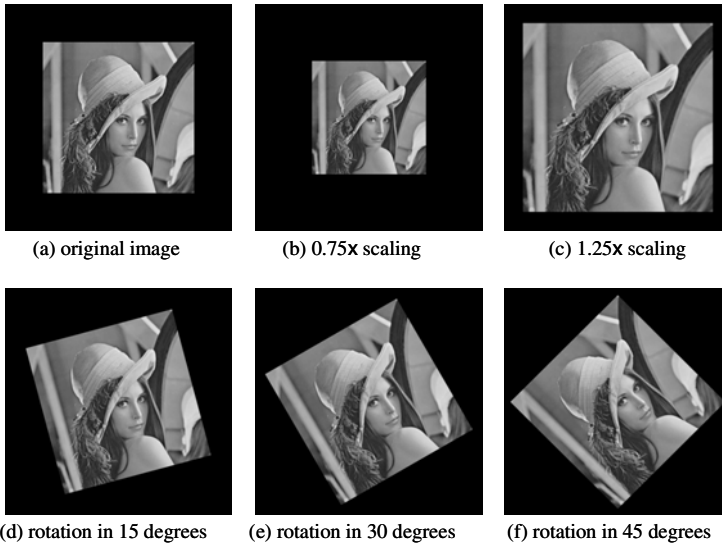
With this in mind, pixel value histograms are used as a feature quantity for image searches because this feature quantity is insensitive to attack by geometric transform, such as rotation and scaling. The correlation coefficient between pixel value histograms normalized by the number of pixels in the image is used to evaluate similarity between two images. The correlation coefficient is defined as follows:

$$r = \frac{\sum_{i=0}^{255} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=0}^{255} (x_i - \bar{x})^2 \sum_{i=0}^{255} (y_i - \bar{y})^2}}, \quad (1)$$

where  $x=\{x_i\}$ ,  $y=\{y_i\}$  denote the normalized frequency at pixel value  $i$ , and  $\bar{x}$ ,  $\bar{y}$  denote arithmetic mean.

Firstly, we verify the correlation coefficient between the original image and a geometrically attacked image as a preliminary experiment. "Lenna" (256x256 pixels, grayscale) is used as a test image, and the geometrically attacked images are transformed by 0.75x or 1.25x scaling or by 15, 30, or 45 degree rotation. Bilinear interpolation is used as the pixel interpolation method in the geometric transform.

Fig. 2 shows the original image and geometrically attacked images, and Table 1 shows the correlation coefficient between these images and the original. From Table 1, the correlation coefficient is high despite the geometric attack. Therefore, it is evident that the correlation coefficient between pixel value histograms is suitable as a feature quantity for image searches involving geometrically attacked images. From the results, the decision threshold for the matching image search is set to 0.95 to consider margin for avoiding missing image match.



**Fig. 2.** Original image "Lenna" and geometrically attacked images

**Table 1.** Correlation coefficient between original image "Lenna" and geometrically attacked images

scaling/rotation attack	correlation coefficient
15 degrees for anticlockwise rotation	0.993
30 degrees for anticlockwise rotation	0.994
45 degrees for anticlockwise rotation	0.991
0.75x scaling	0.992
1.25x scaling	0.994
original	1.000

## 4 Image Search for JPEG Compression Images

Most images on the Internet are compressed by some method for efficient transmission. Hence, it is very important to extract compressed images. JPEG compression is the main coding format worldwide, and so handling JPEG compressed images in particular is considered.

**Table 2.** Correlation coefficient between original image and JPEG compressed images

quality parameter	correlation coefficient without histogram smoothing		correlation coefficient with histogram smoothing	
	Lenna	Bridge	Lenna	Bridge
100	0.999	0.998	1.000	0.998
80	0.995	0.272	0.998	0.824
60	0.986	0.287	0.997	0.824
40	0.964	0.288	0.996	0.825
20	0.760	0.279	0.985	0.821

**Table 3.** Correlation coefficient between JPEG compressed image with quality parameter 60 and JPEG compressed images

quality parameter	correlation coefficient without histogram smoothing		correlation coefficient with histogram smoothing	
	Lenna	Bridge	Lenna	Bridge
100	0.986	0.322	0.995	0.853
80	0.988	0.990	0.997	0.998
60	1.000	1.000	1.000	1.000
40	0.958	0.990	0.994	0.998
20	0.744	0.987	0.933	0.996

Firstly, we measure the correlation coefficient between an original image and JPEG compressed images as a preliminary experiment. In the experiment, the original image is transformed into a JPEG compressed image using GIMP [8], and the parameter  $q$  for image quality control is set from 20 to 100. Table 2 shows the correlation coefficient between original image and JPEG compressed images for "Lenna" and "Bridge" images. The experiment value of left column in Table 2 shows the correlation coefficient when pixel value histogram has been used as it is. From the column in Table 2, the correlation coefficient decreases significantly as JPEG compression image quality is degraded, especially in the case of the "Bridge" image. The reason for this is that JPEG compression causes distinctive changes to the histogram, and the histogram change is large because the pixel resolution of the "Bridge" image is 64 levels essentially.



Therefore, in this paper, two techniques are introduced into the image search as corrective measures for JPEG compressed images. One is  $N$ -level smoothing of the pixel value histograms in order to defuse the effect of the histogram changes. By a preliminary experiment, we selected not three-level but five-level with large effect of smoothing. The experiment value of right column in Table 2 shows the correlation coefficient with five-level smoothing histogram. A large effect of smoothing has been achieved. Another proposed technique is to use some quality JPEG compressed images as the standard image instead of the original image.

Table 3 shows the correlation coefficient between the smoothed histograms of JPEG compressed image with quality parameter 60 (for example) and all JPEG compressed images. Here, 60 is median value from 20 to 100. From the table, it is clear that it is possible to extract the match image using both histograms without and/or with histogram smoothing, using the 0.95 decision threshold value from among JPEG compressed images.

### 5 Pre-search Algorithm for the Proposed Image Search Method

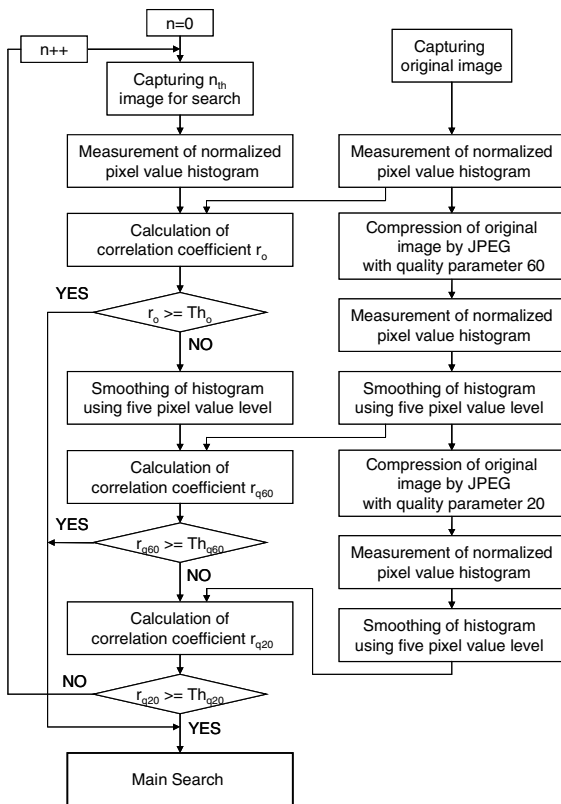


Fig. 3. Pre-search algorithm for images attacked by scaling, rotation, and JPEG compression

Fig. 3 shows the pre-search algorithm in the proposed image search method.

In the pre-search algorithm, firstly, an image on the Internet is input, and the correlation coefficient  $r_o$  between pixel value histograms of the original image and the observed image is calculated. If  $r_o$  is  $Th_o$  or more, the observed image is extracted as a candidate illegally copied image. The stage to here is effective for attack image of scaling and rotation. If  $r_o$  is less than  $Th_o$ , the correlation coefficient  $r_{q60}$  between the smoothed histogram of the JPEG compressed image with quality parameter 60 and the observed image is calculated. If  $r_{q60}$  is  $Th_{q60}$  or more, the observed image is extracted as a candidate. If  $r_{q60}$  is less than  $Th_{q60}$ , the correlation coefficient  $r_{q20}$  between the smoothed histogram of the JPEG compressed image with quality parameter 20 and the observed image is calculated. If  $r_{q20}$  is  $Th_{q20}$  or more, the observed image is extracted as a candidate. The obtained candidate is output to the main search. If  $r_{q20}$  is less than  $Th_{q20}$ , the observed image is excluded from the input of the main search. In this paper,  $Th_o$ ,  $Th_{q60}$  and  $Th_{q20}$  are set to 0.95.

## 6 Simulation and Results

The pre-search algorithm in the proposed method is analyzed by computer simulations. In the experiment, 12 images (256x256 pixels, grayscale) from SIDBA, "Airplane", "Barbara", "Boat", "Bridge", "Building", "Cameraman", "Girl", "Lax", "Lenna", "Lighthouse", "Text", and "Woman", are used as test images.

### 6.1 Identification of Match Image and Extractive Omission

Firstly, we evaluate performance of the match image extraction in the proposed pre-search method using JPEG compressed test images. Table 4 shows the correlation coefficients  $r_o$ ,  $r_{q60}$  and  $r_{q20}$  in the pre-search algorithm. The images for which all of  $r_o$ ,  $r_{q60}$  and  $r_{q20}$  are less than 0.95 are the omitted images of the extraction process.

From Table 4, the proposed pre-search method extracted the match image for all test images.

### 6.2 Excessive Extraction

Secondly, we measure the correlation coefficient between different images for evaluation of excessive extraction. In the experiment, a few pairs of "Building" and "Lighthouse" having white bars at the same position at the bottom of the image were extracted as the match image. These images are distinctive images which include too many white pixels in white bar region. The common white pixels cause a significant increase in the correlation coefficient for these images. However, excessive extraction is not as important as extractive omission in the case of illegally copied images.

As a result, the proposed pre-search method was able to identify all images as the match image except for a few excessive extractions.

**Table 4.** Identification of match image by the proposed pre-search algorithm

quality parameter	Airplane			Barbara			Boat		
	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$
100	1.000	0.998	0.989	0.999	0.998	0.994	1.000	0.998	0.990
80	0.998	0.999	0.990	0.992	0.999	0.995	0.997	0.999	0.990
60	0.982	1.000	0.990	0.989	1.000	0.995	0.969	1.000	0.990
40	0.931	0.998	0.993	0.981	0.998	0.996	0.885	0.996	0.998
20	0.681	0.990	1.000	0.936	0.995	1.000	0.665	0.990	1.000
quality parameter	Bridge			Building			Cameraman		
	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$
100	0.998	0.853	0.851	1.000	0.999	0.567	1.000	0.990	0.952
80	0.272	0.998	0.996	0.997	1.000	0.582	0.984	0.997	0.964
60	0.287	1.000	0.996	0.997	1.000	0.584	0.960	1.000	0.973
40	0.288	0.998	0.996	0.997	1.000	0.582	0.744	0.995	0.985
20	0.279	0.996	1.000	0.316	0.584	1.000	0.450	0.973	1.000
quality parameter	Girl			Lax			Lenna		
	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$
100	0.996	0.984	0.969	1.000	0.998	0.988	0.999	0.997	0.985
80	0.653	0.999	0.990	0.996	0.999	0.990	0.995	0.999	0.988
60	0.659	1.000	0.991	0.995	1.000	0.993	0.986	1.000	0.989
40	0.505	0.998	0.993	0.991	0.999	0.996	0.964	0.998	0.991
20	0.422	0.991	1.000	0.892	0.993	1.000	0.760	0.989	1.000
quality parameter	Lighthouse			Text			Woman		
	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$	$r_o$	$r_{q60}$	$r_{q20}$
100	1.000	0.998	0.623	1.000	0.997	0.774	1.000	0.998	0.990
80	0.991	0.999	0.644	0.995	1.000	0.811	0.994	0.999	0.992
60	0.989	1.000	0.647	0.996	1.000	0.813	0.993	1.000	0.993
40	0.971	0.999	0.646	0.996	1.000	0.807	0.976	0.999	0.995
20	0.316	0.647	1.000	0.595	0.813	1.000	0.827	0.993	1.000

$r_o$ : correlation coefficient between histograms of the original and JPEG image

$r_{q60}$ : correlation coefficient between the smoothed histograms of the JPEG image with quality parameter 60

$r_{q20}$ : correlation coefficient between the smoothed histograms of the JPEG image with quality parameter 20

## 7 Conclusions

In this paper, we proposed an efficient two-stage image search method for extraction of illegal copies from among the large number of images on the Internet. In addition, we proposed a pre-search algorithm which searches for candidate illegally copied images using a simple method. The proposed pre-search method is possible to combine with the current extraction technologies of embedded information. Simulation results

revealed that the proposed pre-search method provides good performance in the image search for a digital watermark, indicating that the future prospects of the proposed image search method are excellent. In the future, we intend to develop a search method for other types attack of against digital watermarks in images.

## References

1. Nocke, T., Schmann, H.: Meta Data for Visual Data Mining. In: Proceeding of Computer Graphics and Imaging, CGIM 2002, Kaua'i, Hawaii, USA (August 2002)
2. Hmida, M.B.H., Slimani, Y.: Meta-learning in Grid-based Data Mining Systems. *International Journal of Computer Network and Communications (IJNC)* 2(5), 215–230 (2010)
3. Ju, H.S., Kim, H.J., Lee, D.H., Lim, J.I.: An Anonymous Buyer-Seller Watermarking Protocol with Anonymity Control. In: Lee, P.J., Lim, C.H. (eds.) *ICISC 2002*. LNCS, vol. 2587, pp. 421–432. Springer, Heidelberg (2003)
4. Birk, D., Gaines, S.: Using Digital Watermarking for Securing Next Generation Media Broadcasts. In: *Advances in Machine Learning and Data Analysis*, pp. 27–41. Springer, Heidelberg (2010)
5. Tang, C.-W., Hang, H.-M.: A Feature-Based Robust Digital Image Watermarking Scheme. *IEEE Transactions on Signal Processing* 51(4), 950–959 (2003)
6. Lin, C.-Y., Wu, M., Bloom, J.A., Cox, I.J., Miller, M.L., Lui, Y.M.: Rotation, Scale, and Translation Resilient Watermarking for Images. *IEEE Transactions on Image Processing* 10(5), 767–792 (2001)
7. Cox, I.J., Miller, M.L., Bloom, J.A., Fridrich, J., Kalker, T.: *Digital Watermarking and Steganography*. Morgan Kaufmann (2008)
8. GIMP: The GNU Image Manipulation Program, <http://www.gimp.org/>

# Measuring Blockiness of Videos Using Edge Enhancement Filtering

Md. Mehedi Hasan, Kiok Ahn, and Oksam Chae

Image Processing Lab, Department of Computer Engineering,  
Kyung Hee University, South Korea  
{mehedi,oschae}@khu.ac.kr, kiokahn@mediachorus.com

**Abstract.** To represent high quality videos or images with low bit rate, an effective compression algorithm removes the redundancy because of statistical correlation and also the insignificant component of image signal. This paper represents a new algorithm to measure the blocking artifacts of videos by analyzing the distortions of local properties of image signals like dominant edge magnitude and direction. Extensive experiments on various videos show that the new algorithm is very much efficient and faster to measure the blocking artifacts in real time video error detection applications.

**Keywords:** Blockiness, Kirsch Mask, Discrete Cosine Transform, Edge Enhancement.

## 1 Introduction

Block transform coding is the most popular approach for image and video coding. There are many standards exist in image and video coding, like JPEG, H.263, MPEG-1, MPEG-2, MPEG-4 etc. The BLOCK-based discrete cosine transform (B-DCT) [1,2,3] is the fundamental component of many image and video compression standards, used in a wide range of applications. The B-DCT scheme takes into account the local spatial correlation property of the images by dividing the image into  $8 \times 8$  blocks of pixels. Then transforming each block from the spatial domain to the frequency domain is done by using the discrete cosine transform (DCT) and quantizing the DCT coefficients. In discrete cosine transform (DCT) each block pixels are treated as single entities and coded separately. But the correlation among spatially adjacent blocks is not taking into account in coding and as a result block boundaries are visible when decoded image is reconstructed. As for example, a slight change of luminance in border area can cause a step in the decoded image if the neighbor block fall into different quantization intervals. Therefore, the decompressed image and video exhibits various kind of artifacts. One of the most obtrusive artifacts is the "Blocking Artifact"[1,2,4,5,6].

There are many potential methods of measuring discrete cosine transform (DCT) based codec degradations involve directly examining the coarseness of the compressed

video stream at the time of quantization scaling. This kind of video streams optionally combined with a measure of the complexity of the original image transmitted by means of outside the video – a form of compressed reference. It is computationally expensive, not accurate and in any case can only make measurements on compressed video that has been already decompressed and potentially passed through other systems, including other additional codecs, prior to end-user delivery.

Objective image quality matrices are divided into Full-reference, Reduced-reference and No-reference. In Full and Reduced reference approach the access to original image is required. By using the original and reproduced image as inputs the system outputs a numerical value to show the quality of the reproduced image. This approach is not useful in applications like image and video communication, where original image and video is not accessible which is called No-reference approach, which is most experimented approach because of its computational efficiency, wider scope of applications, in-service visual monitoring and post-processing of decoded signal. In image coding techniques, Chou et. al. [7] addressed a key concept of perceptual coding considering human visual system, namely just-noticeable-distortion (JND). JND provides each signal being encoded with threshold level or error visibility. Actually JND is a function of local signal properties, such as background intensity, activity of luminance changes, dominant spatial frequency and changes in edge gradients. Once the JND profile of an image is obtained, the energy of the perceptible distortion like blockiness can be measured. But this kind of HVS [8,9,10] measurement system is computationally expensive and cannot be applicable for fast real time cases.

In this paper we actually incorporate light weighted human vision measurement system like edge information[5] to measure blocking artifacts in real time. To detect the distortions in edges we use kirsch mask in eight direction to make the edge detection process more perfect. Then simple bucket filling approach is applied, where the particular bucket contains the maximum value also indicating the block boundaries that are passed to the report module. So, to design an algorithm in this arena, what is desired is a method of monitoring distortions in block-based codecs in a manner that is truly reference-less and it uses minimum amount of hardware acceleration like processing speed to perform such measurements in real time.

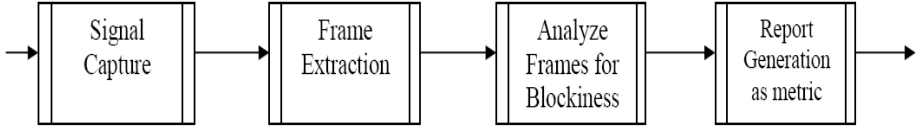
## 2 Proposed Blockiness Metric to Detect Video Artifacts

Most of the compression systems today separate video into blocks of data and separately compress each block using a discrete cosine transform (DCT). Research shows that in video which has been subjected to such kind of compression, image or, picture quality is strongly correlated with the visible blockiness. In our approach we first assume that our video is reference free and so we have to blindly measure the blockiness of videos or images.

### 2.1 Blind Measurement Approach

The measurement is shown in Fig. 1. Where firstly the video signal that has been compressed is to be measured for picture quality degradation is input to a signal

module. Then the frame is captured and converted to component video if necessary. Now the contents of the extracted frames are analyzed to measure the blockiness of the frame and the picture quality is provided to a report results module.



**Fig. 1.** Overview of the proposed system

## 2.2 Edge Enhancement Filtering

The A frame or field of a proposed video signal representing an image or picture is captured and converted, if necessary into luminance and color components. Traditionally, one or more of the components is analyzed by appropriate vertical and horizontal edge enhancement filtering. But if we consider the high gradient information in eight directions and use this information for the distortion measurement, we will gain apparently more accurate results.

So, to consider the different directional information we use Kirsch Masks [22] in eight directions. In Fig. 2, the kirsch masks for detecting eight directional edges information are shown. Taking a single mask and rotating it to eight major compass orientations: N, NW, W, SW, S, SE, E, and NE. The edge magnitude is equal to the maximum value found by the convolution of each mask with the image. The edge direction is defined by the mask that produces the maximum magnitude.

$$\begin{aligned}
 G_1 &= \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} & G_2 &= \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} & G_3 &= \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} & G_4 &= \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \\
 G_5 &= \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} & G_6 &= \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} & G_7 &= \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} & G_8 &= \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}
 \end{aligned}$$

**Fig. 2.** Eight directional Kirsch Masks

The  $g(x, y)$  across the pixel at  $(x, y)$  is determined by calculating the edge changes in eight directions. As shown in Fig. 2. Eight operators,  $G_k(x, y)$ , for  $k = 1, \dots, 8$  and  $i, j = 1, 2, 3$ , are employed to perform the calculation.

$$g(x, y) = \max_{k=1, \dots, 8} \{|grad_k(x, y)|\} \quad (1)$$

$$grad_k(x, y) = \sum_{i=1}^3 \sum_{j=1}^3 P(x - 2 + i, y - 2 + i).G_k(i, j)$$

for  $0 \leq x \leq H, 0 \leq y \leq W,$  (2)

Where  $p(x, y)$  denotes the pixel at  $(x, y)$ .  $H$  and  $W$  are the height and width of the frame consecutively. The resulting edges are correlated with an infinite grid having boundaries corresponding to the block boundaries used in the video compression. Optionally, a second correlation may be made with boundaries slightly different than the block boundaries used in the video compression, with this result being subtracted from the first value. Further the locations of the compression block boundaries may be detected by observing where the maximum correlation value occurs. The resulting correlation results are proposed to generate a picture quality rating for the image which represents the amount of human-perceivable block degradation that has been introduced into the proposed video signal. Fig. 3 shows the consequence after using Kirsch masks to detect the edges.

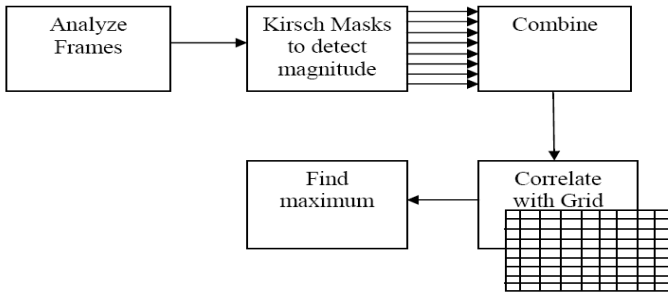


Fig. 3. An illustration of the extraction of blockiness from a frame

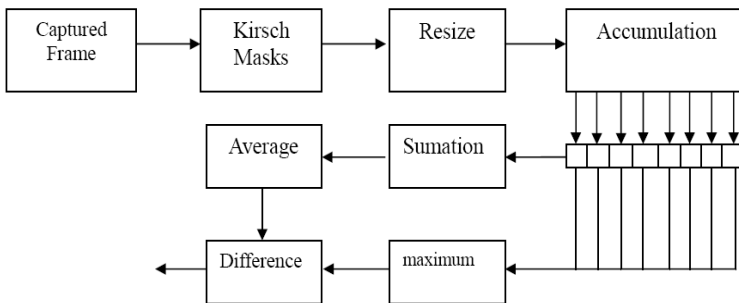


Fig. 4. Schematic diagram to measure blockiness

### 2.3 Generate Blockiness Metric

To generate the blockiness metric for real time systems we have to consider a faster and efficient approach to measure and also detect the location of the blockiness occurred in a frame. Many of the existing algorithms are based on HVS (Human



Visual Sensitivity). But these kinds of algorithms are not faster enough to use in real time. So, to use light weighted metric in real time applications we have develop a simpler approach. Fig. 4 shows a block diagram view of a block detection apparatus for a picture quality measurement system.

Here is the Matlab pseudo code is given to show the algorithm to detect blockiness after using the kirsch masks. After taking the maximum magnitude from the masks we will take the absolute value of the maximum edge direction:

```
AbsB = abs(B)
CB = AbsB(10:1920 , 10:1080) //Optional clipping//

% Here is the efficient cross-correlation algorithm:
Buck = zeroes(1,8); //Create buckets and initialize with zero//
ImgR = length(V(::1)); //Get picture Height//
ImgC = length(V(1:)); //Get Picture Weidth//
ImgC8 = floor(ImgC/8)*8-8; // Round width down to nearest x8//

for i=1:ImgR
    for j=0:8:ImgC8
        for k=1:8
            Buck(k) = Buck(k) + V(i, j+k);
        end
    end
end
end
```

Here we, sum the values into eight buckets with each bucket containing the total of all the columns by modulo eight (bucket #1 contains the sum of columns 1, 9, 17.....etc., bucket #2 contains the sum of columns 2, 10 ,18.....etc.). And finally we achieve the measurement values by the following equations:

$$S = \text{Sum}(Buck) / 8 \quad (3)$$

$$Block\_msr = (\max(Buck) - S) \times msr\_scl \quad (4)$$

The average values which are acquired from equation (3) are subtracted from the maximum values of the buckets and multiplied by *msr\_scl*, a scale factor which is a constant value (4) and which range can be achieved as the Picture Quality Rating (PQR) produced by a device like PQA200. Nevertheless, the bucket which contains the maximum value is also indicates where the compression block boundary is. It is also output to the report module. To improve the algorithm accuracy, measurement can be changed correlating with the block or, macro block spacing and then simultaneous calculation using a non-block correlated kernel size. This significantly reduces false positives in blockiness that a noisy image might otherwise produce. The location of the compressed blocks whether they are human-perceivable or not, may be determined and separately reported.

### 3 Experimental Result

The JPEG image dataset in the LIVE image quality assessment database release 2 [11, 12, 13] and the MPEG-2 video dataset in the LIVE video quality database [14, 15] are used. The JPEG image dataset includes 29 color reference images (typically  $768 \times 512$  in size) and 204 JPEG distorted images. The LIVE Video Quality Database uses ten uncompressed high-quality videos with a wide variety of content as reference videos. A set of 150 distorted videos were created from these reference videos (15 distorted videos per reference) using four different distortion types-MPEG-2 compression, H.264 compression, simulated transmission of H.264 compressed bit streams through error-prone IP networks and through error-prone wireless networks. Distortion strengths were adjusted manually taking care of ensuring that the different distorted videos were separated by perceptual levels of distortion. Each video in the LIVE Video Quality Database was assessed by 38 human subjects in a single stimulus study with hidden reference removal, where the subjects scored the video quality on a continuous quality scale. Please notice, only the luminance component of each image or video sequence is used for blockiness measurement.

#### 3.1 Experiments on JPEG Images

In the following experiments, we used a number of still images, as well as frames from the test video sequences. These images have different resolutions, ranging from  $176 \times 144$  to  $1920 \times 1080$ . We also compared our results with those from other objective quality metrics such as PSNR, the quality metrics  $M_{GBIM}$  of [5], and the NR quality metrics  $S$  of [3]. In order to plot all these metrics in the same figure, we scale PSNR by dividing a factor of 5. According to [5], there is no defined range for the  $M_{GBIM}$  and if  $M_{GBIM}$  values are greater than one, then blocking effect turns out severe. On the other hand, according to [3], the smaller the  $S$  is, the greater the severity of the blocking effect is.

Table 1 shows the Pearson Correlation and Spearman rank order Correction between the proposed blockiness measure and the subjective ratings of QCIF video sequences (obtained from subjective video quality experiments similar to that conducted for the evaluation of the JVT sequences [16]). It can be seen that compared to the metrics of [5] and [3], the *Block\_msr* of this paper has a better correlation with subjective test results. Table 2 shows the Pearson Correlation and Spearman rank-order Correlation between various quality metrics and the subjective ratings of the JEPG database provided by LIVE [11]. Table 1 and Table 2 show that, our metrics have a comparable correlation with other approaches using subjective data. Additionally, the advantages of our algorithm are that, it is locally adaptive, fast response to blocking artifacts and most of all, it is suitable for real-time implementation. These good technicalities of our algorithm can make it a good choice for practical usage and possibly outweigh the slight drop in correlation values [Table 1, Table 2]

**Table 1.** Pearson Correlation and Spearman for FUB database

Algorithm	Pearson Correlation	Spearman Correlation
<i>Block_msr</i>	-.721	.685
$M_{GBIM}$ [5]	-.597	.584
S[3]	.614	.570

**Table 2.** Pearson Correlation and Spearman for LIVE database

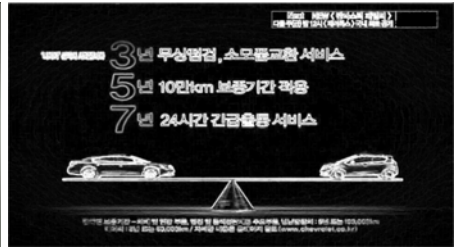
Algorithm	Pearson Correlation	Spearman Correlation
<i>Block _ msr</i>	-.843	.838
$M_{GBIM}$ [5]	-.727	.925
S[3]	.944	.937

### 3.2 Experiments on MPEG-2 Video Frames

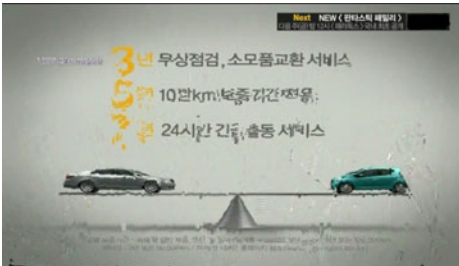
The proposed approach can be applied to a video sequence on a frame-by-frame basis. The blockiness measure for a sequence is defined as the mean value of the blockiness measures over all the video frames in the sequence. Testing results on the MPEG-2 video dataset are given in Table 3. In the first step we are showing experiment results on our own video datasets. For experimental result, we mention different frames of a video, where different frames have blocking artifacts. We can observe the reliability of our algorithm by comparing *Block\_msr* values between them.



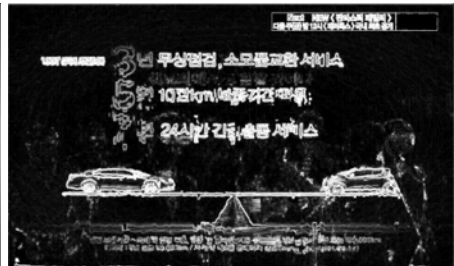
Frame No: 115



*Block\_msr*: 2455



Frame No: 116



*Block\_msr*: 7552



Frame No: 430



Block\_msr: 4485



Frame No: 536



Block\_msr: 10844



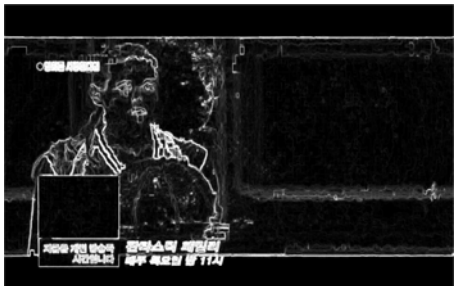
Frame No: 982



Block\_msr: 3409.34



Frame No: 1045



Block\_msr: 9804.73

Fig. 5. Experimental result of picture quality by using blockiness measurement

In Fig. 5 Left column is the color frame and right column is the frame by using kirsch masks. By observing the edge enhanced images we can observe that how the edge direction and magnitude is changed by the blockiness artifacts. Experimental results on the same video dataset using Wu and Yuen's [5], Vlachos' [18], Pan et al.'s [19], Perra et al.'s [20], Pan et al.'s [21], and Muijs and Kirenko's [6] are also reported. From Table 3. we can observe that most of these methods give very satisfactory performance while the proposed outperforms the state of the arts.

**Table 3.** Test blockiness result using different approaches on the MPEG-2 video dataset

Approaches	Pearson Corr.	Spearman Corr.	RMSE
Wu and Yuen's [5]	.6344	.7365	7.1869
Vlachos' [18]	.5378	.7930	7.0183
Pan et al.'s [19]	.6231	.6684	8.4497
Perra et al.'s [20]	.6916	.6531	8.4357
Pan et al.'s [21]	.5008	.6718	8.1979
Muijs & Kirenko's [6]	.7875	.6939	7.9394
Proposed Method	.8627	.7104	7.0236

## 4 Conclusion and Future Work

The proposed metric involves a grid detection phase, which is used to account for a block size change or grid shift, and intrinsically ensures the subsequent local processing of the blocking artifacts. For each predicted blocking artifact the blockiness is individually calculated as there is a signal discontinuity relative to its local content and its visibility because the masking is locally estimated. Combining the results in a simple way yields a metric that shows a promising performance with respect to practical reliability, prediction accuracy, and computational efficiency. Incorporating HVS to blockiness with this method is under implementation to make the algorithm to be more perfect depending on human visual characteristics.

**Acknowledgement.** This work was supported by a grant from the NIPA (National IT industry Promotion Agency) in 2011, (Fostering Global IT Human Resource Project).

## References

1. Liu, S., Bovik, A.C.: Efficient DCT-Domain Blind Measurement and Reduction of Blocking Artifacts. *IEEE Trans.on CSVT* 12(12), 1139–1149 (2002)
2. Wang, Z., Bovik, A.C., Evan, B.L.: Blind measurement of blocking artifacts in images. In: *International Conference on Image Processing*, Vancouver, BC, Canada (2000)
3. Wang, Z., Sheikh, H.R., Bovik, A.C.: No-Reference Perceptual Quality Assessment of JPEG Compressed Images. In: *IEEE International Conference on Image Processing*, pp. 477–480 (2002)
4. Tan, K.T., Ghanbari, M.: Frequency domain measurement of blockiness in MPEG- 2 coded video. In: *International Conference on Image Processing*, Vancouver, BC, Canada (2000)

5. Wu, H.R., Yuen, M.: A generalized block-edge impairment metric for video coding. *IEEE Signal Processing Letters* 4(11), 317–320 (1997)
6. Muijs, R., Kirenko, I.: A no-reference blocking artifact measure for adaptive video processing. In: *European Signal Processing Conference*, Antalya, Turkey (2005)
7. Chou, C., Li, L.: A perceptually tuned subband image coder based on the measure of just-noticeable-distortion profile. *IEEE Trans.on CSVT* 5(6), 467–476 (1995)
8. Karunasekera, S.A., Kingsbury, N.G.: A Distortion Measure for Blocking Artifacts in Images Based on Human Visual Sensitivity. *IEEE Transactions on Image Processing* 4(6), 713–724 (1995)
9. Liu, H., Heynderickx, I.: A no-reference perceptual blockiness metric. In: *International Conference on Acoustics, Speech, and Signal Processing*, Las Vegas, NV, USA (2008)
10. Yu, Z., Wu, H.R., Winkler, S., Chen, T.: Vision-Model-Based Impairment Metric to Evaluate Blocking Artifacts in Digital Video. *Proceedings of the IEEE* 90, 154–169 (2002)
11. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: LIVE image quality assessment database release 2 (2005), <http://live.ece.utexas.edu/research/quality>
12. Sheikh, H.R., Sabir, M.F., Bovik, A.C.: A statistical evaluation of recent full reference image quality assessment algorithms. *IEEE Transactions on Image Processing* 15(11), 3440–3451 (2006)
13. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing* 13(4), 600–612 (2004)
14. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: Study of subjective and objective quality assessment of video. *IEEE Transactions on Image Processing* 9(6), 1427–1441 (2010)
15. Seshadrinathan, K., Soundararajan, R., Bovik, A.C., Cormack, L.K.: A subjective study to evaluate video quality assessment algorithms. In: *SPIE Proceedings Human Vision and Electronic Imaging* (2010)
16. Baroncini, V.: ISO/IEC JTC 1/SC29/WG 11, 4240, Sidney (2001)
17. VQEG: Final report from the video quality experts group on the validation of objective quality metrics for video quality assessment, [http://www.its.bldrdoc.gov/vqeg/projects/frtv\\_phaseI](http://www.its.bldrdoc.gov/vqeg/projects/frtv_phaseI)
18. Vlachos, T.: Detection of blocking artifacts in compressed video. *IET Electronics Letters* 36(13), 1106–1108 (2000)
19. Pan, F., Lin, X., Rahardja, S., Lin, W., Ong, E., Yao, S., Lu, Z., Yang, X.: A locally-adaptive algorithm for measuring blocking artifacts in images and videos. In: *International Symposium on Circuits and Systems*, Vancouver, BC, Canada (2004)
20. Perra, C., Massidda, F., Giusto, D.D.: Image blockiness evaluation based on Sobel operator. In: *International Conference on Image Processing*, Genova, Italy (2005)
21. Pan, F., Lin, X., Rahardja, S., Ong, E.P., Lin, W.S.: Using edge direction information for measuring blocking artifacts of images. *Multidimensional Systems and Signal Processing* 18(4), 279–308 (2007)
22. Pratt, W.K.: *Digital Image Processing*. Wiley, New York (1978)

# An Improved Joint Particle Filter Algorithm for Multi-target Tracking

Jin-Long Yang and Hong-Bing Ji

School of Electronic Engineering, Xidian University,  
Xi'an 710071, P.R. China  
yjlgedegn@163.com

**Abstract.** An improved joint particle filter (JPF) algorithm is proposed by decomposing the joint particle weights. In the proposed method, the joint particle weights are used to estimate target states and the decomposed weights are used to resample particles of each target. Simulation results show that the proposed algorithm can better track closely spaced multi-target and/or other crossing targets, has a better performance than the conventional joint particle filter and the independent particle filter (IPF) algorithm.

**Keywords:** Joint particle filter, Independent particle filter, data association, decomposed weight.

## 1 Introduction

Multi-target tracking techniques for passive multi-sensor have already attracted great interest in the applications of computer vision, traffic monitoring and radar tracking of military aircrafts, etc. However, the increasing complexity of war environments makes multi-target tracking for passive multi-sensor a difficult problem and a hot issue in the tracking field.

Sequential Monte Carlo (SMC) method [1, 2] can solve nonlinear, non-Gaussian element of a dynamic system, which is widely used to track multiple targets in combination with the joint probability data association (JPDA) technique. However, due to the drawbacks of particle filter (PF) algorithm based JPDA, more improved PF-JPDA algorithms have been proposed for different problems. Especially, the problem of closely spaced targets has been presented and studied extensively in [3-5]. Independent PF-JPDA (IPF-JPDA) [5] has a good performance in tracking multi-target of non-crossing tracks; otherwise, track coalescence problem may appear. On the contrary, Joint MC-JPDA, also called Joint Particle Filter JPDA (JPF-JPDA) [5], has a good performance in tracking closely spaced targets, but it may cause the filter divergence when the targets deviate from each other.

In this paper, to solve the aforementioned problem, an improved Joint PF-JPDA algorithm is proposed, which can track closely spaced multi-target and other multi-target of crossing or non-crossing tracks. Simulation results show that the proposed method has a better performance than the IPF-JPDA and JPF-JPDA.

## 2 Joint PF-JPDA Algorithm

### 2.1 Sequential Monte Carlo JPDA (SMC-JPDA) Method

Suppose there are  $c$  targets in clutter, the state equation and measurement equation in two-dimensional case are expressed as follows:

$$\mathbf{x}_{k+1}^i = F\mathbf{x}_k^i + Gw_k^i, \quad (1)$$

$$\mathbf{y}_k^i = h(\mathbf{x}_k^i) + v_k^i. \quad (2)$$

Where,  $\mathbf{x}_k^i = [x_k^i, v_x, y_k^i, v_y]^T$  denotes the state of the  $i$  th target at time  $k$ ,  $i=1, \dots, c$ .  $(x_k^i, y_k^i)$  denotes the target position and  $(v_x, v_y)$  denotes the target velocity.  $w_k^i$  and  $v_k^i$  are system noise and observation noise sequences, which meet independent identical distribution with covariance  $Q$  and  $R$ , respectively,  $F$  and  $G$  denote the state transition matrix and the system process noise matrix, re-

spectively,  $F = \begin{bmatrix} 1 & T & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & T \\ 0 & 0 & 0 & 1 \end{bmatrix}$ ,  $G = \begin{bmatrix} \frac{T^2}{2} & 0 \\ T & 0 \\ 0 & \frac{T^2}{2} \\ 0 & T \end{bmatrix}$ .  $\mathbf{y}_t^j$  denotes the information of measure-

ment and  $h(\cdot)$  is a known non-linear evolution model.

Let  $\mathbf{y}_k = \{\mathbf{y}_k^1, \mathbf{y}_k^2, \dots, \mathbf{y}_k^{M_k}\}$  denotes the measurements at time  $k$ , where  $M_k$  denotes the number of measurements.  $\mathbf{Y}^k = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_k\}$  is the complete measurement history up to time  $k$ . The probability of an individual joint association event  $\lambda$  assigning measurement  $j$  to target  $i$  can be calculated according to the JPDA algorithm [6] as follows

$$p(\lambda | \mathbf{Y}^k) = \frac{1}{C} p_D^{\tau-z^0} (1-p_D)^{z^0} p_{FA}^{M_k-(\tau-z^0)} \prod_{(j,i) \in \lambda} p^i(\mathbf{y}_k^j | \mathbf{Y}^{k-1}). \quad (3)$$

Where,  $p_D$  denotes the detection probability,  $p_{FA}$  denotes the probability of false alarm.  $p^i(\mathbf{y}_k^j | \mathbf{Y}^{k-1})$  denotes the predicted likelihood of target  $i$ , and is given by

$$p^i(\mathbf{y}_k^j | \mathbf{Y}^{k-1}) = \int p(\mathbf{y}_k^j | \mathbf{x}_k^i) p(\mathbf{x}_k^i | \mathbf{Y}^{k-1}) d\mathbf{x}_k^i. \quad (4)$$

where,  $p(\mathbf{y}_k^j | \mathbf{x}_k^i)$  denotes the likelihood of measurement  $j$  and target  $i$ .



## 2.2 Joint PF-JPDA Algorithm

Assume the multi-target states are correlated and coupled. We concatenate the state vectors of all targets as a joint state vector. The idea of JPF-JPDA is to approximate the posterior probability of the joint state vector through the joint particles sampled from the joint state. In the process of resampling, the particles are resampled according to the joint particle weights. The steps of JPF-JPDA are as follows:

Step 1. Initialization of the joint particles of the joint state vector.  
 $\{\mathbf{x}_0^n\}_{n=1}^N = \{\mathbf{x}_0^{1,n}, \dots, \mathbf{x}_0^{c,n}\}_{n=1}^N$ ,  $w_0^n = 1/N$ , where  $\mathbf{x}_0^n$  and  $w_0^n$  denote the joint particle state of the  $n$ th particle and its weight, respectively.

Step 2. Calculation of the joint weights of joint particles when  $k \geq 1$ .

$$\omega_k^{(n)} = \bigcup_{\substack{j=0,\dots,M_k \\ i=1,\dots,c}} \omega_k^{(j,i)(n)}. \quad (5)$$

where  $\omega_k^{(j,i)(n)} = \sum_{\theta \in \Theta_{j,i}} p(\theta | \mathbf{Y}^k)^{(n)}$ ,  $p(\theta | \mathbf{Y}^k)^{(n)} = \frac{1}{C} p_D^{\tau-z^0} (1-p_D)^{z^0} p_{FA}^{M_k-(\tau-z^0)} \prod_{(j,i) \in \theta} p^i(\mathbf{y}_k^j | \mathbf{x}_k^{(n),j})$ ,  
 $p(\mathbf{y}_k^j | \mathbf{x}_k^{i,(n)}) = N(\mathbf{y}_k^j; h(\hat{\mathbf{x}}_k^{i,(n)}), \Sigma_{j,i})$ ,  $\Theta_{j,i}$  is the set of all valid events assigning measurement  $j$  to target  $i$  in the joint state vector for particle  $n$ .  $\Sigma_{j,i}$  is the association innovation covariance matrix,  $\hat{\mathbf{x}}_k^{i,(n)}$  is the estimated state of target  $i$ . Normalization of the joint weights

$$\bar{\omega}_k^{(n)} = \omega_k^{(n)} / \sum_{n=1}^N \omega_k^{(n)}. \quad (6)$$

Step 3. Estimation of each target states by  $\mathbf{x}_k^i = \sum_{n=1}^N \bar{\omega}_k^n \mathbf{x}_k^{i,(n)}$ .

Step 4. Resampling according to the joint weights  $\bar{\omega}_k^{(n)}$ , i.e., resampling  $N$  particles

$$\{\mathbf{x}_k^{n*}\}_{n=1}^N \text{ from } \{\mathbf{x}_k^n\}_{n=1}^N \text{ according to } p(\mathbf{x}_k^{n*} = \mathbf{x}_k^l) = \bar{\omega}_k^{(l)}.$$

Step 5. Prediction of new particles by  $\mathbf{x}_{k+1}^{i,n} = F \mathbf{x}_k^{i,n*} + G \mathbf{v}_k^{i,n}$ . Go back to Step 2.

## 3 Improved Joint Particle Filter Algorithm

As can be seen from Steps 2 and 4 of the JPF-JPDA, the particles of each target of the joint state have the same joint weights, and then are resampled by these weights. However, in practice, each particle of every target has different importance for their states. Especially, when the targets deviate from each other, the distinction of the importance is more evident. It is unreasonable to resample the particles according to the joint weights. Therefore, JPF-JPDA has a good performance in tracking closely

spaced targets, but it may cause the filter divergence for other targets. To solve this problem, the joint weights are decomposed in the proposed method, and are used to resample the particles of each target separately.

The decomposed weights can be obtained by

$$\omega_k^{i,(n)} = \left( l_k^{n,i} / \sum_{i=1}^c l_k^{n,i} \right) \bar{\omega}_k^{(n)}. \quad (7)$$

$$l_k^{n,i} = \frac{1}{\sqrt{2\pi \det(\sigma_k^i)}} \exp\left(-\frac{1}{2}(\hat{y}_k^i - h(x_k^{n,i}))^T (\sigma_k^i)^{-1} (\hat{y}_k^i - h(x_k^{n,i}))\right). \quad (8)$$

where  $l_k^{n,i}$  denotes the likelihood of the  $n$ th particle of the target  $i$  at time  $k$ ,  $\hat{y}_k^i = h(\mathbf{x}_k^i)$ ,  $\sigma_k^i$  is the covariance of the target  $i$ .

Normalize the decomposed weights for  $i=1, \dots, c$  by  $\bar{\omega}_k^{i,(n)} = \omega_k^{i,(n)} / \sum_{n=1}^N \omega_k^{i,(n)}$ .

Then, resample the particles according to the decomposed weights  $\bar{\omega}_k^{i,(n)}$ , which displaces  $\bar{\omega}_k^{(n)}$  in Step 4 of the JPF-JPDA, i.e.,  $p(\mathbf{x}_k^{i,(n^*)} = \mathbf{x}_k^{i,(l)}) = \bar{\omega}_k^{i,(l)}$ ,  $i=1, \dots, c$ .

*Remark 1. Resampling the particles of each targets according to the decomposed weights, which can increase the diversity of the particles, propagates the particles with big weights and restrains the particles with small weights reasonably. So that the proposed method can overcome the disturbance among the targets, and then better track the closely spaced targets and other crossing targets.*

## 4 Simulations

To verify the effectiveness of the proposed algorithm, we take two tracking examples for multiple passive target tracking. The first example considers a tracking case for closely spaced multiple targets, the second example considers a tracking case for crossing multiple targets. The state equation and measurement equation under passive multi-sensor tracking system in two-dimensional cases are the same as Eqs. (1) and

(2), where  $h(\mathbf{x}_k^i) = \arctan\left(\frac{y_k^i - S_y}{x_k^i - S_x}\right)$ ,  $[S_x, S_y]$  denotes the location of the sensor.

Assume that the measurement data of each observation station have been associated and registered, and the centralized fusion strategy is employed.

**Example 1.** There are three targets with a formation parallel flight with the same velocity, whose initial states are  $\mathbf{x}_0^1 = [15500\text{m}, -350\text{m/s}, 9100\text{m}, -100\text{m/s}]^T$ ,  $\mathbf{x}_0^2 = [15000\text{m}, -350\text{m/s}, 9100\text{m}, -100\text{m/s}]^T$ , and  $\mathbf{x}_0^3 = [14500\text{m}, -350\text{m/s}, 9100\text{m}, -100\text{m/s}]^T$ .

There are also three passive sensors, whose locations are  $[-2000\text{m}, -2000\text{m}]$ ,  $[18000\text{m}, -2000\text{m}]$ , and  $[-2000\text{m}, 17000\text{m}]$ , respectively. In this simulation, the sampling interval is  $T = 1\text{s}$ , sampling 40 times in all. Averagely 1 clutter appears at each scan. Take  $Q = \text{diag}[\sigma_x^2 \ \sigma_y^2]$ , where  $\sigma_x = \sigma_y = 10\text{m}$ ,  $R = \text{diag}[\sigma_{s_1}^2 \ \sigma_{s_2}^2 \ \sigma_{s_3}^2]$ , where  $\sigma_{s_1} = \sigma_{s_2} = \sigma_{s_3} = 5 \text{ mrad}$ , detection probability  $P_d = 0.98$ .

Figs.1 and 2 show the true tracks of the three targets and their estimated root mean square errors (RMSE) of position from 100 Monte Carlo runs. As can be seen, the proposed method, called decomposed weight JPF-JPDA (WDJPF-JPDA), has a higher accuracy than the IPF-JPDA method, and slightly superior to the JPF-JPDA method.

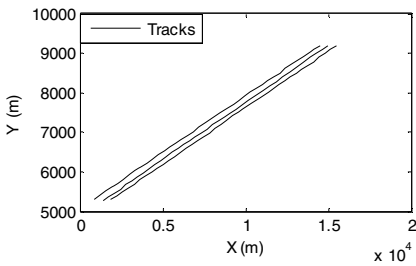


Fig. 1. True tracks of the three targets

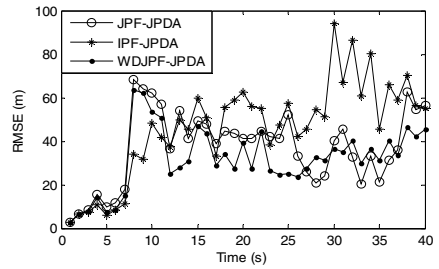


Fig. 2. Comparison of the estimated RMSE

**Example 2.** Assuming there are three crossing targets with uniform motion, whose initial states are  $\mathbf{x}_0^1 = [15000\text{m}, -350\text{m/s}, 4900\text{m}, 100\text{m/s}]^T$ ,  $\mathbf{x}_0^2 = [15000\text{m}, -350\text{m/s}, 9100\text{m}, -100\text{m/s}]^T$ , and  $\mathbf{x}_0^3 = [9100\text{m}, 0\text{m/s}, 13000\text{m}, -300\text{m/s}]^T$ . The other parameters are the same with those of the Example 1.

Figs.3 and 4 show the true tracks of the three targets and their estimated RMSE in position from 100 Monte Carlo runs. As can be seen, the proposed method WDJPF-JPDA has higher accuracy than the JPF-JPDA method and the IPF-JPDA method.

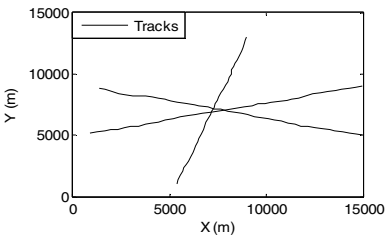


Fig. 3. True tracks of the three targets

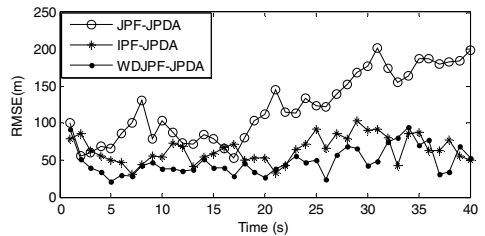


Fig. 4. Comparison of the estimated RMSE

## 5 Conclusion

An improved JPF-JPDA algorithm is proposed in this paper, which can track the closely spaced targets and/or the other crossing targets. Simulations show that the proposed method has better performance than the IPF-JPDA algorithm and JPF-JPDA algorithm. In the future work, this algorithm also can be extended to track the unknown number and varying target tracking.

**Acknowledgment.** This work was supported by the National Natural Science Foundation of China (No. 60871074) and partially by Xidian Graduate Innovation Project.

## References

1. Jaco, V., Simon, J., Patrick, P.: Monte Carlo Filtering for Multi-Target Tracking and Data Association. *IEEE Transactions on Aerospace and Electronic Systems* 41(1), 309–332 (2005)
2. Liu, B., Ji, C., Zhang, Y., et al.: Multi-target tracking in clutter with sequential Monte Carlo methods. *IET Radar Sonar Navig.* 4(5), 662–672 (2010)
3. Boers, Y., Sviestins, E., Driessen, H.: Mixed labelling in multitarget particle filtering. *IEEE Trans. on Aerospace and Electronic System* 46(2), 792–801 (2010)
4. Kreucher, C., Kastella, K., Hero, A.: Multi-target tracking using the joint multitarget probability density. *IEEE Trans. on Aerospace and Electronic System* 41(4), 1396–1414 (2005)
5. Ekman, M., Boers, Y., Driessen, J.N., et al.: Particle filters for tracking closely spaced targets. In: *Proceedings of 10th International Conference on Information Fusion, Quebec, Canada, (2007)*
6. Ekman, M.: Particle filters and data association for multi-target tracking. In: *International Conference on Information Fusion, pp. 1–8 (2008)*

# Enhanced Fuzzy-Based Models for ROI Extraction in Medical Images

Yasser El-Sonbaty<sup>1</sup>, Sherin M. Youssef<sup>2</sup>, and Karma M. Fathalla<sup>2</sup>

<sup>1</sup> College of Computing and IT, Arab Academy for Science and Technology, Egypt

<sup>2</sup> College of Engineering, Arab Academy for Science and Technology, Egypt  
yasser@aast.edu, sherin@aast.com

**Abstract.** Standard Fuzzy C-Means (FCM) clustering has been widely used as an effective method for image segmentation. However, FCM is sensitive to initialization and is easily trapped in local optima. In this paper, several enhanced models for FCM clustering were proposed, namely W\_SS\_FCM, LAWS\_SS\_FCM and H\_FCM, to promote the performance of standard FCM. The proposed algorithms merge partial supervision with spatial locality to increase conventional FCM's robustness. A comparison study was conducted to validate the proposed methods' performance applying well established measures on three datasets. Experimental results show considerable improvement over standard FCM and other variants of the algorithm. It also manifests high robustness against noise attacks.

**Keywords:** FCM clustering, segmentation, spatial locality, partial supervision.

## 1 Introduction

Fuzzy C-Means (FCM) clustering [1] is an unsupervised technique that has a wide popularity over many other clustering algorithms[2-4]. It has been used in many application fields such as medical imaging, geology, astronomy and image segmentation. FCM can be applied on data points in various feature spaces, where it groups similar data points in the feature space

The field of image segmentation has attained wide attention during the past decades. Segmentation is considered an essential preprocessing step in many image-processing systems. The immense importance and criticality of such a process demands the continuity of enhancing the available segmentation FCM clustering is one of the established means of segmentation. A lot of research has been carried out to promote FCM's performance.

In this paper, the standard FCM algorithm has been integrated with several modifications to form enhanced versions of conventional FCM.

This paper is organized as follows. Section 2 represent some variants of FCM that are present in the literature, the problem formulation is illustrated in section 3 while the proposed methods are explained in section 4. Section 5 describes the experimental results. The conclusions are discussed in section 6.

## 2 Related Work

Several researches have been dedicated to promoting FCM clustering performance and overcoming the related problems.

In [5], FCM\_S is introduced where the neighborhood effect is also incorporated within the objective function through a crucial parameter  $\alpha$ . It is used to balance between robustness to noise and effectiveness of preserving the details of the image to act as a regularizing term and produce more homogeneous regions. Chen et al [6] present two variants namely KFCM\_S<sub>1</sub> and KFCM\_S<sub>2</sub>. KFCM\_S<sub>1</sub> and KFCM\_S<sub>2</sub> introduce the mean-filtered image and median filtered image respectively, which can be computed in advance. They also use a kernelized distance measure to elevate the drawbacks of Euclidean distance. A similar approach of computing in advance a linearly weighted sum image is presented in [7]. An approximate mean filtered image is generated where a parameter  $\alpha$  controls the weight of the mean filtered neighborhood. A novel locality factor is introduced in FGFCM [8], local measure  $S$  incorporates both spatial and gray levels similarity measures. It aims to preserve details and provides robustness against different types of noise. It overcomes the limitation of choosing the parameter  $\alpha$ .

## 3 Overview on Semi-supervised FCM

Partial Supervision can be incorporated within Standard FCM algorithm to provide considerable solutions to some of FCM problems. Partial supervision ensures correct choice of the number of clusters and adds meaningful physical labels to them. One of the examples of partially supervised FCM is semi-supervised FCM algorithm (SS\_FCM) which was introduced by [9] which allows the incorporation of expert knowledge into the FCM model. An expert defines a set of crisp labeled pixels that guide the clustering process. We denote the partially labeled pixels as:

$$P = \{ p_1^1, \dots, p_{n_1}^1, p_1^2, \dots, p_{n_2}^2, p_1^c, \dots, p_{n_c}^c \mid p_1^u, \dots, p_{n_u}^u \} = P^l \cup P^u \quad (1)$$

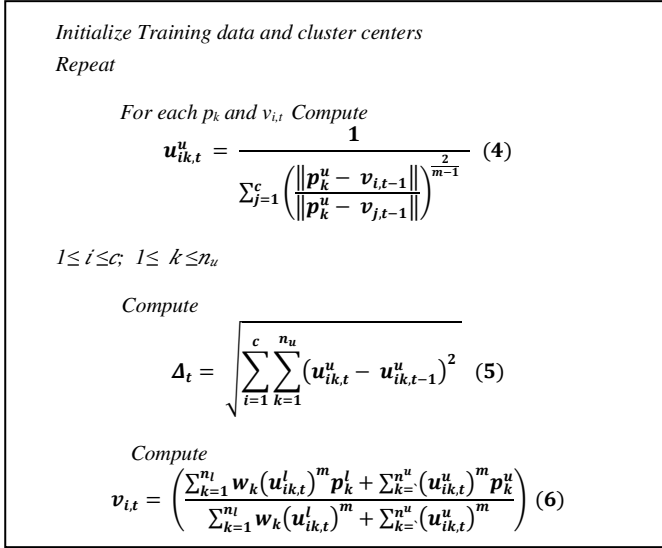
The superscripts represent the class numbers with maximum  $c$  classes specified by the training data,  $u$  represents unlabeled pixels and  $n_i$  denotes the number of pixels belonging to class  $i$ . The total number of pixels equals  $n$  where  $n = n_l + n_u$ . The fuzzy  $c$ -partition matrix of  $P$  has the following form:

$$\left\{ \begin{array}{l} U^l = \{u_{ik}^l\} \\ \text{labeled} \end{array} \right\} \left\{ \begin{array}{l} U^u = \{u_{ik}^u\} \\ \text{unlabeled} \end{array} \right\} \quad (2)$$

$U^u$  is initialized randomly while the membership values in  $U^l$  are hard labeled beforehand. The provided  $u_{ik}^l$  are used to compute an initial set of ‘‘well seeded’’ cluster prototypes  $v_{1 \rightarrow c}^0$ .

$$v_{i,0} = \frac{\sum_{k=1}^{n_l} (u_{ik}^l)^m p_k^l}{\sum_{k=1}^{n_l} (u_{ik}^l)^m} \quad (3)$$

Several parameters need to be set such as the maximum allowable number of iterations  $T_{max}$ ,  $m$  which controls the fuzziness of the resulting partition and  $\varepsilon$  which is the minimum difference required between iterations also  $w_k$  that can be used to weigh the labeled pixels more heavily than the unlabeled ones. The algorithm can be summarized as shown in Fig.1. The described variant shown in Fig.1 can be reduced to conventional FCM by setting  $n_l = 0$  and  $w_k = 0$ .



**Fig. 1.** A Brief Summary of SS\_FCM Algorithm

## 4 The Proposed Models Description

When processing medical images, there is necessitating demand to extract Regions of Interest (ROIs) from the image. The ROI of image is the most important part of the image which needs focusing on. The ROI of the image usually needs to be kept intact from any further processing.

Our suggested work focuses on extraction of ROIs within an image and separating them from non\_ROI and background (BG) layers. The ROIs needs to be easily separated if different type of processing is needed to be applied on the ROIs. The categorization of the ROIs, non\_ROI and BG is based on expert selection which is incorporated in SS\_FCM model.

In this paper, several modifications have been applied that aim to increase robustness to noise. Also, yield more homogeneous regions with less spurious blobs. The presented work was also motivated by the extensive study of incorporating spatial information into the FCM clustering model and its promising results. The modifications applied all aim at including the effect of the local neighborhood and allowing it to influence the member ship value of each pixel. The first modification was to apply 2D adaptive noise removal filtering as a preprocessing step to the semi-supervised FCM (SS\_FCM) clustering process. Two-Dimensional low pass Wiener

filter [10] was used to produce  $W\_SS\_FCM$  variant. Wiener filter performs pixel-wise adaptive filtering based on statistics estimated from a local neighborhood of each pixel.

It is specifically useful in case of an intensity image that has been degraded with constant power additive noise. 2D Wiener filter estimates the local mean and variance around each pixel as in equations (7) and (8) :

$$\mu = \frac{1}{nm} \sum_{i,j \in NB} I(i,j) \quad (7) \quad \text{and} \quad \sigma^2 = \frac{1}{nm} \sum_{i,j \in NB} I(i,j)^2 - \mu^2 \quad (8)$$

where  $i$  and  $j$  belong to  $n \times m$  neighborhood around each pixel of image  $I$ .

An estimate for each pixel in filtered image  $FI$  is computed as follows:

$$FI(r,c) = \mu + \frac{\sigma^2 - \omega}{\sigma^2} (I(r,c) - \mu) \quad (9)$$

where  $\omega$  is the average of all the local estimated variances.

Another adopted approach, is to create weighted-sum intensity image using LAWS [11] level mask. It is used as the input image to  $SS\_FCM$  clustering algorithm creating  $LAWS\_SS\_FCM$  variant. The weighted-sum image ( $WI$ ) is computed by convolving the original  $N \times M$  intensity image ( $I$ ) by LAWS  $5 \times 5$  normalized level mask ( $NL5L5$ ). The 2D convolution is given by:

$$WI(r,c) = \sum_{i=r-2}^{r+2} \sum_{j=c-2}^{c+2} NL5L5(i,j) * I(i,j) \quad (10)$$

$$\text{Where } NL5L5 = \frac{1}{36} \begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix},$$

$r = 0,1,2,\dots,N$ ,  $c = 0,1,2,\dots,M$  and  $*$  denotes 2D convolution.

The resultant image pixel values represent a weighted summation of the values of the neighborhood pixels in  $5 \times 5$  cardinality. The influence of the neighborhood pixels is inversely proportional to their distance to the center pixel in a radial manner. This approach targets reducing the effect of noisy pixels on segmentation but it also leads to blurring of some details.

Moreover, one of the promising methods of incorporating spatial locality was presented by [12]. We integrated this spatial model with the  $SS\_FCM$  model into a Hybrid FCM (H-FCM) algorithm. The main change to the previous  $SS\_FCM$  is the update function of the membership values.

The function accounts for the membership values of the neighborhood pixels. A spatial function defined as:

$$s_{ik}^u = \sum_{j \in NB(p_k)} u_{ij} \quad (11)$$

$NB(p_k)$  represents a square  $5 \times 5$  window centered at pixel  $p_k$

The spatial function is included within the membership function using the following formula:

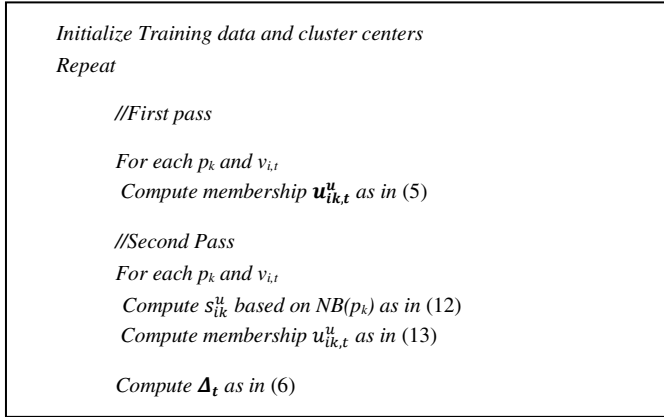
$$u_{ik}^u = \frac{u_{ik}^u s_{ik}^u}{\sum_{j=1}^c u_{jk}^u s_{jk}^u} \quad (12)$$

The clustering process comprises two passes. In the first pass the membership function  $u_{ik}^u$  is calculated as in equation (4) in the spectral domain. The spatial



function  $s_{ik}^u$  is incorporated into  $u_{ik}^u$  in the second pass as in equation (12) to reflect the mapping of the pixels into the spatial domain. A summary of the described H\_FCM is shown in Fig.2.

After the convergence of the clustering process, a defuzzification step is required to provide hard labels to the clustered data points (pixels). Maximum membership method is used to obtain our segmented image.



**Fig. 2.** A Brief Summary of H\_FCM Algorithm

## 5 Experimental Results

Experiments have been carried out to validate the efficiency of the proposed models. In this section we illustrate the characteristics of the proposed technique with several data sets. The performance of the proposed modification models has been compared with FCM, S\_FCM and SS\_FCM clustering algorithm. Results are compared using different validity indices (as shown below). Comparative results are presented for synthetic datasets and various real data collections.

### 5.1 Test Dataset

The datasets used to evaluate the performance of the proposed techniques are divided into three categories. The test set comprises hundred real MR images of patients with brain tumor which is the ROI in the experiments. Additionally, pathology free MR phantoms [13-15] of twenty subjects. A degree of abstraction was applied to segment the phantoms into three layers (clusters) this was obtained by combining several types of tissues into a single layer. Three sections are taken for each subject. The main advantage of experimenting with MR phantoms is the availability of a segmentation ground truth to be able to judge the segmentation accuracy. The images used for testing are 256x256 grayscale images. 50 Lung CT scan [16] images of size 512 x 512 are used in the testing phase. The used images contain nodular lesions considered as our ROI.

## 5.2 Validation Functions

Several validation functions are used to evaluate the clustering algorithms performance. Two of the most commonly used functions are partition coefficient index (*pcIndex*) and partition entropy index (*peIndex*). They measure the degree of fuzziness of partitions. They are defined as:

$$pcIndex = \frac{\sum_{k=1}^n \sum_{i=1}^c u_{ij}^2}{n} \quad (13)$$

and

$$peIndex = \frac{-\sum_{k=1}^n \sum_{i=1}^c [u_{ij} \log u_{ij}]}{n} \quad (14)$$

Another measure that is used is the Xie and Beni validity function [17]. It is based on the feature structure which is an advantage over the previous methods. It is defined as:

$$xbIndex = \frac{-\sum_{k=1}^n \sum_{i=1}^c u_{ik} \|p_k - v_i\|^2}{n * \min_{i \neq j} \{\|v_j - v_i\|^2\}} \quad (15)$$

$u_{ik}$  presents the degree of association or membership function of the  $i$ -th data point to the  $k$ -th cluster.

Segmentation accuracy (*SA*) is the most direct measure to evaluate clustering algorithms as segmentation algorithms. It can be only applied when the ground truth of segmentation is known. *SA* is defined as:

$$SA = \frac{\# \text{ correctly classified pixels}}{\text{Total \# pixels}} \quad (16)$$

## 5.3 Results

Different experiments have been carried out to test and validate the proposed model using the described data sets. The performance indicators described above were used to provide statistical evaluation of performance. Given the specified measures, better performance is indicated by maximization of *pcIndex* as well as segmentation accuracy and minimization of *peIndex* and *xbIndex*. All experimental results are averaged over 5 runs. Experimental study has been applied for parameter settings. After experimental analysis, the parameters  $Tmax$ ,  $m$ ,  $w_k$ ,  $n_l$  and  $\varepsilon$  (described in section IV) are set to 100, 2, 1, 150 and 0.0001, respectively.

In Tables 1, 2 and 3, the performance of the proposed algorithms is compared to standard FCM, SS\_FCM and S\_FCM for different datasets as specified. As observed from the tables, there is slight variation in the performance; however, on average the following observations can be made. The application of adaptive filtering in W\_SS\_FCM leads to slight enhancement in compared to standard FCM and SS\_FCM mechanisms. LAWS\_SS\_FCM has resulted in some deterioration in fuzziness and compactness of the resulting partitions; this is due to the blurring effect of the applied mask. The results also show a promising improvement in performance for H\_FCM algorithm over all the other algorithms. H\_FCM offers about 5.9% increase in *pcIndex* and around 43% decrease in both *peIndex* and *xbIndex*. H\_FCM provides a

small but yet considerable advancement in performance over S\_FCM. This can be explained by the fact that the hard labeled pixels reduce the fuzziness and increase the compactness of partitions. The inclusion of spatial distribution within H\_FCM allows the effect of the training pixels to propagate.

**Table 1.** Average Results of FCM and its Variations applied on the MR Phantoms Dataset

	<i>pcIndex</i>	<i>peIndex</i>	<i>xbIndex</i>
<b>FCM</b>	0.9269	0.0619	-0.0821
<b>SS_FCM</b>	0.9289	0.0607	-0.0812
<b>S_FCM</b>	0.9600	0.0454	-0.0486
<b>W_SS_FCM</b>	0.9299	0.0661	-0.0810
<b>LAWS_SS_FCM</b>	0.9250	0.0637	-0.0819
<b>H_FCM</b>	0.9730	0.0405	-0.0390

Table 1 illustrates the results of the discussed algorithms when applied on MR phantoms. S\_FCM and H\_FCM show the best results. H\_FCM outperforms S\_FCM with an average improvement of 10.2% with the used indicators. Table 2 presents the discussed algorithms results on Real Brain MR images. The ranking of the algorithms is nearly the same as the MR phantoms results with slight degradation in the relative performance of H\_FCM

**Table 2.** Average Results of FCM and its Variations applied on the Real Brain Images Dataset

	<i>pcIndex</i>	<i>peIndex</i>	<i>xbIndex</i>
<b>FCM</b>	0.8992	0.0934	-0.1066
<b>SS_FCM</b>	0.9216	0.0816	-0.1029
<b>S_FCM</b>	0.9424	0.0458	-0.0699
<b>W_SS_FCM</b>	0.9097	0.0804	-0.1026
<b>LAWS_SS_FCM</b>	0.8997	0.0863	-0.1067
<b>H_FCM</b>	0.9534	0.0437	-0.0690

Table 3 illustrates the results applied on Lung CT scans. H\_FCM still exhibits the highest performance. It manifests good performance as it overcomes the problem of equating cluster sizes.

**Table 3.** Average Results of FCM and its Variations applied on the Lung CT Scans Dataset

	<i>pcIndex</i>	<i>peIndex</i>	<i>xbIndex</i>
<b>FCM</b>	0.9003	0.0750	-0.1563
<b>SS_FCM</b>	0.8807	0.0739	-0.1650
<b>S_FCM</b>	0.9504	0.0476	-0.0918
<b>W_SS_FCM</b>	0.8959	0.0729	-0.1613
<b>LAWS_SS_FCM</b>	0.8940	0.0767	-0.1658
<b>H_FCM</b>	0.9608	0.0424	-0.0876

Table 4 demonstrates the segmentation accuracy of the explained algorithms on various MR phantoms sections. Segmentation accuracy shows different results in terms of ordering the relative performance of the clustering algorithms. H\_FCM still presents the highest performance with 4.6% increase in average accuracy over FCM while S\_FCM segmentation accuracy fell slightly below the proposed algorithms. In the proposed algorithms, the cluster centers are better defined than in case of S\_FCM for two reasons. All of the proposed algorithms present better results in terms of accuracy relative to conventional FCM. First, initialization is based on prior knowledge unlike random initialization. The second reason is that during center update the labels of the training pixels are unaltered thus allow centers to remain “well seeded”. Also, the provided supervision reduces the tendency of unsupervised FCM to equate cluster volumes.

**Table 4.** Average Segmentation Accuracy of FCM and its Variations applied on Different MR Phantoms

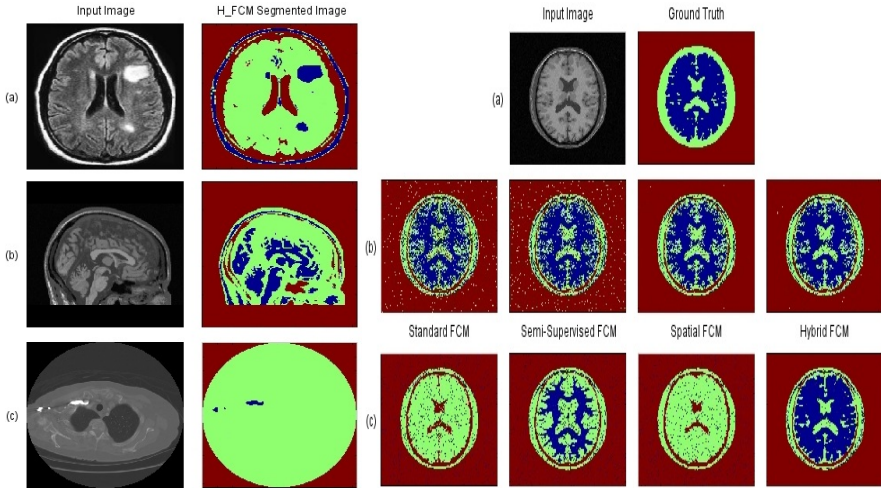
	Set-1 Traversal	Set-1 Coronal	Set-1 Sagittal	Average
FCM	0.8676	0.8650	0.9324	<b>0.8883</b>
SS_FCM	0.8698	0.9242	0.9345	<b>0.9095</b>
S_FCM	0.8719	0.9290	0.9272	<b>0.9094</b>
W_SS_FCM	0.8690	0.9252	0.9349	<b>0.9097</b>
LAW_SS_FCM	0.8686	0.9239	0.9306	<b>0.9077</b>
H_FCM	0.8848	0.9445	0.9579	<b>0.9291</b>

The discussed results revealed that H\_FCM provides the best performance within the presented algorithms. Fig.3. shows the results of applying H\_FCM. Fig.3.(a) shows real MR image with brain tumor, Fig.3.(b) MR phantom and Fig.3. (c) Chest CT scan. Fig.3. (c) displays the extraction of nodular lesion in the image, it is clear that H\_FCM does not suffer from the limitation of equating the cluster sizes since it extracted a very small ROI. It is also evident through visual analysis that it provides highly acceptable results. Further investigations were conducted to validate its robustness against noise. MR Phantoms were used in the coming section of our results. Two types of noise were added which are Gaussian noise and salt & pepper with varying strength. Fig.4. shows the segmentation results of the illustrated algorithms when adding different types of noise. Fig.4.(b) shows results when adding uniform Gaussian noise with variance 0.1 while Fig.4. (c) when adding 10% salt and pepper noise. The ground truth is provided based on the given segmented images in [13] to enable comparison. Visually, it is clear that H\_FCM provides the highest performance compared to other algorithms. Standard FCM and Spatial FCM give very poor results with salt and pepper noise.

Fig.5(a). displays the segmentation accuracy against different levels of Gaussian noise. It is evident H\_FCM performance is stable with increasing levels of noise while FCM and SS\_FCM accuracy is degraded rapidly. S\_FCM manifests similar results to these of H\_FCM with almost 0.2% worsening in results. Since the added Gaussian noise is uniform with coherent gray levels, the consideration of spatial locality affects positively the algorithms' output.

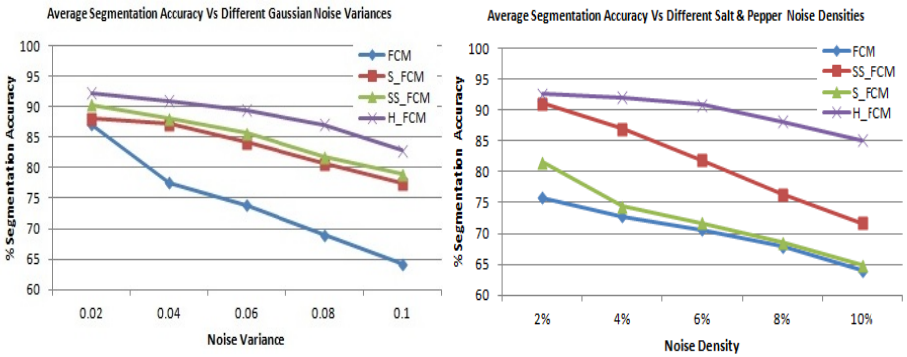
The addition of salt and pepper noise reduces remarkably the efficiency of S\_FCM as illustrated in Fig.5(b) while H\_FCM shows noticeable robustness till 8% salt and

pepper noise. Salt and Pepper noise can be considered as spike noise relative to the gray levels of MR phantoms. S\_FCM treats salt noise as a separate cluster thus exhibit low segmentation accuracy. In H\_FCM clustering centers are guided by expert help thus avoid the effect of randomness and floating centers position which lead to S\_FCM results.



**Fig. 3.** Segmentation Results of H\_FCM on sample images

**Fig. 4.** Segmentation Results of the illustrated algorithms under Gaussian(b) and Salt&Pepper (c)



**Fig. 5.** Average Segmentation Accuracy Vs (a)Gaussian Noise Variances (b) Salt& Pepper Densities

## 6 Conclusions

In this paper, several enhanced models for FCM clustering were proposed namely W\_SS\_FCM, LAWS\_SS\_FCM and H\_FCM. The proposed algorithms merge partial supervision with spatial locality to increase conventional FCM's robustness and improve its performance. The algorithms showed variable results in terms of partition

fuzziness and segmentation accuracy. All the results indicated improved performance compared to standard FCM and other suggested variations in the literature. H\_FCM demonstrated the best results compared to the proposed counterparts. Thus, its performance was validated against Gaussian and Salt and pepper attacks and it revealed considerable robustness. It is evident from the results that the integration of expert knowledge with spatial locality provides satisfactory enhancement.

## References

1. Bezdek, J.C.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum Press, New York (1981)
2. Youssef, S.M., Rizk, M., El-Sherif, M.: Dynamically Adaptive Data Clustering using Intelligent Swarm-like Agents. *International Journal of Mathematics and Computing* 1, 108–118 (2007)
3. Likas, A., Vlassis, N., Verbeek, J.J.: The Global k-means Clustering Algorithm. *Pattern Recognition* 36, 451–461 (2003)
4. Yang, M.-S., Wu, K.-I.: A Similarity-based Robust Clustering Method. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 434–448 (2004)
5. Ahmed, M.N., Yamany, S.M., Mohamed, N., Farag, A.A., Moriarty, T.: A Modified Fuzzy C-means Algorithm for Bias Field Estimation and Segmentation of MRI Data. *IEEE Transactions on Medical Imaging* 21, 193–199 (2002)
6. Chen, S.C., Zhang, D.Q.: Robust Image Segmentation using FCM with Spatial Constraints Based on New Kernel-Induced Distance Measure. *IEEE Transactions on Systems, Man and Cybernetics* 34, 1907–1916 (2004)
7. Szilagyi, L., Benyo, Z., Szilagyi, S.M., Adam, H.S.: MR Brain Image Segmentation Using an Enhanced Fuzzy C-Means Algorithm. In: *25th International Conference of IEEE EMBS*, pp. 17–21 (2003)
8. Cai, W., Chen, S., Zhang, D.: Fast and Robust Fuzzy C-means Clustering Algorithms Incorporating Local Information for Image Segmentation. *Pattern Recognition* 40, 825–838 (2007)
9. Bensaid, A.M., Hall, L.O., Bezdek, J.C., Clarke, L.P.: Partially Supervised Clustering for Image Segmentation. *Pattern Recognition* 29, 859–871 (1996)
10. Lim, J.S.: *Two-Dimensional Signal and Image Processing*. Prentice Hall, Englewood Cliffs (1990)
11. Laws, K.I.: *Textured image segmentation*. University of Southern California, Los Angeles, California (1980)
12. Chuang, K.-S., Tzeng, H.-L., Chen, S., Wu, J., Chen, T.-J.: Fuzzy c-means Clustering with Spatial Information for Image Segmentation. *Computerized Medical Imaging and Graphics* 30, 9–15 (2006)
13. <http://www.bic.mni.mcgill.ca/brainweb>:
14. Aubert-Broche, B., Evans, A.C., Collins, L.: A new improved version of the realistic digital brain phantom. *NeuroImage* 32, 138–145 (2006)
15. Aubert-Broche, B., Griffin, M., Pike, G.B., Evans, A.C., Collins, D.L.: Twenty New Digital Brain Phantoms for Creation of Validation Image Databases. *IEEE Transaction on Medical Imaging* 25, 1410–1416 (2006)
16. <https://public.cancerimagingarchive.net/ncia>:
17. Xie, X.L., Beni, G.A.: Validity Measure for Fuzzy Clustering. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13, 841–847 (1991)

# A Variation of Local Directional Pattern and Its Application for Facial Expression Recognition

Tianwei Xu, Juxiang Zhou, and Yunqiong Wang

College of Information  
Yunnan Normal University  
Kunming, China

xutianwei@ynnu.edu.com, zjuxiang@126.com,  
wyunqiong@sina.com.cn

**Abstract.** In this paper we first present an effective image description method for facial expression recognition, which is a variation of local directional pattern (LDP). Then we introduce weightings on the modular's LDP and investigate the effect on recognition rates with different weightings. Finally, the overlapped block is proposed when using LDP and proposed method. For recognition, this paper adopts PCA+LDP subspace method for feature reduction, and the nearest neighbor classifier is used in classification. The results of extensive experiments on benchmark datasets JAFFE and Cohn-Kanade illustrate that the proposed method not only can obtain better recognition rate but also have speed advantage. Moreover, the appropriately selected weightings and regional overlapping can improve recognition rates for both proposed method and LDP method.

**Keywords:** Facial expression recognition, LBP, LDP.

## 1 Introduction

With incessant growing of image analysis and pattern recognition technology, facial expression recognition has attracted more and more attention. It is not only an important content of the research field of human-computer interaction, but also an integral part of human psychology theory and emotional simulation study [1]. Facial expression recognition generally contains three processes: image acquisition, feature extraction and expression classification, in which feature extraction is the key point. And the performance of an expression recognition method more critically depends on the extracted expression features with better discrimination capability.

In general there are two type of approaches to extract facial features: geometric feature-based methods and appearance-based methods [2-4]. The Geometric feature-based methods extract the shape and locations of facial components (including mouth, eyes, and nose) to represent the face geometry [5, 6]. Appearance-based methods deal with the whole face or specific face-regions to extract appearance changes of face using image filters such as Gabor-wavelet and local binary pattern (LBP) [7, 8]. Some research reveals that geometric feature-based approaches can provide similar or better performance than appearance-based methods [9]. However, the geometric

feature-based methods usually require accurate and reliable facial feature detection and tracking, which is difficult to accommodate in practical applications [8]. Hence, the appearance-based methods are the most widely accepted at present.

As one of another potential appearance-based feature extraction method, LBP has many applications in the field of face detection, face recognition and facial expression recognition [7, 10]. Although LBP has advantages of efficient computation and robustness to monotonic illumination changes, it is sensitive to non-monotonic illumination variation and also shows poor performance in the presence of random noise. In order to overcome this weakness of LBP, Tasked Jabid and others proposed LDP method [11] which was applied successfully in face recognition, object description, gender recognition and facial expression recognition [12-14].

LDP operator encodes images through calculating the edge response values in different directions for each pixel. For LDP descriptor itself, in the process of encoding, the corresponding value on each direction has contribution to many different LDP values (patterns). In certain extent, that scatters the role of a certain direction and weakens important directional information of expression features. Moreover, in the application of LDP method, it needs to divide image region into many blocks, and for the LDP histogram feature vector formed by every block; it is relative to the selection of block numbers. Thus eventually the dimension of feature vector formed of LDP is relatively high. Based on aforementioned issue, this paper proposes a method which is a variation of LDP method. It does not need to change into the corresponding decimal after binary code formation, but calculate histogram statistics for each direction directly. The proposed method not only emphasizes the different contribution of the main edge directions, but the size of feature vectors formed finally is not relative to the value of  $k$  (There need to set the top- $k$  directional bits to 1 when forming LDP code), which greatly reduce the computational efficiency. This is the first contribution of this paper. In addition, paper [9] introduced contrast information to LDP operator, and proposed a weighted LDP method (LDP variance) for facial expression recognition. It adjusts the different contributions of LDP coding using the variance  $\sigma$  of local structure. However, in application we found that the variance  $\sigma$  as weight is so large that it will affect the final recognition rates in our experiments unstably. So this paper will discuss and amend the weight, which is the second contribution. Moreover, paper [2] explored the effect on recognition rates of LBP method with overlapped block. And the result indicated that appropriately overlapping can improve the final recognition rate. Based on such idea, the third contribution of this paper is to research the effect on recognition rate of LDP with overlapped block. Our proposed method is verified on the JAFFE and Cohn-Kanade these two benchmark databases.

The structure of the paper is organized as follows. In section 2, we will introduce LDP and expression features based on LDP first. Then we will describe the proposed method and the facial expression recognition algorithm based on it in section 3. And the investigation of weightings and the overlapping block technique will be investigated in section 4. Then in section 5 we will do extensive experiments on two benchmark datasets and analyze the results, and the conclusions and future work are given in section 6.



## 2 LDP

### 2.1 Local Directional Pattern

The LDP operator calculates the edge response values in different directions and uses these to encode the image texture. Since the edge responses are less sensitive to illumination and noise than intensity values, the resultant LDP features describe the local primitives, including different types of curves, corners, and junctions, in a more stable manner and also retain more information [12].

The LDP descriptor is an eight bit binary code assigned to each pixel of an input image that can be calculated by comparing the relative edge response value of a pixel in different directions. So that eight directional edge response values  $\{m_i\}$ ,  $i = 0, 1, \dots, 7$  of a particular pixel are computed using Kirsch masks in eight different orientations  $M_i$  centered on its own position. These Kirsch masks shown in the Fig.1 and Fig.2 represent eight directional edge response positions and LDP binary bit positions. Because different importance of the response values, the  $k$  most prominent directions are chosen to generate the LDP. So the top  $k$  values  $|m_i|$  are set to 1, and the other positions are set to 0. Finally, the LDP code is derived by formula (1), where  $m_k$  is the  $k$ -th most significant directional response value. Fig.3 shows a standard LDP code with  $k=3$ .

$$LDP_k = \sum_{i=0}^{i=7} b_i(m_i - m_k) \cdot 2^i, \quad b_i(a) = \begin{cases} 1, & a \geq 0, \\ 0, & a < 0, \end{cases} \quad (1)$$

$$\begin{array}{cccc} \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} & \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} & \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} & \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \\ \text{East } M_0 & \text{North East } M_1 & \text{North } M_2 & \text{North West } M_3 \\ \\ \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} & \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix} \\ \text{West } M_4 & \text{South West } M_5 & \text{South } M_6 & \text{South East } M_7 \end{array}$$

Fig. 1. Kirsch edge response masks in eight directions

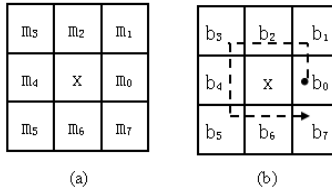
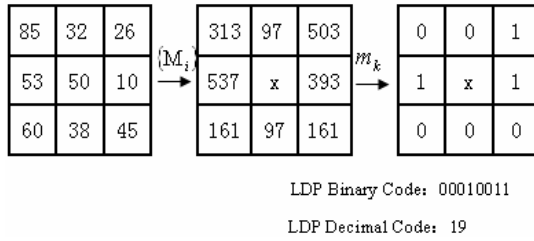


Fig. 2. (a) Eight directional edge response positions; (b) LDP binary bit positions



**Fig. 3.** LDP code with  $k = 3$

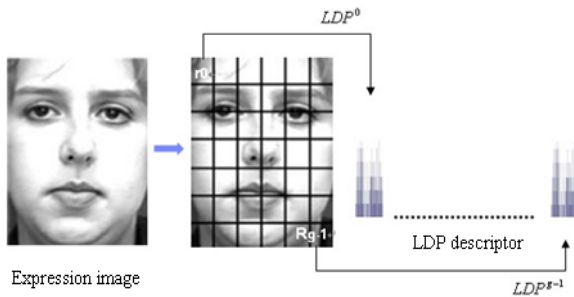
### 2.2 Expression Features Based on LDP

An input image of size  $M \times N$  can be represented by an LDP histogram  $H$  using (2) after computing all the LDP code for each pixel  $(r, c)$ , where  $i$  is the LDP code value.

$$H(i) = \sum_{r=1}^M \sum_{c=1}^N f(LDP_k(r, c), i), \quad f(a, i) = \begin{cases} 1 & a = i, \\ 0 & a \neq i, \end{cases} \quad (2)$$

For a particular value  $k$ , there are  $C_8^k$  different numbers of bins for the histogram  $H$ . In essence, a resulting histogram vector with size of  $1 \times C_8^k$  is produced for the image.

LDP descriptor contains detail information of an image, such as edges, spots, corner, and other local textures. Whereas computing LDP over the whole face image only considers the occurrences of micro-pattern without any information of their locations and spatial relationships which usually represent the image content better. Hence, the image is divided into  $g$  regions  $R_0, R_1, \dots, R_{g-1}$  as shown in Fig.4 when using LDP, so there will be a  $LDP^i$  histogram for every region  $R_i$ . Consequently, the resulting LDP descriptor is obtained via concatenating all these  $LDP^i$  histograms.



**Fig. 4.** Expression image is divided into small regions from which LDP histograms are extracted and concatenated into LDP descriptor

### 3 Expression Recognition Based on Improved LDP

#### 3.1 Expression Features Based on Improved LDP

For LDP descriptor itself, in the process of encoding, the corresponding one value of each direction has contribution to  $C_7^{k-1}$  different LDP values. In certain extent, that scatters the role of a certain direction, thus weakening important direction information of expression features. In addition, through the introduction of LDP previously, we know in the application of LDP method, it needs to divide image region into blocks. Assuming there are  $n_{block}$  blocks, and for the LDP histogram as feature vector formed by every block, it is relative to the selection of k value, namely  $C_8^k$ . So finally the dimension of LDP feature vector formed is  $n_{block} \times C_8^k$ , which is relatively high.

Therefore, this paper proposes a variation method based LDP, where we set the first larger k edge response values of eight neighborhood pixel  $(r,c)$  to 1, other to 0. That is to say, a vector  $LDP'_{(r,c)}$  of  $1 \times 8$  is formed through formula (3).

$$LDP'_{(r,c)}(i) = b_i(m_i - m_k) \quad b_i(a) = \begin{cases} 1, & a \geq 0, \\ 0, & a < 0, \end{cases} \quad (3)$$

where  $i = 0 \dots 7$  are eight different edge directions separately. Here it is worth noting, for different k values, the proposed method will form a vector  $LDP'_{(r,c)}$  with size of  $1 \times 8$ .

Same as LDP characteristics description, for an image of  $M \times N$ , direction histogram H can be formed by using formula (4). That is to sum vector  $LDP'_{(r,c)}$  formed for each pixel  $(r,c)$ , which is equivalent to calculate statistically the number of 1 on every direction.

$$H = \sum_{r=1}^M \sum_{c=1}^N LDP'_{(r,c)} \quad (4)$$

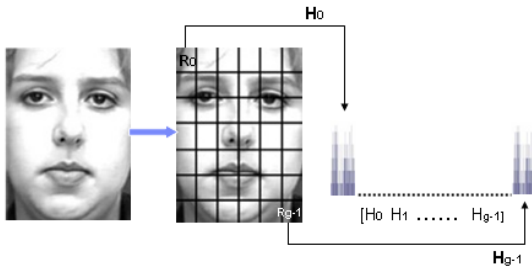


Fig. 5. Feature formation of improved LDP method

### 3.2 Feature Extraction and Recognition Based on the Improved LDP

In order to extract the local characteristics of image, we also divide image into blocks and every block has the same treatment.  $H_p$  is formed using formula (4) for every block, , where  $p = 0 \dots g - 1$ . Then concatenate all these direction histograms and form the final direction histogram  $H = [H_0, H_1 \dots H_{g-1}]$ , as shown in fig.5, as the final expression features.

Despite the dimension of feature vector formed by proposed method eventually is relatively low, but it still can't directly regard as the input of classifier. We still need further reduce its dimension and optimize features. This paper adopts PCA + LDP subspace methods for feature dimension reduction of features. And in classification stage we use the nearest neighbor classifier for expression classification.

## 4 Weighting and Overlapping Techniques

### 4.1 Weighting

The LDP operator encodes images through calculating edge response values for each pixel in different directions. When generating LDP code, a local structure of  $3 \times 3$  with low contrast and a local structure with high contrast have same contributions. However, compared with relative smooth region, these local regions containing edges such as eyes and mouth have more obvious contrast. Therefore, paper [9] introduced the contrast information in LDP operator, and proposed a weighted LDP method for facial expression recognition. It adjusts the different contributions of LDP code in LDP histogram using the variance  $\sigma$  of local structure. Formula (5), (6) and (7) are the calculation process used in [9].

$$LDP_v(\tau) = \sum_{r=1}^M \sum_{c=1}^N w(LDP_k(r, c), \tau) \quad (5)$$

$$w(LDP_k(r, c), \tau) = \begin{cases} \sigma(LDP_k(r, c)) & LDP_k(r, c) = \tau \\ 0 & otherwise \end{cases} \quad (6)$$

$$\sigma(LDP_k(r, c)) = \frac{1}{8} \sum_{i=0}^7 (m_i - \bar{m})^2 \quad (7)$$

where  $\bar{m}$  is the average of eight edge response value of a  $3 \times 3$  local structure centered with pixel  $(r, c)$ .

Inspired by this, we add weighting in our proposed method when conducting histogram statistics, namely weighting based on formula (4) and getting formula (8)

$$H = \sum_{r=1}^M \sum_{c=1}^N w(r,c) \cdot LDP'_{(r,c)} \quad (8)$$

where  $w(r,c)$  is the variance of local structure centered with pixel  $(r,c)$ .

However, in application we found that when using variance as weighting the value obtained is too large. For example, taking  $3 \times 3$  region in fig.3, the variance of local structure centered with pixel  $x$  is  $\sigma \approx 28110$ . So we try to modify weight  $\sigma$  as standard deviation  $\sqrt{\sigma}$  and the fourth root of variance  $\sqrt[4]{\sigma}$ . The experimental results suggest that it will obtain the highest recognition rate when the weighting is the standard deviation. The detailed experimental results are reported in section 5.3.

## 4.2 Overlapping Block

Whether standard LDP method or the proposed LDP method, both need to divide image into appropriate blocks in application. Because facial expression features are always reflected distinctly on local region (such as eyes and mouth area). Therefore appropriately dividing can guarantee the local characteristics of expression well represented. However, the existing methods in the application of LDP didn't discuss the influence of overlapping for recognition algorithm. Paper [3] explored the influence of overlapping of LBP method for its recognition, and the results showed that the appropriate overlap would raise the ultimate recognition rate. Based on such observations, this paper studies the influence of different overlap degrees on LDP method and our proposed method.

## 5 Experiments and Results

We will evaluate the proposed method on two benchmark databases: the JAFFE databases [15] and the Cohn-Kanade databases [16]. The JAFFE database contains 213 gray images (256x256) of individual human subjects with a variety of facial expressions. In this database, 10 different Japanese females performed seven prototypical expressions: anger, disgust, fear, joy, sadness, surprise and neutral. We choose three samples per facial expression for each subject, and a total of 210 images among which every expression has 30 images. The Cohn-Kanade database includes video sequences of 97 subjects displaying distinct facial expression. We create a subset with 10 subjects for our experiments. All the subjects selected have six basic expressions: anger, disgust, fear, joy, sadness and surprise. Every expression is a sequence. From six sequences of expressions for a selected subject, we select the last four frames as six basic expressions and the first frame as neutral expression four of which are used. So there are 280 total images (640x490) in all. After choosing the images, they were cropped from the original one using the position of two eyes according to the distribution proportion of facial organs and resized into  $150 \times 110$  pixels as [14]. The resulting expression images are shown in Fig.6. For following experiments, every group is repeated 10 times for each method, and the average recognition rate is calculated.



**Fig. 6.** Samples preprocessed (a)JAFFE database; (b)Cohn-Kanade database

## 5.1 Comparison of LBP, LDP and Proposed Method

To verify the effectiveness of improved method, we compare with LBP and LDP methods. For JAFFE database there are 30 images per expression, and we select randomly 20 images per expression for training, and the rest 10 images are used for testing respectively. Experimental results are listed in table 1 and table 2. At same time, table 3 lists the highest recognition rate of these three methods and their corresponding parameters (when  $k=3$  LDP method have highest recognition rate in paper [14]).

Contrast to the results listed in table 3, it can be seen that, our proposed method not only obtain the highest recognition rate but also its characteristic dimension is relatively low. Here it is interesting to note that the optimal numbers of blocks for these three methods when getting the best performance are different. From the results listed in table 1 and table 2 we can see that, the number of blocks has certain effect on recognition rate, and the recognition rate of all three method are improved when the number of blocks increases but with a certain disjunctive point. Although the increase of the number of blocks means there incorporate more local and spatial relationship information, some blocks may incorporate unnecessary or counteractive information that will degrade the performance when there are too many small blocks. Meanwhile, the optimal numbers of blocks are various for different methods which have different description of image information. Even for images from different databases they express different local information in the same size of local structure.

## 5.2 Extensional Experiment with Independence of Training Objects Completely

In face expression recognition, the training and testing samples distributions have significant effect on recognition rate. Therefore we have expanded research for our method with different databases and sample distributions. We mainly divide into three groups: in Cohn-Kanade expression database, randomly select 20 images from each expression for training, and the remaining 20 images for testing; in Cohn-Kanade expressions database adopts complete independent of person as 10 fold cross-validation, namely using all expression images of nine person for training, the remaining expression images for testing; all expressions images of Cohn-Kanade database are used for training sample, and all expressions images of JAFFE database are used as testing sample. Experimental results are reported in table 4.

From the results it can be seen that, different samples distribution has significant effect on recognition rate. Contrasting to the results of the first group and second group, when all expressions of all persons in test sample are not appeared in test sample, the recognition rate decreased obviously. Especially, when we adopt cross database for validation, recognition rate is quite low, and even less than 50%. The similar

results are reported in [8], which used the Boosted-LBP characteristics as expressions feature with classifier of SVM electing radial basis function as the kernel function, and its recognition rate is only 41.3%. That's because the characteristics of various expressions are still not completely the same in different database. In this circumstance, although the recognition rate of our proposed method is not lower than LBP and LDP, the performance is still far from satisfactory. And this needs farther research and improvement.

**Table 1.** The performance of LBP and LDP

Blocks	LBP		LDP(k=3)	
	Recognition rate	Vector length of feature	Recognition rate	Vector length of feature
g = 5x5	77.28%	256x25	77.57%	56x25
g = 7x6	82.57%	256x42	82.86%	56x42
g = 9x8	82.29%	256x72	85.43%	56x72
g = 12x9	77.29%	256x108	84.86%	56x108

**Table 2.** The performance of improved LDP method (k=3)

Blocks	Vector length of feature	Recognition rate
g = 7x6	8x42	78.57%
g = 9x8	8x72	83.43%
g = 12x9	8x108	85.86%
g = 16x13	8x208	87.57%
g = 21x18	8x378	88.14%
g = 29x21	8 x588	87.85%

**Table 3.** Comparison of LBP, LDP and the proposed method

Method	Recognition rate	Length of feature vectors	Blocks
LBP	82.57%	6400	7X6
LDP	85.43%	4032	9X8
Proposed method	88.14%	3024	21x18

**Table 4.** Extensional experiment with different databases

Sample distribution	Recognition rate		
	LBP	LDP	Proposed method
Cohn-Kanade: Training : Testing =1:1	93.21%	95.79%	95.79%
Cohn-Kanade: 10-fold cross validation	72.86%	76.28%	77.14%
Training:Cohn-Kanade;Testing:JAFFE	38.57%	41.90%	43.38%

### 5.3 Performance with the Weighting and Overlapping Techniques

In this section we do some experiments based on the weighting and overlapping techniques, and the results are listed in table 5 and table 6.

First, for our proposed method, we can see from table 5 that the recognition rate is reduced when adding variance as the weighting while the best performance can be obtained with the standard variance as weighting. Next, for LDP reported in [9] where the variance is selected as weighting, this method doesn't improve the recognition rate in our experiments using the JAFFE database, and also the best performance is achieved with the standard variance as weighting. This may be due to a fact that the C-K database was used in experiments of paper [9], and in contrast, the images in JAFFE database have stronger contrast that is reflected on some important local structure such as edges of organs. So, in JAFFE database, it is too large to use the variance of local structure as weighting.

Table 6 shows the comparison of LDP method and proposed method with or without weighting and overlapping block. From the results we can conclude the following observations. First, only slight improvement of recognition rate appears for these two methods with or not with weighting. In addition, when the weighting is chosen as the standard variance and regional overlap is 1/2, it will obtain the best recognition rate. Finally, in the best case, the recognition rate of our method is still slightly higher than standard LDP method.

**Table 5.** Comparison of LDP and proposed method with different weight

Weight	Proposed method	LDP
No weight	88.14%	85.43%
Variance	84.43%	84.71%
Standard deviation	90.71%	90.42%
Forth root of Variance	89.57%	88.29%

**Table 6.** Comparison of LDP and improved LDP with overlapping block

	LDP		Proposed method	
	No weight	Standard deviation	No weight	Standard deviation
Overlap=0	85.43%	90.42%	88.14%	90.71%
Overlap=1/2	86.57%	91.14%	89.71%	92.57%
Overlap=3/4	85.86%	90.29%	89.00%	91.14%

## 6 Conclusions

This paper presented a variation LDP method for facial expression recognition, by introducing the weighting to proposed method and investigated the effect on



recognition rates with different weightings, and then proposed the overlapped block technique when using LDP and proposed methods. Extensive experiments illustrate the following conclusions. First, the proposed method not only has low dimension of feature vector with simple calculation, but also achieve better recognition rates. Moreover, the appropriate weighting and regional overlapping have certain help to improve recognition rates of both proposed method and LDP method. Finally, in the case of independence of persons, there always has lower recognition rate, and this implies that we need to improve existing feature extraction methods continually in order to express substantive characteristics of expressions. In future, we will make further studies on the choice of the number of blocks, the size of weighting and the degree of overlap for general databases, especially the relationship and mutual effect between these parameters and different databases. Also, we will try to use some complex classifier with strong learning capability, such as support vector machine and neural networks.

**Acknowledgement.** This paper is supported by a Chunhui Planning project from China Education Committee with grant number Z2009-1-65001 and a Science and Technology Planning project in Yunnan Province with grant number2007F202M.

## References

1. Niu, Z., Qiu, X.: Facial expression recognition based on weighted principal component analysis and support vector machines. In: *Int. Conf. on Advanced Computer Theory and Engineering*. IEEE (2010)
2. Gritti, T., Shan, C., Jeanne, V., Braspenning, R.: Local Features based Facial Expression Recognition with Face Registration Errors 978-1-4244-1/08/2008. IEEE
3. Murthy, G.R.S., Jadon, R.S.: Effectiveness of Eigenspaces for Facial Expressions Recognition. *International Journal of Computer Theory and Engineering* 1(5) (December 2009)
4. Fasel, B., Luttin, J.: Automatic Facial Expression Analysis: a survey. *Pattern Recognition* 36(1), 259–275 (2003)
5. Zhang, Z., et al.: Comparison between Geometry-Based and Gabor-wavelet-based Facial Expression Recognition Using Multi-layer Perception. In: *Proc. IEEE Int. Conf. Auto. Face Gesture Recog.*, pp. 454–459 (April 1998)
6. Praseeda Lekshmi, V., Sasikumar, M.: Analysis of Facial Expression using Gabor and SVM. *International Journal of Recent Trends in Engineering* 1(2) (May 2009)
7. Xue, W.: Facial Expression Recognition Based on GaborFilter and SVM. *Chinese Journal of Electronics* 15(4A) (2006)
8. Shan, C., Gong, S., McOwan, P.W.: Facial expression recognition based on Local Binary Patterns:A comprehensive study. *Image and Vision Computing* 27, 803–816 (2009)
9. Kabir, M.H., Jabid, T., Chae, O.: A Local Directional Pattern Variance (LDPv) based Face Descriptor for Human Facial Expression Recognition. *IEEE* (2010), doi:10.1109/AVSS.2010.9978 -0-7695-4264-5/10 © 2010 IEEE
10. Ojala, T., Pietikainen, M.: Multiresolution Gray-Scale and Rotation with Local Binary Patterns and Linear Programming. *IEEE Trans.Pattern Anal. Mach.Intell.* 29(6), 915–928 (2007)

11. Jabid, T., Kabir, M.H., Chae, O.S.: Local Directional Pattern (LDP) for Face Recognition. In: IEEE Int. Conf. Consum. Electron. 2010, pp. 329–330 (2010)
12. Jabid, T., Kabir, M.H., Chae, O.S.: Local Directional Pattern(LDP)-A robust Descriptor for Object Recognition. In: IEEE Int. Conf. on AVSS (2010) 978-0-7695-4264-5/10
13. Jabid, T., Kabir, M.H., Chae, O.S.: Gender Classification using Directional Pattern(LDP). IEEE Pattern Recognition (2010)
14. Jabid, T., Kabir, M.H., Chae, O.S.: Robust Facial Expression Recognition Based on Local Directional Pattern. ETRI Journal 32(5) (October 2010)
15. Jaffe dataset, <http://www.kasrl.org/jaffe.html>
16. Kanade, T., Cohn, J., Tian, Y.: Comprehensive Database for Facial Expression Analysis. In: Proc. IEEE Int'l. Conf. Face and Gesture Recognition (AFGR 2000), pp. 46–53 (2000)

# Application of Fourier Transform to Get an Original Satellite Image without Applying the SRM or 180 Degree Rotation

Subhasis Kesh and Srishty Chakravarty

B.Tech Geo Informatics Engineering (2009-2013),  
University of Petroleum and Energy Studies Dehradun- 248007,  
Uttarakhand, India  
subhasiskesh@aol.com, srishty7@gmail.com

**Abstract.** It is well established that Optical remote sensing data from Sun-synchronous satellites of a rugged terrain always suffer from topographic effects. The result of this, people frequently perceive valleys as ridges and vice versa and hence sometimes incorrect image interpretations. The appearance of inverse topography has been termed as False Topographic Perception Phenomena (FTPP) by Dr A.K Saraf et al. (1996).

For getting the correct image, the process of SRM, 180° rotation or image negative is applied. But, it will be easy to us for studying the satellite image, if we can get the corrected image directly from the satellite. For this we have to apply the Fourier Transform in the satellite image system. It is observed that it would be much to us advantage if we could rotate an image in Fourier space instead of having to rotate the image in real space.

At first, take the real part or logarithm of the Fourier transform of the image function for getting the magnitude of the image and then take the imaginary part of Fourier transform of the image function for getting the phase of the image. After getting the phase of the image, take the inverse function of the same and then only we will get the original satellite image. It will be possible to get everything all together when we will compile these applications together in a computer programming by the representation of software. Here, we are only going to give a brief idea about the application of Fourier transform to get the original Image from satellite.

**Keywords:** Optical remote sensing data, Image interpretation, Inverse topography, FTTP, SRM, Fourier Transform, Fourier space, Real space, Inverse function, Magnitude, Phase.

## 1 Introduction

From the very first day of satellite engineering, we got the optical remote sensing data from sun-synchronous satellite. But, after long days it was observed that all the data in the form of image is incorrect and not perfect for studying. And this phenomenon is

termed as FTTP by Dr A.K Saraf and et al. (1996). To remove this phenomenon he proposed for the SRM,  $180^\circ$  rotation of image and negative of image.

But, these methods are only applicable after getting the image from satellite. These methods can't remove the FTTP at the time of satellite operation. But, right now we have to come out such new technical feasibility which will be helpful to save our important time and getting the corrected image directly from the satellite operation. This idea will be applicable when only it will be possible to compile a program along with the satellite operation programing. For this programing we have to apply the Simple Fourier Transform, Inverse Fourier Transform and Discrete Fourier Transform. At first, we would like to give an introduction about the above mentioned three types of Fourier Transform in the form of mathematical expression.

### 1.1 What Is Simple Fourier Transform?

Let,  $f(t)$  is a function of  $t$  defined for all positive values of  $t$ . Then, the Fourier Transform of  $f(t)$  is defined by,

$$F(s) = \int_{-\infty}^{\infty} f(t)e^{ist} dt \quad (1.1)$$

### 1.2 What Is Discrete Fourier Transform?

In mathematics, the **Discrete Fourier transform (DFT)** is a specific kind of discrete transform, used in Fourier analysis. It transforms one function into another, which is called the frequency domain representation, or simply the DFT, of the original function (which is often a function in the time domain). But the DFT requires an input function that is discrete and whose non-zero values have a limited (finite) duration. Such inputs are often created by sampling a continuous function, like a person's voice. Unlike the discrete-time Fourier transform (DTFT), it only evaluates enough frequency components to reconstruct the finite segment that was analysed.

The sequence of  $N$  complex numbers  $x_0 \dots x_{N-1}$  is transformed into the sequence of  $N$  complex numbers  $X_0, \dots, X_{N-1}$  by the DFT according to the formula:

$$X_k = \sum_{n=0}^{N-1} x_n e^{-\frac{2\pi i}{N}kn} \quad k = 0, \dots, N - 1 \quad (1.2)$$

Where  $i$  is the imaginary unit and  $e^{-\frac{2\pi i}{N}}$  is a primitive  $N$ 'th root of unity. (This expression can also be written in terms of a DFT matrix; when scaled appropriately it becomes a unitary matrix and the  $X_k$  can thus be viewed as coefficients of  $x$  in an orthonormal basis.)

The transform is sometimes denoted by the symbol  $F$ , as in  $X = F\{x\}$  or  $F(x)$  or  $F_x$ . The **Inverse Discrete Fourier Transform (IDFT)** is given by

$$x_n = \frac{1}{N} \sum_{k=0}^{N-1} X_k e^{\frac{2\pi i}{N}kn} \quad n = 0, \dots, N - 1. \quad (1.3)$$

A simple description of these equations is that the complex numbers  $X_k$  represent the amplitude and phase of the different sinusoidal components of the input "signal"  $x_n$ . The DFT computes the  $X_k$  from the  $x_n$ , while the IDFT shows how to compute the  $x_n$  as a sum of sinusoidal components  $(1/N)X_k e^{\frac{2\pi i}{N}kn}$  with frequency  $k/N$  cycles per sample. By writing the equations in this form, we are making extensive use of Euler's formula to express sinusoids in terms of complex exponentials, which are much easier to manipulate. In the same way, by writing  $X_k$  in polar form, we obtain the sinusoid amplitude  $A_k/N$  and phase  $\varphi_k$  from the complex modulus and argument of  $X_k$ , respectively:

$$A_k = |X_k| = \sqrt{\text{Re}(X_k)^2 + \text{Im}(X_k)^2}, \quad (1.4)$$

$$\varphi_k = \arg(X_k) = \text{atan2}(\text{Im}(X_k), \text{Re}(X_k)), \quad (1.5)$$

Where **atan2** is the two-argument form of the arctan function. Note that the normalization factor multiplying the DFT and IDFT (here 1 and  $1/N$ ) and the signs of the exponents are merely conventions, and differ in some treatments. The only requirements of these conventions are that the DFT and IDFT have opposite-sign exponents and that the product of their normalization factors be  $1/N$ . A normalization of  $\frac{1}{\sqrt{N}}$  for both the DFT and IDFT makes the transforms unitary, which has some theoretical advantages, but it is often more practical in numerical computation to perform the scaling all at once as above (and a unit scaling can be convenient in other ways).

### 1.3 Different Form of Discrete Fourier Transform

There is a different form of Discrete Fourier Transform. That is,

$$I(r, c) = \sum_{v=0}^{R-1} \sum_{u=0}^{C-1} Q(v, u) e^{-i2\pi \left( \frac{rv}{R} + \frac{cu}{C} \right)} \quad (1.6)$$

Where,

$$Q(v, u) = \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} I(r, c) e^{-i2\pi \left( \frac{rv}{R} + \frac{cu}{C} \right)} \quad (1.7)$$

## 2 Methodology

To remove the FTTP from the Sun-Synchronous satellite Image we have to follow the following steps. But, those steps should be compiled within a computer programming or within software. After the methods description we are describing the program code for the following steps. These program codes should be compiled within a computer program or software.

## 2.1 Steps

### 2.1.1. First, We Are Going to Do the 2D Fourier Transform of a Satellite Image

Let  $I(r,c)$  be a single-band (intensity) Satellite Image with  $R$  rows and  $C$  columns. Then,  $I(r,c)$  has Fourier representation,

$$I(r,c) = \sum_{v=0}^{R-1} \sum_{u=0}^{C-1} Q(v,u) e^{+i2\pi\left(\frac{vr}{R} + \frac{uc}{C}\right)} \quad (2.1)$$

Where,

$$Q(v,u) = \sum_{r=0}^{R-1} \sum_{c=0}^{C-1} I(r,c) e^{-i2\pi\left(\frac{rv}{R} + \frac{cu}{C}\right)} \quad (2.2)$$

In the equation (2.1) and (2.2), the complex exponentials are 2D sinusoids.

To simplify the situation assume  $R=C=N$ . Then,

$$e^{\pm i2\pi\left(\frac{vr}{R} + \frac{uc}{C}\right)} = e^{\pm i\frac{2\pi}{N}(vr+uc)} = e^{\pm i\frac{2\pi\omega}{N}(r\sin\theta + c\cos\theta)} \quad (2.3)$$

$$v = \omega\sin\theta, u = \omega\cos\theta, \omega = \sqrt{v^2 + u^2}, \text{ and } \theta = \tan^{-1}\left(\frac{v}{u}\right)$$

Where,

$$\text{Put, } \lambda = \frac{N}{\omega}$$

Then by Euler's relation,

$$e^{\pm i2\pi\frac{1}{\lambda}(r\sin\theta + c\cos\theta)} = \cos\left[\frac{2\pi}{\lambda}(r\sin\theta + c\cos\theta)\right] \\ \pm i\sin\left[\frac{2\pi}{\lambda}(r\sin\theta + c\cos\theta)\right] \quad (2.4)$$

It is notified that since images are indexed by row & column with  $r$  down and  $c$  to the right,  $\theta$ , is positive in the counter clockwise direction.

Both the real part of the equation (2.4),

$$\text{Re}\{e^{\pm i2\pi\frac{1}{\lambda}(r\sin\theta + c\cos\theta)}\} = +\cos\left[\frac{2\pi}{\lambda}(r\sin\theta + c\cos\theta)\right] \quad (2.5)$$

And the imaginary part of the equation (2.4)

$$\text{Im}\{e^{\pm i2\pi\frac{1}{\lambda}(r\sin\theta + c\cos\theta)}\} = \pm\sin\left[\frac{2\pi}{\lambda}(r\sin\theta + c\cos\theta)\right] \quad (2.6)$$

Equations (2.5) and (2.6) are sinusoidal “gratings” of unit amplitude, period  $\lambda$  and direction  $\theta$ . Then  $\frac{2\pi\omega}{N}$  is the radian frequency, and  $\frac{\omega}{N}$  the frequency, of the wave front and  $\lambda = \frac{\omega}{N}$  is the wavelength in pixels in the wave front direction.

Then the 2D sinusoids equation is,

$$I(r, c) = \frac{A}{2} \left\{ \cos \left[ \frac{2\pi}{\lambda} (r \sin \theta + c \cos \theta) + \varphi \right] + 1 \right\} \quad (2.7)$$

### 2.1.2. Then, Get the Fourier Phase of the Image

To get the phase of the Image we are taking the imaginary part of 2D Fourier Transform of the Image. Then only we will get an image which will not understandable to us. So, to get an understandable image we will reconstruct the phase of the image. After reconstruct the image we will get an image which contains the FTTP. To remove the FTTP we will follow the Inverse and a special Fourier Transform of the Phase of the Image.

### 2.1.3. Get the Rotated Satellite Image or Actual Satellite Image

Looking for ways to speed up a particular process, We discovered that it would be much to us advantage if we could rotate an image in Fourier space instead of having to rotate the image in real space and then taking the Fourier transform again. I convinced myself that the operation of rotation commutes with the Fourier transform operation:

Let  $g(\mathbf{x}) (\mathfrak{R}^2 \rightarrow \mathfrak{R})$  represent the image  $f(\mathbf{x})$  rotated by an angle  $\Theta$ . This can be expressed as a simple change of coordinates  $g(\mathbf{x}) = f(\mathbf{R}^{-1}\mathbf{x})$  where  $\mathbf{R} = \mathbf{R}(\Theta)$  is the rotation matrix (or, in general, any orthogonal transformation). Let  $G(\mathbf{k})$  be the Fourier Transform of  $g(\mathbf{x})$ . Then,

$$G(\mathbf{k}) = \int g(\mathbf{x}) \exp[-2\pi i \langle \mathbf{k}, \mathbf{x} \rangle] d\mathbf{x}. \quad (2.8)$$

Let  $\mathbf{k}' = \mathbf{R}\mathbf{k}$  and substitute  $f(\mathbf{R}^{-1}\mathbf{x})$  for  $g(\mathbf{x})$ , then

$$G(\mathbf{k}') = \int f(\mathbf{R}^{-1}\mathbf{x}) \exp[-2\pi i \langle \mathbf{R}\mathbf{k}, \mathbf{x} \rangle] d\mathbf{x}. \quad (2.9)$$

Because  $\mathbf{R}$  (and  $\mathbf{R}^{-1} = \mathbf{R}^T$ ) is  $\mathbf{x}, \mathbf{k} = \langle \mathbf{R}^{-1}\mathbf{R}\mathbf{x}, \mathbf{R}^{-1}\mathbf{k} \rangle = \langle \mathbf{x}, \mathbf{R}^{-1}\mathbf{k} \rangle$ . Thus we have

$$G(\mathbf{R}\mathbf{k}) = \int f(\mathbf{R}^{-1}\mathbf{x}) \exp[-2\pi i \langle \mathbf{k}, \mathbf{R}^{-1}\mathbf{x} \rangle] d\mathbf{x} \quad (2.10)$$

Because the integral is over all  $\mathbf{x}$ , and thus over all  $\mathbf{R}^{-1}\mathbf{x}$  we can perform the change of variable  $\mathbf{R}^{-1}\mathbf{x} \rightarrow \mathbf{x}$  to get

$$G(\mathbf{R}\mathbf{k}) = \int f(\mathbf{x}) \exp[-2\pi i \langle \mathbf{k}, \mathbf{x} \rangle] d\mathbf{x} \quad (2.11)$$

This is just  $F(\mathbf{k})$ , the Fourier Transform of  $f(x)$ . Finally we have

$$G(R\mathbf{k}) = F(\mathbf{k}) \text{ or } G(\mathbf{k}) = F(R^{-1}\mathbf{k}).$$

Thus we see that a rotation in  $\mathbf{x}$ -space has the same effect as a rotation in  $\mathbf{k}$ -space, i.e. Rotation commutes with the Fourier Transform.

## 2.2 Programing Code for Discrete Fourier Transform

```

/*
    Direct fourier transform
*/
int DFT(int dir,int m,double *x1,double *y1)
{
    long i,k;
    double arg;
    double cosarg,sinarg;
    double *x2=NULL,*y2=NULL;

    x2 = malloc(m*sizeof(double));
    y2 = malloc(m*sizeof(double));
    if (x2 == NULL || y2 == NULL)
        return(FALSE);

    for (i=0;i<m;i++) {
        x2[i] = 0;
        y2[i] = 0;
        arg = - dir * 2.0 * 3.141592654 * (double)i /
(double)m;
        for (k=0;k<m;k++) {
            cosarg = cos(k * arg);
            sinarg = sin(k * arg);
            x2[i] += (x1[k] * cosarg - y1[k] * sinarg);
            y2[i] += (x1[k] * sinarg + y1[k] * cosarg);
        }
    }

    /* Copy the data back */
    if (dir == 1) {
        for (i=0;i<m;i++) {
            x1[i] = x2[i] / (double)m;
            y1[i] = y2[i] / (double)m;
        }
    } else {
        for (i=0;i<m;i++) {
            x1[i] = x2[i];

```



```
        y1[i] = y2[i];  
    }  
}  
  
free(x2);  
free(y2);  
return(TRUE);  
}
```

### 3 Results

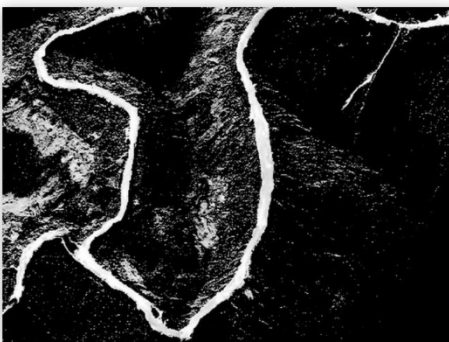
After applying the 2D Fourier Transform of the image we are taking the imaginary part of that to get the Phase of the Image. In the Image 1, it is representing the original satellite image and in the Image 2, it is showing the Phase of the Image. It is notified that both of the images are full of FTTP.



**Image 1.** Tsangpo River Basin Image Taken From Satellite



**Image 2.** Fourier Phase of the Satellite Image of Tsangpo River Basin



**Image 3.**



**Image 4.**

### **This Satellite Image Is of the TANSGPO River Basin in China**

To remove the FTTP, we applied the Inverse Fourier and the one another form of Transform. We represent the correction of FTTP by the Image 3 and Image 4. Image 3 is representing the rotated Fourier Phase of the Image after applying the Fourier Transform and Image 4 is representing the reconstruction of Inverted Phase Image. Take the Fourier transform, rotate it by  $180^\circ$ , and then take the inverse transform. This is an exact operation, so there will be no problems with interpolation. We expect that  $F^{-1}\{R(F\{\text{reference}\}, 180^\circ)\} = R(\text{reference}, 180^\circ)$  i.e. we expect that the final result will just be the original reference image rotated by  $180^\circ$ .

Here, first of all we have taken the Fourier Phase because, otherwise after applying the Inverse Function the image will get the low Phase and the low magnitude as well and it is observed by us.

We have only given the Idea to remove the FTTP in an another way except the SRM and  $180^\circ$  rotation, but, the above process should be compiled in a computer program, then only this process will be successful. In this paper we have only given the programing code for the 2D Fourier Transform of a Satellite Image but not for the inverse transform. For the inverse transform program code will be mostly same as the above.

## **4 Conclusion**

We can conclude that this proposal can be helpful for the Satellite Image Processing Department of all over the World for removing the FTTP from the Satellite Image. But, we want to notified one thing that the above proposal should be compile within a software which will be installed along with the Satellite Process system. This proposal is easier than the Dr A.K Saraf's given proposal. It will be more convenient for using in the Satellite Image processing.

## **References**

1. Holingsworth, M.: Application of Fourier Transform
2. Saha, S.K.: Discrete Fourier Transform, Fast Fourier Transform & their Application in Signal Processing
3. Implementation of Fast Fourier Transform for Image Processing in DirectX 10 by Intel
4. Frankovich, J.S.: Unsupervised Classification of Spectrally Enhanced Landsat TM Data of Midland, MI (December 1999)

# Evaluating Inpainting Methods to the Satellite Images Clouds and Shadows Removing\*

Ana Carolina Siravenha, Danilo Sousa, Aline Bispo, and Evaldo Pelaes

Signal Processing Laboratory, Federal University of Para (UFPA), Belem, PA, Brazil  
{siravenha,danilofrazao,pelaes}@ufpa.br, aline.bispo@itec.ufpa.br

**Abstract.** This paper presents the evaluation of two approaches widely used in the inpainting literature, applied in the context of atmospheric noise removal, such as fog, dense and sparse clouds and shadows, which often occurs in remote sensing images. One approach uses the technique of nearest neighbor interpolation for the information dissemination by a DCT-based smoothing method, and the other is based on second-order partial differential equations methods that uses the heat diffusion and thin-plate spline methods, achieving their solutions by using the finite-difference method. Finally, the evaluation uses the Kappa coefficient and the PSNR index. The metrics indicate the effectiveness of the nearest neighbor interpolation strategy, which produces higher quality images, specially when comparing the results obtained by the use of differential equations modeled by thin-plate spline.

**Keywords:** Inpainting, Cloud removal, Nearest neighbor interpolation, Heat diffusion.

## 1 Introduction

Commonly, remote sensing images are susceptible to the undesirable presence of the atmospheric interferences, such as fogs, clouds and haze. These occurrences, that are largely present in equatorial and tropical warm-to-hot regions, changes the brightness values of pixels at different levels of saturation, which can corrupt the regions of the land surface covered by clouds. Experts worldwide carried out many studies in order to find an effective method to reduce or eliminate the impact of clouds.

The method used in treatment of the atmospheric interferences vary according to the clouds density. It can be dense or thin. The dense clouds removal is mainly related to the use of reference images or by the estimative of the covered areas. The first strategy can use a multi-temporal analysis [24], or can use an other image of the same scene, captured by a different sensor, as a SAR image [12], in order to make the interpolation of the affected area. This interpolation can

---

\* This work was supported by *Fundação de Amparo à Pesquisa do Estado do Pará*, grant no. 021/2008, *Desenvolvimento de um sistema de interpretação de imagens de satélite baseado em modelos híbridos*.

use similar techniques to those described in [6,16,17]. These works developed derivations of the *inpainting* method in order to fill in some damaged regions or to remove large objects.

In this work, two *inpainting* based approaches are evaluated. We are interested in the removal of the dense clouds and shadows present in satellite images. The presence of dense clouds are related to the presence of the shadows, formed principally by the sun position in the moment of scene capture and the capture in the off-nadir angle. The first approach is based in Partial Differential Equations, that are modeled by two methods: the heat diffusion and thin-plate spline [4,3,18], whilst the other approach uses nearest neighborhood interpolation (NNI) and Multidimensional DCT [8,11,2].

In the sequence, Section 2, are described the techniques used in the approaches. The results achieved and its discussions are presented in Section 3. Lastly, we present the work conclusions (Section 4).

## 2 Methodology

In this section will be described the methodology adopted in this work. Firstly, a region detection method is presented, proceeding with the proposed approaches.

### 2.1 The Region Detection Method

Before applying the method of cloud removal, the image is submitted to a region detection method. This method is based on statistical measures, similar to the method used in [11]. In that reference, was not detected shadows regions, feature added during this work. Therefore,

$$l(x, y) = \left\{ \begin{array}{ll} i(x, y) < i_{m-dp}, & i(x, y) \in 0; \\ i_{m-dp} < i(x, y) < i_m, & i(x, y) \in 1; \\ i_m < i(x, y) < i_{m+dp}, & i(x, y) \in 2; \\ i(x, y) > i_{m+dp}, & i(x, y) \in 3. \end{array} \right. \quad (1)$$

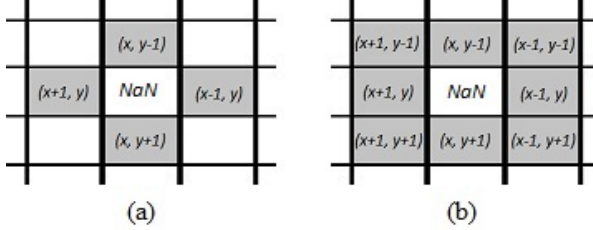
where  $i(x, y)$  is an original image pixel,  $i_m$  and  $i_{dp}$  represents the mean and standard deviation values of the image pixels, respectively. The mapped matrix  $l(x, y)$  have the same dimension of the original image.

The matrix  $l(x, y)$  have four values (0, 1, 2 and 3), where the regions mapped as 0 corresponds to the shadows areas; the pixels mapped as 1 represents regions without atmospheric damages detected; the regions damaged with clouds are divided in regions with thin clouds (2) and dense clouds (3). In this work, the regions of shadows and dense clouds will be interpolated by the inpainting methods.

### 2.2 First Approach

The interpolation process depends on the neighborhood of the pixels labeled as 0 and 3. These pixels, redefined to  $p = NaN$ , have four neighbors, being two

neighbors in the horizontal direction and two neighbors in the vertical direction, identified as  $(x + 1, y)$ ,  $(x - 1, y)$ ,  $(x, y + 1)$  and  $(x, y - 1)$ , respectively. Besides these, there are other four diagonal neighborhoods  $(x - 1, y - 1)$ ,  $(x - 1, y + 1)$ ,  $(x + 1, y - 1)$  and  $(x + 1, y + 1)$ . These neighborhood are illustrated by the Fig. 1(a) and Fig. 1(b).



**Fig. 1.** *NaN* pixels neighborhood. (a) 4-neighborhood (b) 8-neighborhood [9].

**A.1 Method.** In this method, the regions with dense clouds and shadows are reconstructed based on a formulation of a Partial Differential Equation (PDE). The PDE is modeled as a heat diffusion, where its approximation is given by a finite differences. The system of linear equations can be solved by applying a two-dimensional Laplacian elliptic equation [21]. According [9], the most simple isotropic derivative operator is the Laplacian, expressed by the Eq. 2, where the function  $f(x, y)$  describes a two-dimensional matrix.

$$\nabla^2 f = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} = 0. \quad (2)$$

To an arbitrary number of undefined pixels  $n$ , there are  $n$  system of equations. The Laplacian operator is linear to any derivative order, and in discrete form, the equation, on the  $x$  direction, is defined as:

$$\frac{\nabla^2 f}{\nabla x^2} = f(x + 1, y) + f(x - 1, y) - 2f(x, y), \quad (3)$$

and, on the  $y$  direction, as:

$$\frac{\nabla^2 f}{\nabla y^2} = f(x, y + 1) + f(x, y - 1) - 2f(x, y). \quad (4)$$

Replacing the Eq. 3 and 4 on the Eq. 2 is obtained the discrete Laplace equation:

$$\begin{aligned} \nabla^2 f(x, y) &= f(x + 1, y) + f(x - 1, y) \\ &+ f(x, y + 1) + f(x, y - 1) - 4f(x, y). \end{aligned} \quad (5)$$

Considering the coefficients of the Eq. 5 in its spatial form, as showed in Fig. 2, the central pixel (the pixel to be interpolated) will be redefined by the average of the its neighbors weights [7].

0	1	0
1	-4	1
0	1	0

**Fig. 2.** Coefficients in spatial domain, where the central pixel must be interpolated [9]

The linear equations systems are formulated by the undefined pixel neighborhood, according to the illustrated in Fig. 1(a). The system solution gives the values attributed to the undefined pixel.

This method propagates the smooth information of the edge to the regions identified as containing shadow or dense cloud. To small regions, this method is more efficient, supposing that the high frequency elements (typical of the edges) does not totally damaged by the atmospheric noise.

**A.2 Method.** The second method follows the same lines of the method A.1, but in this case, the PDE uses the neighborhood illustrated in Fig. 3.

		$(x, y-2)$		
	$(x+1, y-1)$	$(x, y-1)$	$(x-1, y-1)$	
$(x+2, y)$	$(x+1, y)$	<i>NaN</i>	$(x-1, y)$	$(x-2, y)$
	$(x+1, y+1)$	$(x, y+1)$	$(x-1, y+1)$	
		$(x, y+2)$		

**Fig. 3.** Neighborhood of the undefined pixel (*NaN*) in method A.2

According [7], the differential equation model, as described in this work, is more appropriated to a thin-plate spline model than the heat diffusion model. Spline interpolation technique estimates the unknown data among known data. Thin-plate spline PDE is a scalar function expressed as a linear combination of shifted samples of the kernel function added to the polynomial term, and can be described as:

$$f(x) = \sum_{i=1}^N c_i \Phi(|x - x_i|) + p_{m-1}(x), \quad (6)$$

where  $N$  is a number of the pixels in an image,  $\Phi$  is the radial kernel function  $r^2 \log(r)$  associated to the PDE, where  $r = |x - x_i|$  [5]. The coefficients  $c_i (i = 1, \dots, N)$  are restricted to the conditions [5]:

$$\sum_{i=1}^N c_i q_{m-1}(x_i) = 0 \quad \forall q_{m-1} \in \prod_{m-1} \quad (7)$$

The two-dimensional function  $\Phi$  is defined as:

$$\Phi(r) = \left[ \begin{array}{c} \frac{1}{\alpha} \partial_{xx} + \frac{1}{\beta} \partial_{yy} \quad (\frac{1}{\alpha} - \frac{1}{\beta}) \partial_{xy} \\ (\frac{1}{\alpha} - \frac{1}{\beta}) \partial_{xy} \quad \frac{1}{\alpha} \partial_{yy} + \frac{1}{\beta} \partial_{xx} \end{array} \right] K(r), \quad (8)$$

where  $\alpha$  and  $\beta$  are the model parameters that controls the divergence values to the obstructed regions reconstruction.  $K(r) = r^4 \log(r)$  is the three-harmonic scalar kernel function.

### 2.3 Second Approach

In this approach, the known pixels information, in an undefined region, spreads its information from the surrounding areas throughout the contour direction [1]. It uses a nearest neighbor interpolation, where the defined pixel value is assigned to an undefined pixel [13]. The purpose is to minimize the influence of the noise on estimating the direction of the isophotes [1]. After the assignment step, is applied an interactive smoothing method based in a nearest zero parameter. The parameter values are changed according to the application. Is used Discrete Cosine Transform (DCT) to the inpainting process. In each iteration an interpolation is made, and this process repeats until a steady state is achieved, according to the criteria presented in [8].

The nearest neighbor interpolation takes a point  $x$  to be interpolated and a few other neighbor points  $x_k$ , which are a sample portion of an image  $f_k$ . It is calculated the Euclidean distance among the value to be estimated and its neighbors, according Eq. [9].

$$d_k = ||x - x_k|| \quad (9)$$

The  $x_k$  value that produces the lower  $d_k$  value has its intensity information attributed to  $x$  point. This measure repeats until to fulfill all undefined pixels [10].

In [8] is defined the smoothing parameter using a multidimensional DCT and [2] describes in detail the DCT broach as follows. Firstly, is calculated the smoothed image  $\hat{y}$  (related to the original one  $y$ ):

$$\hat{y} = DCT^{-1}(\Gamma^N \circ DCT(y)), \quad (10)$$

where  $DCT$  and  $DCT^{-1}$  refers to the  $N$  dimension  $DCT$  and its inverse, respectively, and  $\circ$  represents the *Schur* product (element-by-element).  $\Gamma^n$  is a  $N$  class tensor, defined by:

$$\Gamma^n = \frac{1^N}{(1^N + s \wedge^n \circ \wedge^n)}. \quad (11)$$

This tensor executes an element-by-element division and the value 1 is a  $N$  class tensor composed by 1's. The  $\wedge$  tensor is also a  $N$  class tensor, defined by:

$$\wedge_{i_1, \dots, i_N}^N = \sum_{j=1}^N \left( -2 + 2 \cos \frac{(i_j - 1)\pi}{n_j} \right), \quad (12)$$

where  $n_j$  is the length of  $\wedge^n$  in the  $j^{\text{th}}$  dimension [2]. The automatic definition of  $s$  requires the minimization of the Generalized Cross-Validation ( $GCV$ ) index [20,22]. This index is a weighted average of all  $n$  squared error points of a value prediction at a given point (by interpolation), and can be described as:

$$GCV = \frac{\|\hat{y} - y\|_F^N}{(1 - \|\Gamma^n\|_1)^2}, \quad (13)$$

where  $\|\cdot\|_F$  denotes the *Frobenius* norm,  $\|\cdot\|_1$  is the 1 norm, and  $n = \prod_{j=1}^N n_j$  is the number of the elements in  $y$ . Then, the  $GCV$  index, without weighted and without undefined values, can be described as:

$$GCV = \frac{n \|\Gamma^n - 1^N \circ DCT(y)\|_F^2}{(n - \|\Gamma^n\|_1)^2}. \quad (14)$$

This procedure can also be applied to other weighted or undefined data. The interactive process have a form:

$$\hat{y}_{k+1} = DCT_{-1}(\Gamma^N \circ DCT(W \circ (y - \hat{y}_k) + \hat{y}_k)), \quad (15)$$

where  $\hat{y}_k$  refers to the  $\hat{y}$ , calculated on  $j^{\text{th}}$  iteration. It may be appropriate enter with lower weights to the data, because high weights to high quality data should represent a bad choice.

The main diagonal of the matrix  $W$ ,  $diag(w_i)$ , contains the weights  $w_i \in [0, 1]$  that corresponds to the  $y_i$  data. To the undefined pixels  $w_i$  is equals to 0.

### 3 Results and Discussions

The evaluation methods has been done by its application on, at least, 9 satellite images from different sensors, with visible areas damaged with dense clouds and shadows. Following, the methods efficiency was numerically measured according to the Kappa coefficient, which is based of confusion matrix of the sample image [11]. To complete the evaluation was computed the Peak Signal-to-Noise (PSNR) index, which is based on the information similarity between the reference image and the interpolated image [19,14].

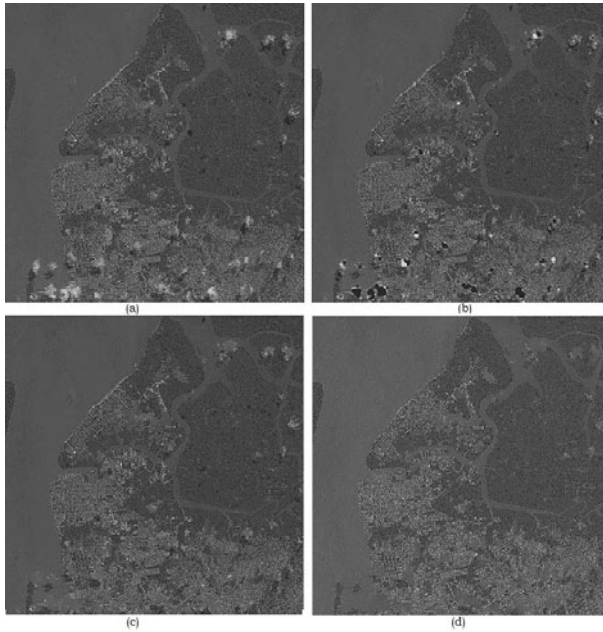
To produce the reference images or ground-truth images, the original image was manually manipulated in order to remove the noise, simulating an ideal image, allowing that the methods evaluation were applied correctly.

Firstly, was compared the A.1 method and A.2 method, the both methods that use PDE. The Table 1 shows that the A.1 method was superior than A.2 method, according the adopted criteria. The execution time of the method based



**Table 1.** Numerical evaluation of the methods. The PSNR index is given in *dB*.

Image	A.1 method		A.2 method		B method	
	Kappa	PSNR	Kappa	PSNR	Kappa	PSNR
1	0.7608	24.62	0.3917	17.29	0.7566	23.67
2	0.6941	25.02	0.5805	20.27	0.6935	24.98
3	0.4449	21.15	0.3493	19.09	0.4435	20.95
4	0.8329	30.25	0.3873	19.52	0.8232	29.41
5	0.8173	28.94	0.8105	26.38	0.8120	28.61
6	0.6442	25.68	0.6603	24.54	0.6414	25.56
7	0.7440	28.40	0.6901	24.87	0.7375	28.30
8	0.7572	22.36	0.6854	20.40	0.7531	22.53
9	0.7624	28.34	0.7183	26.39	0.7622	28.34

**Fig. 4.** (a)Image 2 of the Table 1 (b)Visual aspect produced by A.2 method (c)Visual aspect produced by A.1 method (d)Visual aspect produced by B method

in heat diffusion is considerably lower than thin-plate spline based method. As showed in Fig. 4(c)<sup>1</sup> and in Fig. 5(c)<sup>2</sup>, is evidenced the visual quality produced by the A.1 method over the A.2 method results. Whilst, the images produced by the A.2 method presented noisily results, as showed in Fig. 4(b) and Fig. 5(b).

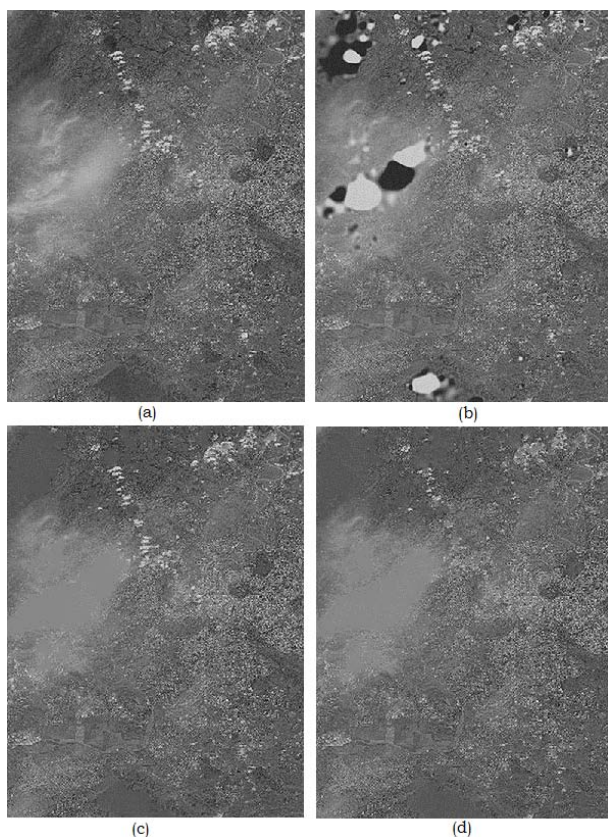
<sup>1</sup> Refers to the image 9 in Table 1.

<sup>2</sup> Refers to the image 2 in Table 1.

When combined, the numerical and visual factors, the  $A.2$  method shows as an unfeasible algorithm to atmospheric noise removal, presenting white and black spots, that represents integer values too much larger or smaller than their neighbors. It concludes that the Laplacian function of the image intensity becomes too much larger around the noisily pixels, furthermore, the function associated to the diffusion coefficient is set to decrease quickly to preserves its edges.

The function related to the finite differences equation solution has small values around the noisily pixels, therefore, the spots tends to damage the images, as well observed in [23].

Finally, comparing the  $A.1$  and  $B$  methods, observing the Table II, both methods have similar performance. The methods efficiency is also similar in the visual results, as showed in Fig. 4(c) and (d) and Fig. 5(c) and (d).



**Fig. 5.** (a)Image 9 of the Table II (b)Visual aspect produced by  $A.2$  method (c)Visual aspect produced by  $A.1$  method (d)Visual aspect produced by  $B$  method

As well the other inpainting methods, the approaches studied in this work do not fulfill larger damaged regions, these reconstructed areas tend to become blurred and without texture information [15,6]. This fact is explained by the only use of the edge pixels information in diffusion. An example, is the Fig. 5 which have a large area damaged with dense clouds, in contrast with the Fig. 4 that presents smaller undefined regions. It is possible to note that, in larger damage areas homogeneous, spots without texture information are created.

## 4 Conclusion

In this work was evaluated two different inpainting approaches applied in the clouds removal field. The first approach is based on differential partial equations modeled by heat diffusion and thin-plate spline. Smoothed by a method based in DCT, the second approach uses nearest neighbor interpolation to the information diffusion.

As showed in Table 1, the use of the thin-plate spline model did not achieved satisfactory results. Moreover, this method presented an unacceptable execution time. The *A.1* method, based on heat diffusion, had similar results achieved by the approach based in nearest neighbor interpolation. The visual aspect of these methods was also similar, where the *B* method was slight more efficient than *A.1* method, generally.

These methods were applied to other images, which followed the same interpretation. In some instances, the *A.1* method was superior to the *B* method, but most of the opportunities the *B* method presented better results in terms of PSNR and Kappa values, as well as more consistent visual appearance.

## References

1. Bertalmio, M., Sapiro, G., Caselles, V., Ballester, C.: Image inpainting. In: Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques, pp. 417–424 (2000)
2. Buckley, M.J.: Fast computation of a discretized thin-plate smoothing spline for image data. *Biometrika* 81, 247–258 (1994)
3. Caselles, V., Haro, G., Sapiro, G., Verdera, J.: On geometric variational models for inpainting surface holes. *Comput. Vis. Image Underst.* 111, 351–373 (2008)
4. Chan, T.F., Shen, J.: Morphologically invariant pde inpaintings (2001)
5. Chen, F., Suter, D.: Fast multipole method for accelerating the evaluation of splines. *IEEE Comput. Sci. Eng.* 3, 24–31 (1998)
6. Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image Inpainting. *IEEE Transactions on Image Processing* 13, 1200–1212 (2004)
7. D'Errico, J.: Re interpolating over nans newsgroup comp soft-sys matlab (2003)
8. Garcia, D.: Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational Statistics & Data Analysis* 54(4), 1167–1178 (2010)
9. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Publishing Company (2008)

10. Hale, D.: Image-guided blended neighbor interpolation of scattered data. In: 79th Annual International Meeting, Society of Exploration Geophysicists, vol. 28, pp. 1127–1131 (2009)
11. Hau, C.Y., Liu, C.H., Chou, T.Y., Yang, L.S.: The efficacy of semi-automatic classification result by using different cloud detection and diminution method. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2008)
12. Hoan, N.T., Tateishi, R.: Cloud removal of optical image using SAR data for ALOS applications. Experimenting on simulated ALOS data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* (2008)
13. Htwe, A.N.: Image interpolation framework using non-adaptive approach and nl means. *International Journal of Network and Mobile Technologies 1* (2010)
14. Kresimir, D., Mislav, G.: *Handbook Of Data Compression*. Springer, Heidelberg (2009)
15. Kwok, T., Sheung, H., Wang, C.: Fast query for exemplar-based image completion. *IP 19*, 3106–3115 (2010)
16. Liu, H., Wang, W., Bi, X.: Study of image inpainting based on learning. In: *Proceedings of The International MultiConference of Engineers and Computer Scientists*, pp. 1442–1445 (2010)
17. Maalouf, A., Carre, P., Augereau, B., Fernandez Maloigne, C.: A bandelet-based Inpainting technique for clouds removal from remotely sensed images. *IEEE Transactions on Geoscience and Remote Sensing 47(7)*, 2363–2371 (2009)
18. Paragios, N., Chen, Y., Faugeras, O.: *Mathematical models in computer vision: the handbook*. Springer, Heidelberg (2005)
19. Salomon, D., Motta, G.: *Handbook Of Data Compression*. Springer, Heidelberg (2009)
20. Wahba, G.: *Spline models for observational data*. SIAM (1990)
21. Wang, Z., Zhou, F., Qi, F.: Inpainting thick image regions using isophote propagation. In: *Proceedings of International Conference on Image Processing - ICIP*, pp. 689–692 (2006)
22. Whatmough, R.: Applying generalised cross-validation to image restoration. In: *1994 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1994*, vol. 5, pp. V/453 –V/456 (1994)
23. You, Y.L., Kaveh, M.: Fourth-order partial differential equations for noise removal. *IEEE Transactions on Image Processing 9(10)*, 1723–1730 (2000)
24. Zhang, X., Qin, F., Qin, Y.: Study on the thick cloud removal method based on multi-temporal remote sensing images. In: *International Conference on Multimedia Technology (ICMT)*, pp. 1–3 (2010)

# Comparing Different High-Pass Filters to Improve the Accuracy of Classification of Satellite Imagery Obstructed by Clouds and Fog

Danilo Sousa, Ana Carolina Siravenha, and Evaldo Pelaes

Signal Processing Laboratory,  
Federal University of Para (UFPA),  
Belem, PA, Brazil  
{danilofrazaos,siravenha,pelaes}@ufpa.br

**Abstract.** This work has as main objective to overcome a frequent problem in remote sensing, which is the undesirable presence of atmospheric constituents as scattered clouds, fogs and mists. The presence of such elements can affect the urban and environmental monitoring, as well as subsequent steps of the digital image processing such as segmentation and classification, main responsible for extracting information of the image. Therefore, is presented a technique to detect these elements, which uses statistical measures and morphological filters. To the removal or smoothing these atmospheric elements is applied a homomorphic filter. Motivating the present work, is presented a comparative analysis of the widely used high-pass filters in homomorphic filtering, Ideal and Butterworth, with the alternative High-boost filter. The results are evaluated by the Kappa coefficient and PSNR index, pointing to the High-boost filter as the best approach to use.

**Keywords:** Image enhancement, High-boost filtering, Homomorphic filtering, thin clouds removal

## 1 Introduction

The visual quality of multi-spectral satellite images can be affected according to the atmospheric conditions at the time of capture by sensors aboard remote platforms. The most common interference is the presence of clouds that changes the brightness values of the pixels at different levels of saturation, corrupting the land surface covered by clouds. Activities such as environmental or urban monitoring, and any other study for extraction of relevant information from remote sensory images are directly affected by these atmospheric components which can become unusable a set of damaged images. In order to find an effective method to reduce or eliminate the impact of clouds, experts worldwide carried out many studies, some specifically for thin clouds [10].

Some techniques have been used for the removal or, at least, the mitigation of these effects. An efficient method to do such task is the substitution method, which replaces image cloudy regions with corresponding regions of another image

of the same site obtained at other time. This method can eliminate cloud effect completely. But images to select non-cloud regions must have the same or approximate land surface characters; precise image registration and hue adjustment have to be done before substitution; obvious boundaries often exist in joint positions. Moreover, appropriate substitution images are difficult to obtain [6]. Nevertheless, when it comes to the removal of scattered clouds, the most used method is the homomorphic filtering [1,8,6].

In [3], homomorphic filtering is also applied, besides, an evaluation was made among five High-pass filters used in that filtering process, highlighting one that gives the best results, the Butterworth filter. Thus, this paper is encouraged to do a comparative analysis of the widely used high-pass filters in homomorphic filtering, Ideal and Butterworth, with the alternative High-boost filter. Similar to each other, the Ideal and the Butterworth filters emphasize the high frequencies completely suppressing the low frequencies. Alternatively, the High-boost filter does not entirely suppress the low frequencies, besides emphasizes the high frequencies. This feature provides less loss of information increasing the accuracy of the classifier [4,9]. At this work too, a more efficient way to detect regions, that was applied by [3], is proposed. It uses a constant  $cc$  for cloud detection and performs a morphological operation of opening on the image.

The paper is divided as follows. Section 2 describes the theory of all techniques used in this work as well as how the algorithm works; section 3 shows the results after the application of the algorithm and the comparison of the applied methods; and finally, section 4 shows the conclusions about the evaluation made in this work.

## 2 Methodology

In this section we present the basic techniques employed to apply the scattered cloud removal algorithm. Each technique is presented and its contribution to the proposed method are described.

### 2.1 Cloud Detection

The basis of the clouds detection is described in [3], where is made a separation of regions with different characteristics in the image, in order to improve results. This division is made considering statistical measures of the image, detecting scattered clouds, fogs and mists areas.

The difference on the lighting conditions, on the sensor characteristics and on the surface, make that the images have different ranges of gray levels representing a class. For example, under opposite lighting conditions a same region can be easily labeled as dense and non dense vegetation, or an urban region can be mislabeled as a cloud region. In order to improve the technique described in [3], making the regions separation more flexible, we added one constant to the detection algorithm, called  $cc$ , cloud constant. Such constant have default value equals to 1, and after the tests, was noted that the range from 0 to 3 returns

optimized results. Therefore, by the change of these values we can differentiate more precisely the classes of a scene.

The process of regions separation is expressed by

$$f(x, y) = \begin{cases} i(x, y) < i_m, & i(x, y) \in 1; \\ i_m < i(x, y) < (cc \times i_{m+dp}), & i(x, y) \in 2; \\ i(x, y) > (cc \times i_{m+dp}), & i(x, y) \in 3. \end{cases} \quad (1)$$

where  $i(x, y)$  is equal to the pixel value of the contaminated image by cloud,  $i_m$  represents the average value of the noisy image,  $i_{m+dp} = i_m + \sigma_i$ . Thus, regions defined as 1 are free from any kind of noise, while the regions defined as 2 are labeled as containing scattered clouds (that work is only interested in this region) and finally those defined as 3 are detected as dense clouds.

## 2.2 Image Opening

The morphological filters refer to the study of geometric structure of the entities present in an image. Being the filters of dilation and erosion basic morphological operations and based on set theory, this technique involves the interaction between an image  $A$  (the object of interest) and a structuring element  $B$ . In general, most of the morphological operations are based on simple operations of expansion and shrinkage [2].

Opening generally smooths the contours of an image, breaking narrow isthmus and eliminating thin protrusions, and is defined as follows: the opening of  $A$  by  $B$  is given by the erosion of  $A$  by  $B$ , followed by dilation by  $B$ , that is,  $A \circ B = (A \ominus B) \oplus B$ .

Therefore, after opening the image, small objects inside a larger tend to be extinct. Our interest in applying such a morphological transformation, relies on the fact of small objects contained in large blocks can be misclassified in a different region. Given this, the opening operation can eliminate this problem, generating a better detection of image pixels at a scene.

## 2.3 Homomorphic Filter

The images usually consist of light reflected from objects. The basic nature of an image can be characterized by two components: (1) the amount of light coming from the incident source on the scene and (2) the amount of reflected light by objects in the scene. These bands of light are called components of *illumination* and *reflection*, and are denoted by  $l(x, y)$  and  $r(x, y)$ , respectively. The functions  $l$  and  $r$  combine multiplicatively to produce the  $F$  image:

$$F(x, y) = l(x, y)r(x, y), \quad (2)$$

where  $0 < l(x, y) < \infty$  e  $0 < r(x, y) < 1$ .

The Fourier transform of the product of two functions is not separable [3] and it is desirable to manipulate the image in frequency domain, then we apply the

natural logarithm function that approximates the function  $F(x, y)$  to the form of a sum of the components  $l$  and  $r$

$$z(x, y) = \ln F(x, y) = \ln l(x, y) + \ln r(x, y). \tag{3}$$

In frequency domain, we can manipulate the image in terms of low and high components, or illumination and reflection, separately. Then, we apply a Fast Fourier Transform on  $z(x, y)$ , the natural logarithm of  $F(x, y)$

$$\mathcal{F}(z(x, y)) = \mathcal{F}(\ln F(x, y)) = \mathcal{F}(\ln l(x, y)) + \mathcal{F}(\ln r(x, y)). \tag{4}$$

Therefore, considering that  $Z$ ,  $L$  and  $R$  are Fourier Transforms of  $z$ ,  $\ln l$  and  $\ln r$ , respectively, then

$$Z(w, v) = L(w, v) + R(w, v). \tag{5}$$

The advantage of the use of high pass High-boost filter to the cloud removal applications is that it does not completely suppress the lower frequencies. Thus, the filter  $H(\cdot)$  is applied to the function  $Z$  and assumes

$$S(w, v) = H(w, v)Z(w, v) = H(w, v)L(w, v) + H(w, v)R(w, v) \tag{6}$$

We can take the inverse Fourier transform of Eq. [6](#)

$$s(x, y) = \mathcal{F}^{-1}(H(w, v)L(w, v)) + \mathcal{F}^{-1}(H(w, v)R(w, v)), \tag{7}$$

and finally, as  $z$  was obtained using the logarithm of the original image  $F$ , the reverse process produces the desired image  $\hat{F}$

$$\hat{F} = \exp s(x, y) = \exp (l'(x, y)) \exp (r'(x, y)) = l_0(x, y)r_0(x, y). \tag{8}$$

By applying a multiplicative factor of amplification ( $A$ ) before subtraction of the low frequencies (LPF), we obtain a filter HPF High-boost. Thus,

$$\begin{aligned} \text{High - boost} &= (A)(\text{original}) - \text{LPF} \\ \text{High - boost} &= (A - 1)(\text{Original}) + \text{HPF} \end{aligned} \tag{9}$$

where  $\text{HPF} = \text{Original} - \text{LPF}$ . If  $A = 1$ , we have a simple filter high-pass. When  $A > 1$ , a part of the original image is maintained in the output. A Butterworth HPF used in the formulation of this proposed High-boost filtering.

Homomorphic filtering using High-boost is applied only in the areas labeled as 2, i.e., areas where there is presence of scattered clouds, mists and fogs. The filtering in these regions is suitable, because the region of cloud is a low frequency region due to the homogeneity of its pixels.

Thus, the homomorphic filtering process can be summarized by the Fig. [11](#)



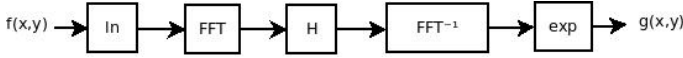


Fig. 1. Homomorphic Filter Schema [118]

## 2.4 Algorithm of the Utilized Method

The steps of the proposed algorithm are shown in sequence:

1. The regions of the image  $F$  are mapped as described in Subsection 2.1. The result of this step is a mapped image among the three categories described, called  $FM$ .
2. The  $FM$  mapped image passes through an opening operation, generating a new image  $FMO$ .
3. The the original image  $F$  has regions mapped as 2 in  $FMO$  corresponding image, passed through homomorphic filter (according Subsection 2.3), removing scattered clouds, and producing an image called  $FH$ , which is the resultant image.

The flowchart of the process described above is shown in the diagram of Fig. 2.

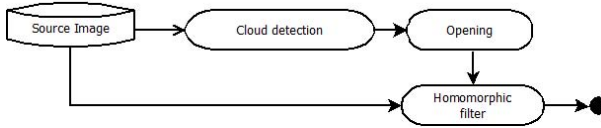


Fig. 2. Flowchart of the algorithm

## 3 Results and Discussions

This section aims to illustrate the procedure previously described, applying it in scenes from different orbital sensors contaminated with scattered clouds, mists and fogs.

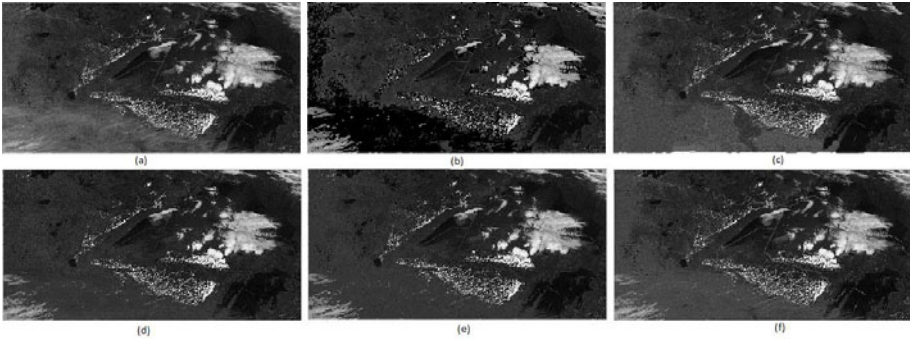
The algorithms to remove these elements will be evaluated according to the image Kappa coefficient [3] and Peak Signal to Noise Ratio (PSNR) [75]. These metrics are based on similarity and compatibility of information between the process resulting image and the reference image, and as larger indexes values, better the outcome. The called reference image is the original image after going through a manually process of atmospheric noise removal, simulating an ideal image, free of all types of damage caused by the atmosphere.

The comparison that motivated this work is done based on what was proposed by [3], where is applied a certain algorithm for areas detection and the Butterworth high-pass filter as the best approach among other high-pass filters during homomorphic filtering.

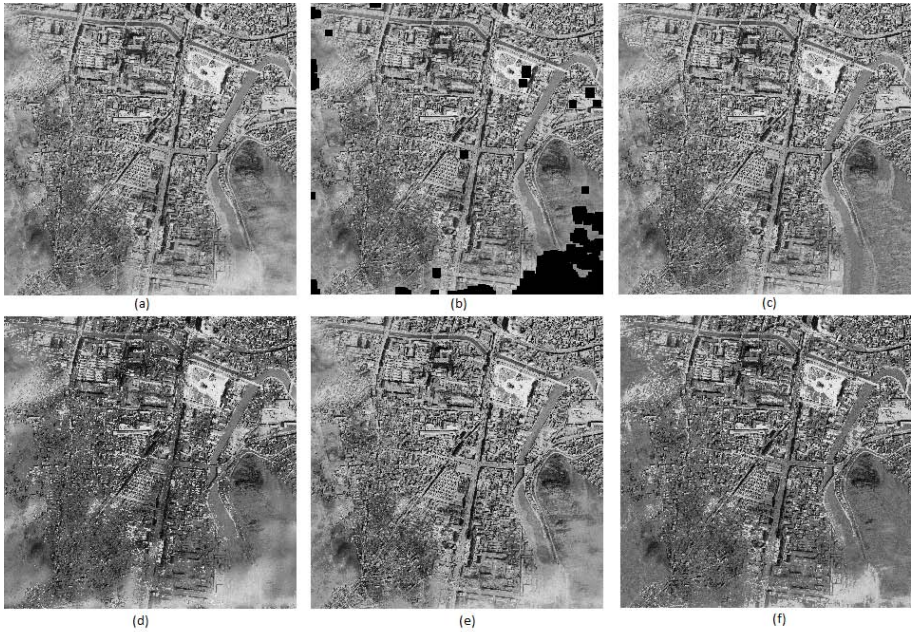
In this article, the process of removing scattered clouds is proposed using an alternative approach, which uses an improved detection of regions, from addition of one constant ( $cc$ ) as well as the opening morphologic operation and using the High-boost filter in homomorphic filtering. This alternative approach obtains similar results and even better than the previous one, as seen in Table II, where are listed the results of the comparison between different approaches among several images and their associated  $cc$  coefficient. Combined with numerical results, one can evaluate the visual aspect resultant of two scenes at Figures 3 and 4, which show the original images, areas of scattered clouds (illustrated by black regions in image), reference image free of noise and images filtered by Ideal, Butterworth and High-boost filters.

**Table 1.** Evaluation of filters Ideal, Butterworth and High-boost. The PSNR index is given in dB.

Tested Images	Original		Ideal		Butterworth		High-boost		cc
	Kappa	PSNR	Kappa	PSNR	Kappa	PSNR	Kappa	PSNR	constant
Italy	0.8662	24.32	0.4806	20.18	0.6866	20.32	0.7491	23.67	1.25
Hawaii	0.8812	25.24	0.5158	16.98	0.5307	19.53	0.7465	23.36	0.75
Tucurui	0.8392	23.91	0.7426	23.8	0.7611	23.9	0.8036	24.6	2.25
Ikonos 1	0.2052	18.17	0.1690	18.53	0.1240	18.37	0.1712	19.38	1.25
Ikonos 2	0.8472	23.63	0.4205	23.24	0.4685	23.55	0.5009	23.63	0.75
Nasa	0.9329	23.24	0.4836	21.02	0.5235	21.19	0.8277	23.17	0.75
Lake	0.6237	17.81	0.5587	18.44	0.5039	17.10	0.5770	18.57	0.9
Andes	0.7843	21.42	0.5897	19.19	0.3973	19.82	0.7788	21.63	3



**Fig. 3.** Tested image 'Lake': (a) Original image (b) Area of scattered clouds (illustrated by black regions in image) (c) Reference image free of noise. (d) Ideal ( $Kappa = 0.5587$  and  $PSNR = +18.44dB$ ), (e) Butterworth ( $Kappa = 0.5039$  and  $PSNR = +17.10dB$ ) and (f) High-boost ( $Kappa = 0.5770$  and  $PSNR = +18.57dB$ ).



**Fig. 4.** Tested image 'Ikonos 2': (a) Original image (b) Area of scattered clouds (illustrated by black regions in image) (c) Reference image free of noise. (d) Ideal ( $Kappa = 0.4205$  and  $PSNR = +23.24dB$ ), (e) Butterworth ( $Kappa = 0.4685$  and  $PSNR = +23.55dB$ ) and (f) High-boost ( $Kappa = 0.5009$  and  $PSNR = +23.63dB$ ).

It is known that the filtering process causes data compression and losses, which explains the decrease of the PSNR index in the output images in comparison to the original ones [3]. The Kappa coefficient follows the behavior of the PSNR index. The computational cost is equivalent among the approaches and, although the method destroys the original pixels, the affected regions become more homogeneous than in the background. In this study, the obstructions of thin clouds are visually reduced, but the systematic reduction of noise and the random distribution of clouds in the image limit the use of filters.

## 4 Conclusion

Therefore, this work deals with the removal of undesirable elements in the atmosphere such as fog, mist and scattered clouds, that damage and lead to loss of image data. Thus, we propose an improvement in clouds detection (applying image opening) and the use of an alternative high-pass filter (High-boost) in Homomorphic filtering to aid removing such elements, being after compared with the approaches applied in [3], mainly with Butterworth filter.

Analyzing the test results, it was observed that the Ideal filter, in most of the images, does not maintain the texture information of the damaged area. The

Butterworth and High-boost filters produce similar results, but the High-boost filter outperforms the Butterworth filter in most of the tests, in terms of Kappa and PSNR parameters. The visual appearance is an important aspect to be analyzed, and in this evaluation, the High-boost also presents better results.

## References

1. Delac, K., Grgic, M., Kos, T.: Sub-image Homomorphic filtering technique for improving facial identification under difficult illumination conditions. In: International Conference on Systems, Signals and Image Processing, pp. 95–98 (2006)
2. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Publishing Company (2008)
3. Hau, C.Y., Liu, C.H., Chou, T.Y., Yang, L.S.: The efficacy of semi-automatic classification result by using different cloud detection and diminution method. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences (2008)
4. Kekre, H.B., Athawale, A., Halarnkar, P.N.: High payload using High Boost filtering in Kekre's Multiple LSB's algorithm. In: 2nd International Conference on Advances in Computer Vision and Information Technology (2009)
5. Grgic, M., Delac, K.: Handbook Of Data Compression. Springer, Heidelberg (2009)
6. Ma, J., Gu, X., Feng, C., Guo, J.: Study of thin cloud removal method for CBERS-02 image. Science in China Series E 48 2(2005-03), 72–90 (2005)
7. Salomon, D., Motta, G.: Handbook Of Data Compression. Springer, Heidelberg (2009)
8. Seow, M., Asari, V.: Ratio rule and homomorphic filter for enhancement of digital colour image. Proceedings of Neurocomputing, 954–958 (2006)
9. Tasdizen, T., Whitaker, R., Burchard, P., Osher, S.: Geometric surface processing via normal maps. Proceedings of ACM Trans. Graph, 1012–1033 (2003)
10. Zhang, X., Qin, F., Qin, Y.: Study on the thick cloud removal method based on multi-temporal remote sensing images. In: International Conference on Multimedia Technology (ICMT), pp. 1–3 (2010)

# A Fast Implementation of Semi-Markov Conditional Random Fields

La The Vinh, Sungyoung Lee, and Young-Koo Lee

Dept. of Computer Engineering,  
Kyung Hee University  
{vinhlt,sylee,yklee}@oslab.khu.ac.kr

**Abstract.** Recently, Conditional Random Fields (CRF) model has been used and proved to be a good model for sequential modeling. It, however, lacks the capability of duration modeling. Therefore, some researchers introduced semi Markov Conditional Random Fields (semi-CRF) to take into account the duration distribution and showed some improvements. Nevertheless, the training algorithms for semi-CRF require quite a high complexity making semi-CRF impractical in some large-scale problems. Therefore, in this work we propose a fast implementation of the training algorithm in order to reduce the complexity required by semi-CRF. Our theoretical analysis as well as experiments' result show a noticeable improvement in computation time, which is about ten times less than that of the original algorithm.

**Keywords:** Conditional Random Fields, Semi-Markov Model.

## 1 Introduction

Recently, Conditional Random Fields(CRF) was introduced as a discriminative model for sequential data and produced a much better result than the well-known existing generative model so-called Hidden Markov Model [3]. Nonetheless, both conventional HMM and first order linear chain CRF are limited to the Markovian property, which assumes that the current state depends only on the previous state. Because of this assumption, the model is not able to capture the duration distribution as well as the long-range transitions of states. To solve the problem, Sarawagi and Cohen proposed a semi Markov Conditional Random Fields (semi-CRF) model in [5]. However, Sarawagi and Cohen's model increases the complexity of forward/backward as well as gradient estimation algorithms by  $D$  times, where  $D$  were the maximum duration length. Although there is some other work about semi-CRF [2], [6], none of them shows any improvement in the computation complexity of the model. In [4] the author proposed a method to make semi-CRF scalable. In that work, the author utilized a Naive Bayes classifier to filter out some candidates making the computation much more faster. However, there may be a trade-off since the removed candidates may affect the final accuracy of the model in some applications. Therefore, we are going to overcome the above limitations by developing fast gradient estimation algorithms for semi-CRF while keeping the model's behavior [5] unchanged.

## 2 Semi-Markov Conditional Random Fields

Hereafter, we assume that the observation and the corresponding label sequence of length  $T$  are given in the form

$$X = \{x_1, x_2, \dots, x_T\}, \quad (1)$$

$$Y = \{y_1, y_2, \dots, y_T\}. \quad (2)$$

In our work, each state of a semi-CRF is defined as

$$s_i = (y, b, e) \quad i = 1, 2, \dots, P, \quad (3)$$

where  $P$  is the length of sequence  $S = s_1 \dots s_P$ , which is constructed from input labels  $Y = (y_1, y_2, \dots, y_T)$ .  $y, b, e$  in turn are label, beginning time, and ending time of state  $s_i$ . The beginning and ending time must satisfy the following constraints.

$$s_i.b \leq s_{i+1}.e \quad i = 1, 2, \dots, P, \quad (4)$$

$$s_i.e + 1 = s_{i+1}.b \quad i = 1, 2, \dots, P - 1, \quad (5)$$

$$s_1.b = 1, \quad (6)$$

$$s_P.e = T. \quad (7)$$

Now, the likelihood of  $S$  given  $X$  is estimated by

$$P(S|X) = \frac{\prod_{i=1}^P \Psi(s_{i-1}, s_i, X)}{Z_X}, \quad (8)$$

$$Z_X = \sum_{S'} \prod_{i=1}^{P'} \Psi(s'_{i-1}, s'_i, X), \quad (9)$$

where  $\Psi(s_{i-1}, s_i, X)$  encodes the potential of the transition from  $s_{i-1}$  to  $s_i$ . In equations (8) and (9), we can consider the product of potential functions  $\Psi$  over all transitions of a sequence as the potential of the sequence, then we can rewrite (8) in the form

$$P(S|X) = \frac{Pol(S)}{\sum_{S'} Pol(S')}, \quad (10)$$

where

$$Pol(S) = \prod_{i=1}^P \Psi(s_{i-1}, s_i, X) \quad (11)$$

is the potential of the sequence  $S = s_1, s_2, \dots, s_P$ . The potential function can be defined as followings

$$\Psi(s_{i-1}, s_i, X) = \begin{pmatrix} e^{Q^T(s_{i-1}, s_i, X)} \times \\ e^{Q^D(s_{i-1}, s_i, X)} \times \\ e^{Q^O(s_{i-1}, s_i, X)} \end{pmatrix}, \quad (12)$$

where  $Q^T$ ,  $Q^D$  and  $Q^O$  are the weighted transition, duration and observation potential functions, respectively. These functions may have different forms in different applications. Nevertheless, we present below examples of the functions which could be applied in some well-known applications, namely Name Entity Recognition (NER), Activity Recognition (AR).

$$Q^T(s_{i-1}, s_i, X) = \sum_{y', y} w^T(y', y) \delta(s_{i-1}.y = y', s_i.y = y), \quad (13)$$

where  $w^T(y', y)$  is the weight of transition from  $y'$  to  $y$  and  $\delta$  is given by

$$\delta(X) = \begin{cases} 1 & \text{if } X \text{ is true} \\ 0 & \text{if } X \text{ is false} \end{cases}. \quad (14)$$

$$Q^D(s_{i-1}, s_i, X) = \sum_{y, d} G^D(y, d) \delta(s_i.y = y, d = s_i.e - s_i.b + 1) \quad (15)$$

$$= \sum_{y, d} w^D(y) \frac{(d - m_y)^2}{2\sigma_y^2} \delta(s_i.y = y, d = s_i.e - s_i.b + 1), \quad (16)$$

where  $w^D(y)$  is the duration weight of  $y$ .  $m_y$  and  $\sigma_y$  are the average and the standard deviation of state  $y$ 's duration respectively, which can be easily extracted from training data.

$$Q^O(s_{i-1}, s_i, X) = \sum_{y, t_1, t_2} \left( \delta(s_i.y = y, s_i.b = t_1, s_i.e = t_2) \times G_y(y, t_1, t_2) \right), \quad (17)$$

where

$$G_y(y, t_1, t_2) = \sum_{t=t_1}^{t_2} \sum_o w^O(y, o) \delta(x_t = o), \quad (18)$$

where  $w^O(y, o)$  is the weights of the observation given that input symbol  $o$  is observed in state with label  $y$ . For our convenience in the presentation of the following equations, we denote  $G(y, t_1, t_2) = G_y(y, t_1, t_2) + G^D(y, t_2 - t_1 + 1)$  as a combined potential function.

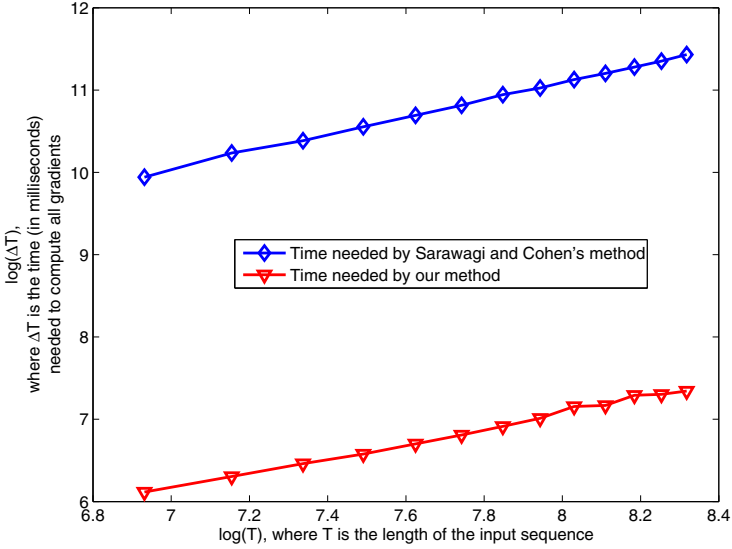
As we noted above that the potential functions may have different definitions. The above equations are just examples of them, which generally can be applied to a large number of applications including language processing [1], gene prediction [2], activity recognition [7], etc.

## 2.1 Forward Algorithm

Forward algorithm is used to compute the normalization factor  $Z_X$  efficiently by using the dynamic programming method.

To compute  $Z_X$  in (9) let we denote

$$\alpha(y, t) = \sum_{S^t \in \Gamma_t^y} Pol(S^t), \quad (19)$$



**Fig. 1.** Average time needed for computing all the gradients. Herein, the number of labels ( $M$ ) is 4, the maximum duration ( $D$ ) is 16, the number of input values (or often known as size of codebook) ( $V$ ) is 128, the length of the input sequence ( $T$ ) changes from 1024 to 4096 with a step of 256. Therefore, the number of gradients is  $M + M^2 + MV = 532$ .

where  $\Gamma_t^y = \{S = s_1, s_2, \dots, s_q\}$  is a set of all semi-Markov sequences, which have an original label sequence  $(y_1, y_2, \dots, y_t)$  with the last label is  $y$ . Thus, every  $S^t = s_1, s_2, \dots, s_q \in \Gamma_t^y$  satisfies  $s_q.e = t$ ,  $s_q.y = y$ . Our forward algorithm is implemented in the following equations

$$Z_X = \sum_y \alpha(y, T), \quad (20)$$

$$\alpha(y, t) = \sum_{d=1}^D (\lambda(y, t-d) e^{G(y, t-d+1, t)}), \quad (21)$$

$$\lambda(y, t) = \sum_{y'} \alpha(y', t) e^{w^T(y', y)}. \quad (22)$$

## 2.2 Backward Algorithm

Similarly to the forward algorithm, let we denote

$$\beta(y, t) = \sum_{S^{T-t+1} \in \Omega_t^y} Pol(S^{T-t+1}), \quad (23)$$



where  $\Omega_t^y = \{S = s_1, s_2, \dots, s_q\}$  is a set of all semi-Markov sequences, which have an original label sequence  $(y_t, y_{t+1}, \dots, y_T)$  with the first label is  $y$ . The backward algorithm is described in the below equations

$$Z_X = \sum_y \beta(y, 1), \quad (24)$$

$$\beta(y, t) = \sum_{d=1}^D (\zeta(y, t+d) e^{G(y, t, t+d-1)}), \quad (25)$$

where

$$\zeta(y, t) = \sum_{y'} \beta(y', t) e^{w^T(y, y')}. \quad (26)$$

### 2.3 Parameter Estimation

The goal of parameter estimation is to choose appropriate values for the model weights ( $w^T$ ,  $w^D$  and  $w^O$ ) so that the likelihood of the observation data  $P(S|X)$  is maximized. Take the logarithm form of  $P(S|X)$  we have

$$L(S|X) = \sum_{i=1}^P \left( \frac{Q^T(s_{i-1}, s_i, X) + Q^D(s_{i-1}, s_i, X) + Q^O(s_{i-1}, s_i, X)}{Q^O(s_{i-1}, s_i, X)} \right) - \log(Z_X). \quad (27)$$

To find the optimal parameter values  $w$  we have to solve  $\frac{dL}{dw^*} = 0$ . From (27) we know that

$$\frac{dL}{dw^*} = \sum_{i=1}^P \frac{dQ^*(s_{i-1}, s_i, X)}{dw^*} - \frac{1}{Z_X} \frac{dZ_X}{dw^*}. \quad (28)$$

Herein we use  $Q^*$  and  $w^*$  with the superscript to refer to any kind of the potential function and weight (\* can be D, T, or O). Computing the first term of the right side is trivial,  $Z_X$  is calculated by using forward/backward variables. Therefore, here we mainly focus on evaluating  $\frac{dZ_X}{dw^*}$  for different kind of weights. From (9) and (12) we have

$$\frac{dZ_X}{dw^*} = \sum_{S^T} \left( \left( \sum_{i=1}^P \frac{dQ^*(s_{i-1}, s_i, X)}{dw^*} \right) \prod_{i=1}^P \Psi(s_{i-1}, s_i, X) \right). \quad (29)$$

**Gradient of the Transition Weight.** Since

$$\frac{dQ^T(s_{i-1}, s_i, X)}{dw^T(y', y)} = \delta(s_{i-1}.y = y', s_i.y = y), \quad (30)$$

we have

$$\frac{dZ_X}{dw^T(y', y)} = \sum_{t=1}^T \alpha(y', t) \beta(y, t+1) e^{w^T(y', y)}. \quad (31)$$

**Gradient of the Duration Weight.** From the definition of the duration potential function, it is clear that

$$\frac{dQ^D(s_i, s_{i-1}, X)}{dw^D(y)} = \sum_{y,d} \delta(s_i.y = y, s_i.e - s_i.b + 1 = d) \frac{(d - m_y)^2}{2\sigma_y^2}. \quad (32)$$

As a result

$$\frac{dZ_X}{dw^D(y)} = \sum_{d=1}^D \sum_{t=1}^T \frac{(d - m_y)^2}{2\sigma_y^2} \theta(y, t, d), \quad (33)$$

where

$$\theta(y, t, d) = \lambda(y, t - 1) \zeta(y, t + d) e^{G(y, t, t+d-1)} \quad (34)$$

can be considered to be the sum of potential values of all sequences  $Y = y_1, y_2, \dots, y_T$ , which have  $d$  labels  $y$  from time  $t$ , or equivalently  $y_t = y_{t+1} = \dots = y_{t+d-1} = y$ .

**Gradient of the Observation Weight.** From (17) and (18) we have

$$\frac{dQ^O(s_{i-1}, s_i, X)}{dw^O(y, o)} = \sum_{k=s_i.b}^{s_i.e} \delta(s_i.y = y, x_k = o). \quad (35)$$

Combining (35) and the definition of  $\theta$  in (34) leads to

$$\frac{dZ_X}{dw^O(y, o)} = \sum_{\substack{k, t, d \\ k \in [t, t + d - 1]}} \theta(y, t, d) \delta(x_k = o). \quad (36)$$

### 3 Evaluation

In this section, we show that our proposed algorithm achieves a remarkable improvement in terms of the complexity in comparison with the previous work theoretically as well as practically. In the following experiments, we use C++ programming language with Microsoft Visual Studio IDE to implement the algorithms.

#### 3.1 Complexity Analysis

As it is described in [5], the required complexity for computing each gradient is  $O(TM^2D)$ , where  $T$ ,  $M$ ,  $D$  are the length of the input sequence, the number of label values, and the maximum duration of a label, respectively. Because of this, the estimation of gradients for all  $N$  model's parameters takes  $O(NTM^2D)$ .

In our solution, gradients are computed by using (31), (33), and (36). It is obvious that if  $\alpha$ ,  $\lambda$ ,  $\beta$ ,  $\zeta$ , and  $\theta$  are cached then the maximum time needed is about  $O(TD)$ . Therefore, for optimizing  $N$  parameters, our algorithm needs only

$O(NTD)$  to complete calculating all gradients. Nevertheless, we need to take into account the extra time of estimating the cached variables. As it is shown in the forward and backward algorithm,  $\alpha$ ,  $\lambda$ ,  $\beta$ , and  $\zeta$  can be computed with  $O(2TM(M+D))$ , meanwhile from (34) we see that  $\theta$  takes  $O(TMD)$ . Totally, for caching of these variables, we need a complexity of around  $O(2TM^2 + 3TMD)$ . Herein, we take a numerical example to compare  $O(NTM^2D)$ , which is the estimated complexity in [5] and  $O(NTD + 2TM^2 + 3TMD)$ , our algorithm complexity. Let  $T = 1000$ ,  $D = 100$ ,  $M = 8$  and  $N = 10000$  then the former is about  $64 \times 10^9$ , the latter is about  $10^9$ .

In addition, Fig. 1 illustrates another comparison of the two complexities with  $N=532$ ,  $M=4$ ,  $D=16$  and  $T$  changes from 1024 to 4096, time is measured in millisecond. The blue represents the amount of time which is needed by the method proposed by Sarawagi and Cohen in [5], time consumed by our algorithm is marked by the red. Undoubtedly, there is a remarkable improvement in our complexity.

## 4 Conclusion

In this paper, we have presented a fast implementation for estimating the gradients of semi-CRF in the training phase, increasing the computation performance of the model. Although, semi-CRF is a powerful discriminative model for sequential data. It, however, is limited in use because of the high computation complexity. Therefore, we believe that our contribution will make semi-CRF more practical especially for large-scale applications.

**Acknowledgments.** This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency) (NIPA-2009-(C1090-0902-0002)). This work was also sponsored by the IT R&D program of MKE/KEIT, [10032105, Development of Realistic Multiverse Game Engine Technology], and by the Basic Science Research Program funded by National Research Foundation (2009-0076798).

## References

1. Andrew, G.: A hybrid markov/semi-markov conditional random field for sequence segmentation. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing, EMNLP (2006)
2. Doherty, M.K.: Gene Prediction with Conditional Random Fields. Master's thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology (2007)
3. Lafferty, J., McCallum, A., Pereira, F.: Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceeding of International Conference on Machine Learning, ICML (2001)

4. Okanohara, D., Miyao, Y., Tsuruoka, Y., Tsujii, J.: Improving the scalability of semi-markov conditional random fields for named entity recognition. In: Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual Meeting of the Association for Computational Linguistics, pp. 465–472 (2006)
5. Sarawagi, S., Cohen, W.: Semi-markov conditional random fields for information extraction. In: Proceedings of Advances in Neural Information Processing Systems, NIPS (2004)
6. Truyen, T., Phung, D., Bui, H., Venkatesh, S.: Hierarchical semi-markov conditional random fields for recursive sequential data. In: Proceeding of International Conference on Advances in Neural Information Processing, NIPS (2008)
7. Vail, D.L., Veloso, M.M., Lafferty, J.D.: Conditional random fields for activity recognition. In: Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems (2007)

# Adapted Scan Based Lossless Image Compression

Tarek Ouni, Arij Lassoued, and Mohamed Abid

National Engineering School of Sfax  
Road Sokkra, Km 3 Sfax,  
Tunisia  
{tarek.ouni,arij.lassoued}@gmail.com,  
Mohamed.abid@enis.rnu.tn

**Abstract.** This paper deals with the application of lossless compression algorithms to two-dimensional curves scanned images. An image is scanned along a space filling curve (SFC) so as to exploit inherent coherence in the image. The used SFC is determined by a gradient based method allowing the detection of global pixel's change direction for each image block. The resulting one-dimensional representation of the image has improved auto-correlation compared with universal scans such as the Peano-Hilbert space filling curve. Combined with conventional coding algorithms the proposed algorithm shows significant compression efficiency improvement. The new algorithm used for SFC determination is presented and used as an input to conventional coding schemes.

**Keywords:** Gradient, global, SFC, image, Lossless coding.

## 1 Introduction

Most conventional lossless image compression schemes, such as GIF and PNG, are based on entropic coding. In such schemes, image's pixels are scanned from left to right top to bottom providing pixels sequences further compressed using entropic coding. In such coder, redundancies in directions others than horizontal one, are omitted [1]. However, redundancy in these directions could be more relevant than horizontal one, and so, scanning the image in suitable direction may provide a more correlated signal and image redundancy that could be more explored.

For this reason, many researches attempt to explore more efficient scan methods, called SFC or Space Filling Curves, able to explore redundancy in different directions.

Such approaches aim to translate the intra-frame correlation in the image to a favorable autocorrelation within the pixel-sequence.

In fact, the scanning process transforms the 2-dimensional image into 1-dimensional representation focusing on the nearby pixel similarity in the image source. They are designed to exploit this characteristic to improve the autocorrelation in the resulting 1-dimensional image representation. Image scanning using SFCs is a typical example of such an algorithm [1-6].

They define a continuous scan that traverses through every image pixel exactly once.

The resulting sequence of pixels is processed as required by the particular application, like lossless or lossy compression, halftoning, analysis, pattern recognition or texture analysis.

Indeed, such approaches have been vastly emphasized in the 90s. Then they were submerged by transform based compression methods and practically abundant since the beginning of this century. However, some researchers still believe that such approaches were not really exhausted, and can still be useful.

First SFC based approaches were based on statistic image characteristics. The most popular one is the Peano-Hilbert curve, which has been considered, for its strong locality property, for numerous applications [1][4][5][6]. Lempel and Ziv [9] showed that, for images generated by suitably random sources, the entropy of the pixel-sequence obtained using the Peano-Hilbert curves converges asymptotically to the two-dimensional entropy of the image. Hence, compressing the sequence using the Lempel-Ziv encoder [8] results in an image compression scheme that is optimal in the information theoretic sense. Matias and Shamir considered [7] the relationship between the two-dimensional autocorrelation of an image and the one-dimensional autocorrelation of the pixel-sequence. They showed that, for first-order Markov isotropic images, the autocorrelation of the pixel-sequence is a function of the fractal-dimension  $n$  of the SFC. These studies support the approach that recursive SFCs, such as the Peano-Hilbert curve would be a good choice as a universal SFC that would work well (statistically) for large families of images [10]. Next tendency are turned towards the application this SFC with dynamic approaches which react with the image semantic content.

In [12] authors propose the use of context-based space filling curves that are to be computed so as to exploit inherent coherence in the image. That is, rather than relying on a universal SFC that works statistically well, this approach consists in selecting a SFC that would work well for the particular image or group of images. A favorable SFC would traverse the pixels within the principal shapes in the considered image for as long as possible before moving across the border edges to traverse the pixels outside the shapes. Thus the context-based SFC is tailored to avoid edge crossings to a considerable extent, resulting in a smoother pixel-sequence. However, the problem of finding such a SFC is NP-hard, especially in image with complex texture. Another consideration is that, unlike with universal SFCs, the selected context-based SFC needs to be encoded along with the pixel sequence, to enable retrieval at a later stage which demands costly additional information to be transmitted. The same inconvenient persists in [14], where authors utilize a new algorithm based on correlation optimization used for Space Filling Curve (SFC) determination in image compression. In [13], authors have proposed a bit-plane processing system which utilizes an image scanning language called SCAN. If an 8-bit gray scale image matrix is separated into its eight binary components or bit planes, and each of these 2D arrays is scanned with the appropriate SCAN language algorithm, then pixels with the same binary value can be grouped together and run-length encoding would yield very good compression. This methodology works very well for the more significant bit

planes, but compression becomes impossible for the less significant bit planes due to the random nature of these image patterns.

In this context, our research in image compression is based on finding specific SFC that will be adapted for the image content. Our approach consists in applying the gradient concept for analyzing the image local pixels activity in order to select the direction in where minimal pixels change is estimated. According to this direction, we assign, for each considered group of pixels, the suitable SFC (the one which advantages the selected direction). The resulting 1-dimensional image representation provides a higher neighbour pixel similarity. The proposed algorithm can be considered as a preprocessing which transforms the image source into some strongly correlated representation before applying coding algorithms. We have demonstrated the contribution of integrating this method in the conventional coding algorithms.

The Organization of the rest of the paper is as follows: In Section 2 we introduce the proposed gradient-based SFC approach. In Section 3 we discuss the autocorrelation improvement in the proposed scan method. In Section 4 we discuss compression efficiency improvement of the proposed approach. We conclude with a short summary.

## 2 Proposed Approach Description

The basic idea is to rank the pixels in such a way that adjacent pixels in the resulting bit sequence will have the highest similarity. To overcome this objective we analyze the image pixels activity in order to find the suitable scan. The best way is to scan the image according to the direction where minimal pixel's change is recorded. To do so, we use a gradient based to detect image pixel's change direction.

First, the proposed scan methodology splits the original image into isometric squared blocks. The SFC is computed for each block individually because each one has its own local auto-correlation characteristic and nearby pixel similarity. The SFC is determined by the decisional image gradient - based algorithm presented below.

The gradient of a function of two variables,  $F(x, y)$ , is defined by:

$$\vec{G} = \nabla F = \frac{\partial F}{\partial x} \vec{i} + \frac{\partial F}{\partial y} \vec{j} \quad (1)$$

and can be viewed as a collection of vectors pointing in the direction of increasing values of  $F$ .

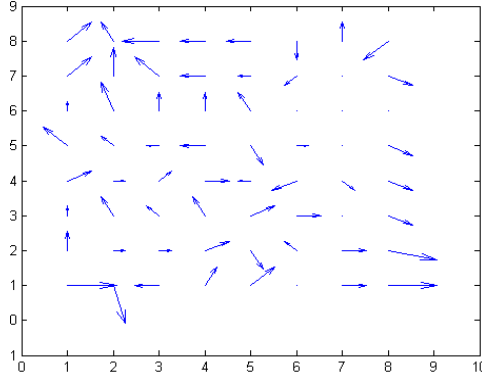
The gradient of one two-dimensional bloc  $F$  is defined by:

$$[G_x, G_y] = \text{gradient}(F) \quad (2)$$

where  $G_x$  (resp.  $G_y$ ) corresponds to  $\frac{\partial F}{\partial x}$  (resp.  $\frac{\partial F}{\partial y}$ ), the differences in  $x$  (resp.  $y$ ) direction. The spacing between two adjacent points in each direction is assumed to be one.

Each point  $(i, j)$  in the block matrix  $F$  is presented by a local gradient (cf. figure 1):

$$\vec{g}(i, j) = g_x(i, j) \vec{i} + g_y(i, j) \vec{j} \quad (3)$$



**Fig. 1.** Example of 8x8 Block pixel's gradient

The global gradient is:

$$\vec{g} = \sum_{i=0}^{H-1} \sum_{j=0}^{L-1} g_x(i, j) \vec{i} + g_y(i, j) \vec{j} \quad (4)$$

$$\vec{g} = \sum_i \sum_j g_x(i, j) \vec{i} + \sum_i \sum_j g_y(i, j) \vec{j} \quad (5)$$

Where H and L are respectively the block height and the width.

Let:

$$\vec{G}_x = \sum_i \sum_j g_x(i, j) \vec{i} \quad \text{and} \quad \vec{G}_y = \sum_i \sum_j g_y(i, j) \vec{j}$$

Then, each image block is represented by its own gradient components  $\vec{G}_x$  and  $\vec{G}_y$ . Therefore, the direction of the block pixel's change will be approximated by the orientation of the global gradient vector:

$$\vec{g} = \vec{G}_x + \vec{G}_y \quad (6)$$

The global gradient orientation is:

$$\theta = \tan^{-1} \left( \frac{G_x}{G_y} \right) \quad (7)$$

Because of the infinite number of values of  $\theta$ , it should be approximated to one of the useful orientation values after mentioned. This approximation aims to limit the used SFC's for two major reasons; the first is reducing the method processing complexity, the second is reducing the additional information to be transferred with the scanned image to perform correct reconstruction on the receiver side.

We only consider 4 approximate directions:



- Horizontal:  $\vec{i}$  **direction** ( $\theta = 0$ )
- Vertical:  $\vec{j}$  **direction**  $\theta = 90$ ,
- First diagonal  $(\vec{i} - \vec{j})$  **direction**  $\theta = -45$  ,
- Second diagonal  $(\vec{i} + \vec{j})$  **direction**  $\theta = 45$ .

According to the aforementioned considered directions, the coder chooses between 4 proposed SFC's:

- Horizontal snake scan (c.f. Figure 2)
- Vertical snake scan (cf. Figure 3)
- First Zigzag (c.f. Figure 4)
- Second Zigzag (c.f. Figure 5)

We only need 2 bits per bloc to code the selected scan:

- 00: horizontal snake scan
- 01: vertical snake scan
- 10: first zigzag scan
- 11: second zigzag scan

For each image bloc, the appropriate scan is selected as follows:

- If  $(|G_x| \ll |G_y|)$  then  $\vec{G} \cong 0\vec{i} + G_y\vec{j}$

The minimal pixel's change direction is the horizontal one ( $\vec{i}$  direction), consequently the selected SFC will be the horizontal snake scan.

- If  $(|G_x| \gg |G_y|)$  then  $\vec{G} \cong G_x\vec{i} + 0\vec{j}$

The minimal pixel's change direction is the vertical one ( $\vec{j}$  direction), consequently the selected SFC will be the vertical snake scan.

- If  $|G_x| \cong |G_y|$  and  $G_x \times G_y > 0$  then:

$$\begin{aligned}\vec{G} &\cong G_x\vec{i} + G_y\vec{j} \\ &\cong G_x(\vec{i} + \vec{j}) + 0 \times (\vec{i} - \vec{j})\end{aligned}$$

The minimal pixel's change direction is the first diagonal one ( $(\vec{i} - \vec{j})$  direction). Consequently, the selected SFC will be the first ZigZag scan

- If  $|G_x| \cong |G_y|$  and  $G_x \times G_y < 0$  then

$$\begin{aligned}\vec{G} &\cong G_x\vec{i} - G_y\vec{j} \\ \vec{G} &\cong G_x(\vec{i} - \vec{j}) + 0 \times (\vec{i} + \vec{j})\end{aligned}$$

The minimal pixel's change direction is the second diagonal one ( $(\vec{i} + \vec{j})$  direction). Consequently, the selected SFC will be the second ZigZag scan.

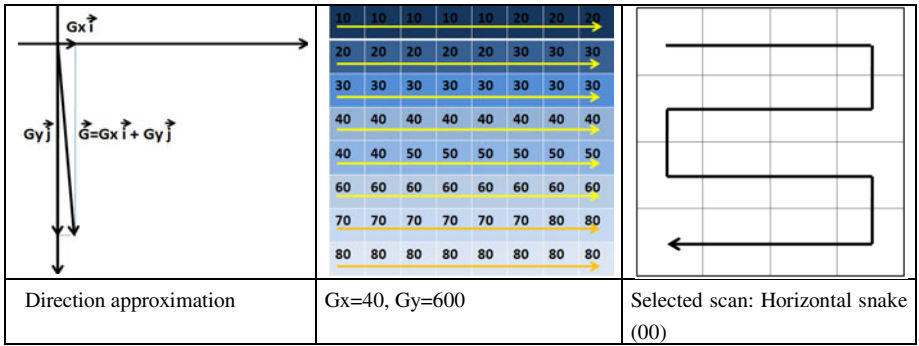


Fig. 2. Horizontal snake scan selection example

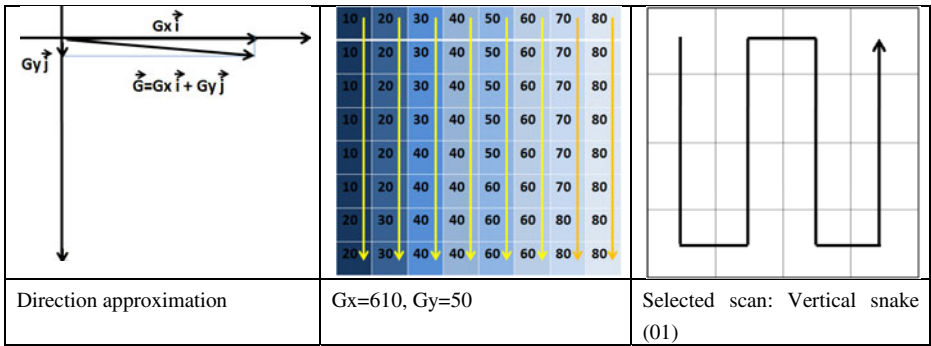


Fig. 3. Vertical snake scan selection example

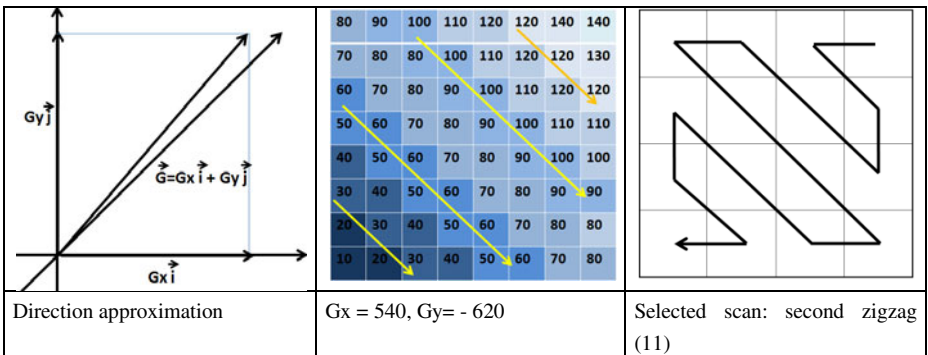


Fig. 4. Second zigzag scan selection example

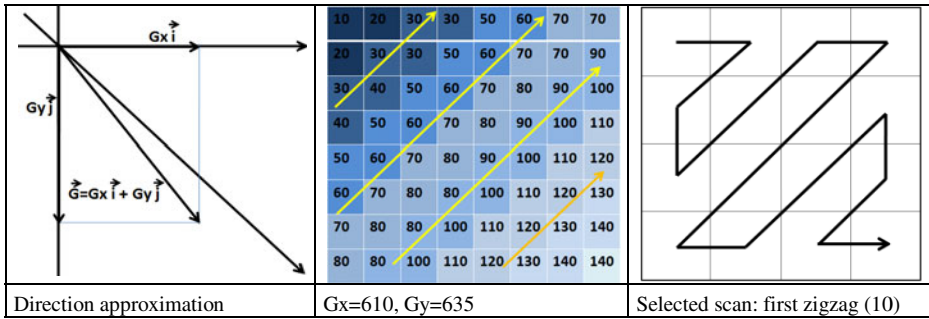


Fig. 5. First zigzag scan selection example

### 3 Autocorrelation Improvement

The advantage of gradient-based SFC is illustrated by the three 8x8 blocks-scans of “badoo” image in Figure 3. The pixel-sequences are displayed as a line scan into the displayed frame, and consist of the following SFCs: a scan line-hence showing the original image- (figure 3-a), hilbert scan (figure 3-b), and the gradient-based scan (figure 3-c). As can be observed, the pixel-sequence resulting from the proposed scan is smoother than the one based on the hilbert curve.

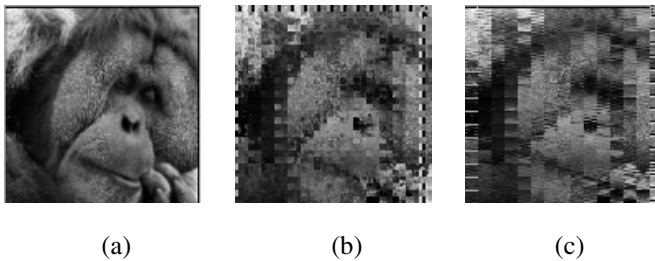


Fig. 6. Images scanning using a scan line (a), Peano scan (b), and gradient-based scan (c)

The proposal for image gradient-based space filling curve is primarily motivated by the proposition that a curve tailored for a given image would better exploit its spatial coherence than a universal curve. To support this proposition, we compare the autocorrelation of 1-D pixel-sequences generated by both gradient-based and Hilbert space filling curves. The pixel-sequences were generated for the four pictures in Figure 4, and their average autocorrelation is displayed.

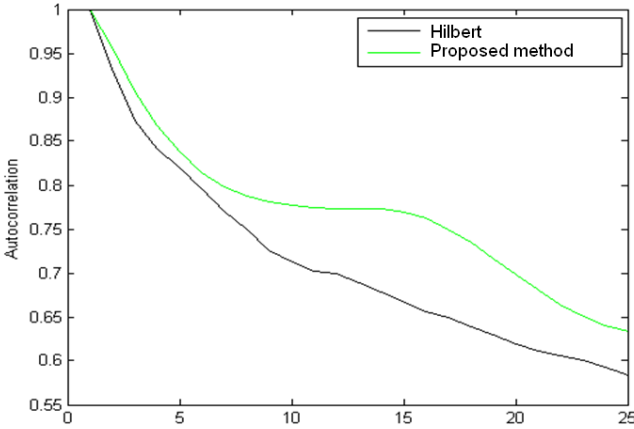


Fig. 7. Comparative results on autocorrelation

## 4 Compression Efficiency Improvement

As a complementary approach to evaluate the redundancy of the pixel-sequences, images are scanned, block by block, respectively with Hilbert scan (the best scan in statistic point of view), and gradient based scan. The resulting pixel-sequences, firstly DPCM encoded, were LZW encoded. The results, depicted in figure 8, show a noticeable improvement in compression efficiency compared to Hilbert scan.

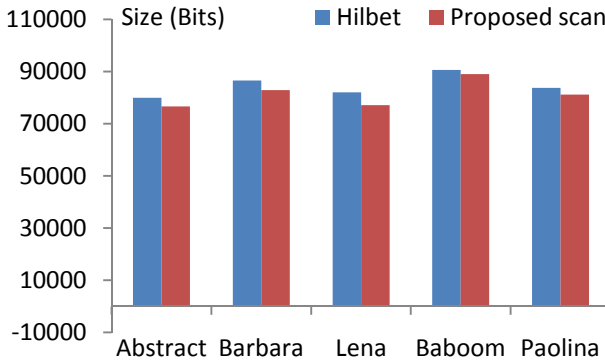
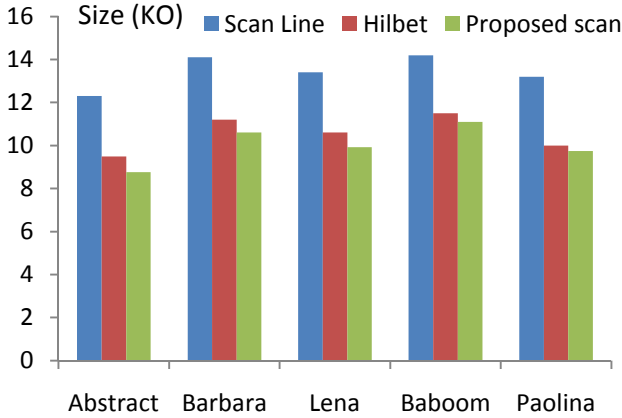


Fig. 8. LZW compression results of the line scanned, Hilbert scanned and gradient-based scanned 2D-images

In one second experiment, images are scanned, block by block, respectively with linear scan, Hilbert scan, and gradient based scan. The resulting pixel-sequences were positioned (block by block) using linear scan to reconstruct the 2D pictures (as illustrated in Figure 3), and were compressed using PNG encoder. The results, depicted in figure 9, show a perceptible improvement in compression efficiency compared to Hilbert scan.



**Fig. 9.** PNG compression results of the line scanned, Hilbert scanned and gradient-based scanned 2D-images

This improvement is primarily due to two facts factors; First of all, the proposed scan takes its advantages from the local autocorrelation, which is accentuated in the block by block 2D image reconstructing. Second, the proposed method has insignificant additional data for the further reconstruction of the image unlike other state of the art approaches which require considerable additional data to code the non standard used scan trajectory.

## 5 Conclusion

Space filling curves are standard means for scan based image processing approaches; they translate the 2-D spatial coherence in the image into a 1-D autocorrelation in the sequence. The increased autocorrelation provided by state of the art SFC based compression methods such as context-based space filling curves approach [12] and correlation optimization based SFC described in [14] are done at the cost of computational processing and additional data information required for further image reconstruction.

The presented gradient-based scanning approach used here in 2D-lossless image compression presents many advantages; First it exploits the spatial coherence of the images better than the conventional scan method (linear and Hilbert scans). Second, it presents less computational processing and least additional data information compared to related works. This work is a first step in the re-exploring of the space filling curves utility for image compression. Actually, we are working on other decisional approaches aiming to better select the appropriate SFC. Future researches will reveal to what degree the proposed technique can be advantageous in lossy compression.

## References

1. Ansari, A., Fineberg, A.: Image data compression and ordering using peano scan and lot. *IEEE Transactions on Consumer Electronics* 38(3), 436–445 (1992)
2. Velho, L., Gomes, J.M.: Digital halftoning with space filling curves. *Computer Graphics* 25(4), 81–90 (1991)
3. Witten, I.H., Neal, R.M.: Using peano curves for bilevel display of continuous-tone images. *IEEE Computer Graphics and Applications* 2(5), 47–52 (1982)
4. Kamata, S., Eason, R.O., Kawaguchi, E.: An implementation of the hilbert scanning algorithm and its application to data compression. *IEICE Transaction Information and Systems* E-76(4), 420–427 (1993)
5. Thyagarajan, K.S., Chatterjee, S.: Fractal scanning for image compression. In: *Conference Record of the Twenty-Fifth Asilomar Conference on Signals, Systems and Computers*, pp. 467–471 (June 1992)
6. Nguyen, P.T., Quinqueton, J.: Space filling curves and texture analysis. In: *IEEE Intl. Conf. Pattern Recognition*, pp. 282–285 (October 1982)
7. Matias, Y., Shamir, A.: A Video Scrambling Technique Based on Space Filling Curves. In: Pomerance, C. (ed.) *CRYPTO 1987*. LNCS, vol. 293, pp. 398–417. Springer, Heidelberg (1988)
8. Ziv, J., Lempel, A.: Compression of individual sequences via variable-rate coding. *IEEE Transactions on Information Theory* IT-24(5), 530–536 (1978)
9. Ziv, J., Lempel, A.: Compression of two-dimensional data. *IEEE Trans. on Information Theory* 32(1), 2–8 (1986)
10. Max, N.: Visualizing hilbert curves. *IEEE Visualization*, 447–450 (1998)
11. Memon, N., Neuhoff, D., Shende, S.: An Analysis of Scanning Techniques for Lossless Image Coding. *IEEE Transactions on Image Processing* 9, 1837–1848 (2000)
12. Dafner, R., Cohen-Or, D., Matias, Y.: Context-based Space Filling Curves. In: Gross, M., Hopgood, F.R.A. (eds.) *EUROGRAPHICS 2000*, vol. 19(3) (2000)
13. Drosta, G.W., Bourbakisa, N.G.: A hybrid system for real-time lossless image compression. *Microprocessors and Microsystems* 25, 19–31 (2001)
14. Pračko, R., Polec, J.: DPCM Application to images scanned by SFC method. *Journal of Electrical Engineering* 58(3), 161–164 (2007)

# Localization of Passengers Inside Intelligent Vehicles by the Use of Ultra Wideband Radars

Philipp Galdia<sup>1</sup>, Carsten Koch<sup>2</sup>, and Anthimos Georgiadis<sup>1</sup>

<sup>1</sup> Leuphana University Lueneburg, Institute of Product and Process Innovation,  
Volgershall 1, 21339 Lueneburg, Germany  
{philipp.galdia, anthimos.georgiadis}@leuphana.de

<http://www.leuphana.de>

<sup>2</sup> University of Applied Sciences Emden/Leer, Institute of Informatics,  
Automation and Robotics  
Constantiaplatz 4, 26725 Emden, Germany  
koch@i2ar.de

<http://www.i2ar.de>

**Abstract.** In this paper an approach for the localization of passengers inside a vehicle by the use of ultra wideband technology (UWB) is presented. Established approaches as static background subtraction for the separation of fore- and background show good performance in static scenes. But in the case of a dynamic surrounding such as a car's interior, former techniques are limited, since significant errors are produced as soon as the scene gets dynamic.

Instead of trying to separate fore- and background, the presented approach involves the complete data set in a first step. Since human breathing shows certain behaviour such as a typical frequency and amplitude, the breathing movement can be separated from other movement or static parts of the radar scan. The respiration and by that the passenger inside a vehicle is located and tracked.

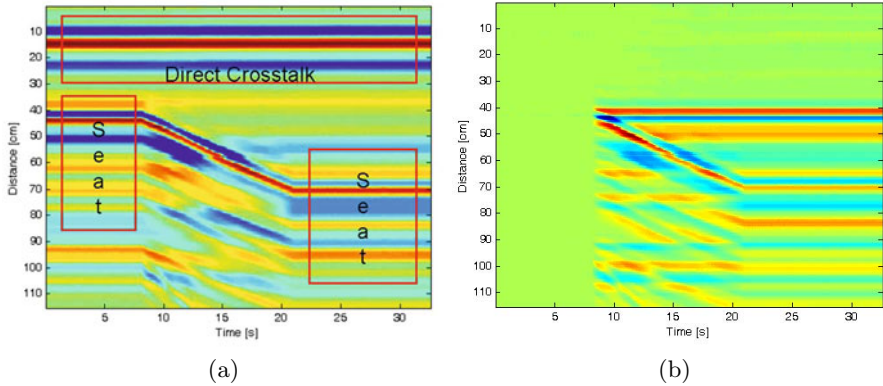
**Keywords:** Ultra wideband, radar, vital sign analysis, localization, airbag.

## 1 Introduction

It is essential to detect human presence and to locate passengers for safety purposes inside a car. An example for an application is a smart airbag control, which inflates airbags depending on the position of the passengers [1].

An established technique of human being detection via ultra wideband is the use of a static background subtraction [2,3]. A reference frame is taken from the scene with no passengers. This frame is subtracted from the incoming data to remove the background. If the scene does not change, the person is always seen as foreground. But as soon as the scene includes dynamic elements, the reference frame includes outdated values. The subtraction results in erroneous foreground objects, which interfere a subsequent object analysis.

In practice the scene of a car interior will always be dynamic. Passengers will move their seat to other positions, the seat is deformed by the person's weight or the car is shaking from motion caused by driving. This shows that an approach using UWB with static background subtraction is very limited for applications inside vehicles.



**Fig. 1.** UWB radar scan: (a) of a dynamic scene. The seat is moved to another position beginning after 8 seconds and ending 20 seconds after start of measurement. The color represents the amount of reflected energy at a given distance (ordinate) and measurement time (abscissa). (b) Static background subtraction applied on the dynamic scene: As long as the scene stays static (first 8 seconds) antenna crosstalk and reflections caused by the car interior are completely removed.

The influence of static background subtraction on a dynamic scene is illustrated in figure 1(a) and (b). At first the scene is static and the direct crosstalk between the antennas and the reflections caused by the seat (see fig. 2) can be seen. Starting at second 8 the seat is moved and gets to its final position after 20 seconds. To apply a static background subtraction a reference frame is taken at the beginning of the measurement. For the static part of the scene the results are correct. But when the seat is moved a 'shadow' of the seat gets visible at its former position and the reflections of the seat itself gets also visible at its new position. So the whole background is now seen as foreground and besides that there is an erroneous shadow of the seat.

In this paper an approach is presented in which human vital functions are used as main indicators for locating passengers, optimized for an application in cars.

## 2 Ultra Wideband Basics

An UWB radar is comparable to an impulse radar and is defined by a minimum absolute bandwidth which is greater than 500 MHz and a relative bandwidth



which is greater 20 %, see Thomä *et al.* [2] for further details. The wide bandwidth allows the creation of extremely short pulses. Thereby a very good spatial resolution can be achieved for localization applications. Further applications are communication (wireless USB [11]) or health care [6].

These short pulses are sent out through directional antennas and after the signal is reflected at dielectric borders the received signal is cross correlated with a reference signal.

The amplitude of the cross correlation function represents the amount of energy which is reflected at a dielectric border. By using the time of flight and speed of the signal the position of objects causing reflections can be determined. Besides its good spatial resolution the wide bandwidth allows good penetration of material, e. g. human beings, clothing or blocking objects such as cases or backpacks [4], contrary to camera based systems.

### 3 System Setup

This study for passenger detection in vehicles is based on a UWB demonstrator with characteristics as shown in table 1.

**Table 1.** Hardware specification. The high spatial resolution is archived by using swept threshold and stochastic resonance sampling as described by Hjortland *et al.* [10].

Parameter	
Bandwidth	4.8 GHz
Center frequency	3.4 GHz
Range resolution	4 mm
Sampling frequency (whole scan)	30 Hz
Antennas	Bow-tie-geometry, 1 Rx, 1 Tx
Interface	USB

The antennas were mounted on the top of the dashboard inside a car in front of the passenger’s seat.

Our study focuses on two kinds of test scenarios. In one case a person was entering the car with the engine turned off. The person sat down on the passenger seat for several minutes. During the test the seat was moved and the person was bending forward at certain point in time. Finally the back rest was lowered to an extreme position (see fig. 3). In the other case a person was sitting on the passenger seat, driving through town with different kinds of road surface including rough, copped and smooth road. The seat position was not altered in this case, but the person was also moving (see fig. 4).

The region of interest was set to 1.15 meter in front of the antennas, which were mounted on the dashboard, so that the front of the car and therefore the area important for airbag control is covered.

## 4 Approach

Instead of separating fore- and background, the whole scene including reflections of a person as well as the reflections of the car's interior is analyzed. A background subtraction is not applied at first.

When the signal enters the human body a significant part of the energy is reflected due to its high relative permittivity [6]. The distance between body and antenna  $d(t)$  alters periodically around a static value  $d_o$ , since vital signs cause periodic movement  $m(t)$  of the body surface:

$$d(t) = d_o + m(t) = d_o + m_b \cdot \sin(2\pi f_b t) + m_h \cdot \sin(2\pi f_h t) \quad (1)$$

where  $m_b$  and  $m_h$  are the displacements caused by breathing and heartbeat with their frequencies  $f_b$  and  $f_h$  [7].

Since the amplitude of the heartbeat is very small, it is hard to identify with the proposed mechanical setup. While driving the signal generated by heartbeat is often overlapped by movement of the car interior. Since it is not robust enough for the first step of vital sign localization, heartbeat is not used for analysis. With additional bandwidth and further data processing it might be possible to detect the heartbeat reliably inside a car.

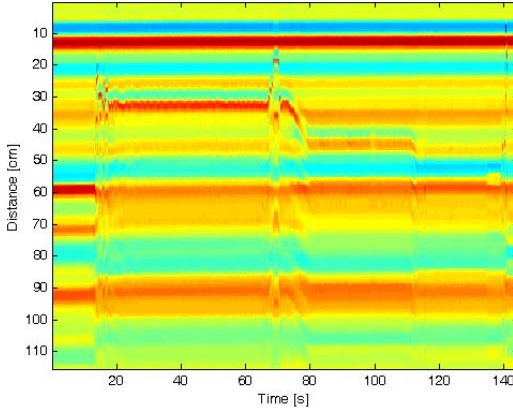
The typical frequency and amplitude for breathing can be assumed to be in certain ranges. For example the frequency for the respiration of a relaxed adult is typically between 0.2 and 0.33 Hz. Depending on the physical strain the frequency can raise to slightly more than 1 Hz [8]. This allows discrimination between breathing and other movements within the scene.

To analyse the influence of movement of the environment, several scenes in which the car was driving with an empty passenger seat were analysed. The Fourier transform showed that the amplitudes in the range of the breathing frequency were insignificant.

The received reflections are appended in a vector depending on their runtime which represents a scan over space. The time of flight of the UWB-signal



Fig. 2. Antenna setup mounted in a passenger's seat



**Fig. 3.** Raw data scan of a passenger sitting in a standing vehicle. The person enters the car at approximately 18 seconds and leaves it at 140 seconds.

represents a distance in space inside the vehicle. It was calculated by using the travelling speed of the signal in air. To evaluate the data over a given period the scans are appended which results in a matrix where the ordinate represents space and the abscissa represents time. To illustrate this, an UWB example scan over time has already shown in figure [1](#)

It is necessary to receive data over a certain period of time to give a reliable response about the breathing movement. Assuming the minimal breathing frequency of 0.25 Hz and 2 periods of movement as a minimum, 8 seconds have to be recorded. With a sampling frequency of 30 Hz, 240 scans have to be recorded to cover the necessary time window. For better usability concerning the Fourier transform the number of scans is increased to the next power of two. So one time window consists of 256 scans which cover 8.53 seconds. This window shifts every time a new scan is received.

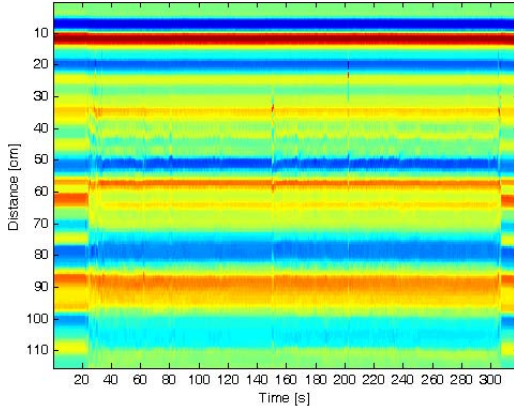
The window size is a trade off between being fast and being robust. The smaller the window the faster is the system, but it is also more sensitive to disturbing influences. A wider window makes it easier and more robust to identify breathing, because more periods of breathing movement can take place in one window. But in this case more time is needed before breathing is found and strong body movement will disturb the system longer since the movement is longer present in data. A shifting window is shown in fig. [5](#) and the result after a Fourier transform over time in fig. [6](#).

The reflections received include static background  $r(\tau)$  which is not dependent on long time, disturbing movement  $m(t, \tau)$  and the movement  $b(t, \tau)$  caused by breathing.

$$x(t, \tau) = n(t, \tau) + b(t, \tau) \quad (2)$$

where

$$n(t, \tau) = r(\tau) + m(t, \tau) \quad (3)$$



**Fig. 4.** Raw data scan of a passenger sitting in a driving vehicle. The person enters the car at approximately 23 seconds and leaves it at 320 seconds.

where  $t$  is the overall time while  $\tau$  is the runtime of the signal. The data of interest is  $b(t, \tau)$ . Since all parts of the signal have a different behaviour in frequency domain, a Fourier transform is applied over  $t$ .

$$X(\omega, \tau) = N(\omega, \tau) + B(\omega, \tau) \quad (4)$$

where  $N(\omega, \tau)$  and  $B(\omega, \tau)$  are the Fourier transforms of the noise and the breathing movement.

Assuming that there are also parts of the signal which are static over a period of 8.53 seconds, there is an offset in time domain which will result in an amplitude for the first data point in frequency domain. Direct cross talk between the antennas is one example for static parts of the signal. Setting this amplitude to zero in frequency domain will remove static background in time domain.

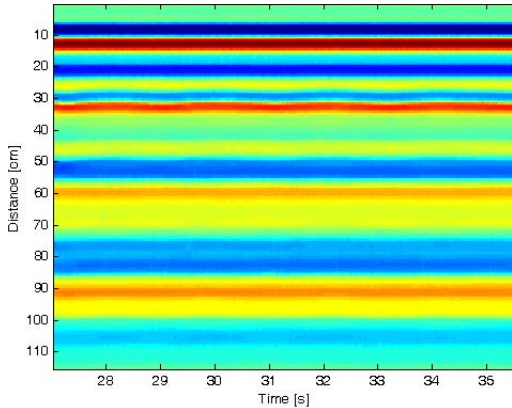
The same procedure is used to remove unwanted frequencies from the scene. As stated before breathing takes place in certain ranges of frequencies. Setting all other frequencies to zero will leave only the movement taking place in the desired frequency range.

After removing the unwanted parts of the signal an inverse Fourier transform is applied on the data.

$$\begin{aligned} m(t, \tau) &= \int_{-\infty}^{\infty} N(\tau) \cdot e^{i\omega t} \partial t \\ &+ \int_{-\infty}^{\infty} B(\tau) \cdot e^{i\omega t} \partial t \end{aligned} \quad (5)$$

Since  $N = 0$  as stated above, the resulting movement  $m(t, \tau)$  represents the movement of breathing.

$$m(t, \tau) = \int_{-\infty}^{\infty} B(\tau) \cdot e^{i\omega t} \partial t = b(t, \tau) \quad (6)$$



**Fig. 5.** Shifting window with a person sitting still and breathing

The significant part of the breathing movement takes place at the border between air and the human body. Also a significant part of the signal is reflected at this border because both mediums have a great difference in their relative permittivity [9]. As a result the breathing movement gets clearly visible at this border and thus, when extracting breathing movement the global maximum will arise at this border. Fig. 7 shows the result of the algorithm on the shifting window as presented in fig. 5. Let  $m(t, \tau)$  be a  $M \times N$  matrix with  $1 \leq o \leq M$  and  $1 \leq p \leq N$ .

$$P_{op} = \max \{m | a_{op} \in m\} \quad (7)$$

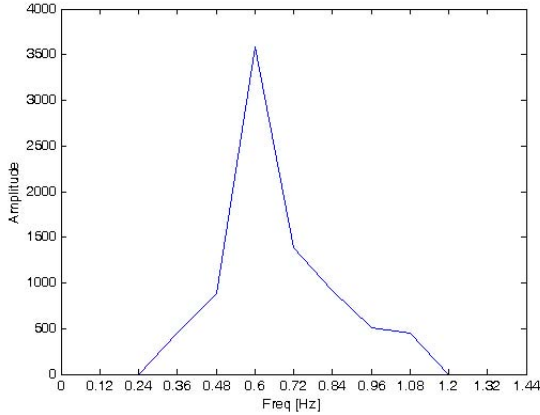
where  $o$  and  $p$  are the coordinates for the maximum breathing movement. By tracking the local maximum in the next scans, the location of the passenger's chest is extracted over time. The results of an example scan are illustrated in fig. 8(a) and (b).

## 5 Tests and Results

Overall 30 persons including children and person with different size and weight were recorded. In one case the car was driving through town, as described before. The system performance was evaluated with regard to the precision of the calculated location. The results were split into different classes:

- Correct layer
- Small error. Less or equal than 15 centimetres distance to correct layer
- Significant error. More than 15 centimetres distance to correct layer
- No detection

The correct layer is the first significant reflection showing breathing movement, representing the border between the human body and air in this scenario.



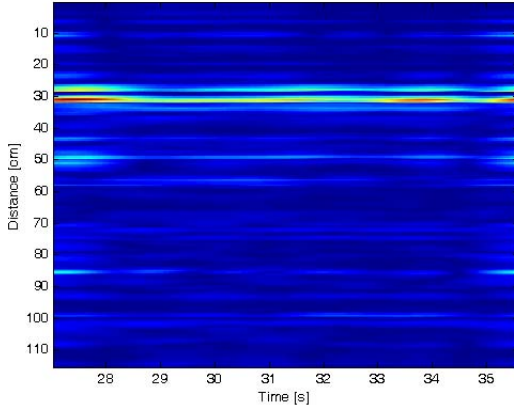
**Fig. 6.** Part of the Fourier transform showing the breathing frequency

In case of misuse by setting the backrest to an extreme angle, the angle between the person and the send out signal was increasing. As a result the quality of the reflected signal is decreasing significantly.

The angle between the direction of the signal and the person's body is no longer approximately  $90^\circ$  as before and so a big part of the signal is no longer reflected directly to the antennas. Additionally the distance between the antennas and the person is increased slightly, which decreases the amount of energy which is received by the antennas. Even more considerable is the fact that the breathing movement does no longer take place in horizontal direction only. A part of it is now also vertically directed, which is not captured by the radar system, because the direction of the signal is horizontal. These influences make it very hard to identify breathing with the proposed setup when the angle of the backrest exceeds a certain value. Considering this the evaluation data was split up into 'normal' case and 'unusual' case with extreme positioning of the backrest.

As stated before the algorithm works with windows including 256 scans. So the algorithm needs 8.53 seconds to initialize. Before locating a person the frequency space is monitored over several scans, so that not the first window with the right amplitude and the correct frequency is interpreted as breathing. In all cases an initialization time of 12 seconds is used.

After applying the algorithm on all 30 test cases the results shown in table 2 were achieved. One promising result is, that there was no localisation of breathing with an error greater than 15 centimetres. If the breathing could not be located due to heavy body movement or other disturbing influences, the class was set to no detection instead of locating an uncertain position. In two cases the body could not be located at all due to heavy body movement. In two other cases it took more than one minute until the person calmed down and the breathing was located. In all other cases the detection rate is near 100 percent. As a result, there was no detection in 11.28 percent of all measurements.



**Fig. 7.** Extracted breathing movement after inverse Fourier transform. Antenna crosstalk, reflections caused by the vehicle interior and moving objects such as the passenger seat have been removed and human breathing gets clearly visible at a distance of approximately 31 centimetres.

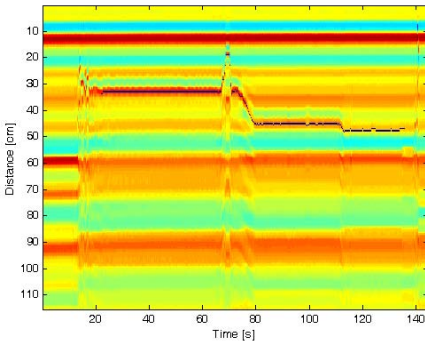
**Table 2.** Test results with regular (a) and unusual (b) back rest setup

Class	%
Correct	76.23
Small Error	12.48
Small Error + Correct	88.71
Significant Error	0
No detection	11.28

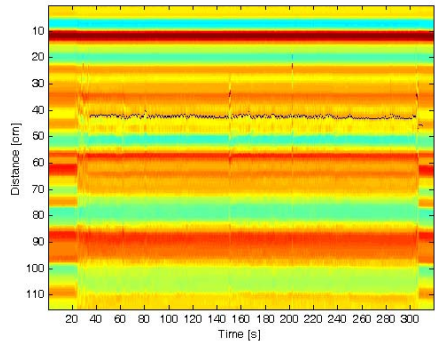
(a)

Class	%
Correct	17.28
Small Error	48.81
Small Error + Correct	66.09
Significant Error	0.49
No detection	33.42

(b)



(a)



(b)

**Fig. 8.** Algorithm result: The tracked person, indicated by the blue line, in a standing (a) and a driving (b) car

If breathing was detected the localization worked precisely. In 76.23 percent of all measurements, the correct body layer was located. In 12.48 percent a reflection in the range of 15 centimetres or less away from the correct reflection was located. The reason is that the breathing movement causes shadows, which also circulate with breathing frequency. In very few cases this shadow is located, resulting in a small error.

As expected the system performance for the setup with a greater angle between the seat and the back rest differs from the results for the initial setup. The overall performance drops due to the weaker signal received by the antenna. The detailed results are shown in table 2.

To sum up the results it can be said, that the system and the algorithm show a good opportunity to locate a person inside a car in case of normal use. For the presented setup a lowered back rest significantly reduced the power of the received signal, which led to uncertainties in analysis. In case of a weak signal or strong body movement the algorithm does not produce false positives.

If the person was stable and the setup of the seat was 'usual' the system showed very good performance and the breathing movement was located precisely.

## 6 Conclusion and Outlook

The discussed UWB system shows promising performance on locating a person inside a vehicle. In contrast to other approaches, the presented algorithm allows localization of vital signs in dynamic surroundings such as the interior of a car. The percentage of the location of correct layers at least with a small error is near 90 percent. The rest of the measurements represents no detection. There was no localization with an error greater than 15 centimetres.

Limitations of the system got clear when the back rest was lowered. The amount of the received energy decreased significantly and also the movement of respiration was directed unfavourable. In this case the rate for detection and location of breathing was lowered.

There still is potential to improve the algorithm. An additional feature to decide if there is a human being present could be the signature of the signal around the point where breathing was located. This would make the algorithm less sensitive to movement which has a similar frequency and amplitude as breathing.

Furthermore there is an opportunity to use multi channel setups, to give detailed information about subjects in the car. Particularly when considering additional airbags, not only the distance between the person and the dashboard but also the distance to the side airbags is critical. This is subject of further research.

## References

1. Yoon, J.J., Koch, C., Ellis, T.J.: Vision based occupant detection system by monocular 3D surface reconstruction. In: IEEE Conference on Intelligent Transportation Systems, Washington, USA (October 2004)



2. Thomä, R.S., Hirsch, O., Sachs, J., Zetik, R.: UWB Sensor Networks for Position Location and Imaging of Objects and Environments. In: ISAP 2007, Niigata, JP (2007)
3. Harker, B.J., Chadwick, A.D., Harris, G.L.: Ultra-Wideband 3-dimensional Imaging (UWB 3D Imaging). Roke Manor Research Limited, UK (2008)
4. Nezirovic, A., Yarovoy, A.G., Ligthart, L.P.: Experimental study on human being detection using UWB radar. In: IRS 2006, Poland (2006)
5. Allen, B., Dohler, M., Okon, E.E., Malik, W.Q., Brown, A.K., Edwards, D.J.: Ultra-wideband Antennas and Propagation for Communications, Radar and Imaging. John Wiley & Sons, West Sussex (2007)
6. Staderini, E.M.: UWB radars in medicine, Aerospace and Electronic Systems Magazine. IEEE (2002)
7. Lazaro, A., Girbau, D., Villarino, R.: Analysis of Vital Signs Monitoring using an IR-UWB Radar, Progress in Electromagnetics Research, PIER 100, Cambridge, Massachusetts (2010)
8. McArdle, W.D., Katch, F.I., Katch, V.L.: Exercise Physiology: Energy, Nutrition and Human Physiology. Lippincott Williams & Wilkins (2006)
9. Gabriel, C.: Compilation of the dielectric Property of Body Tissues at RF and Microwave Frequencies. King's College London, London (1996)
10. Hjortland, H.A., Wisland, D.T., Lande, T.S.: Thresholded samplers for UWB impulse radar. In: IEEE International Symposium on Circuits and Systems, ISCAS 2007, New Orleans, Louisiana (2007)
11. Aiello, R., Batra, A.: Ultra wideband systems: technologies and applications, Newnes, Burlington, USA (2006)

# Spectral Density Analysis: Theta Wave as Mental Stress Indicator

Saidatul Ardeenawatie Awang<sup>1</sup>, Paulraj Murugesu Pandiyan<sup>1</sup>, Sazali Yaacob<sup>1</sup>,  
Yusnita Mohd Ali<sup>1</sup>, Fadzly Ramidi<sup>2</sup>, and Fauziah Mat<sup>1</sup>

<sup>1</sup> School of Mechatronic Engineering, University Malaysia Perlis, 02600,  
Arau, Perlis, Malaysia

<sup>2</sup> Hospital Tuanku Fauziah, Jalan Kolam, 01000, Kangar, Perlis, Malaysia  
saidatul@unimap.edu.my

**Abstract.** This paper aims to investigate the correlation between mental tasks and the emanated theta and alpha brain waves. Electroencephalogram (EEG) signal is a highly random, complex and non linear in nature. The analysis of Electroencephalogram remains problematic due to the limited understanding of the signal origin. In spite of these shortcomings, the EEG is valuable tool to evaluate the overall cerebral activity. Advanced signal processing technique is required to extract hidden features carried by the signal. These features are useful to determine the mental state condition. Analysis was carried out using five adult right handed subjects, healthy and non smokers. Subjects were asked to perform mental arithmetic (MA) task in a period of time to induce the mental stress. To ensure the subjects performed all the MA tasks with full effort, a honorarium was given for every correct response. Using Burg's Method, power spectral density (PSD) were extracted from the emanate brain signal. Statistical analysis using correlation was performed and a positive relationship ( $r=0.625$ ,  $p<0.05$ ) has been observed between the highest PSD theta waves to the mental stress condition.

**Keywords:** EEG signal, Mental Arithmetic, Burg's Method, Statistical Analysis.

## 1 Introduction

The electroencephalogram (EEG) is a highly complex signal. It is one of the most common sources of information used to study the brain function and brain condition. However, the large amount data recorded by EEG machine for encephalographic changes and detection of the changes for EEG waves by visual inspection is an impossible task. Computer is the most suitable tool and thus, automated system to recognize the EEG behavior and pattern have been under study for several years. There is a high demand to the automated system since the analysis of the EEG signal can be long term recording and a large amount of data is also involved for each session.

Stress is well known as an important factor that induces various diseases. Recent papers reported that a strong stress factor can cause not only reversible but also irreversible biological symptoms such as cardiovascular diseases [1] and reducing hippocampus volume [2]. Besides that, it has been reported that stress and relaxation have

some effects on salivary amylase activity [3-4]. Most people react to stress in a similar body/physical manner (i.e changes in heart rate, skin conductance, blood volume). Previous studies using EEG signal have focused on spectral changes in sub-band frequency (alpha (8-13 Hz), beta (14-30 Hz), delta (0.5-4 Hz) and theta (4-8 Hz) [5]. In the EEG, increased EEG alpha wave activity has often been associated with an increased level of relaxation, otherwise it decrease during stress condition [6].

The entire process of methodologies developed for automated diagnosis can generally be divided into 4 stages namely data acquisition, preprocessing, feature extraction, and classification. The most crucial part is the experimental protocol. A variety of experimental mental stress model have been proposed and employed in the laboratory to examine the physiological response to mental stress [7-9]. The mental arithmetic primarily causes a beta adrenergic activation of the sympathetic nervous system and induces mental stress [11-12].

In preprocessing, elliptic bandpass filter is adopted to extract the signal into sub-band frequency namely alpha, beta, theta and delta. This study is focused on the probability of adopting alpha or beta waves as mental stress indicator. Previous literatures suggest that frontal midline (Fz) theta power is useful to assess mental stress during mental task and is a good indication of the subject concentration [12-14]. In feature extraction, parametric method (Burg Method) is applied to solve the spectral leakage problem due to windowing in non parametric method (Welch or FFT method). This method assumes the signal is a stationary random process in nature.

There are several types of classifier such as Artificial Neural Network (ANN), Support Vector Machine (SVM), Genetic Algorithmn (GA) and etc that can be adopted in order to classify the features of the signal. However, as a preliminary study, statistical analysis has been performed in this study. In order to show the relationship between changes of highest peak of PSD to stress state, correlation test has been adopted in this paper.

The rest of this paper is organized as follows. Section 2 briefly introduces methodology including subject preparation, experimental set up and experimental protocol. In Section 3, the EEG signal processing methods namely preprocessing and PSD feature extraction were discussed. Experimental results justifying the highest PSD method are reported in Section 4. Finally, conclusion of the work is given in Section 5.

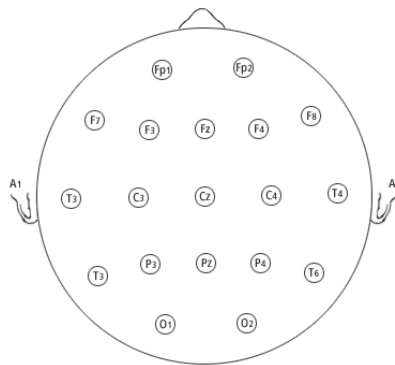
## 2 Experiment

### 2.1 Subject

Five healthy male subjects aged between 21 and 22 years old have participated in this experiment. All subjects are right handed and free from any health problems. Furthermore, exclusion criterion excluded: (i) left-handers (ii) smokers (iii) non native Malay speakers (iv) those with vision that non corrected to normal (v) those with high blood pressure problem (vi) those with exposure to general anesthesia in the last year. This criterion of subject selection is necessary in order to standardize the subject's condition.

## 2.2 Experiment Set Up

EEG signals were collected by using standard Mindset 24 Topographic Neuro Mapping Instrument by Nolan Computer Systems LLC. A 19 Electro cap channel was used to record brain signal. EEG signals were measured with the electrodes fixed on the scalp at 19 sites namely Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1 and O2 according to the international 10-20 system as shown in Figure 1. EEG signals were recorded by a digital EEG machine with a 16 bits A/D resolution. The signals were recorded in a quiet and dingy room where the subjects were asked to sit comfortably on a chair. The subjects were given 5-10 minutes to relax before starting the experimental procedures. The sampling frequency was set to 256 Hz and the impedances were kept below 5 kilo $\Omega$ . The cutoff range signal was set as -80 to 80.



**Fig. 1.** The position of electrodes according to 10-20 International System [15]

## 2.3 Experiment Protocol

In the beginning, the subjects were asked to fill in the demographic form along with Perceived Stress Scale (PSS) questionnaire. The subject is given 10 minutes to relax after filling the form and the questionnaire. Then, the subjects were asked to sit comfortably on the chair in front of a LCD computer screen. The distance between LCD screen and subject was 50 cm. The protocol used for the mental arithmetic task to induce mental stress is shown in Figure 2. In the baseline reading, the relax mood was induced by displaying a very calm picture. The baseline reading was recorded for 10 seconds before and after performing mental arithmetic task. During mental arithmetic stage, ten arithmetic questions were displayed on the screen. The subject needs to solve all ten arithmetic questions within 20 seconds and write the answer on a piece of paper. After this procedure, the subjects were asked to determine the highly stress stages namely before, during or after performing the whole trial. Furthermore, Likert Scale was used to rank the questions difficulty level. The scale is fixed as 1 to 5 where 1 denotes the very easy level and 5 denotes a very difficult level.

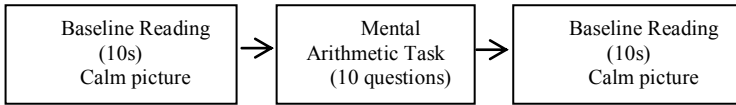


Fig. 2. The procedure of experiment

### 3 EEG Signal Processing

The signal processing involves main processes as namely preprocessing and feature extraction. Raw EEG signal consists of noise signals due to power line interference, electromyography (EMG), Electrocardiography (ECG), subject’s movement and etc. In order to filter out the noise signal, the preprocessing technique is applied. Features in filtered signal are then extracted to characterize the signal behavior. The process of data analysis is illustrated in Figure 3.

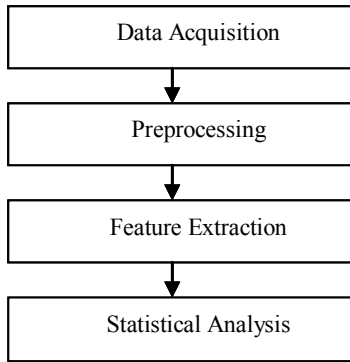


Fig. 3. Flowchart of signal processing method

#### 3.1 Preprocessing Method

Recorded raw EEG signals have amplitudes of the order of microvolts and contain frequency components up to 100 Hz. Hence, preprocessing is required which includes filtering, and artifact removals. Notch filter with 60 Hz cutoff frequency has been used to remove line noise. Different types of noise such as breathing frequencies, dc drift, and high frequency noise affect the EEG signal. An elliptic band pass filter is applied to extract the sub-band frequency components such as alpha, beta, theta and delta.

#### 3.2 Burg Method

Burg method is based on minimizing forward and backward prediction errors while satisfying the Levinson Durbin recursion [16]. This method calculates the autocorrelation

function and reflection coefficient can be estimated directly. Burg method ensures a stable AR model and is computationally efficient. The accuracy for Burg methods is lower with high order, long data records and high signal to noise ratio (SNR). The spectral density calculated by Burg Method is also susceptible to frequency shifts (relative to the true frequency) resulting from the initial phase of noisy sinusoidal signals. This effect is magnified when analyzing short data sequences.

Power Spectral Density by using Burg Method can be calculated using the following equation:

$$P_{xx}^{BU}(f) = \frac{\hat{e}_p}{|1 + \sum_{l=1}^p \hat{a}_p(l)e^{-2jfl}|^2} \quad (1)$$

where  $\hat{a}_p(l)$  is the estimates of the AR parameter obtained from the Levinson-Durbin recursion and  $\hat{e}_p$  denotes the reflection coefficients in an equivalent lattice structure, which are chosen to obtain total least mean square error and it is the summation of forward and backward prediction errors.

$$\hat{e}_{p,i}(n) = x(n) + \sum_{i=1}^p \hat{e}_{p,i} x(n-1), n = p+1, \dots, U \quad (2)$$

$$\hat{e}_{b,p}(n) = x(n-p) + \sum_{i=1}^p \hat{e}_{p,i} x^*(n-p+i), n = p+1, 2, 3, \dots, U \quad (3)$$

The order of model,  $p$  was chosen based on the Akaike Information criterion (AIC) figure of merit [17-18].

### 3.3 Akaike Information Criterion

For both Yule Walker and Burg Method, order of model,  $p$  was chosen as the one that minimizes AIC figure. The equation of AIC is derived in equation 4:

$$AIC(p) = N \cdot \ln(\hat{\lambda}^2) + 2p \quad (4)$$

where  $N$  denotes the number of samples and  $\hat{\lambda}^2$  is the estimated white noise variance. To reduce the computational complexity, we assumed as optimal the value of  $p$  that fulfilled the AIC criterion in the first two epochs. In EEG signal modelling, the  $p$  value can be taken as  $p=20$  [19].

In general, the power of the EEG signal will decrease with the increasing of the frequency and increase with the decreasing of the frequency. It is known that excitation process will be strengthen when the wave is transformed from high amplitude slow wave to low amplitude fast wave. On the other hand, inhibitory process will be enforced when the low amplitude fast wave is transformed to high amplitude slow wave. For this study, highest PSD's peak for pre, during and post was considered for every subject.

### 4 Result and Discussion

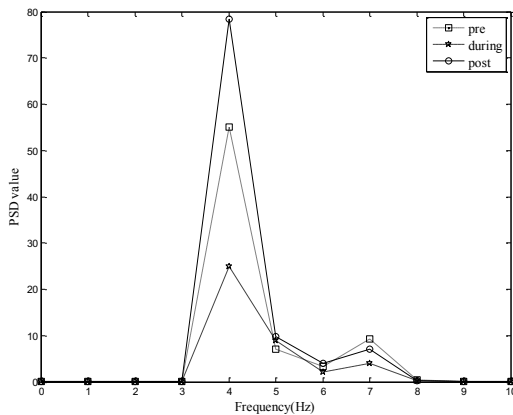
Figure 4 and 5 show the graph plotted highest peak of PSD value for theta and alpha waves, respectively, of the Subject 1. Based on the Figure 4, the highest PSD value for theta was occurred after performing mental arithmetic task, followed by before performing task and the lowest was occurred during task performance. Theoretically, alpha waves acts in opposite direction against theta waves. Alpha wave will increase when the people in relaxation mood and decrease when in stressed condition [6]. However, PSD value for alpha waves in Subject 1 does not show the opposite reaction against theta waves. Based on the self- reported subject evaluation, the subject notified that high stress was occurred after performing the task. The results for all subjects are summarized in Table 1.

**Table 1.** The highest PSD peak value (theta and alpha waves) for five subjects before, during task performance and after task performance

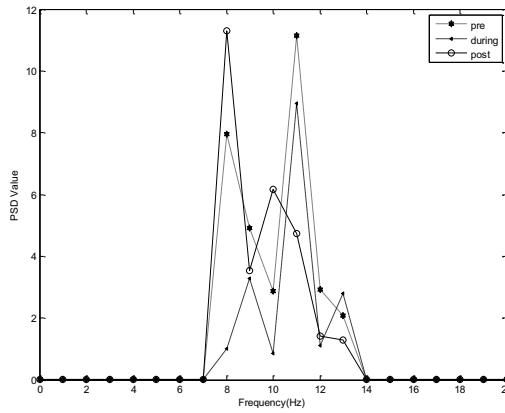
Subject	Theta waves				Alpha waves				Highest stress state
	Before	During	After	Percentage of changes during to post	Before	During	After	Percentage of changes during to post	
1	55.17	24.92	78.38	68.20 <sup>a</sup>	24.61	6.37	4.80	32.71 <sup>b</sup>	After
2	78.80	128.36	97.90	31.22 <sup>b</sup>	2.55	5.96	6.84	12.86 <sup>a</sup>	During
3	46.47	4.89	25.01	80.46 <sup>a</sup>	3.61	10.70	1.61	25.29 <sup>b</sup>	Before
4	13.88	25.93	12.99	99.64 <sup>b</sup>	3.70	2.15	1.47	46.26 <sup>b</sup>	During
5	12.54	33.20	87.11	88.43 <sup>a</sup>	11.14	8.94	11.30	20.88 <sup>a</sup>	After

<sup>a</sup> The increment of percentage changes during task performance to after task performance

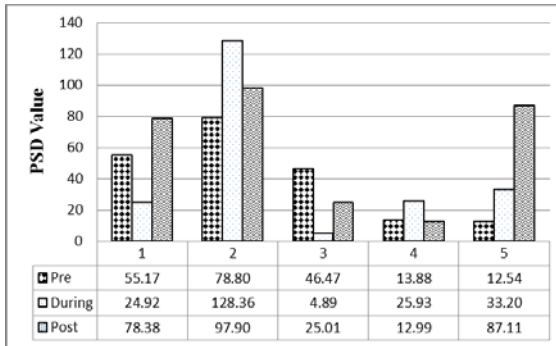
<sup>b</sup> The decrement of percentage changes during task performance to after task performance



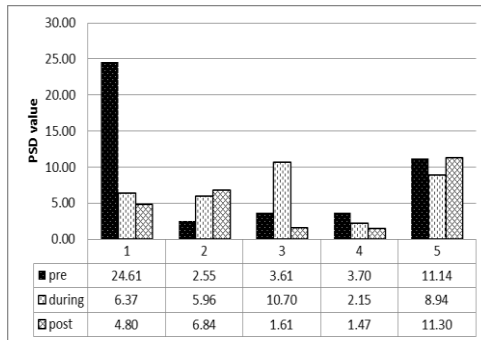
**Fig. 4.** PSD Value for theta waves of Subject 1 (pre, during and post)



**Fig. 5.** PSD Value for alpha waves of Subject 1 (pre, during and post)



**Fig. 6.** PSD value theta waves for 5 subjects



**Fig. 7.** PSD value alpha waves for 5 subjects



**Table 2.** Correlation between highest peak of PSD theta waves changes prior to stress state

		Changes	State
Changes	Pearson Correlation	1	0.625
	Sig. (1-tailed)		0.013
	N	5	5
Stress State	Pearson Correlation	0.625	1
	Sig. (1-tailed)	0.013	
	N	5	5

**Table 3.** Correlation between highest peak of PSD alpha waves changes prior to stress state

		Changes	State
Changes	Pearson Correlation	1	0.280
	Sig. (1-tailed)		0.324
	N	5	5
Stress State	Pearson Correlation	0.280	1
	Sig. (1-tailed)	0.324	
	N	5	5

Based on Table 1, it can be summarized that subject had various reactions to mental stress. In subject 3, he claimed that high stress was occurred before performing mental arithmetic. This is because of his anxiousness towards the type of questions that might be answered. Two subjects (Subject 2 and 4) show the highest peak of PSD occurred while performing the mental arithmetic. Both of them had notified that they felt highly stressed while solving mathematical questions. Furthermore, two subjects (Subject 1 and Subject 5) have recorded highest PSD’s peak after performing mental arithmetic task. By referring to their self-reported evaluation, high stressed was occurred after performing MA task. In order to show the correlation between highest PSD’s peak of theta and alpha waves to the experimental stages (pre, during and post), statistical analysis using Pearson correlation have been performed (Table 2 and Table 3).

A bivariate correlation was undertaken between the highest peak changes for theta waves and the stress state (pre, during and post). It was hypothesized that a positive relationship exists between these two variables. Result of the correlation indicate that changes in highest peak of theta wave are associated with stress state ( $r=0.625$ ,  $p<0.05$ ). The strong relationship between these variables ( $r>0.8$ ,  $p<0.05$ ) cannot be achieved since the size of the sample is too small (Table 2). On the other hand, result of the correlation indicate that changes in highest peak of alpha wave are associated with stress state ( $r=0.280$ ,  $p>0.05$ ). Based on the  $r=0.280$  which is less than 0.5, it shows that a very weak relationship exists between PSD alpha waves to stress state.

## 5 Conclusion

Theta waves is the potential mental stress indicator, hence the statistical results have shown the existing of positive relationship between stress and PSD value. Advanced signal processing technique is acquired in order to extract the features from the signal which carried 'stress' behavior. The reaction of the mental stress is various for every subject due to the different history background.

## References

- [1] Kazuomi, G.P.T.K., Bruce, S.M.: Disasters and the heart: a review of the effects of earthquake-induced stress on cardiovascular disease. *Hypertens. Res.* 26(5), 355–367 (2003)
- [2] Bremner, R.B.I.D., Randall, P., Scott, T.M., Bronen, R.A., Seibyl, J.P., Southwick, S.M., Delaney, R.C., McCarthy, G., Charney, D.S.: MRI-based measurement of hippocampal volume in patients with combat-related posttraumatic stress disorder. *Am. J. Psychiatry* 152(7), 973–981 (1995)
- [3] Groza, P., Zamfir, V., Lungu, D.: Postoperative salivary amylase changes in children. *Rev. Roum Physiol.* 8(4), 307–312 (1971)
- [4] Lewis, R.S., Weekes, N.Y., Wang, T.H.: The Effect of Naturalistic Stressor on Frontal Asymmetry, Stress and Health. *Biology Psychology* 75, 239–247 (2007)
- [5] Trejo, L.J., Kochavi, R., Kubitz, K., Montgomery, L.D., Rosipal, R., Matthews, B.: EEG-based estimation of cognitive fatigue. In: *Proceedings of SPIE*, pp. 105–115 (2005)
- [6] Reisman, S.: Measurement of stress. In: *Proceedings of the IEEE 23rd Northeast, Bioengineering Conference*, pp. 21–23 (1997)
- [7] Lin, C.-T., Wu, R.-C., Liang, S.-F., Chao, W.-H., Chen, Y.-J., Jung, T.-P.: EEG-Based Drowsiness Estimation for Safety Driving Using Independent Component Analysis. *IEEE Transactions on Circuits and Systems-I:Regular Papers* 52(12), 2726–2738 (2005)
- [8] Brookhuis, K.A., Waard, D.D.: Monitoring drivers' mental workload in driving simulators using physiological measures. *Accident Analysis and Prevention* 42(3), 896–901 (2010)
- [9] Yang, Q., Jiang, D.: Cortical Synchrony Change under Mental Stress due to Time Pressure. *Data Processing (Bmei)*, 2004–2007 (2010)
- [10] Yu, X., Zhang, J., Xie, D., Wang, J., Zhang, C.: Relationship between scalp potential and autonomic nervous activity during a mental arithmetic task. *Autonomic Neuroscience: Basic & Clinical* 146(1-2), 81–86 (2009)
- [11] Boha, R., et al.: The acute effect of low-dose alcohol on working memory during mental arithmetic I. Behavioral measures and EEG theta band spectral characteristics. *International Journal of Psychophysiology* 73(2), 133–137 (2009)
- [12] Yamada, F.: Frontal midline theta rhythm and eyeblinking activity during a VDT task and a video game: useful tools for psychophysiology in ergonomics. *Ergonomics* 41, 678–688 (1998)
- [13] Kubota, A.H.Y., Sato, W., Toichi, M., Murai, T., Okada, T., Sengoku, A.: Frontal midline theta rhythm is correlated with mance of an attention demanding meditation procedure. *Cognitive Brain Research* 11, 281–287 (2001)
- [14] Sasaki, T.I.K., Tsujimoto, T., Nishikawa, S., Nishitani, N.: Frontal mental theta wave recorded simultaneously with magnetoencephalography and electroencephalography. *Neuroscience Research* 26, 79–81 (1996)

- [15] Jasper, H.H.: The ten-twenty electrode system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 371–375 (1958)
- [16] Subha, D.P., Joseph, P.K., Acharya, R., Lim, C.M.: EEG Signal Analysis: A Survey. *Journal of Medical Systems* 34(2), 195–212 (2008)
- [17] Akaike, H.: Fitting autoregressive models for prediction. *Ann. Ins. Stat. Math.* 21, 243–247 (1969)
- [18] Akaike, H.: A new look at statistical model identification. *IEEE Trans. Automat. Contr.* 17, 716–723 (1974)
- [19] Faust, O., Acharya, R.U., Allen, A.R., Lin, C.M.: Analysis of EEG signals during epileptic and alcoholic states using AR modeling techniques. *ITBM-RBM* 29, 44–52 (2008)

# Leaf Image Analysis towards Plant Identification

Debnath Bhattacharyya<sup>1</sup>, Tai-hoon Kim<sup>2,\*</sup>, and Gang-soo Lee<sup>3</sup>

<sup>1</sup> Information Technology Department,  
MCKV Institute of Engineering, Liluah,  
Howrah – 711204, India  
debnathb@gmail.com

<sup>2</sup> GVSA and School of Information Science  
University of Tasmania, Australia  
taihoonk@utas.edu.au

<sup>3</sup> Department of Computer Engineering  
Hannam University, Daejeon, Korea  
gslee@hnu.kr

**Abstract.** Plants can be classified and identified by naturally or artificially as per the botanists. Plants can be identified by their leaves also. There are different varieties of plants grown throughout the world. Their identifications are studied using various laboratory methods. The morphological and genetically characteristics are employed to classify different leafs as well as plants. However, the presence of wide morphological varieties through evolution among the various leaf cultivars made it more complex and difficult to classify them. Leaf structures play a very crucial role in determining the characteristics of a plant. The broad and narrow shaped leaves, leaf arrangement, leaf margin characteristics features which differentiate various leaf of a plant. In this paper, we propose the methods to identify the leaf using an image analysis based approach.

**Keywords:** image processing, pattern recognition, image segmentation, classification.

## 1 Introduction

The Plants are basically identified by their leaves. There are different varieties of trees grown throughout the world. Some are important cash crop. Some are used in medicine. The tree identification is very important in day to day life. Their identifications had been studied using various laboratory methods [2, 3, 4]. The morphological and genetically characteristics were employed to classify different leafs. However, the presence of wide morphological varieties through evolution among the various leaf cultivars made it more complex and difficult to classify them. Therefore manual identification as well as classification of these leaves is a tedious task. During the last few decades computational biologists have studied various diversities among leaf due to huge number of evolutionary changes [1, 5, 6]. Leaf structures play a very crucial role in determining the

---

\* Corresponding author.

characteristics of a plant. The broad and narrow shaped leaves, leaf arrangement, leaf margin etc. characteristics features which differentiate various leaf of a tree. This project proposed the methods to identify the leaf using an image analysis based approach.

The Leaf structure: To identify a leaf, first we have to know the structure of the leaf. The following is the description of the structure of the leaf. An example of a sample jpeg image of a leaf structure is shown in Fig. 1.

An entire leaf has the following parts:-

- a. The Leaf Apex: The Apex is the tip of the leaf.
- b. The Leaf Blade: The blade is the entire leaf unit. Sometimes this is made up of several smaller leaflets.
- c. The Leaf Margin: The margin is the term used to describe the edge of the leaf.
- d. The Leaf Base: Base is the name given to the part of the blade that is closest to the stem.
- e. The Petiole: The leaf parts that is closest to the stem. The Petiole is the thin stalk that connects the leaf blade to the stem.
- f. The Stipule: The leaf parts that is closest to the stem. Stipules are tiny leaf-like structures that may or may not be present at the base of the petiole.
- g. Petiolate and Sessile: Leaves which have petioles are called Petiolate. If a leaf has no petiole it is called Sessile.

Importance of Identifying Leaf:

Identify Plants: If we can able to identify leaf, we can easily able to identify plants.

Disease identification: The structure of the leaf will be changed, if the leaf is affected by some disease. So, we can quickly identify the disease of the plant [7, 8, 9]. If the disease is early detected then we can take the step to prevent the diseases that occur in the plant. In this way we prevent the production fall in agriculture system. In this way we can achieve green revolution which helps our country to grow rapidly.

Plant leaves classification: If we can able to identify leaf, we can easily able to classify the plants.

Pollen leaves classification: If we can able to identify leaf, we can easily able to classify the pollen [10]. It is found that pollen grain of some plant families is responsible for producing allergy or asthma of human being. So this type of pollen should be identified.

Seed classification: If we can able to identify leaf, we can easily able to classify the seed [11]. This classification helps the farmer very much.

Crop Productivity: If we can able to identify the leaves, it helps to identify the different varieties of seeds, diseases, changing the color of leaves in heat, etc. This helps in crop productivity.

Identification of individual crop plants in the field: If we can able to identify the leaves, it helps to identify of individual crop plants in the field and locating their exact position of the crop [12]. Thus it helps in modern automated farming.

**Recreation:** Many people enjoy gardening or home landscape work. They need to identify which plants are useful and which ones are weeds. Other people simply like being able to name the plants they see when they walk through a city park or down a country road.

**Food and Fiber:** It is useful to be able to recognize plants which are sources of food (fruits, vegetables, spices, etc.) and fiber (lumber, firewood, paper pulp, etc.). This ability can also be profitable.

**Career Skill:** The ability to identify plant materials is an important skill for people who are interested in horticulture, forestry, landscape architecture, parks and recreation, botany, and a number of other career areas.

## 2 Previous Works

In this paper several methods [9] for recognizing of leaf are proposed. They are described below:-

- Leaf Blade Shape
- Simple and Compound Leaf Grouping
- Leaf Arrangement
- Leaf Venation
- Leaf Margin
- In this paper we will discuss all the above points.

### 2.1 Leaf Blade Shape

The first step in identifying a plant is to look at its leaves and determine their shape, and grouping.

One way to identify a plant is by the shape of its leaf blade (Fig. 2). The main part of a leaf is the blade. There are nine major shape categories:

- Linear: Narrow, with parallel or nearly parallel sides, (Fig. 3).
- Lanceolate: Lance shaped, longer than broad and tapering to a point at the tip (Fig. 4).
- Oblong: Much longer than wide, with nearly parallel sides (Fig. 5).
- Elliptic: Shaped like an ellipse, wider at the center and tapering to equal tips (Fig. 6).
- Ovate: Egg-shaped, broader at the base than the tip (Fig. 7).
- Cordate: Heart-shaped, either the leaf shape or base (Fig. 8).
- Reniform: Kidney-shaped (Fig. 9).
- Spatulate: In the shape of a spatula or spoon (Fig. 10).
- Orbicular: Round (Fig. 11).

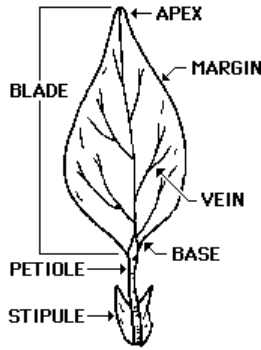


Fig. 1. Entire leaf structure

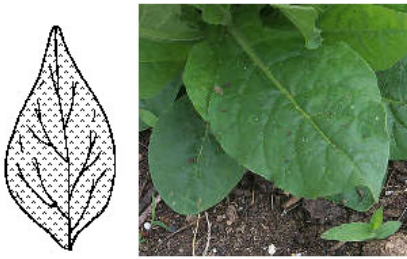


Fig. 2. Leaf blade shape



Fig. 3. Leaf with linear blade shape

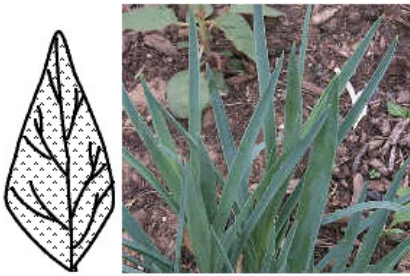


Fig. 4. Leaf lanceolate blade shape

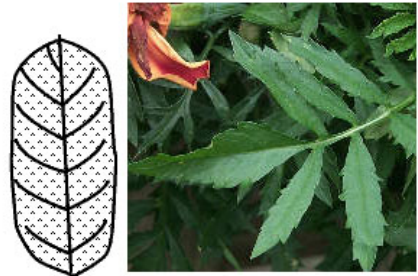


Fig. 5. Leaf with oblong blade shape

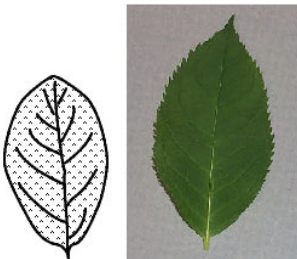


Fig. 6. Leaf elliptic blade shape

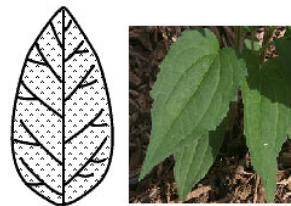


Fig. 7. Leaf with ovate blade shape

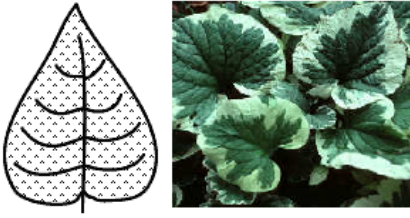


Fig. 8. Leaf cordate blade shape

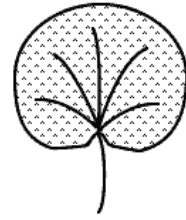


Fig. 9. Leaf with reniform blade shape



Fig. 10. Leaf spatulate blade shape

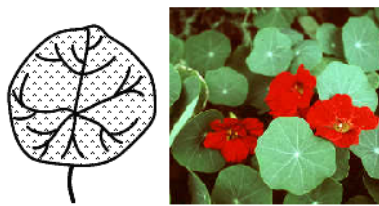


Fig. 11. Leaf with orbicular blade shape

## 2.2 Leaf Grouping

After determination of the leaf shape, we need to examine how the leaves are grouped on the stem. This is called Leaf Grouping. There are two types of leaf grouping.

**Simple Grouping:** Some leaves are simple, meaning they appear alone (Fig. 12). There are three main types of Simple leaf groupings as follows:

- Entire
- Palmately lobed
- Pinnately lobed

**Compound Grouping:** Other leaves are compound, meaning that they appear in groups and are made up of leaflets. Three kinds of compound leaves are as follows:

- Palmate: Here leaflets form a fan shape (Fig. 13).
- Pinnate: Leaflets are opposite each other on the stem (Fig. 14).
- Bipinnate: Leaflets are in pairs of pinnate groups (Fig. 15).

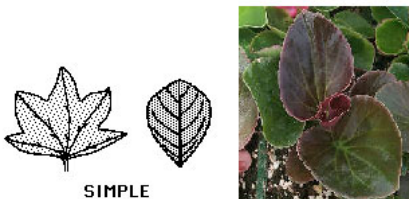


Fig. 12. Leaf with simple grouping



Fig. 13. Leaf with palmate compound grouping



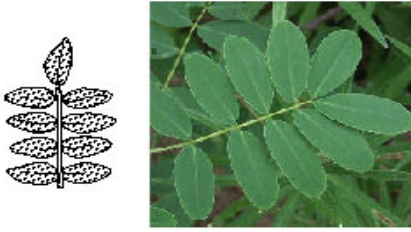


Fig. 14. Leaf with pinnate compound grouping

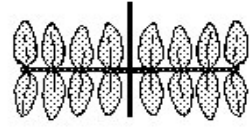


Fig. 15. Leaf with orbicular blade shape

### 2.3 Leaf Arrangement

After we have determined the shape and grouping of the plant's leaves, we need to examine the arrangement of the leaves on the stem. Here are the three principle ways leaves are arranged on the stem.

- Opposite: Leaves are directly opposite each other on the stem (Fig. 16).
- Alternate: Leaves are on both sides, but not directly across from one another (Fig. 17).
- Whorled: Leaves are arranged alternately around the stem.



Fig. 16. Opposite leaves arrangement



Fig. 17. Alternate Leaves arrangement

### 2.4 Leaf Venation

The lines that appear on the surface of a leaf and look like blood vessels are called Veins. The two main types are Palmate and Pinnate (Fig. 18 and Fig. 19, respectively).

### 2.5 Leaf Margin

After we have determined the shape and grouping of the plant's leaves, the arrangement of the leaves on the stem, and the structure of the leaf veins, we need to examine the leaf margin characteristics.

The shape of the edge or margin of a leaf is another way it can be identified. Three kinds of margins are Entire, Toothed, and Lobed.

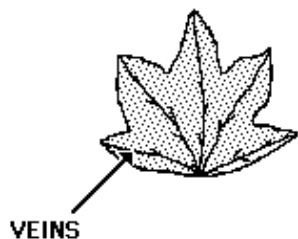


Fig. 18. Leaf with palmate leaf venation

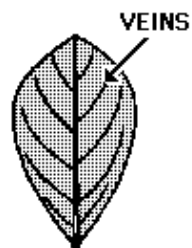


Fig. 19. Leaf with pinnate leaf Venation

**Entire Margins:** Margin smooth no bumps (Fig. 20).

**Toothed Margins:** A number of margins are named by the kinds of "teeth" they have (Fig. 21 and Fig. 22). Leaf Margins are as follows:

- Toothed: shallow bumps;
- Dentate: pointing outward;
- Serrate: pointing to the leaf's tip;
- Crenate: broad and round; and
- Incised: deeply cut with sharp, irregular teeth.



Fig. 20. Leaf with smooth margin



Fig. 21. Toothed leaf margin

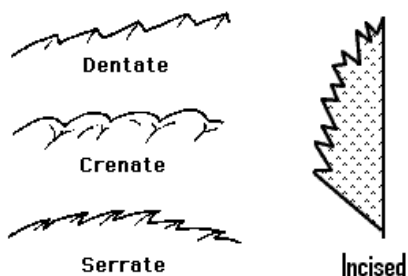
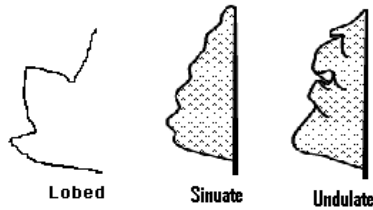


Fig. 22. Leaves with various types of toothed margins

**Lobed Margins:** Leaves with lobed margins are as follows (Fig. 23):

- Lobed: deep indents or sinuses;
- Sinuate: very wavy margin; and
- Undulate: wavy margin, but not as wavy as a sinuate margin.



**Fig. 23.** Leaves with lobed margins

### 3 Our Work

In this paper we applied some methods and algorithms to identify the leaf. These are explained as follows:

#### 3.1 Algorithms for Database Creation

The leaf samples are stored in the database. The images are in .bmp format. Here we consider only the leaf margins. There are nine types of leaf margins. These nine types of leaves are stored in the database. Algorithm is stated hereunder:

- Step 1: Start
- Step 2: List all .bmp files from the folder
- Step 3: Initialize the variable Numfids with total no. of listed .bmp files
- Step 4: Declare an array to store the loaded .bmp files
- Step 5: Read the .bmp files from the folder
- Step 6: Store the loaded .bmp file in an array for future use
- Step 7: Stop

#### 3.2 Algorithms for Preprocessing of Images in the Database

The leaf samples which are stored in the database are scaled in standard size and then converted into gray scale image. From these gray scale images edge are detected using 'sobel' technique. Algorithm is given hereunder:

- Step 1: Start
- Step 2: Initialize the variable Numfids with total no. of listed .bmp files
- Step 3: Read the images from the stored array
- Step 4: Scale the images with size 186,200
- Step 5: Convert the color images to gray scale images
- Step 6: Identify the edge of the gray images using sobel technique
- Step 7: Store the properties of the detected edge in an array
- Step 8: Stop

**3.3 Finding Out the Mathematical Characteristics of the Stored Image in the Database**

The leaf samples which are stored in the database edges are identified. These edges are logical in nature. The edges area will store '1' in the array. From that array we calculate the correlation coefficient for each image and store in another array. The correlation coefficient is calculated with the following formula mentioned below.

The correlation coefficient of x and y is:

$$\text{Correlation coefficient} = \text{cov}(x,y) / \sigma_x * \sigma_y$$

Where cov(x,y) is covariance of x and y,

$\sigma_x$  and  $\sigma_y$  are standard deviation of the values of x and y respectively.

$$\sigma_x^2 = 1/n \sum x_i^2 - (1/n \sum x_i)^2 \text{ and } \sigma_y^2 = 1/n \sum y_i^2 - (1/n \sum y_i)^2$$

The covariance of x and y is

$$\text{Cov}(x,y) = 1/n \sum_{i=1}^n (x_i - \bar{X}) (y_i - \bar{Y})$$

where  $\bar{X}$  and  $\bar{Y}$  are the arithmetic means of the value assumed by x and y respectively,

$$\bar{X} = 1/n(x_1 + x_2 + \dots + x_n)$$

$$\text{and } \bar{Y} = 1/n(x_1 + x_2 + \dots + x_n)$$

The algorithm is given hereunder:

- Step 1: Start
- Step 2: Initialize the variable Numfids with total no. of listed .bmp files
- Step 3: Read the array which store the detected edge properties of the stored images
- Step 4: Find out the correlation coefficient from the array
- Step 5: Store the correlation coefficient value in an array
- Step 6: Stop

**3.4 Algorithms for Preprocessing of Input Image and Mathematical Characteristic**

Same algorithms are used to process the input image for identification, from the algorithm stated in 3.2 to 3.3.

**3.5 Algorithms for Comparison of Characteristics of the Input Image with the Stored Images**

After the mathematical calculation we get the value of correlation coefficient of the stored images and the input images. Then we calculate the matching percentage with the following formula:

Percentage = (Correlation coefficient of each stored image/ Correlation coefficient of the input image)\*100;

The matching percentage if greater than 80% then we say 'image is fully matched'. If it is greater than 60 but less than 80 then we say 'image is partially matched' else we say that 'image is not matched'.

- Step 1: Start
- Step 2: Initialize the variable Numfids with total no. of listed .bmp files
- Step 3: Read the array which store the correlation coefficient value of the stored images
- Step 4: If Correlation coefficient of the stored image=correlation coefficient of the input image?  
Then Print “Image is found” Else Print “Image is not found”
- Step 5: Stop

**3.6 Another Algorithms for Comparison of Characteristics of the Input Image with the Stored Images**

- Step 1: Start
- Step 2: Initialize the variable Numfids with total no. of listed .bmp files
- Step 3: Read the array which store the correlation coefficient value of the stored images
- Step 4: Calculate Matching percentage= (Correlation coefficient of the individual stored image/correlation coefficient of the input image)\*100
- Step 5: Store the matching percentage in an array
- Step 6: Initialize the variable Per with the maximum matching value from the array which stores the matching percentage value
- Step 7: if per>=80 then Print “image is fully matched”
- Step 8: if (per>=60) && (per<80) Print “image is partially matched, Else Print “image is not matched”
- Step 9: Stop

**4 Result and Analysis**

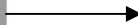
Table 1, shows the data of input file and Table 2 shows the detail experimental result. Fig. 24 to Fig. 27 shows the Algorithm output. Fig. 28 shows the Matlab implementation shot.

**Table 1.** Detail information of input leaf

Serial no	Name of input file	Value of the correlation coefficient of the input file
1.	toothed_large.bmp	-0.0911

**Table 2.** Detail of output information after matching

Serial no	Name of stored files	Value of the correlation coefficient of the stored file	Matching percentage with the input image file	Output
1.	Crenate.bmp	0.0891	-97.7624	image is not found
2.	Dentate.bmp	0.0619	-67.9869	image is not found
3.	Entire.bmp	0.0344	-37.7459	image is not found
4.	Irregular.bmp	0.1592	-174.8101	image is not found
5.	Lobed.bmp	0.0159	-17.4670	image is not found
6.	Serrate.bmp	0.1006	-110.4112	image is not found
7.	Sinuate.bmp	-0.1114	81.8079	image is not found
8.	Undulate.bmp	0.2544	-279.2548	image is not found
9.	Toothed.bmp	-0.0911	100	Image is matches up to 100 % image is fully matched



**Fig. 24.** Original Toothed Leaf

**Fig. 25.** Toothed Leaf after scaling

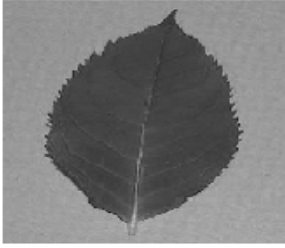


Fig. 26. Leaf after gray conversion



Fig. 27. After Edge Detection

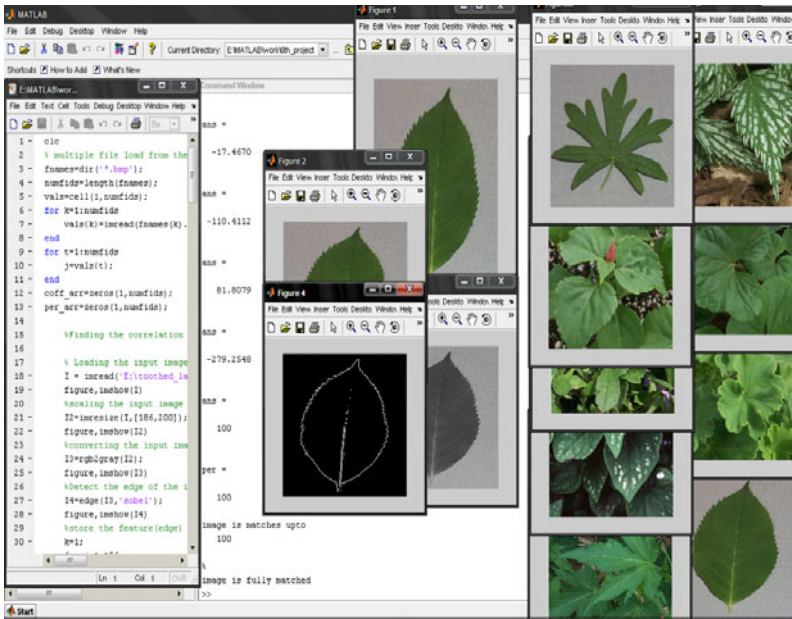


Fig. 28. Matlab Implementation Screen shot

## 5 Conclusion

This approach is useful in MLP classifier that helps in identification of various plant based on its morphological or genetically characteristics. Since the leaf structure varies in different cultivars, the outcome of the present work clearly shows the exact characteristics present in each variety of leaves. The figure of percentage required for the presence of similar characteristics of different leaves. For a more detailed and accurate classification other morphological features like fruits, stem, flower, and root can be considered for further processing.

**Acknowledgement.** This work was supported by the Security Engineering Research Center, granted by the Korea Ministry of Knowledge Economy.

## References

1. Sanyal, P., Bhattacharya, U., Bandyopadhyay, S.: Analysis of SEM Images of Stomata of Different Tomato Cultivars Based on Morphological Features. In: IEEE International Conference, AMS 2008 (2008)
2. Sabino, D.M.U., da, L., Costa, F., Rizzatti, E.G., Zago, M.A.: A texture approach to leukocyte recognition. *Real-Time Imaging* 10, 205–216 (2004)
3. Soille, P.: Morphological image analysis applied to crop field mapping. *Image and Vision Computing* 18(13), 1025–1032 (2000)
4. Stojanovic, R., Mitropoulos, P., Koulamas, C., Karayiannis, Y., Koubias, S., Papadopoulos, G.: Real-time vision-based system for textile fabric inspection. *Real-Time Imaging* 7(6), 507–518 (2001)
5. Polder, G., van der Heijden, G.W.A.M., Young, I.T.: Hyperspectral Image Analysis For Measuring Ripeness Of Tomatoes. In: ASAE International Meeting (2000)
6. Tzionas, P., Papadakis, E.S., Manolakis, D.: Plant leaves classification based on morphological features and a fuzzy surface selection technique. In: Fifth International Conference on Technology and Automation, Thessaloniki, Greece, pp. 365–370 (2005)
7. Zhao-yan, L., Fang, C., Yi-bin, Y., Xiu-qin, R.: Identification of rice seed varieties using neural network. *Journal of Zhejiang University SCIENCE*, 1095–1100 (2005)
8. Neuman, M., Sapirstein, H.D., Shwedyk, E., Bushuk, W.: Wheat grain color analysis by digital image processing: I. Methodology. *J. Cereal Sci.* 10, 175–182 (1989)
9. Damian, M., Cernadas, E., Formella, A., Sa-Otero, P.M.: Pollen classification of three types of plants of the family Urticaceae,  
<http://trevinca.ei.uvigo.es/~formella/inv/aaa/formella-2002-pollen.pdf>,  
<http://cas.psu.edu/docs/CASDEPT/Hort/LeafID/Arrangement.html>
10. Slaughter, D.C., Harrell, R.C.: Discriminating fruit in a natural outdoor scene for robotic harvest. *Trans. Of the ASAE* 32(2), 757–763 (1989)
11. Tian, L., Slaughter, D.C., Norris, R.F.: Machine Vision Identification Of Tomato Seedlings For Automated Weed Control,  
<http://www.age.uiuc.edu/faculty/lft/papers/tomato.pdf>
12. Sanyal, P., Bhattacharya, U., Parui, S.K., Bandyopadhyay, S.K.: Color Texture Analysis of Rice Leaves for Detection of Blast Disease. In: 20th CSI Conference Proceedings, pp. 45–48 (2006)



# Watermarking Using Multiresolution Cosine Transformation: A Review

Debnath Bhattacharyya<sup>1</sup>, Tai-hoon Kim<sup>2,\*</sup>, and Gang-soo Lee<sup>3</sup>

<sup>1</sup> Information Technology Department,  
MCKV Institute of Engineering, Liluah,  
Howrah – 711204, India  
debnathb@gmail.com

<sup>2</sup> GVSA and School of Information Science  
University of Tasmania, Australia  
taihoonk@utas.edu.au

<sup>3</sup> Department of Computer Engineering  
Hannam University, Daejeon, Korea  
gslee@hnu.kr

**Abstract.** In this paper, an image accreditation technique by embedding digital watermarks in images is proposed. The proposed method for the digital watermarking is based on the cosine transform. This is unlike most previous work, which used a random number of a sequence of bits as a watermark and where the watermark can only be detected by comparing an experimental threshold value to determine whether a sequence of random signals is the watermark. The proposed approach embeds a watermark with visual recognizable patterns, such as binary, gray, or color image in images by modifying the frequency part of the images. This thesis discusses the issues regarding data hiding and its application to multimedia digital information have also daily lives of the peoples in this security and communication, addressing both theoretical and practical aspects, and tackling both design and attack problems. Data hiding is modeled as a communication problem where the embedded data is the signal to be transmitted. Various embedding mechanisms target different robustness-capacity tradeoffs. The trade-off for different major categories of embedding mechanisms has been done. In this approach, an original image is decomposed into wavelet coefficients. Then, multi-energy watermarking scheme based on the qualified significant wavelet tree (QSWT) is used to achieve the robustness of the watermarking. Unlike other watermarking techniques that use a single casting energy, QSWT adopts adaptive casting energy in different resolutions. The performance of the proposed watermarking is robust to a variety of signal distortions, such as JPEG, image cropping, sharpening, median filtering, and incorporating attacks. These algorithms are covering applications such as annotation, tamper detection, copy/access control, fingerprinting, and ownership protection.

**Keywords:** Image authentication, watermark, security, private key, public key.

---

\* Corresponding author.

## 1 Introduction

Digital watermarking includes a number of techniques that are used to imperceptibly convey information by embedding it into the cover data. There has always been a problem in establishing the identity of the owner of an object in this digital world where the unlimited number of copies of the digital data can be produced with zero to no error. In case of a dispute, identity was established by either printing the name or logo on the objects. But in the modern era where things have been patented or the rights are reserved i.e. copyrighted, more modern techniques to establish the identity and leave it unhampered have come into picture. Unlike printed watermarks, digital watermarking is a technique where bits of information are embedded in such a way that they are completely invisible. Actually, digital watermarking exploits the limitation the human visual system (HVS). The problem with the traditional way of printing logos or names is that they may be easily tampered or duplicated. In digital watermarking, the actual bits are scattered in the image in such a way that they cannot be identified and show resilience against attempts to remove the hidden data [1].

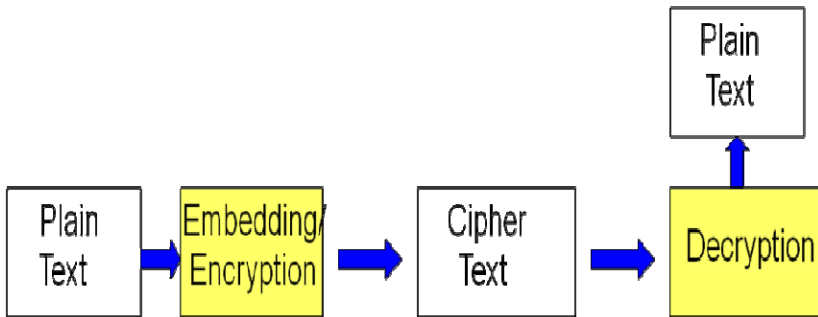
Digital Watermarking can be defined as a technology of embedding watermark with intellectual property rights into images, videos, audios and other multimedia data by a certain algorithm. This kind of watermark contains the author and the user's information, which could be the owner's logo, serial number or control information. In fact, it's making use of the ubiquitous redundancy and randomness in data, and adding to the data information which is difficult to be detected but can be distinguished to protect product copyright and data integrity. The difference between digital watermarking and other technology is of three important aspects: First of all, unlike encryption, watermarking is imperceptible so that the image will not be distracted from the aesthetic sense. Second, the watermarks and the cover data they in which they are embedded should be inseparable. Even if the data were displayed or converted into other file formats, the watermarks will not be eliminated. This is different from the first section. Finally, the watermarks will have exactly the same transformation experience as the data that means you can get the information of transformation by looking at the watermarks.

In addition to Digital watermarking, the general idea of hiding some information in digital content has a wider class of applications that go beyond mere copyright protection and authentication. The techniques involved in such applications are collectively referred to as information hiding. For example, an image printed on a document could be annotated by information that could lead an user to its high resolution version. Metadata provides additional information about an image. Although metadata can also be stored in the file header of a digital image, this approach has many limitations. Usually, when a file is transformed to another format (e.g., from TIFF to JPEG or to bmp), the metadata is lost. Similarly, cropping or any other form of image manipulation destroys the metadata. Finally, the metadata can only be attached to an image as long as the image exists in the digital form and is lost once the image is printed. Information hiding allows the metadata to travel with the image regardless of the file format and image state (digital or analog). Metadata information embedded in an image can serve many purposes. For example, a

businessman can embed the website URL for a specific product in a picture that shows an advertisement for that product. The user holds the magazine photo in front of a low-cost CMOS camera that is integrated into a personal computer, cell phone, or a palm pilot. The data is extracted from the low-quality picture and is used to take the browser to the designated website. Further, Before Understanding the watermarking it is very essential to understand the subtle difference between the most popular forms of data hiding techniques. Therefore, an attempt has been made to differentiate cryptography, Steganography and Watermarking as follows:

### 1.1 Cryptography

Cryptography as the study of secret (crypto) writing (graphy) can be defined as the science of using mathematics to encrypt and decrypt data back [2]. It allows two people, commonly known as Alice and Bob, to communicate with each other securely. This means that an eavesdropper known as Eve will not be able to listen in on their communication.



**Fig. 1.** Cryptography for secure communication

Cryptography also enables Bob to check that the message sent by Alice was not modified by Eve and that the message he receives was really sent by Alice [3]. A message is known as a plaintext or clear text. The method of disguising the plaintext in such a way as to hide its information is encryption and the encrypted text is also known as a ciphertext. The process of reverting ciphertext back to its original text is decryption, shown in Fig. 1 [4].

### 1.2 Steganography

While cryptography is about protecting the content of the messages, steganography is about concealing their very existence. Steganography comes from a Greek word that means covered writing (*stego* = covered + *graphy* = writing) [5]. Examples can be thought as messages exchanged between drug dealers via emails in encrypted forms, or messages exchanged by spies in covert communication. Steganography hides the fact that the communication ever occurred as shown in Fig. 2 [6].

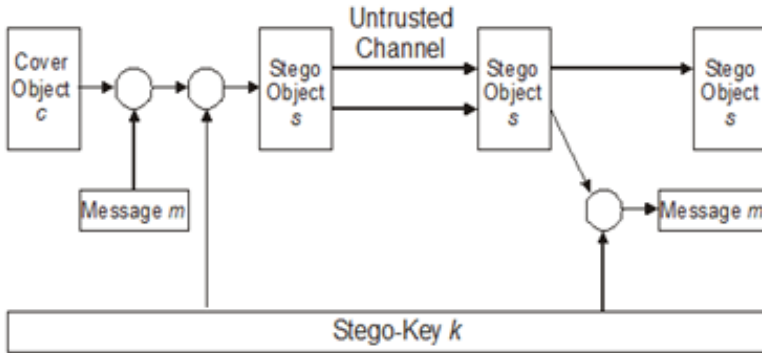


Fig. 2. General Stegnographic process

## 2 General Framework for Watermarking

Although steganography and watermarking both describe the techniques used for covert communication, steganography typically relates only to covert point to point communication between two parties [4]. Steganographic methods are not robust against attacks or modification of data that might occur during transmission, storage or format conversion [1]. Watermarking, as opposed to steganography, has an additional requirement of robustness against possible attacks. An ideal steganographic system would embed a large amount of information perfectly securely, with no visible degradation to the cover object.

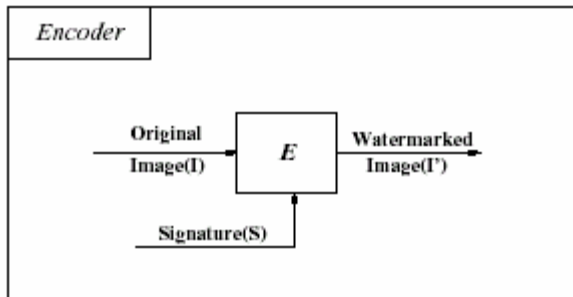


Fig. 3. General embedding process

For the embedding process the inputs are the watermark, cover object and the secret or the public key. The watermark used can be text, numbers or an image. The resulting final data received is the watermarked data  $I'$ . The inputs during the decoding process are the watermark or the original data, the watermarked data and the secret or the public key. The output is the recovered watermark  $I$ . Fig. 3 and Fig. 4 explain embedding and extraction of watermark [2].

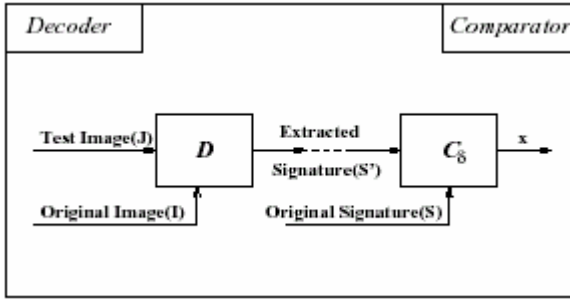


Fig. 4. Watermark extraction process

### 3 Basic Principles of Watermarking

Since this research field is still relatively young and has contributors from several disciplines with varying traditions, the terminology used is still quite diverse. This section provides a formal introduction to watermarking systems and the terms used in this context for their presentation.

The basic principle of current watermarking systems is comparable to symmetric encryption as to the use of the same key for encoding and decoding of the watermark. Each watermarking system consists of two subsystems: a watermarking encoder and a respective decoder. Formally, a watermarking system can be described by a tuple  $(O, W, K, EK, DK, C_\tau)$  where  $O$  is the set of all original data the set of all watermarks, and  $K$  the set of all keys. The two functions describe:

$$E_K: O \times W \times K = O \tag{1}$$

$$D_K: O \times K = W \tag{2}$$

the embedding and detection process, respectively. The comparator function:

$$C_\tau: W^2 \rightarrow \{0,1\} \tag{3}$$

Compares the extracted with the really embedded watermark using the threshold  $\tau$  for comparison. The input parameters of the embedding process are the carrier object (or original  $C_0$ ), the watermark  $w$  to be embedded, as well as a secret or public key  $K$ :

$$E_K(C_0, W) = C_0 \tag{4}$$

The output of the encoder forms the marked data set. Fig. 5 and Fig. 6 show the structure of generic watermark encoder and decoder respectively.

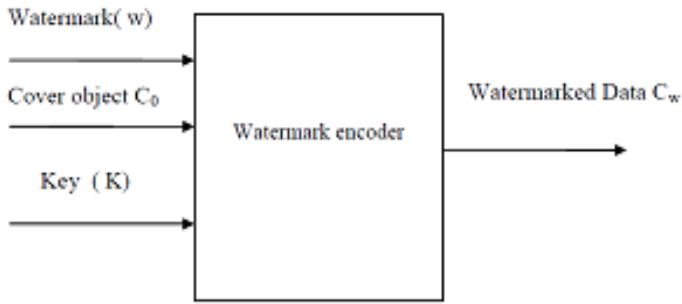


Fig. 5. Generic Watermark Encoder

In the detection process, the marked and possibly manipulated data set, the original  $co$ , the watermark, and the key  $K$  used during the embedding process form the maximal set of input parameters. The various types of watermarking systems differ in the number of input parameters in the reading process. The extracted watermark differs in general from the embedded watermark  $w$  due to possible manipulations. In order to judge the correspondence of both watermarks, the comparator function  $C_\tau$  compares the suspected watermark with the retrieved one against a threshold  $\tau$ :

$$C_\tau(\hat{w}, w) = \begin{cases} 1, & c \geq \tau \\ 0, & c < \tau \end{cases} \quad (5)$$

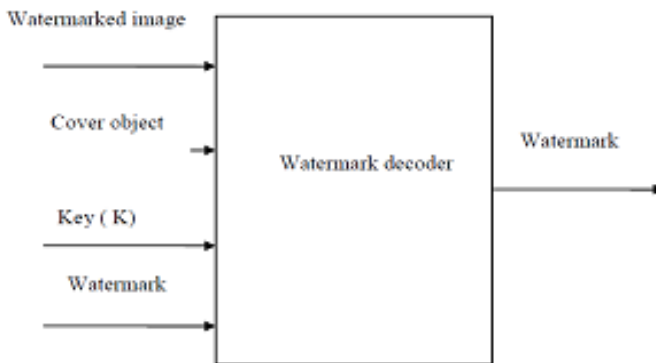


Fig. 6. Generic Watermark Decoder

The threshold  $\tau$  depends on the chosen algorithm and should in a perfect system be able to clearly identify the watermarks. This formal analysis of the watermarking systems can also be used to develop a geometric interpretation of the watermarking algorithms [7].

Using more information at the detector site increases the reliability of the whole watermarking system but limits the practicability of the watermarking approach on the embedder side. The side information in the detection process can be the original and the watermark  $w$  itself.

## 4 Digital Watermarking Techniques

### 4.1 General

The digital watermarking is technique of embedding the hidden information into original data. There may be different way of embedding the watermark (hidden information) into host data such as embedding in spatial domain, transform domain and fractional domain. The concept of digital watermarking consists of inserting information into the host signal under the condition that the modifications are not perceptible. In addition, it is desirable to put maximum energy into the watermark in order to achieve high robustness. This is a well known concept from communication theory: to decrease the error rate, the signal energy must be maximized. In mathematical formulation, the watermark embedding process can be considered as a constrained maximization problem i.e. to maximize the watermark energy under the visibility constraint. Although the problem is straightforward to formulate, it is extremely difficult to implement because of the visibility constraint, which is usually based on a non-linear model of the human visual system.

### 4.2 Spatial Domain Technique

In the spatial domain, watermarking the watermark is embedded by directly modifying the pixel values of the original image. The technique under this category is Least Significant Bit Substitution. Least significant bit (LSB) insertion is a common, simple approach to embedding information in a cover image. *Least significant bit (LSB) insertion* is a common, simple approach to embedding information in a cover file. Unfortunately, it is vulnerable to even a slight image manipulation. Converting an image from a format like GIF or BMP, which reconstructs the original message exactly (*lossless compression*) to a JPEG, which does not (*lossy compression*), and then back could destroy the information hidden in the LSBs. The least significant bit (in other words, the 8th bit) of some or all of the bytes inside an image is changed to a bit of the secret message. It can be understood by the following example when using a 24-bit image, a bit of each of the red, green and blue colour components can be used, since they are each represented by a byte.

### 4.3 Frequency Domain Technique

In frequency domain watermarking the cover object is first transformed in frequency domain and then the watermark is added into the frequency domain. Spatial domain watermarking has the advantage that they can be easily applied to any image; regardless of subsequent processing (whether they survive this processing however is a different matter entirely). A possible disadvantage of spatial techniques is they do not allow for the exploitation of this subsequent processing in order to increase the robustness of the watermark.

In addition to this, adaptive watermarking techniques are a bit more difficult in the spatial domain. Both the robustness and quality of the watermark could be improved if the properties of the cover image could similarly be exploited. For instance, it is generally preferable to hide watermarking information in noisy regions and edges of

images, rather than in smoother regions. The benefit is two-fold; Degradation in smoother regions of an image is more noticeable to the HVS, and becomes a prime target for lossy compression schemes. Considering these aspects, working in a frequency domain of some sort becomes very attractive. The classic and still most popular domain for image processing is that of the Discrete Cosine Transform (DCT).

The discrete cosine transform (DCT) helps separate the image into parts (or spectral sub-bands) of differing importance (with respect to the image's visual quality). The DCT is similar to the discrete Fourier transform: it transforms a signal or image from the spatial domain to the frequency domain as shown in Fig. 7.

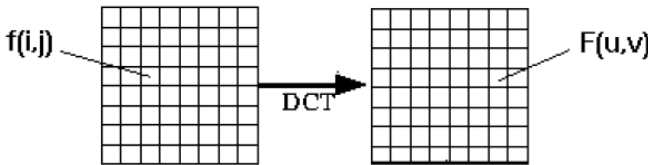


Fig. 7. Spatial domain to frequency domain

DCT Encoding, the general equation for a 1D ( $N$  data items) DCT is defined by the following equation:

$$C(u) = \alpha(u) \sum_{x=0}^{N-1} f(x) \cos \left[ \frac{(2x+1)u\pi}{2N} \right] \tag{6}$$

where  $u=0,1,2,\dots,N-1$

The corresponding *inverse* 1D DCT transform is defined as:

$$f(x) = \sum_{u=0}^{N-1} \alpha(u) C(u) \cos \left[ \frac{(2x+1)u\pi}{2N} \right]$$

where

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{for } u = 0 \\ \sqrt{\frac{2}{N}} & \text{for } u = 1,2,3,\dots,N-1 \end{cases} \tag{7}$$

If one interpret the  $N$ -size sequences as column vectors  $f = f(x) = [f_0, f_1, \dots, f_{N-1}]$  and

$C = C(u) = [C_0, C_1, \dots, C_{N-1}]$  denote the  $N \times N$  matrix by

$$D = \left[ \frac{1}{\sqrt{N}} \alpha(u) \cos \left( 2\pi \frac{(2x+1)u}{4N} \right) \right] \tag{8}$$



The general equation for a 2D ( $N$  by  $N$  image) DCT is defined by the following equation:

$$C(u, v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x, y) \cos \left[ \frac{u(2x+1)\pi}{2N} \right] \cos \left[ \frac{v(2y+1)\pi}{2N} \right] \quad (9)$$

where  $u, v = 0, 1, 2, \dots, N-1$

and the corresponding inverse 2D DCT transform is defined as:

$$f(x, y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} C(u, v) \alpha(u)\alpha(v) \cos \left[ \frac{u(2x+1)\pi}{2N} \right] \cos \left[ \frac{v(2y+1)\pi}{2N} \right] \quad (10)$$

where  $x, y = 0, 1, 2, \dots, N-1$

## 5 Application of Digital Watermarking

**Owner Identification:** The owner identification can be printed on the covers or mentioned somewhere on the item. Examples are the identification mark of an audio company on the CD case or the mark of the paper manufacturer on top corner of the paper. These types of watermarks can be easily removed by cropping the image or by tearing the part that has the identification. Digital watermarking helps to overcome this problem by embedding the watermark in the form of bits and forming an integral part of the content. [1].

**Copy Protection:** To prevent the data from being copied a watermark can be introduced in the data with a copy protect bit. When the copying device reads the data, the watermark detecting circuitry should detect the watermark and stop recording. This would need all the copying machines to have the watermark circuitry to identify the watermark and act accordingly [4].

**Data Authentication:** A given set of data (images) can be easily tampered without even being detected. To avoid this and maintain the originality of the image a watermark like signature, a set of words, may be embedded into the image. If the image is now being tampered it can be easily detected as the pixel values of the embedded data would change and not match the original pixel values. If the image is being copied it would lose its authentication as the embedded data would not be copied along with the image [8].

## 6 Attacks on Watermarking

The purpose of watermarking is to survive the authenticity of the original data against threats or attacks. Attacks can be of different types such as: removal attacks, geometrical attacks, cryptographic attacks, and protocol attacks [1]. The aim of removal attacks, as

the word indicates, is to remove the embedded watermark. The result will be better system performance and higher security due to an optimized design of the watermark extractor. Thus the possible attacks on watermarks can be defined as:

## 6.1 JPEG Compression

JPEG stands for Joint Photographic Experts Group. JPEG is a standardized image compression mechanism. JPEG is designed for compressing either full-color (24 bit/pixel) or gray-scale images of natural, real-world scenes. It works well on photographs, naturalistic artwork and similar material; not so well on lettering, simple cartoons or line drawings. JPEG handles only still images, but there is a related standard called MPEG for motion pictures. A useful property of JPEG is that the degree of lossiness can be varied by adjusting compression parameters. This means that the image maker can trade off file size against output image quality.

## 6.2 Geometric Transformations

In geometric transformation the position and value of pixels are changed by using the different transformation such as:

**Rotation:** Small angle rotation, often in combination with cropping, does not usually change the commercial value of the image but can make the watermark undetectable. Rotations are used to realign horizontal features of an image after it has been scanned. For benchmarking rotated image is proposed to be cropped so that there is no need to add a fixed border to it.

**Cropping:** In some cases, infringers are just interested by the “central” part of the copyrighted material moreover more and more Web sites use image segmentation, which is the basis of the “Mosaic” attack. This is of course an extreme case of cropping.

**Scaling:** This happens when a printed image is scanned or when a high resolution digital image is used for electronic application such as Web publishing. Scaling can be divided into two groups, uniform and non-uniform scaling. Very often digital watermarking methods are resilient only to uniform scaling.

# 7 Result and Discussion

The algorithms are tested on the reference Lena image shown in Fig. 8.

## 7.1 Spatial Domain Watermarking

Digital Watermarking using Least Significant Bit (LSB) Substitution: Results from LSB substitution are shown in Fig. 9 and Fig. 10. The watermarked image shows little but not noticeable degradation. Although the watermark is recovered perfectly in the ideal case, the addition of any amount of noise or compression of the image using JPEG fully destroys the embedded watermark. Even worse, the watermark can be removed with no perceivable change to the watermarked image. The message capacity of LSB embedding however is quite good, a 1:1 correlation with the size of the image. The Fig. 9, gives the Watermarked Image using LSB substitution, that is the embedded image, and Fig. 10, gives the extracted watermark from the Fig. 9.



**Fig. 8.** Leena Image taken for experiment



**Fig. 9.** Watermarked Image using LSB Technique



**Fig. 10.** Recovered Image

## 7.2 Frequency Domain Watermarking

Digital Watermarking using Discrete Cosine Transform (DCT):

DCT is a de - facto image transform theorem in frequency domain

Transform coding relies on the premise that pixels in an image exhibit a certain level of correlation with their neighbouring pixels and consequently these correlations can be exploited to predict the value of a pixel from its neighbours.

DCT transform therefore maps these spatial (correlated) data into transformed (uncorrelated) coefficients as given below:

$$C(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \cos \left[ \frac{\pi(2x+1)u}{2N} \right] \cos \left[ \frac{\pi(2y+1)v}{2N} \right],$$

The input image is  $N \times N$

$f(x,y)$  is the intensity of the pixel in  $x$ th row and  $y$ th column.

$C(u,v)$  is the DCT coefficient in the row  $u$  and column  $v$  of the DCT matrix.

The embedded watermark can be retrieved by applying the inverse DCT algorithm as given below:

$$f(x,y) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \alpha(u)\alpha(v) C(u,v) \cos \left[ \frac{\pi(2x+1)u}{2N} \right] \cos \left[ \frac{\pi(2y+1)v}{2N} \right]$$

Due to the efficacy of DCT to pack input data in few coefficients as possible, it provides excellent quality of reconstructed image with minimum distortion.

The reconstructed watermark has a high PSNR.

So now the Watermarked Image is given below in Fig. 11 and the reconstructed watermark after applying the inverse DCT algorithm is given below in Fig. 12.



**Fig. 11.** Watermarked Image using Frequency Domain Technique

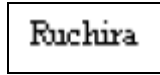


Fig. 12. Recovered Message

## 8 Conclusion

This study has introduced a number of techniques for the watermarking of digital images, as well as discussing the limitations and attributes of each technique.

LSB substitution is not a very good algorithm for digital watermarking due to its lack of even a minimal level of robustness. LSB embedded watermarks can easily be removed using techniques that do not visually degrade the image to the point of being noticeable. Furthermore, if one of the more trivial embedding algorithms is used, the encoded message can be easily recovered and even altered by an adversary. The LSB algorithm remains in the domain of steganography due to its tremendous information capacity. In spatial domain the watermark is embedded directly in to pixels of the image. Thus, the techniques in this domain proved fairly robust and carrying good capacity if small watermark is being used. In spread spectrum algorithm the watermark is converted into a sequence of PN codes known as gold sequence. This PN sequence have the attributes of the spread spectrum communication and very resistant to cropping or other attacks. An advantage of the spatial techniques in the thesis is that they can be easily applied to any image, regardless of subsequent processing (whether they survive this processing however is a different matter entirely). A possible disadvantage of spatial techniques is they do not allow for the exploitation of this subsequent processing in order to increase the robustness of the watermark.

Another observation is that transform domain techniques are typically better for watermarking than spatial, for both reasons of robustness as well as visual impact. Embedding in the DCT domain proved to be highly resistant to JPEG compression as well as significant amounts of random noise as upto 50 dB of PSNR value has been achieved using varying processing gain  $k$  upto 30. By anticipating which coefficients would be modified by the subsequent transform and quantization, it is possible to produce a watermarking technique with moderate robustness, good capacity, and low visual impact. This holds true in general for watermarking; robustness can be improved significantly when the subsequent degradation techniques are known. This holds particularly true in the case of compression techniques, where the compression algorithms are well known.

**Acknowledgement.** This work was supported by the Security Engineering Research Center, granted by the Korea Ministry of Knowledge Economy.

## References

1. Bloom, J.A., Cox, I.J., Kalker, T., Linnartz, J.-p., Miller, M.L., Traw, B.: Copy Protection for DVD Video. Proceedings of the IEEE 87(7), 1267–1276 (1999)
2. Mohanty, S.P.: Digital watermarking: A tutorial review, <http://www.cse.unt.edu/~smohanty/research/OtherPublications/MohantyWatermarkingSurvey1999.pdf> (as on May 13, 2010)

3. Wong, P.: A watermarking for image integrity and ownership verification. In: Proc. Final Program and Proceedings of the IS&-T PICS 1999, Savana, Ga, USA, pp. 374–379 (April 1999)
4. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley Publishing company, Inc. (1993)
5. Jain, A.K.: Fundamentals of Digital Image Processing. Prentice-Hall of India Pvt. Ltd. (1995)
6. Koblitz, N.: A Course in Number Theory and Cryptography. Springer, Heidelberg
7. Petitcolas, F.A.P.: Introduction to information hiding. In: Katzenbeisser, S.C., et al. (eds.) Information Techniques for Steganography and Digital Watermarking, pp. 1–11. Artec House, Northwood (1999)
8. Podilchuk, C.I., Zeng, W.: Image-adaptive watermarking using visual models. IEEE Journal on Selected Areas in Communications 16(4), 525–539 (1998)

# Use of Artificial Neural Network in Bengali Character Recognition

Debnath Bhattacharyya<sup>1</sup>, Tai-hoon Kim<sup>2,\*</sup>, and Gang-soo Lee<sup>3</sup>

<sup>1</sup> Information Technology Department,  
MCKV Institute of Engineering, Liluah,  
Howrah – 711204, India  
debnathb@gmail.com

<sup>2</sup> GVSA and School of Information Science  
University of Tasmania, Australia  
taihoonk@utas.edu.au

<sup>3</sup> Department of Computer Engineering  
Hannam University, Daejeon, Korea  
gslee@hnu.kr

**Abstract.** This paper describes how the Bengali characters are processed, trained and then recognized with the use of a back propagation Artificial neural network. Recognition is done on isolated Bengali character. The size and the font used for the characters are similar in both training and classification of the network. The images are first converted into grayscale and then to binary images. These images are then scaled to a fit a pre-determined area with a fixed but significant number of pixels. By extracting the characteristics points we formed the feature vectors, which in this case is simply a series of 0s and 1s of fixed length. Finally, an Artificial Neural Network is chosen for the training and classification process. The steps are simple, and a network is chosen for the training and recognition process. Researchers involved in recognition of good quality printed text in different scripts around the world have reported drastic decrease in recognition accuracy due to presence of touching characters in the text. So recognition is done here with isolated printed characters with size independent.

**Keywords:** Recognition, OCR, image processing, MQDF, allograph.

## 1 Introduction

Character recognition is a process of recognizing a character by the computer. The initial development of computers was for military, scientific purposes where very limited amount of data had to be entered while a huge amount of computations were to be done and for business application, the amount of data to be entered is fairly high while the amount of computations performed is low. Even today, we use keyboard as input method. So human involvement at the time of input can introduce errors and slows the process. A natural solution would be to let the computers do the jobs for us (the machine would transform the original document into a more suitable form in

---

\* Corresponding author.

lesser time and having few errors and process it). At this point, the idea of Optical Character Recognition (OCR) evolved. The principal motivation of OCR systems is the need to cope up with the enormous flood of papers such as bank cheques, commercial forms, government records, credit card imprints, bill processing systems, airline ticket readers, passport readers and mail sorting systems etc. So analyzing the documents and recognize them is an important role for this problem. The general idea is to develop a fully automate process for entering and understanding printed or handwritten data into the computer.

Character recognition was initially considered as an aid to the visually handicapped and the early successful attempts were made by the Russian scientist Tyurin in 1900. Character recognition have two modes- online (using tablet PC, PDA with stylus for input in the system) and offline (using scanner). Recognition can be performed on printed text or handwritten text. Some degradations in any printed script includes touching characters, broken characters, heavy printed characters, faxed documents, typewritten documents and backside text visible documents, or may be the paper quality is not good. In handwritten text different people have different font style and size. The designing of a recognition system requires careful attention to the following issues: definition of pattern classes, sensing environment, pattern representation, feature extraction and selection, cluster analysis, classifier design and learning, selection of training and test samples, and performance evaluation. In my project, main goal is to recognize Bengali offline printed characters.

## 1.1 Overview of Bengali Scripts

Bengali script was derived from ancient Brahmi script through various transformations. It is the second most popular language in India and the fourth in the world. About 200 million people in the eastern part of Indian subcontinent speak in this language. Modern Bengali script consists of 11 vowels and 39 consonants. These characters are called basic characters shown in Fig. 1. We are dividing a Bengali character into three zones – upper zone, middle zone, and lower zone. Relative height of upper zone is 27%, middle zone is 52% and lower zone is 21%. A horizontal line is present at the upper zone known as head line or matra. From a statistical analysis it has been notice that the probability of a Bengali word will have head line or matra is 0.994. Upper zone denotes above the head line, middle zone is in between head-line and base line and lower zone is below the base line, these three zones are shown in Fig. 2.

## 1.2 Properties of Bengali Scripts

There are some basic features or properties of any Bengali printed script:

- i. Writing style of Bengali is from left to right.
- ii. The concept of upper and lower case (as in English) is absent here.
- iii. A vowel following a consonant takes a modified shape, depending on the position of the vowel: to the left, right (or both) or bottom of the consonant, called *modified characters or allograph* (in Bengali ‘Kaar’) shown in Table 1.

And using these modifiers with a character  $\text{L}$  is shown in Table 2.



অ আ ই ঈ উ ঊ ঋ ঌ ঐ ঔ ঐ ঔ ঐ  
 ক খ গ ঘ ঙ চ ছ জ ঝ ঞ ট ঠ ড ঢ ণ ত থ দ ধ ন প ফ  
 ব ভ ম য র ল শ ষ স হ ঙ ট য় ং ঃ ং

Fig. 1. A set of Bengali Characters



Fig. 2. Three zones of a character

Table 1. Bengali vowels and their modifier forms

Vowel	Corresponding Vowel Modifier
আ	।
ই	ঁ
ঈ	ে
ঊ	ে
ঋ	ে
ঌ	ে
ঐ	ে
ঔ	ে
ঐ	ে
ঔ	ে

**Table 2.** Consonant attached with a modifier

Vowel	আ	ই	ঈ	উ	ঊ	ঋ	এ	ঐ	ও	ঔ
<b>Modified Shape With number</b>	১	২	৩	৪	৫	৬	৭	৮	৯	১০
<b>When attached to Consonant ক</b>	কা	কি	কী	কু	কূ	ক্	কে	কৈ	কো	কৌ

- iv. Exceptional case that some time vowel following a consonant form a compound character not a modifier.
- v. ৩ one part sits to the left and one to the right of the consonant. Thus, a delayed decision is necessary during recognition of these characters.

### 1.3 Artificial Neural Network

An Artificial Neural Network (ANN) is an information processing unit that is inspired by biological nervous systems, such as the brain that process information. In brain there are number of neurons that are highly connected to process information. The key element of this network is the structure of the information processing system. It is composed of a large number of highly interconnected processing elements (neurons) working in parallel to solve a specific problems. People can learn from examples. Learning in biological systems involves adjustments to the synaptic connections that exist between the neurons. And ANN is also doing this. The Back propagation algorithm is the most commonly used ANN learning technique. We can solve problems of non-linearly separable units with the help of Back propagation algorithm. Here, we are using delta rule which is converge towards best fit approximation to the target concept .The key idea behind delta rule is to use gradient descent to search the possible weight vector that best fit the training examples. Each processing unit performs a simple job that is to receive input from neighbors or external sources and use this to compute an output signal which is propagated to other units. Apart from this processing, a second step is to adjustment of the weights. The system is parallel in the sense that many units can carry out their computations at the same time. Within neural systems it is useful to distinguish three types of units: input units which receive data from outside the neural network, output units which send data out of the neural network, and hidden units whose input and output signals remain within the neural network. Neural networks has been trained to perform complex functions in various fields, including pattern recognition, identification, classification, speech, vision, and control systems.

Advantages of neural network are – it can perform task that cannot be performed by any linear program, when an element of the neural network fails, it can continue without any problem by their parallel nature, a neural network learns and does not need to be reprogrammed, it can be implemented in any application without any problem.

## 2 Literature Review

Various strategies have been proposed by different authors. U. Pal, K. Roy and F. Kimura [1] proposed a system, where first pre processing and then pre-segmentation is applied using water reservoir concept. And merging these primitive components and finding optimal character segmentation, Dynamic Programming (DP) is applied. And this DP is applied using total likelihood of characters as objective function. To compute likelihood of a character Modified Quadratic Discriminant Function (MQDF) is applied. Feature used in MQDF are based on directional feature of contour points.

A new approach to segment and recognize Printed Bangla Text using Characteristic functions and Hamming network was proposed by Md. Al Mehedi Hasan, Md. Abdul Alim, Md. Wahedul Islam [2]. A new algorithm has been proposed to detect and separate text lines, words and characters from printed Bangla text. The algorithm uses a set of characteristic functions for segmenting upper portion of some characters and characters that come under the Base line. It also uses a combination of Flood-fill and Boundary-fill algorithm for segmenting some characters that cannot be segmented using traditional approach.

Segmentation of Bangla text is a challenging problem since there is a variation of the same character due to change of fonts or introduction of noise. Various strategies have been proposed by different authors. A two phase approach is applied by S.M. Milky Mahmud, Nazib Shahrier, A.S.M Delowar Hossain Md. Tareque Mohmud Chowdhury, Md. Abdus Sattar [3] in order to overcome the common problems related to the segmentation of printed Bangla characters. B.B. Chaudhuri and U. Pal [4] combined primitive analysis with template matching to detect compound Bengali characters. Different methodologies in OCR development as well as research work are done on 12 Indian scripts [5]. Md. Abdul Hasnat [7] represent the training and recognition mechanism of a Hidden Markov Model(HMM) based multi-font Optical Character Recognition (OCR) system for Bengali character.

Existence of touching characters in the scanned documents is a major problem to design an effective character segmentation procedure. So by Utpal Garain and Bidyut B. Chaudhuri [6] has suggested a new technique that is used for identification and segmentation of touching characters. The technique is based on fuzzy multifactorial analysis.

## 3 Our Work

### 3.1 Pre-processing

In most of the cases we are assuming that whatever the data is given must be up to the mark i.e. data is good quality, size is perfect for recognition, etc. However, these characteristics are commonly not satisfied by off-line printed data. The low quality of the data is due to basically the combination of some facts. One is the addition of noise during scanning (Skew angle). Skew generation at the time of scanning (due to less concentration at the time of scanning this happens). Quality of the printed document is bad. To overcome these problems, we are using pre-processing, which involves scaling and binarization, and extraction of the data. Pre-processing converts raw data into feature vectors.

### 3.2 Data Collection

We have collected different samples of Bengali Characters by using various Bengali font like Avro keyboard, Tanmatra, Bongolipi, Bhasha Shoinik etc. for our input purpose. Now we have to convert the image into binary form. Then clipping them to desired size and then scaling them. Finally we have kept the characters separately within a folder. The system was trained with both computer-generated images and scanned images of text. It should be noted that no skew correction was done, so the scanning process is expected to be of high quality.

### 3.3 Recognition Process

Here we proposed a system for character recognition. The obtained characteristics vectors are used for classification with Artificial Neural Networks. The input of the experiment is a set of common Bangla characters. The features vectors are obtained from these images are the basis for further classification. At first it was used for recognition of single characters. Then we will apply it to recognize whole words. And after that recognition of numerals. I have divided my project into two phases:

- a. Reading a windows image format and pre process them.
- b. Development of Artificial Neural Network model.

Second phase is further divided into two sub-phases:

- i. Training phase of neural network, and
- ii. Testing phase of neural network.

We have implemented 'Bengali Character Recognition' using Matlab for its ease in image acquisition and neural network applications. The first step of my project consists of images acquisition. Then the picture size is manipulated in order to be able to subsequently process them. Once the preprocessing step is achieved, it is necessary to put them in neural network.

For pre-processing, first reading each character then convert to binary image and scanning them from four directions(left, right, top and bottom) and clipping the extra part of the character and then we have to resize it (Fig. 5).

Algorithm:

STEP 1: Calculate the size of the binary image (width\*height).

STEP 2: for i=1 to width

for j=1 to height

if image I(i, j) is equal to 1 then note the row number and copy the image from that row to another matrix(II).

STEP 3: Scan the image in two direction (because in step 2 we have clipped two sides) find the intersection point and clip them.

STEP 4: Scale the image getting from step3 to desired size (assuming 15\*18).

In our work, we have decided to take resize window  $15 \times 18$  (Fig. 5). If after clipping the character size is greater than  $15 \times 18$  then sample down to resize window and if clipped character is less than resize window then sample it up. A network is to be designed and trained to recognize the 50 letters of the alphabet set. In my project work I have used Multi-Layered Perceptron: feed-forward neural network, with one hidden layer and logistic sigmoid transfer function (because its output range 0-1 is perfect for learning to output boolean values), trained by the back propagation method.

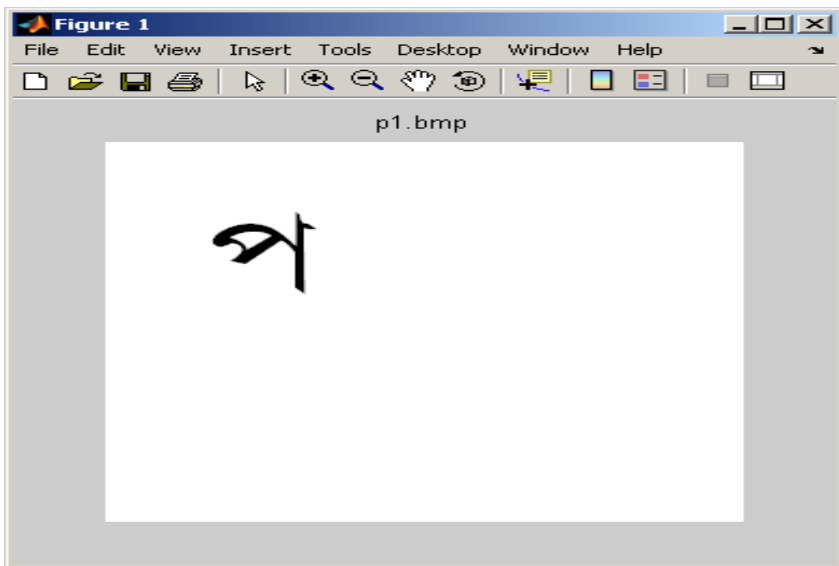
The neural network had three layers: an input layer consisting of 270 nodes (for the  $15 \times 18$  character input), a hidden layer consisting of 38 nodes, and an output layer with 50 nodes (one node for each letter). The network uses back propagation in addition to bias weights and momentum. In hidden layer I have considered 38 neurons. And in output layer 50 (basic character) neurons. The fifty 270 element input vectors are defined in the function "data" as a matrix of input vectors called "alphabet". The target vectors are also defined in this file with a variable called "target". Each target vector is a 50-element vector with a 1 in the position of the letter it represents, and 0's everywhere else. For example, the letter A is to be represented by a 1 in the first element (as A is the first letter of the alphabet), and 0's in rest of the elements (forty-nine). Using the program a feed-forward network is trained to recognize character bit maps, in the presence of noise.

At first, initializes the feed-forward network. Then train a feed-forward network with faster back propagation algorithm. Then simulate the feed-forward network. And finally apply competitive transfer function to calculate a layer's output from its net input. In many network paradigms it is useful to have a layer whose neurons compete for the ability to output a 1. The result is that the only neuron that can respond with appreciable output (neuron whose net input is highest). All other neurons are inhibited so strongly by the winning neuron that their outputs are negligible. To model this type of layer efficiently on a computer, a competitive transfer function is often used. Such a function transforms the net input vector of a layer of neurons so that the neuron receiving the greatest net input has an output of 1 and all other neurons have outputs of 0. It is then required to identify the letter by responding with a 50-element output vector where each represents a letter. To operate correctly, the network should respond with a 1 in the position of the letter being presented to the network. All other values in the output vector should be 0. In addition, the network should be able to handle noise. In practice, the network does not receive a perfect Boolean vector as input. In Fig. 6, it is a network which is trained without noise and its x-axis represents number of iterations and y-axis denotes performance of the network. Its goal is assumed to be 0.1 at  $308^{\text{th}}$  iterations. It is a network which is trained with and without noise (Fig. 7) and x-axis represents number of iterations and y-axis denotes performance of the network. In this network goal is assumed to be 0.2 and it has achieved in  $38^{\text{th}}$  iterations. Initially I have taken one data set and added random number to generate different no of sets. Initialize the network by single data and train it by iterating the network for 8000 epoch. Then copy the same initial network and trained them with noise using fifty different data set (created by adding random

numbers) for 450 epoch and also train without noise for 500 epoch. Then plot a graph between network trained without noise and network trained with and without noise in Fig. 8. After that I have simulate the network and finally compete the network for better results. The solid line on the graph (Fig. 8) shows the reliability for the network trained with and without noise. The reliability of the same network when it had only been trained without noise is shown with a dashed line. Thus, training the network on noisy input vectors greatly reduces its errors when it has to classify noisy vectors. The network did not make any errors for vector of noise of mean 0.00 to 0.3. When noise of mean 0.5 was added to the vectors both networks began making errors. Then we are testing the network by giving the position of the character and it is recognizing correctly in Fig. 9.

## 4 Result

The experimental evaluation of the above technique was carried out using isolated Bengali characters. The data was collected from different Bengali softwares like Avro keyboard, Tanmatra, Bongolipi, etc. Different fonts like Bongon Rupali, Ekushey Azad, Ekushey Mohua etc. are used in the data sample of Bengali characters. First read the image (Fig. 1) and convert it to binary image (Fig. 2). Then using a program clip the extra part (Fig. 3) scales it to our desired window size (Fig. 4).



**Fig. 3.** Original image of single character

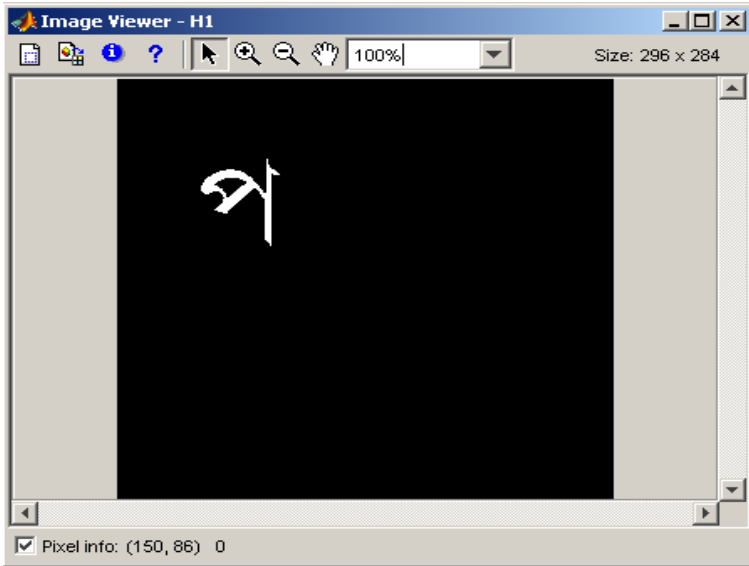


Fig. 4. Original image is converted to binary image

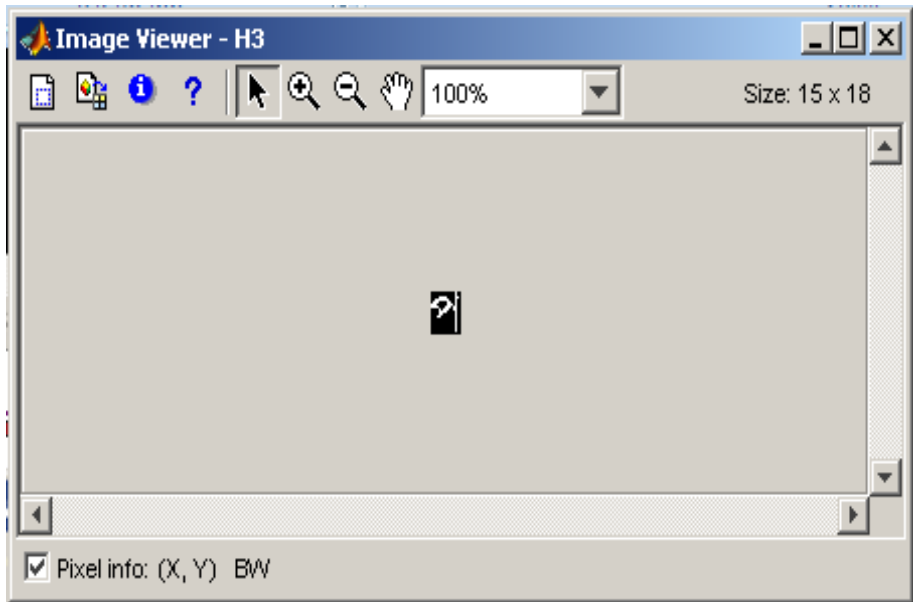


Fig. 5. Clipping and resize the clipped image to desired size

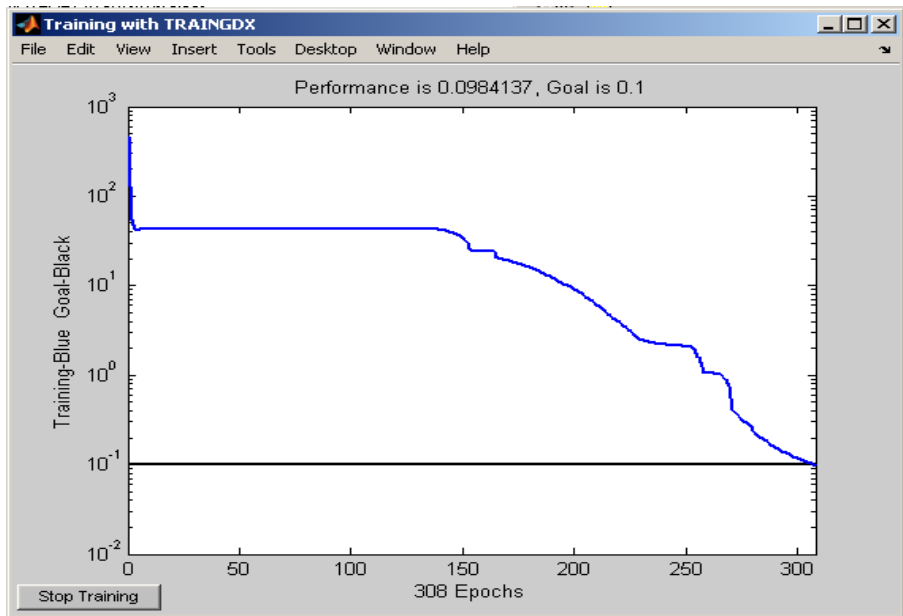


Fig. 6. Training the network without noise

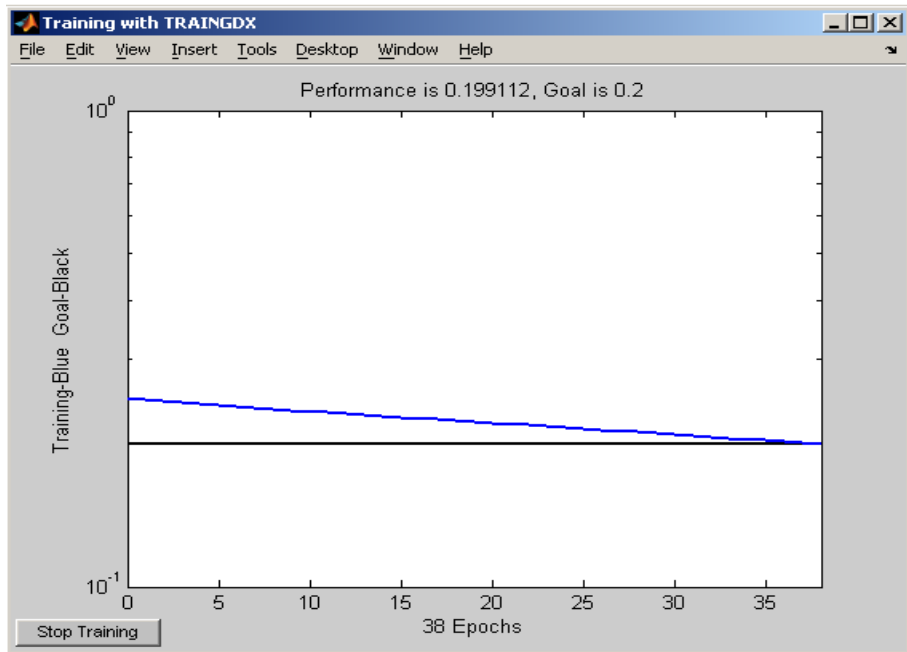


Fig. 7. Training the network with and without noise



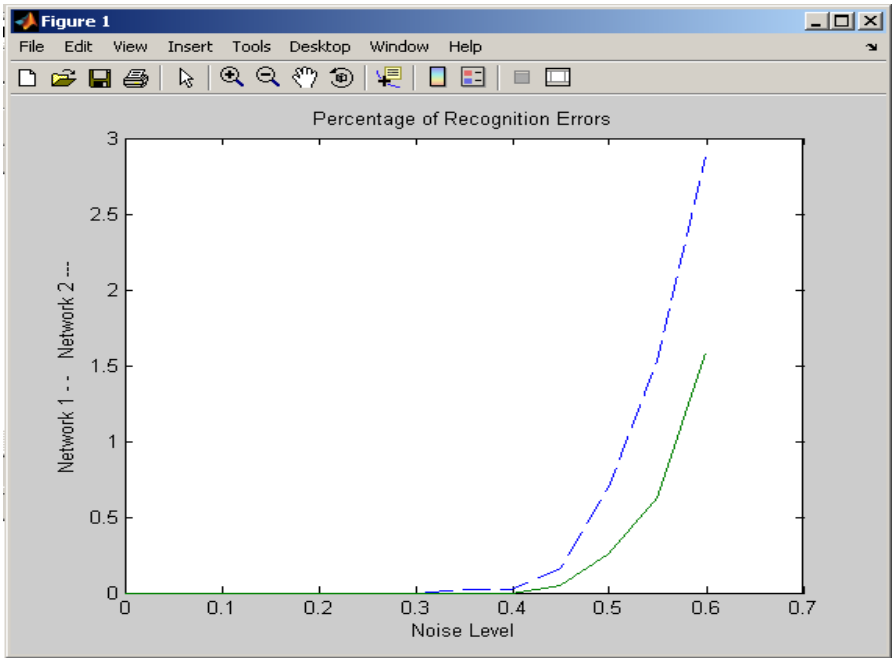


Fig. 8. Percentage of error recognition

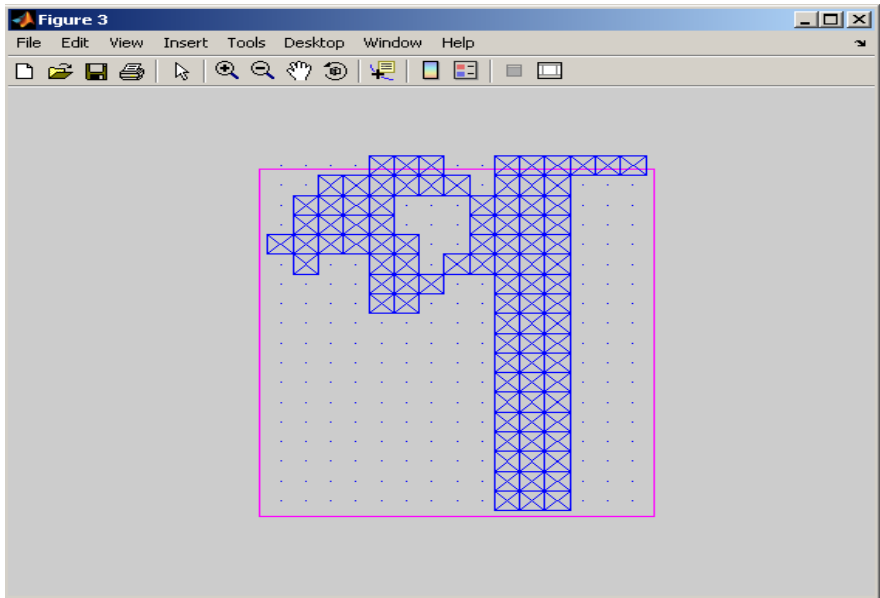


Fig. 9. Recognizing a character after giving the position of basic character

The experiments were carried out in Matlab 7.0, 2 GHz processor, with 1.99 GB RAM. The basic Image Processing operations were performed using Matlab's Image Processing Toolbox. Five Bengali fonts were used here. Different fonts are collected from different software like Bongolipi, Avro keyboard, Tanmatra etc. The entire data set consisted of 50 characters for each of the fonts.

**Table 3.** Recognition accuracy

No. of characters	No. of character's recognized	Character not recognized	Percentage
250	245	5	98%

## 5 Conclusion

The advantage of the proposed system is its efficiency. Table 3 shows the accuracy rate of our method. In computerization of any language, one of the vital tasks is to develop an efficient and effective Optical Character Recognition (OCR) system for the respected language. In order to store million pages of paper documents into electronic form, OCR is the key tool. Otherwise, if those are entered by typing manually, the efficiency, effectiveness and correctness will drastically fall down. For Bengali, there is no good OCR solution till now. Since, the written form of Bengali documents is more complex than that of many other languages, Bengali script segmentation is of great importance for creating a Bengali OCR system.

**Acknowledgement.** This work was supported by the Security Engineering Research Center, granted by the Korea Ministry of Knowledge Economy.

## References

1. Pal, U., Roy, K., Kimura, F.: A Lexicon Driven Method for Unconstrained Bangla Handwritten Word Recognition. In: International Workshop on Frontiers in Handwriting Recognition, pp. 601–606 (2006)
2. Al Mehedi Hasan, M., Abdul Alim, M., Wahedul Islam, M.: A New Approach to Bangla Text Extraction and Recognition From Textual Image. In: International Conference on Computer and Information Technology (ICCIT 2005), Islamic University of Technology (IUT), Dhaka, Bangladesh, December 28-30 (2005), [http://research.banglacomputing.net/iccit/ICCIT\\_pdf/8th%20ICCIT\\_2005\\_564.pdf](http://research.banglacomputing.net/iccit/ICCIT_pdf/8th%20ICCIT_2005_564.pdf)
3. Milky Mahmud, S.M., Shahrier, N., Delowar Hossain, A.S.M., Tareque Mohmud Chowdhury, M., Abdus Sattar, M.: An Efficient Segmentation Scheme for the Recognition of Printed Bangla characters. In: International Conference on Computer and Information Technology (ICCIT 2003), Jahangirnagar University, Dhaka, Bangladesh, December 19-21, pp. 283–286 (2003)

4. Chaudhuri, B.B., Pal, U.: A Complete Printed Bangla OCR System. *Journal of Pattern Recognition* 31(5), 531–549 (1998)
5. Pal, U., Chaudhuri, B.B.: Indian script character recognition: a survey. *Journal of Pattern Recognition* 37(9), 1887–1899 (2004)
6. Garain, U., Chaudhuri, B.B.: Segmentation of Touching Characters in Printed Devnagari and Bangla Scripts using Fuzzy Multifactorial Analysis. *IEEE Transactions on Systems, Man and Cybernetics* 32, 449–459 (2002)
7. Abul Hasnat, M.: Research Report on Bangla OCR Training and Testing Methods, [http://www.bracu.ac.bd/research/crb1p/technical\\_report/2004-2007/BAN25.pdf](http://www.bracu.ac.bd/research/crb1p/technical_report/2004-2007/BAN25.pdf)

# New Algorithm for Skewing Detection of Handwritten Bangla Words

Rajib Ghosh<sup>1</sup>, Debnath Bhattacharyya<sup>2</sup>, Tai-hoon Kim<sup>3,\*</sup>, and Gang-soo Lee<sup>4</sup>

<sup>1</sup> Computer Science and Engineering Department  
Heriatge Institute of Technology  
Kolkata-700107, India  
grajib1@gmail.com

<sup>2</sup> Information Technology Department,  
MCKV Institute of Engineering, Liluah, Howrah – 711204, India  
debnathb@gmail.com

<sup>3</sup> GVSA and School of Information Science  
Unviersity of Tasmania, Australia  
taihoonk@utas.edu.au

<sup>4</sup> Department of Computer Engineering  
Hannam University, Daejeon, Korea  
gslee@hnu.kr

**Abstract.** Segmentation of a word into basic characters or strokes is an essential and necessary preprocessing step for character recognition in many handwritten word recognition systems, especially in case of handwritten bangla words. The major difficulty in character segmentation is the cursive script. This is because different person have different styles for their handwriting. Here, in this article a novel approach for skew detection followed by skew correction has been presented for online handwritten Bangla words. Here, we have used a slight variation of the projection profile method to calculate the amount of skew in an online Bangla handwritten word. The algorithm has been verified on a database of words collected from different people.

**Keywords:** Image processing, segmentation, recognition, database.

## 1 Introduction

Segmentation of online Bangla handwritten words is the most important step required for recognition of the words. This is because of the fact that, the basic strokes or characters that we get as a result of the segmentation procedure (to be discussed in the next chapter), must be recognized first to recognize the whole word successfully. But the most difficult challenge in segmentation is the cursive script. Sometimes it has been seen that, while writing online Bangla words, the words tend to be rotated or skewed by some arbitrary angle. This may happen due to the writing style of different people. The segmentation procedure requires de-skewed or un-skewed word as its input to provide good results. This is where we need *Skew detection* and *Skew correction*.

---

\* Corresponding author.

Since a word can be skewed at any arbitrary angle while writing, *Skew detection* refers to the problem of finding out the angle in which an image is rotated or skewed. Fig. 1 shows the example of two different Bangla skewed handwritten words.



**Fig. 1.** Example of two different handwritten Bangla online skewed words

If we try to segment these types of skewed words, then it will result in incorrect segmentation. So, to correctly segment these words we have to first de-skew the words. To de-skew a skewed image, we have to find out the angle in which the word is rotated as shown in Fig. 2.



**Fig. 2.** Skew angle  $\theta$  for the skewed word shown in Fig. 1

## 2 Literature Review

There are many works available in the literature describing the Skew detection and correction techniques for printed text (offline) till now. A skew detection algorithm is given the digital image of a word or document as input and the algorithm finds out the skew angle for the word or document image. Once this skew angle has been found, then the image can be corrected (Skew correction). These techniques can be broadly classified into several categories such as Projection profile methods, Hough-transform methods, nearest-neighbor methods etc. Projection profile is one of the most popular methods for these purposes. This method tries to find out the profiles at a number of different angles to find out the skew angle of the image [1]. A slightly modified version of this projection profile method has been proposed by Baird [2]. This method

improves the accuracy and speed of skew estimation. The skew-estimation error for this method comes out to be  $\pm 0.5^\circ$ . Hough-transform methods also yield high skew estimation frequency [3]. The estimation accuracy of the nearest-neighbor method is also very high but the problem is that these methods are costly especially when many connected components are involved.

Apart from these popular techniques, there are some other methods too in the literature for skew detection. These approaches include estimation using Neural networks [4], Fourier transform, Cross-correlation methods [5, 6] etc. Another approach uses the concept of Block adjacency graph to find out the skew angle [7]. Once the skew is estimated, the image can be rotated by that angle to correct the skew. But most of these works are there for detecting skew of Offline handwritten words/texts or printed words/texts. No such significant work is yet available for Online Bangla handwritten words. Here in this work, we have proposed an efficient algorithm for skew detection of online handwritten Bangla words using projection profile analysis.

### 3 Projection Profile Analysis

Projection profile analysis is one of the most popular methods to determine the skew-angle of an image. Here in our work, we have used the concept of horizontal projection profile. There is a one-dimensional array which stores the number of locations equal to the number of rows in an image. Each location in that array (projection profile) stores the total number of black pixels in the corresponding row of the input image. This histogram has the maximum peak when the text in the image is skewed at zero degrees i.e. when the image is not skewed at all. This is because of the fact that, the number of co-linear black pixels is maximum in this condition. It has been found that, the peaks in the profile calculated from the un-skewed image are taller than those computed from the skewed image. This happens because when the word is linear then we will get maximum peak within a small region but in case of skewed word all the pixels are evenly distributed over a taller region. If we get many black pixels within a small region then it is evident that the horizontal projection of that word will have more peak than others.

### 4 Skew Detection Algorithm

Several research scholars have modified the basic projection profile method and proposed their own algorithm based on the modification. For example, Postl proposed a method similar to that described above. He calculated the horizontal projection profile at a number of different angles and he used a modified directional criterion to find out the projection profile. Another approach includes a method that calculates the projection profiles over a number of circular windows. In another proposed mechanism, a document image is first divided into a number of vertical “swaths” and then the projection profile is calculated within each of them. But most of these algorithms are for handwritten or digital text images. Below, we have introduced an algorithm which will work on Online Handwritten Bangla words.

In this proposed algorithm, we have taken the skewed word image as input. Then the word image is rotated clockwise and for each resulting word image the horizontal projection profile is calculated. According to our discussion previously, the horizontal histogram of the word image should provide maximum peak when the word is skewed at  $0^\circ$  i.e. when the word is un-skewed. So while rotating the word clockwise, whenever the horizontal histogram of the rotated word will provide maximum peak, at that angle the word is skewed. At this moment the angle in which the maximum peak occurs, is our desired Skew-angle.

In most of the cases we are assuming that whatever the data is given must be up to the mark i.e. data is good quality, size is perfect for recognition, etc. However, these characteristics are commonly not satisfied by off-line printed data. The low quality of the data is due to basically the combination of some facts. One is the addition of noise during scanning (Skew angle). Skew generation at the time of scanning (due to less concentration at the time of scanning this happens). Quality of the printed document is bad. To overcome these problems, we are using pre-processing, which involves scaling and binarization, and extraction of the data. Pre-processing converts raw data into feature vectors.

### SkewDetect (Algorithm)

Step - 1: For each angle  $\alpha$  starting from  $2^\circ$  to  $40^\circ$  do the following steps:

Step - 1.1: Rotate the word image clockwise by the angle  $\alpha$ .

Step - 1.2: Take the horizontal projection profile of the rotated word image.

Step - 1.3: Store the maximum number of pixels found in the horizontal projection profile along with the corresponding angle  $\alpha$ .

Step - 2: Calculate the maximum of all the numbers found in Step - 1.3. we denote it as 'maxh'.

Step - 3: The angle corresponding to 'maxh' is the desired skew-angle  $\theta$ .

Once the skew-angle is found from the previous algorithm, then correcting the skew of the word image is pretty straightforward. To correct the skew of the word, we simply need to rotate the word image anti-clockwise by angle  $\theta$ .

## 5 Result

In this work, we have collected sample skewed words using several datasheets containing different words shown in Fig. 3.

The figure consists of three screenshots of a software window titled 'Bangla Characters Training'. Each screenshot shows a grid of handwritten words in Bengali. The first screenshot shows a 7x3 grid of words. The second screenshot shows a 6x3 grid of words. The third screenshot shows a 5x7 grid of words.

আসফিলাম	যোগসাজশ	সাংবাদিকরা
মতবাদের	উওমকুমার	মানুষগুলোকে
বৈঠকখানায়	আমিনপুরের	প্রশাসনের
সমন্বয়ের	কেলওডাবেই	নববারাকপুর
মানুষজনের	কলকাতায়	লোকসভায়
জনসংযোগের	সমন্বয়গুলিতে	পরিকার্ষে
বেরিয়েছিলাম		

মহলয়া	গবেষকরা	যোগাযোগের
জানিয়েছেন	থাইতেছিলাম	রাজনৈতিক
বিজয়ওয়াদয়	রমাপ্রসাদ	বিধানসভায়
শ্রদ্ধাশ্রী	থাইতেছিলাম	জনগণের
তুলসীদাস	সংগঠনের	
পরিবালের	লালগড়ের	

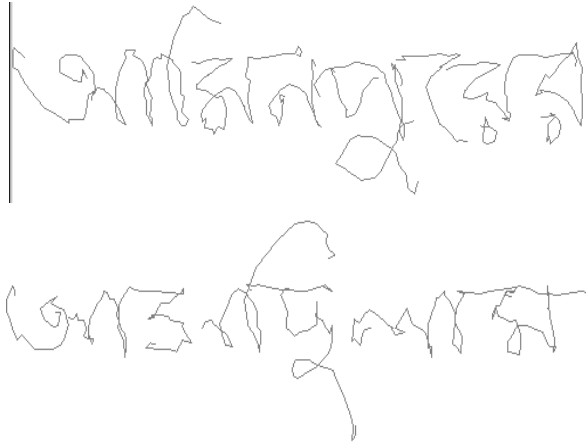
ক্রম	অবল	পাহ	সাহা	বাষ্	চোর	টাকা
বাপ	দাদাম	হাবরা	শ্রাম	শাধা	ভয়	যশ
জল	ঠালা	তুল	পবন	হেলে	ওল	ইম
ঢাকা	ইকল	বালদা	থাল	ভাব	এগরা	ব্য
ঐরাকত	ঐর	ঐম				

Fig. 3. Datasheets used for online handwritten word collection

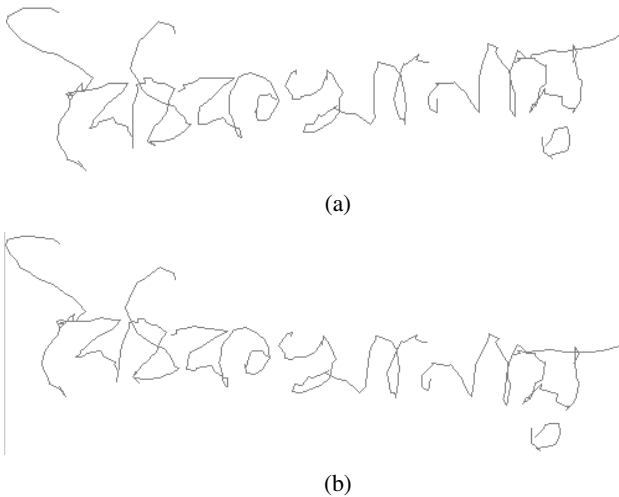
Here we have collected 50-60 online handwritten words for each of the words in the datasheet from different person. All those words are fed into the skew detection algorithm. After finding the skew, they are corrected. The accuracy of skew detection and correction are shown in table – 1. The sample results are shown in Fig. 4 As it can be seen from the table, if the whole word is skewed, the skew detection and correction accuracy of the above mentioned algorithm for online Bangla handwritten words came



out to be approximately 81.33%. The error rate of 18.67% has come out due to the wide variety of the writing styles of different people. Sometimes it has been observed that, while writing some part of the whole word is skewed and the rest of the word is not skewed at all. This makes it difficult to estimate the skew angle for the part of the word (that is skewed). In most of the cases, these problems make up the above mentioned error rate. The proposed algorithm is limited to those words which are skewed in anti-clockwise direction only. Fig. 5 shows an example of erroneous output of the algorithm.



**Fig. 4.** Corrected Words (shown in Fig. 1), after successfully detecting and correcting skew



**Fig. 5.** (a) Original skewed word, (b) The word after skew correction (erroneous)

**Table 1.** Result of skew detection and correction

Skew detection and correction accuracy	Error rate
81.33%	18.67%

## 6 Conclusion

This paper presents a scheme for the Skew Detection followed by the Skew Correction of Online Bangla Handwritten words. Till date, no such significant work is performed for the skew estimation of online Bangla handwritten words. Most of the previous schemes were based on printed words/texts. So, although the result came out to be approximately 65%, but this work will pave the way for further works in case of skew estimation of online bangla handwritten words and this will in turn result in the development of a more and more accurate Online Bangla Handwritten word recognition.

**Acknowledgement.** This work was supported by the Security Engineering Research Center, granted by the Korea Ministry of Knowledge Economy.

## References

1. Postl, W.: Detection of linear oblique structures and skew scan in digitized documents. In: Proceedings of the Eighth International Conference on Pattern Recognition, Paris, pp. 687–689 (1986)
2. Baird, H.S.: The skew angle of printed documents. In: Proceedings of the SPSE Fortieth International Symposium on Hybrid Imaging Systems, New York, pp. 21–24 (1987)
3. Srihari, S.N., Govindaraju, V.: Analysis of textual images using the Hough transform. *Machine Vision and Applications* 2, 141–153 (1989)
4. Rondel, N., Burel, G.: Cooperation of multi-layer perceptrons for the estimation of skew angle in text document images. In: Proceedings of the Third International Conference on Document Analysis and Recognition, Montreal, pp. 1141–1144 (1995)
5. Yan, H.: Skew correction of document images using interline cross-correlation. *CVGIP: Graphical Models and Image Processing* 55(6), 538–543 (1993)
6. Avanindra, S.C.: Robust detection of skew in document images. *IEEE Trans. Image Processing* 6(2), 344–349 (1997)
7. Kwag, H.K., Kim, S.H., Jeong, S.H., Lee, G.S.: Efficient skew estimation and correction algorithm for document images. *Image and Vision Computing* 20, 25–35 (2002)
8. Hull, J.J.: Document image skew detection: Survey and annotated bibliography. *Document Analysis Systems II*, pp. 40–64. World Scientific (1998)

# An Extended Set of Haar-like Features for Bird Detection Based on AdaBoost

Chih-Cheng Huang, Chun-Yi Tsai, and Horng-Chang Yang

Department of Computer Science and Information Engineering,  
National Taitung University, Taiwan, R.O.C  
u9711101@ms99.nttu.edu.tw, {cytsai, hcyang}@nttu.edu.tw

**Abstract.** As Adaboost is an efficient method to select features, we investigate its capability in selecting effective features from a set of Haar-like features for bird detection. Moreover, Adaboost is also used to construct a strong classifier for the task of detecting birds from a set of weak classifiers. We propose to add specially designed new Haar-like features to increase the detection rate of interested objects, here, namely birds. Our experiment shows that this method can increase the TPR (True Positive Rate) and decrease the FPR (False Positive Rate). In addition, a motion detection algorithm is used to detect moving objects, which are segmented from background. Using the strong classifier trained from Adaboost algorithm, segmented objects are classified as bird or not. The advantage of this approach is that higher performance and real-time detection can be achieved due to the fact that the strong classifier needs not to examine all the possible sub-windows of the input image. Thus, this method can further decrease the rate of false positive.

**Keywords:** bird detection, motion detection, Adaboost algorithm, Haar-like features.

## 1 Introduction

Adaboost is an efficient machine learning algorithm which has been successfully applied to the problem of object detection. In their seminal paper, Viola and Jones demonstrated the use of Adaboost on feature selection and classifier training in a real time system for face detection [6]. Adapting their approach for detecting objects, such as birds, this paper examines the effectiveness of a set of special designed Haar-like features, extended from the four features used by Viola and Jones. Such a technique can be a part of bird warning system for civil aviation, or be useful for some ecological survey systems.

The first issue needs to be addressed in this study is to specify appropriate features for birds. Apparently, this is not an easy task, since different kinds of birds vary in size, shape and color. Fortunately, as demonstrated by Viola and Jones, Adaboost can be used to select effective features from a vast number of candidate features efficiently. Hence, even the four Haar-like features used by Viola and Jones may be insufficient for bird-detection, it is feasible to extend them with more Haar-like

features of which some may represent the essential features for birds. It is a simple and straightforward solution and our experiment result shows that this approach does work to some extent.

The second issue is that though Adaboost can be used to select effective features from a vast of number of candidate features, it still takes a long time for such a task. Hence it is not feasible to try out all the possible features. And using only a few more number of features as in our experiment, there is still room to improve the performance of the system, in particular, of the false negative detection rate. Indeed, as we are mainly interested in living birds which are usually in motion, it is reasonable to segment moving parts detected by a simple motion detection method from frame images first. Therefore, by focusing on these moving parts, it can reduce computation time significantly and improve the false positive rate.

The rest of the paper is organized as following. Section 2 reviews some related work. The proposed method is presented in Section 3. A brief experiment is described in Section 4. Finally a short conclusion and future works are given.

## 2 Related Work

Adaboost which stands for Adaptive Boosting proposed by Freund and Schapire [8] is a machine learning algorithm which conducts a finite set of weak classifiers to form a committee that performs two-class pattern recognition through a linear (weighted) combination of the set of weak classifiers. Viewing face detection as a two-class classification problem, Viola and Jones [5,6] consolidated AdaBoost, four Haar-like features, integral images, and cascade classifiers into a robust real-time system mainly for human face detection.

In the case of detecting special objects, based on the empirical results, several methods which either modify the Adaboost algorithm or extend the set of Haar-like features were proposed to improve detection accuracy. For example, Gambs et al.[7] and Rasolzadeh et al.[1] proposed some new weak classifiers to increase the detection rate. Focusing on the representation capacity of feature set, Lienhart et al.[3,4] designed an extended set of Haar-like features which results in a better accurate rate when compared with the original feature set used by Viola and Jones . The same kind of result was reported by Weber[2], who used the same four Haar-like features to detect air planes. However, in Weber's experiment, the reported accuracy was not satisfying. Apparently, the work by Lienhart et al. [7] and that by Weber both indicate that it is necessary to modify or extend the Haar-like features aptly in order to achieve a higher detection rate for some special object.

## 3 Experiment Method

### 3.1 Haar-like Features

This work is based on the set of four Haar-like features which Viola and Jones[5] proposed. The set consists of vertical and horizontal Two-rectangle features, Three-rectangle feature, and Four-rectangle feature as shown in Fig. 1 (A), (B), (C), and (D), respectively.

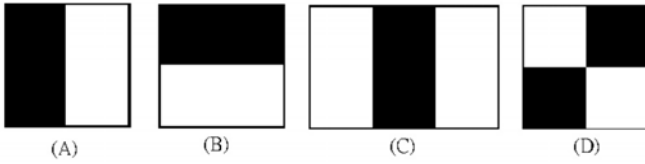


Fig. 1. Four Haar-like features

The value of a Haar-like feature is defined as the difference between the sum of the pixels' values of the black region(s) and that of the white region(s). Keeping the four Haar-like features, we add three more Haar-like features specifically adaptive to bird detection as shown in Fig. 2 and Fig. 3. One is for the detection of eyes and the other two are for diagonal line segments. In a way, the total number of features is kept to a reasonable size for the sake of computation time.



Fig. 2. New Haar-like features adaptive to the birds

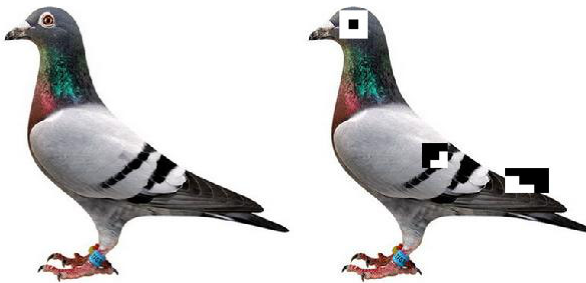


Fig. 3. Examples for the new Haar-like features adaptive to the birds

### 3.2 Integral Image

The total number of Haar-like features of an image can be very huge. Calculating all the feature values for each position in the image without any strategy will considerably decrease the speed of training and detection. Therefore, Viola and Jones [5] use an efficient intermediate representation — Integral Image, to pre-compute the accumulated sum of pixel values at the beginning only once. After that, any Haar-like feature value can be computed very easily and quickly in constant time.

The definition of Integral Image is  $ii(x, y) = \sum_{x' < x, y' < y} i(x', y')$ , where  $i(x', y')$  denotes the pixel value at the point of  $(x', y')$  which is above and to the left of  $(x, y)$  inclusively. As shown in Fig. 4, any Haar-like feature value can thus be computed by this definition. For instance, the sum within D in Fig. 4 can be computed as  $ii(4) - ii(2) - ii(3) + ii(1)$ .

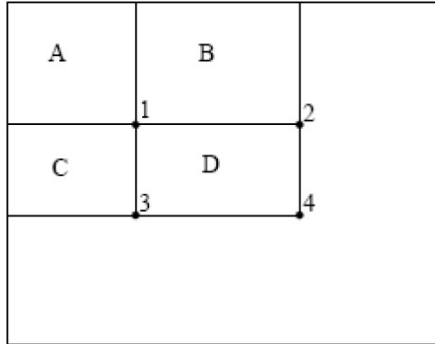


Fig. 4. An example of Integral Image

### 3.3 AdaBoost Algorithm

The purpose of Adaboost algorithm as shown in Table 1 is to find out a finite set of weak classifiers and thus forms a strong classifier after a series of training rounds with a finite set of positive and negative samples of some special object. After training, the strong classifier can be applied to detect the objects.

Table 1. Adaboost Algorithm

<p>A. A set of samples <math>(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)</math>. <math>y_i = 0</math> denotes a negative sample. <math>y_i = 1</math> denotes a positive sample. <math>n</math> is the total number of samples.</p> <p>B. Normalize the weights <math>w_{1,i} = D(i)</math>.</p> <p>C. For <math>t = 1, \dots, T</math> :</p> <p>a. Normalize the weights so they sum to one: <math>q_{t,i} = \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}</math></p> <p>b. For each feature <math>f</math>, train a weak classifier <math>h(x, f, p, \theta)</math>.</p> <p>c. For each feature <math>f</math>, find the error with respect to the distribution <math>D</math>: <math>\epsilon_f = \sum_i q_{t,i}  h(x, f, p, \theta) - y_i </math></p> <p>d. Choose the weak classifier <math>h_t(x)</math> with the lowest error <math>\epsilon_t</math> :</p>
---

$$\varepsilon_t = \min_{f,p,\theta} \sum_i q_i |h(x_i, f, p, \theta) - y_i|$$

e. Update the weights:  $w_{t+1,j} = w_{t,j} \beta_t^{1-\varepsilon_t}$ ,  $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

D. The strong classifier is:  $H(x) = \begin{cases} 1 & \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{otherwise} \end{cases}$ ,  $\alpha_t = \log \frac{1}{\beta_t}$

### 3.4 Motion Process

In our system, the motion process is a simple algorithm to segment areas with objects in motion. For simplicity, each color frame image is converted from RGB representation to its corresponding YUV format and then into a gray level image. The difference of the Y values of each consecutive pair of images is calculated, and stored in an array with the equivalent size as the image, (see Fig. 5.)

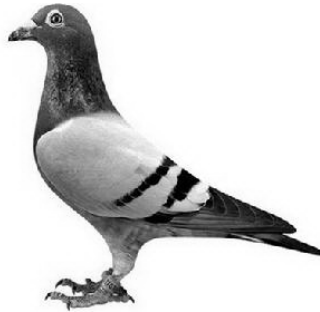


Fig. 5. The example of a transformed gray level image

With the difference image of Y values between two consequent frame images, we can further transfer the difference image into a binary one according the following formula:

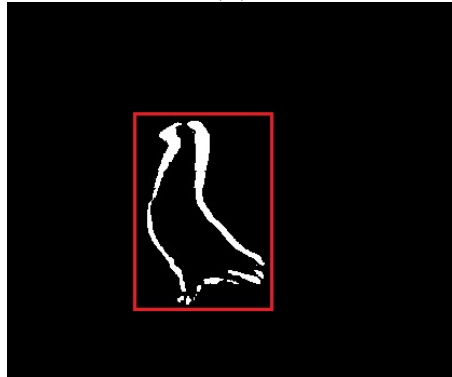
$$I(x, y) = \begin{cases} 1 & \text{if } |Y(x, y, t_i) - Y(x, y, t_{i-1})| > T \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

where  $I(x,y)$  denotes the pixel value at coordinates  $(x,y)$ ,  $Y(x, y, t_i)$  denotes the  $i$ th image of Y value and  $Y(x, y, t_{i-1})$  denotes the previous image of Y value. By thresholding the difference between  $Y(x, y, t_i)$  and  $Y(x, y, t_{i-1})$  according to an appropriate  $T$ , we can make the edge of the object highlighted as can be seen in Fig. 6.

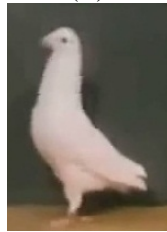
With the binary-transferred image, we can record the coordinate  $(x,y)$ , the width and height  $(w,h)$  of each 1-valued pixel. Thus, the information can be used to extract the objects in motion. Figure 6 shows an example of the motion segmentation process.



(A)



(B)



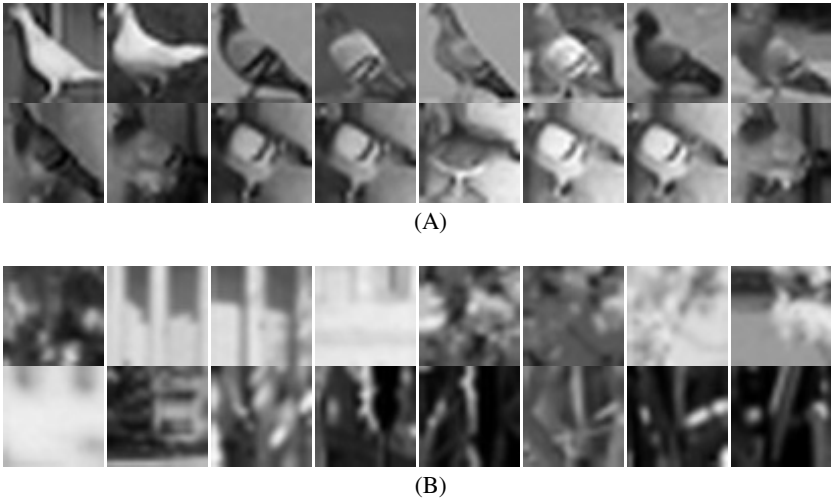
(C)

**Fig. 6.** (A) A sample image when  $i=10$  (B) The binary image when  $i=10$  with the highlighted edge of the object (C) The result of object extraction from the binary image

## 4 Experimental Result

The training process in our experiment totally uses 600 samples, including 200 positive and 400 negative samples where the main positive samples are profiles of birds, and negative samples are common backgrounds as shown in Fig. 7. The image size of all samples is fixed to  $19 \times 19$  pixels.





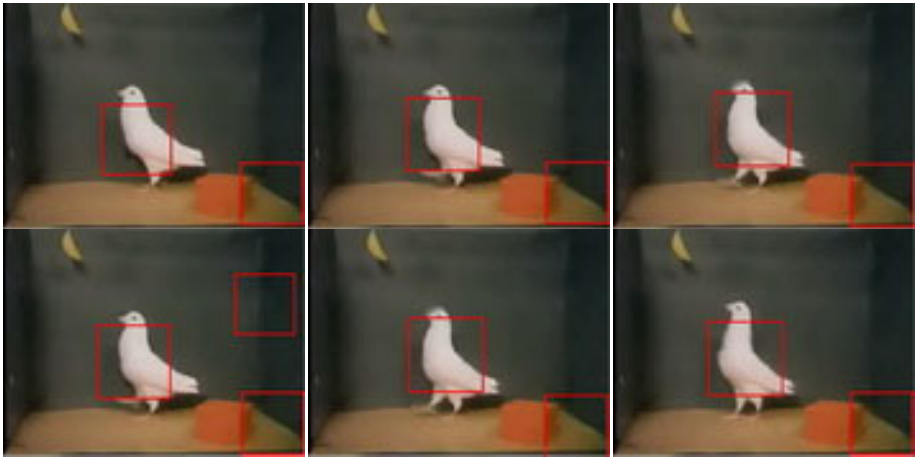
**Fig. 7.** (A) Positive samples (B) Negative samples

The test video with its resolution 320x240 is captured by a fixed camera, so that no calibration is needed. There are about 2000 frames used in this experiment as Fig. 8 shows parts of this video.

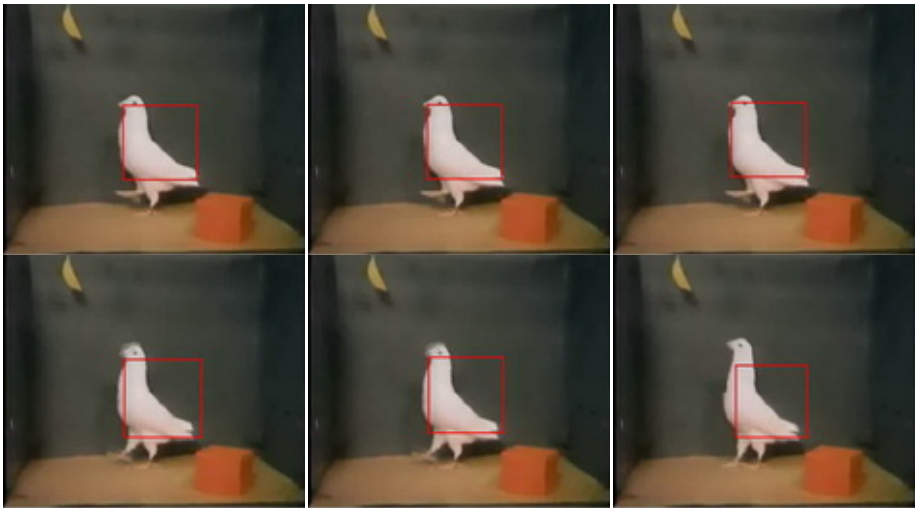


**Fig. 8.** 12 frame images used as part of the input

The detection is carried out with or without the motion detection preprocess. As shown in Fig. 9, the first part (A) is the detection result without the motion detection preprocess, while the second part (B) is that with the motion detection preprocess. In part (A) there exists a false positive detection, but this mistake is corrected when the strong classifier is applied to the preprocessed images.



(A)



(B)

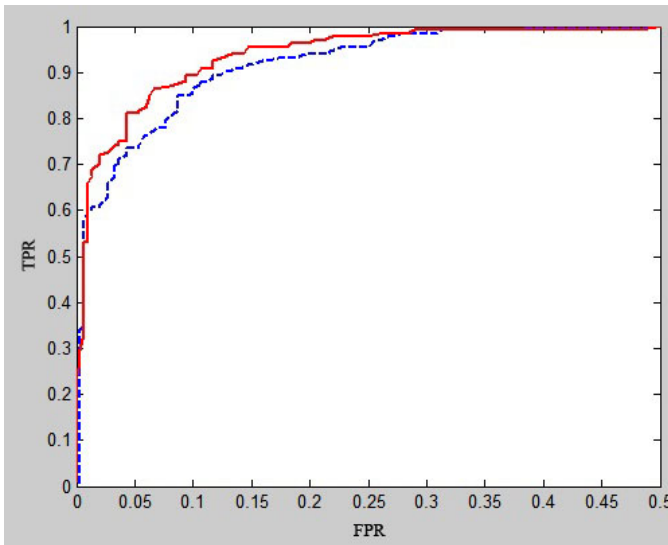
**Fig. 9.** (A) Detection results without motion preprocess (B) Detection results with motion preprocess

In order to evaluate the detection performance of the proposed scheme, two standard metrics, namely TPR (True Positive Rate) and FPR (False Positive Rate) are defined as follows.

$$TPR = \frac{TP}{TP + FN} \quad (2)$$

$$FPR = \frac{FP}{TN + FP} \quad (3)$$

Here, TP(True Positive) denotes the number of positive samples which are classified as birds, while FP(False Positive) denotes the number of negative samples which are classified as birds. TN(True Negative) denotes the number of negative samples which are classified as background, and FN(False Negative) denotes the number of positive samples which are classified as background. According to equation (2) and (3), the values of TPR and FPR can be calculated for each input test sample, and represented in Y-axis and X-axis respectively. The trajectories of the (TPR, FPR) of (A) and of (B) are ROC(Receiver Operating Characteristic) curves as shown in Fig. 10. The blue curve is based on the four basic Haar-like features and the red curve represents the performance of the proposed extended set of Haar-like features. The more the curve deflects from the center and approaches to the top-left, the higher the detection rate is.



**Fig. 10.** Detection performance ROC curves. (The blue-dotted curve is for basic Haar-like features, and the red-solid curve is for adding the extended features.).

## 5 Conclusion

This work focuses on special object detection using extending Haar-like features and a motion detection process to improve, in terms of detection rate, the performance of the approach of Viola and Jones. According to the results of the experiment, it is observed that the proposed Haar-like features adaptive to bird detection obviously achieves a better detection rate as the ROC curves shows. With this experiment, we conclude that the four basic Haar-like features which Viola and Jones defined can generically be used for detecting varieties of objects, but for more accurate object detection, our suggestion is to design some Haar-like features aptly that can capture individual features or outlooks of this object.

In addition, the preprocessing on the input image by the motion process separates objects from background. Hence, if the separation is accurate, the strong classifier needs not to scan the whole image. Consequently, the computation time will be less and the FP number can be reduced significantly. Finally, for a particular object, it is difficult to come up with an ‘optimal’ set of features. Thus, it is interesting to further study the capacity of AdaBoost on feature selection, which may provide an insight into the design of apt feature set for object detection.

## References

1. Rasolzadeh, B., Petersson, L., Pettersson, N.: Response Binning: Improved Weak Classifiers for Boosting. In: IEEE Intelligent Vehicles Symposium, Tokyo, Japan (2006)
2. Weber, B.: Generic Object Detection using AdaBoost. UCSC Technical Report (2008)
3. Lienhart, R., Kuranov, A., Pisarevsky, V.: Empirical Analysis of Detection Cascades of Boosted Classifiers for Rapid Object Detection. MRL Technical Report (2002)
4. Lienhart, R., Maydt, J.: An Extended Set of Haar-like Features for Rapid Object Detection. In: IEEE ICIP (2002)
5. Viola, P., Jones, M.J.: Robust Real-Time Face Detection. *International Journal of Computer Vision* (2004)
6. Viola, P., Jones, M.J.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: IEEE CVPR (2001)
7. Gams, S., Kégl, B., Esma Aïmeur, E.: Privacy-preserving boosting. *Data Mining and Knowledge Discovery* 14, 131–170 (2007)
8. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences* (1997)

# A Non-blind Digital Image Watermarking Method Based on the Dyadic Wavelet Transform and Interval Arithmetic

Teruya Minamoto\* and Ryuji Ohura

Department of Information Science, Saga University, Saga, Japan  
{minamoto,ohura}@ma.is.saga-u.ac.jp

**Abstract.** We propose a new non-blind digital image watermarking method based on the dyadic wavelet transform (DYWT) and interval arithmetic (IA). Since the DYWT has a redundant representation, the amount of information that the watermark must carry is higher than in discrete-wavelet-transform-based methods. Our watermark is a binary logo that is embedded into the high-frequency components by using DYWT and IA. We describe our watermarking procedure in detail and show experimental results demonstrating that our method gives watermarked images that have better quality and that are robust against attacks such as marking, clipping, contrast tuning (MATLAB `histeq` and `imadjust` commands), addition of Gaussian white noise, addition of salt & pepper noise, and JPEG and JPEG2000 compressions.

**Keywords:** Digital watermarking, Dyadic wavelet transform, Interval arithmetic.

## 1 Introduction

Digital watermarking is one of the main security techniques for embedding secret information in images or audio streams, and research into digital image watermarking has attracted a great deal of interest in recent years. Watermarking schemes can also be roughly categorized into two types: non-blind methods, which require the original image in the detection process, and blind ones, which use neither the original image nor the watermark in the detection process. In general, blind methods are more useful than non-blind ones because the original image may not be available in actual applications. However, blind methods have lower robustness and are more complicated to implement than non-blind ones. Moreover, non-blind methods sometimes lead to the development of blind ones. Therefore, it is important to develop not only blind methods but also non-blind ones.

Many non-blind methods based on the transform domain, such as the Discrete Cosine Transform (DCT) and Discrete Wavelet Transform (DWT), or schemes

---

\* Corresponding author.

based on Singular Value Decomposition (SVD) have been proposed [1–3]. However, methods based on the Dyadic Wavelet Transform (DYWT) are rare, and only a few audio watermarking methods have been proposed [9], to the best of the authors' knowledge. Since DYWT has a redundant representation and can be implemented with a fast algorithm, the amount of information that the watermark must carry is higher than in DWT-based methods, and the execution time is almost the same as DWT-based methods. Therefore, DYWT is a promising approach for developing new robust watermarking techniques.

This paper proposes a new non-blind digital image watermarking method based on DYWT. Our method is a combination of DYWT and interval arithmetic (IA) [7] and is an extension of the watermarking method proposed in Ref. [5] where the DWT and IA are used. Experimental results demonstrate that our method gives watermarked images that have better quality and that are robust against attacks, such as marking, clipping, contrast tuning (e.g., MATLAB `histeq` and `imadjust` commands), addition of Gaussian white noise and salt & pepper noise, and JPEG and JPEG2000 compressions.

The remainder of this paper is organized as follows: in Section 2, we briefly describe the basics of IA. In Section 3, we introduce the DYWT based on IA, and in Section 4, we propose a new non-blind digital watermarking method. Experimental results are presented in Section 5, and Section 6 concludes the paper.

## 2 Interval Arithmetic (IA)

An interval  $A$  is a connected subset of  $\mathbb{R}$ , where  $\mathbb{R}$  denotes the set of real numbers. Recall that the closed interval denoted by  $[a_1, a_2]$  is the set of real numbers given by  $\{t | a_1 \leq t \leq a_2, a_1, a_2 \in \mathbb{R}\}$ . In this paper, the term “interval” will mean closed interval. We denote the lower and upper bounds of interval  $A$  by  $\inf(A) = a_1$  and  $\sup(A) = a_2$ , respectively, and the width of any non-empty interval  $A$  is defined by  $w(A) = a_2 - a_1$ . The four basis operations, namely, addition (+), subtraction (−), multiplication (\*), and division (/), on two intervals  $A = [a_1, a_2]$  and  $B = [b_1, b_2]$  are defined as follows:

$$\begin{aligned} A + B &= [a_1 + b_1, a_2 + b_2], \quad A - B = [a_1 - b_2, a_2 - b_1], \\ A * B &= [\min\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}, \max\{a_1b_1, a_1b_2, a_2b_1, a_2b_2\}] \quad (1) \\ A/B &= [a_1, a_2] * [1/b_2, 1/b_1], \quad 0 \notin B. \end{aligned}$$

For interval vectors and matrices whose elements consist of intervals, these operations are executed at each element.

From the basis operations (1), in general, the width of an interval expands in proportion to the number of computations.

## 3 Dyadic Wavelet Transform (DYWT) Based on IA

Let us denote the original images by  $C^0$ . It is well known from Ref. [4, 6] that the DYWT for images with a quadratic spline wavelet and scaling function is given by:

$$\begin{aligned}
 C^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} h[k]h[l]C_{k,l}^j[m, n], & D^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} g[k]h[l]C_{k,l}^j[m, n] \\
 E^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} h[k]g[l]C_{k,l}^j[m, n], & F^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} g[k]g[l]C_{k,l}^j[m, n].
 \end{aligned}
 \tag{2}$$

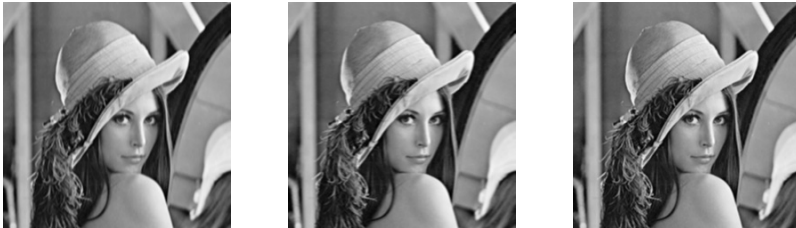
Here,  $C_{k,l}^j[m, n] = C^j[m+2^j k, n+2^j l]$ ,  $h$  is a low-pass filter,  $g$  is a high-pass filter, and  $\tilde{h}$  and  $\tilde{g}$  are these dual filters. More precisely,  $C^j[m, n]$ ,  $D^j[m, n]$ ,  $E^j[m, n]$ , and  $F^j[m, n]$  indicate the low-frequency components and high-frequency components in the horizontal, vertical, and diagonal directions, respectively. The indices  $m$  and  $n$  are the locations in the horizontal and vertical directions, respectively.

To accelerate the interval expansion, we define the interval Dyadic wavelet transform (IDYWT) based on (2) by

$$\begin{aligned}
 IC^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} \tilde{\Delta}_{k,l} \tilde{C}_{h_k, h_l}^j[m, n], & ID^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} \tilde{\Delta}_{k,l} \tilde{C}_{g_k, h_l}^j[m, n] \\
 IE^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} \tilde{\Delta}_{k,l} \tilde{C}_{h_k, g_l}^j[m, n], & IF^{j+1}[m, n] &= \sum_{k,l=-\infty}^{+\infty} \tilde{\Delta}_{k,l} \tilde{C}_{g_k, g_l}^j[m, n],
 \end{aligned}
 \tag{3}$$

where  $\tilde{C}_{h_k, g_l}^j[m, n] = h[k]g[l]C_{k,l}^j[m, n]$ ,  $\tilde{\Delta}_{k,l} = [1 - \Delta_{k,l}, 1 + \Delta_{k,l}]$  and  $\Delta_{k,l}$  are positive real numbers. All operations are executed using IA, and then  $C^{j+1} \subset IC^{j+1}$ ,  $D^{j+1} \subset ID^{j+1}$ ,  $E^{j+1} \subset IE^{j+1}$ , and  $F^{j+1} \subset IF^{j+1}$  hold. Thus, the components obtained by using IDYWT contain the ones obtained by using DYWT. In this sense, each component with IDYWT is a redundant set from the original signal.

Moreover, high-frequency components obtained by using IDYWT contain a low-frequency component in our experience. Fig. 1 demonstrates this situation. Therefore, we may expect that the robustness is maintained even if we embed the watermark into the high-frequency components.



**Fig. 1.** Low-frequency components contained in high-frequency components (From the left :  $\sup(ID^1) - \inf(ID^1)$ ,  $\sup(IE^1) - \inf(IE^1)$  and  $\sup(IF^1) - \inf(IF^1)$ )

## 4 Watermarking Algorithm

We assume that the binary-valued watermark  $W$  consists of values  $-1$  and  $1$ . To simplify the remainder of the discussion, we discuss only the DYWT from level 0 to 1.

The embedding procedure is as follows:

1. Apply IDYWT to the horizontal and vertical directions. Then, obtain the four frequency components corresponding to  $C^1$ ,  $D^1$ ,  $E^1$ , and  $F^1$  in (2) and represent these interval components as  $IC^1$ ,  $ID^1$ ,  $IE^1$ , and  $IF^1$ , respectively.
2. Choose some components of the four interval components with IDYWT, and represent the chosen components as  $I(S_i)(i = 1, 2, \dots, N, 1 < N < 4)$ .
3. Set  $S'_i = \sup(I(S_i))$ .
4. Replace other components with floating point ones.
5. Embed the watermark  $W$  by computing

$$\tilde{S}_i(m, n) = S'_i(m, n) \left( 1 + \alpha W(m, n) x(m, n) \right), \quad (4)$$

where  $0 < \alpha < 1$  is a given hiding factor that adjusts the robustness, and  $x(m, n)$  is a pseudo-random binary sequence.

6. Reconstruct the image using  $\tilde{S}_i$ , the components that were replaced with floating point ones, and the inverse DYWT. Then, the watermarked image  $\tilde{C}^0$  is obtained.

In the extraction procedure, we need the original image  $C^0$ , the width of interval  $\Delta_k$ , the pseudo-random binary sequence, and information about the components where the watermark is embedded. The extraction procedure is as follows:

1. Decompose  $\tilde{C}^0$  into the four components  $\sup(\tilde{IC}^1)$ ,  $\sup(\tilde{ID}^1)$ ,  $\sup(\tilde{IE}^1)$ , and  $\sup(\tilde{IF}^1)$  by using IDYWT and choose the watermarked components  $\tilde{S}_i$ .
2. Decompose  $C^0$  into the four components  $\sup(IC^1)$ ,  $\sup(ID^1)$ ,  $\sup(IE^1)$ , and  $\sup(IF^1)$  by using IDYWT and choose the watermarked components  $S'_i$ .
3. Compute  $\tilde{W}_i = |\tilde{S}_i| - |S'_i|$ ,  $\tilde{W}_e = \sum_{i=1}^N \tilde{W}_i$ , and  $\tilde{W} = \text{sgn}(\tilde{W}_e) \cdot x$ , where  $\text{sgn}(a)$  is the usual signum function of a real number  $a$ .

## 5 Experimental Results

To evaluate the performance of the proposed method, we adopted standard 256-grayscale images of size  $256 \times 256$  pixels, namely Airplane, Building, Lenna, and Woman, and a binary watermark of size  $256 \times 256$  pixels. These images are not included in the present paper due to space limitations. We implemented our method using INTLAB [8], which is a MATLAB toolbox that supports interval arithmetic.



In this experiment, we set the parameters  $\Delta_k = 0.01$  in (3) and  $\alpha = 0.9$  in (4) so as not to considerably decrease the peak signal to noise ratio (PSNR), expressed in decibels, which is computed by

$$PSNR = 20 \log_{10} \left( 255 / \sqrt{\frac{1}{N_x N_y} \sum_{i=1}^{N_x} \sum_{j=1}^{N_y} (C^0(i, j) - \widetilde{C}^0(i, j))^2} \right),$$

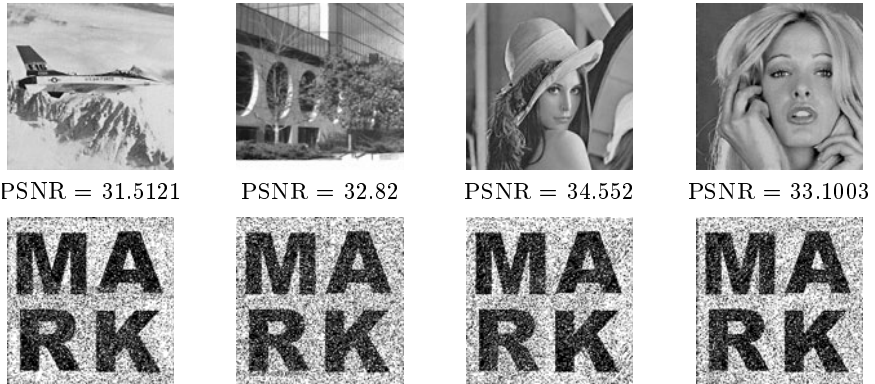
where  $N_x$  and  $N_y$  are the sizes of the image in the horizontal and vertical directions, respectively. Table 1 shows some of preliminary experimental results to decide these parameters, and also does the influence on the width of intervals.

**Table 1.** PSNRs when  $\alpha = 0.9$

	$\Delta_{k,l}=0.006$	$\Delta_{k,l}=0.008$	$\Delta_{k,l}=0.01$	$\Delta_{k,l}=0.012$	$\Delta_{k,l}=0.014$
Airplane	31.9949	31.7769	31.5121	31.2123	30.8849
Building	33.1742	33.0127	32.82	32.5987	32.3534
Lenna	35.0621	34.831	34.552	34.2384	33.8928
Woman	33.5192	33.3276	33.1003	32.8398	32.5537

Moreover, we embedded the watermark into two components,  $ID^1$  and  $IE^1$ .

Fig. 2 shows the watermarked images, together with PSNRs, obtained by the proposed method and the watermarks extracted from the watermarked images without any attack. These watermarked images have better quality, and the inserted watermark is invisible to the naked eye.



**Fig. 2.** Watermarked images and extracted watermarks without any attack

Figs. 3-8 illustrate watermarked images and the extracted watermarks under attacks such as marking, clipping, contrast tuning (MATLAB histeq and

imadjust commands), and addition of Gaussian white noise and salt & pepper noise. (The histeq command enhances the contrast using histogram equalization, and the imadjust command adjusts the image intensity values.) Although the extracted images were degraded, we were able to identify the existence of the watermark at a single glance. For clipped images, we decompose the original image and the clipped image using the DYWT to extract the watermark from the clipped image. Let  $C(m, n)$  and  $CC(m, n)$  be the low-frequency components of the original image and the clipped image, respectively. At the position  $(ox, oy)$ , we compute the mean square error  $\delta(ox, oy) = \sum_{x=0}^{m-1} \sum_{y=0}^{m-1} (C(ox + x, oy + y) - CC(x, y))^2$  and find the minimum point  $(mx, my)$ . After that, we extract the pseudo-random binary sequence corresponding to the clipped image based on  $(mx, my)$  and carry out the extraction procedure using this pseudo-random binary sequence.

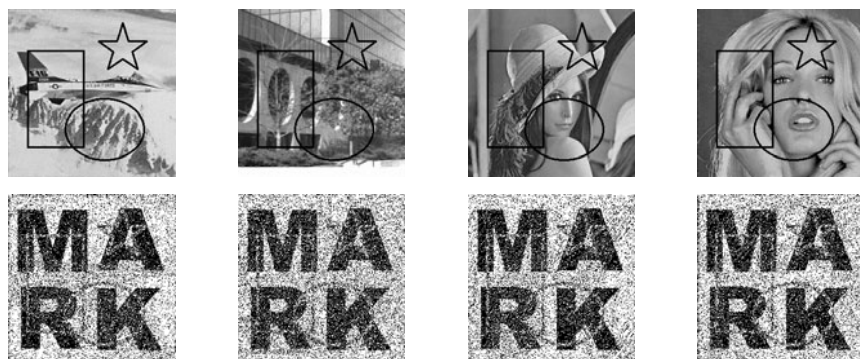


Fig. 3. Watermarked images with marked areas and extracted watermarks



Fig. 4. 190×190 fragment of watermarked images and extracted watermarks



Fig. 5. Images after applying MATLAB histeq command and extracted watermarks



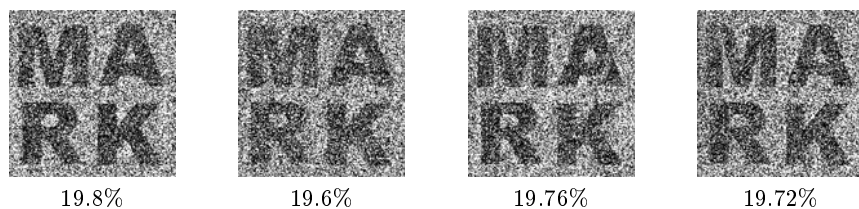
Fig. 6. Images after applying MATLAB imadjust command and extracted watermarks



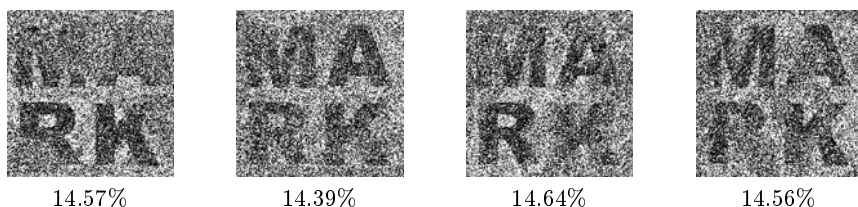
Fig. 7. Watermarked images with added Gaussian white noise (mean: 0; variance: 0.0005) and extracted watermarks



**Fig. 8.** Watermarked images with added salt & pepper noise (noise density: 0.005) and extracted watermarks



**Fig. 9.** Watermarks extracted from watermarked JPEG images (from the left: Airplane, Building, Lenna, and Woman)



**Fig. 10.** Watermarks extracted from watermarked JPEG2000 images (from the left: Airplane, Building, Lenna, and Woman)

Figs. 9 and 10 illustrate the watermarks extracted from the watermarked images under JPEG and JPEG2000 attacks. We are able to barely identify the existence of the watermark. These results are superior to those of the method described in Ref. [5], which does not use a pseudo-random binary sequence.

## 6 Conclusion

We proposed a new non-blind watermarking method based on the DYWT and IA. Experimental results demonstrate that our method gives better-quality watermarked images and is robust against marking, clipping, contrast tuning, addition of Gaussian white noise, addition of salt & pepper noise, and JPEG and JPEG2000 compressions. These results show that the proposed method is superior to our previous methods [5] and the methods proposed in Ref. [2]. It is beyond the scope of this paper to examine the effect of rotation, translation, or scaling attacks, and the development of a blind watermarking technique that is robust against these attacks remains a topic for future study. Our method is expected to find uses in image authentication (Ref. [1], Chapter 11), because DYWT is shift-invariant and the size of each frequency component is the same as in the original image. These properties will make it easier to detect tampered sections of data with higher precision.

**Acknowledgments.** This work was supported by KAKENHI (23540145).

## References

1. Cox, I.J., Miller, M.L., Bloom, J.A., Fridrich, J., Kalker, T.: Digital Watermarking and Steganography. Morgan Kaufmann Publishers (2008)
2. Devapriya, M., Ramar, K.: Non-Blind Image Watermarking Scheme using DWT-SVD Domain. International Journal of Computer Science and Information Security 8(1), 222–228 (2010)
3. Frattolillo, F.: Web-oriented nonblind image watermarking procedure. Journal of Electronic Imaging 15(3), 033011-1–033011-14
4. Mallat, S.: A wavelet tour of signal processing. Academic Press (2009)
5. Minamoto, T., Yoshihara, M., Fujii, S.: A digital Image Watermarking Method Using Interval Arithmetic. IEICE Trans. Fundamentals E90-A(12), 2949–2951 (2007)
6. Minamoto, T., Tsuruta, K., Fujii, S.: Edge-preserving Image Denoising Method Based on Dyadic Lifting Schemes. IPSJ Transactions on Computer Vision and Applications 2, 48–58 (2010)
7. Moore, R.E., Kearfott, R.B., Cloud, M.J.: Introduction to Interval Analysis. SIAM (2009)
8. Rump, S.M.: INTLAB - interval Laboratory, <http://www.ti3.tuharburg.de/rump/intlab/>
9. Wang, Y., Wu, S., Huang, J.: Audio Watermarking Scheme Robust against Desynchronization Based on the Dyadic Wavelet Transform. EURASIP Journal on Advances in Signal Processing 2010, Article ID 232616, 17 pages (2010)

# Song Classification: Classical and Non-classical Discrimination Using MFCC Co-occurrence Based Features

Arijit Ghosal<sup>1</sup>, Rudrasis Chakraborty<sup>2</sup>,  
Bibhas Chandra Dhara<sup>3</sup>, and Sanjoy Kumar Saha<sup>4</sup>

<sup>1</sup> CSE Dept., Institute of Technology and Marine Engg.  
24 Parganas (South), West Bengal, India  
ghosal.arijit@yahoo.com

<sup>2</sup> Indian Statistical Institute, Kolkata, India  
rudrasischa@gmail.com

<sup>3</sup> IT Dept., Jadavpur University, Kolkata, India  
bcdhara@gmail.com

<sup>4</sup> CSE Dept., Jadavpur University, Kolkata, India  
sks\_ju@yahoo.co.in

**Abstract.** In the context of music information retrieval, genre based classification of song is very important. In this work, we have presented a scheme for automatic classification of song signal into two categories like classical and non-classical/popular song. Strong presence of beat and rhythm in the popular songs forms a distinctive pattern and high frequency sub bands obtained after wavelet decomposition bear the signatures. We have computed MFCC based features corresponding to the decomposed signals. Co-occurrence of individual Mel frequency co-efficient computed over a small period are studied and features are obtained to represent the signal pattern. RANSAC has been utilized as the classifier. Experimental result indicates the effectiveness of the proposed scheme.

**Keywords:** Song Classification, Music Retrieval, Audio Classification, MFCC, RANSAC.

## 1 Introduction

Automatic classification of music based on genre is a fundamental step for music retrieval system. It is quite subjective and difficult enough to define a genre. However, music signals of a genre share similar rhythmic patterns and pitch distributions [1]. Broadly a music signal can be classified into two categories like *classical* and *popular/ non-classical*. A non-classical category may include rock, jazz or pop songs. Thus, each high level category may be further classified leading to a hierarchical system. A lot of work have been reported in the context of speech processing, recognition, speech-music discrimination. But, much less work has been reported genre based classification of song or music.

An audio classification system consists of two important steps like *feature extraction* and *classification*. Characteristics like musical structure, rhythm, tempo, melody can be used to discriminate the genres. But, it is not easy to extract the features representing the characteristics. MFCC and Gaussian Mixture Model have been tried by Pye. Solatu et al. [2] have used temporal structures and HMM. Tempo and beat tracking have been explored in [3,4]. As a global representation for rhythm, beat spectrum has been presented in [5]. Tzanetakis [6] et al. have worked with three feature sets representing timbre, rhythm and pitch content. Grimaldi et al [7] have dealt with 143 dimensional features computed based on beat histogram and spectrum of 16 level wavelet decomposed signals. Octave-based spectral contrast have been tried in [8,9]. It considers spectral peak, spectral valley and their difference in different sub-bands. Statistical spectrum descriptors based on 24 critical bands and rhythm histogram features have been used by Lidy and Rauber [10]. For classification, various techniques have been tried by the researchers. Threshold based techniques [11], Neural Network [12] have been deployed for audio classification. A heuristic rule based scheme has been proposed in [13]. K-Nearest Neighbors and GMM have been used in [6,8]. Bayes decision rules [14], SVM based scheme [15] have also been tried.

It has been observed that comparatively less work has been done on genre based classification of music. In this work, we have assumed that music signal has already been classified as music without voice *i.e.* instrumental and music with voice *i.e.* song [16] and we concentrate on classifying the song as classical and non-classical one. The organization of the paper is as follows. The brief introduction is followed by the description of proposed methodology in section 2. Experimental results are presented in section 3. Concluding remarks are put in to section 4.

## 2 Proposed Methodology

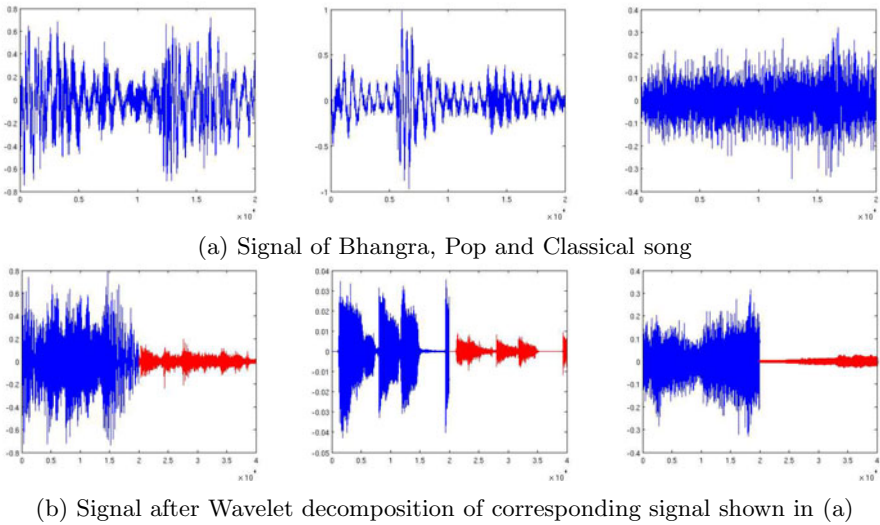
The proposed scheme deals with recorded signal of songs of several genre. We try to classify them into two categories namely, *classical* and *non-classical*. It has been observed that non-classical song like Pop, Jazz, Bhangra (a north Indian genre) are characterized by the frequent and strong presence of beat and it is well reflected in corresponding signals. We have tried to discriminate classical and non-classical song relying on this observation.

### 2.1 Extraction of Features

In the early efforts, it has been observed that researchers have considered collection of variety of features comprising mostly of ZCR, STE based measure and FFT/MFCC based spectral measures. A high dimensional feature vector is used as the signal descriptor. On the contrary, in our approach, we strongly rely on our observation regarding the beat pattern present in the genre. Thus, the features are also designed to capture the same. The feature extraction steps are as follows.

- Perform wavelet decomposition on the signal
- Compute MFCC for high sub-band
- Compute features based on MFCC co-occurrence matrix
- Perform feature reduction

In non-classical/popular songs, occurrence of beats give rise to sudden change of amplitude in the signal. Depending on the beat strength and frequency, change in amplitude forms a prominent signature for popular genres whereas such pattern is very weak in case of classical song as shown in Fig. 1(a). In order to focus on the pattern of amplitude variation, we have decomposed the signal using Haar wavelet transform [17]. The decomposed signals are shown in Fig. 1(b) where left part in each represents the low sub-band and high sub-band is in the right.



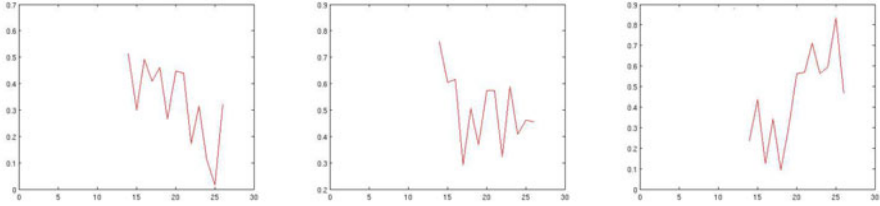
**Fig. 1.** Signal of different types of song and corresponding signal after wavelet decomposition

It may be noted that the low sub-band represents a smoothed version of the original signal whereas the amplitude variation pattern is well emphasized in the high sub-band. It has motivated us to focus only on the high sub-band for feature extraction.

Mel-frequency cepstral co-efficients are short term spectral based features used for speech recognition [18], speech/music discrimination [19], discrimination of music with voice or without voice [16], instrument type identification [20]. MFCC being a compact representation of amplitude spectrum, it can well capture the pattern expressed in the high sub-band signal. The high signal is first broken into number of equal sized frames and for each frame 13 dimensional MFCC



coefficients are computed following the methods elaborated in [21]. MFCC obtained after taking average over all frames have been shown in Fig. 2 for different types of song.



**Fig. 2.** MFCC plot for high sub-band signal of different song type: a) Bhangra b) Pop c) Classical

From Fig. 2, it is apparent that the co-efficient plots are different for classical and non-classical song. In case of classical song, strength of the co-efficients vary over a wide range and higher order co-efficients are quite prominent. For the non-classical songs the co-efficients are of moderate strength and confined within a range. It seems that the intervention of significant and periodic beats have smoothened the co-efficients which was absent in case of classical song. 13 dimensional MFCC is taken as the feature. But it provides a general description. In order to obtain a detailed representation, we have further designed additional features based on the co-occurrences of each co-efficient over the frames.

The occurrence pattern of each co-efficient in MFCC in different frames provides detailed description of the signal. Depending on the beat frequency, the strength of a co-efficient may vary in the frames. Thus, the variation pattern of the same have a correspondence with the beat pattern. It has motivated us to look for co-occurrence of the co-efficient values in the neighboring frames. First of all, the co-efficients values are quantized considering  $\mu_i \pm k * \sigma_i$  as different quantization levels.  $\mu_i$  and  $\sigma_i$  are the mean and standard deviation of  $i$ -th co-efficient of all the frames. In our experiment  $k$  varies from  $-2$  to  $+2$  with  $.25$  as step size. Thus, co-efficient values are mapped on to 16 different bins. For each co-efficient, a  $16 \times 16$  co-occurrence matrix,  $M_i$  is formed where an element denotes number of occurrences of a particular value (bin) pair of the  $i$ -th co-efficient in two successive frames. Finally, co-occurrence matrix based statistical measures [22] namely *energy*, *entropy*, *homogeneity*, *contrast*, *correlation* are computed which represent the pattern of the co-efficients. Thus, for 13 co-efficients altogether 65 such features are obtained. These features along with the co-efficients form 78 dimensional features.

The feature vector of dimension 78 is high enough. Moreover, all the elements may not contribute significantly towards the discriminating process. Hence, Principal Component Analysis(PCA) is carried to reduce the dimensionality of the feature vector. Top 20 features are considered and used in subsequent classification technique.

## 2.2 Classification

In audio classification, the variety present in the database poses the major challenge. Even though, we are dealing with only songs, a single genre may offer sufficient variation. In this work, the non-classical category includes multiple genre that further adds to the variety.

Neural network or SVM based classification schemes are quite popular. But, in the problem under consideration, variation present even within a class poses problem for NN based classification. SVM is robust but the tuning of parameters for optimal performance is very critical. It has motivated us to look for a robust estimator capable of handling the diversity of data and can model the data satisfactorily. RANdom Sample And Consensus (RANSAC) [23] appears as a suitable alternative to fulfill the requirement.

RANSAC is an iterative method to estimate the parameters of a certain model from a set of data contaminated by large number of outliers. The major strength of RANSAC over other estimators lies in the fact that the estimation is made based on inliers *i.e.* whose distribution can be explained by a set of model parameters. It can produce reasonably good model provided a data set contains a sizable amount of inliers. It may be noted that RANSAC can work satisfactorily even with outliers amounting to 50% of entire data set. Using a subset of each category of data, model parameters are first estimated and the evolved model is used for the subsequent classification.

## 3 Experimental Results

In order to carry out the experiments, we have formed a song database consisting of 232 audio files. Each file contains audio of 30-45 second duration. 113 files correspond to non-classical song like Bhangra (Indian genre), Pop, Jazz. Remaining 119 files correspond to Indian and Western classical song. The database thus reflects considerable variety. Sampling frequency for the data is 22050 Hz. Samples are of 16-bits and of type mono.

Apart from RANSAC, we have also considered multi-layer perceptron (MLP) network based and SVM based classification. In MLP we have considered a single hidden layer with 11 nodes. There are 20 nodes and 2 nodes in the input and output layer respectively. For SVM, We have considered RBF kernel. For all the classification schemes, 50% of each category of data set has been used for training (estimating the model in case of RANSAC) and remaining data has been used for testing. The experiment is repeated once again by reverting the test and training data. Average classification accuracy has been shown in Table 1.

Table 1 clearly shows the inclusion of MFCC co-occurrence matrix based features improves the accuracy and the improvement is quite pronounced in case of SVM and RANSAC based classification. It may be further noted that RANSAC has outperformed the rest. The experiment clearly indicates the effectiveness of the proposed methodology.

**Table 1.** Song Classification Accuracy (in %)

Classific. Scheme	Only MFCC Co-eff.		MFCC Co-eff. and Co-occ. Matrix based features (after PCA)	
	Non-Classical	Classical	Non-Classical	Classical
MLP	70.80	74.89	75.22	76.67
SVM	72.57	83.19	88.50	89.09
RANSAC	84.96	85.71	93.81	90.76

## 4 Conclusion

In this work, We have presented a novel scheme to automatically classify classical and Non-classical song. Presence of strong beat in non-classical song has been utilized in classifying the song. Features have been derived based on high sub-band signal obtained after wavelet decomposition as it prominently carries the discriminating pattern. MFCC of the high sub-band signal taken as the broad features and detailed features are computed based on the co-occurrence pattern of the co-efficients over the neighboring frames. PCA is performed to bring down the feature dimension to 20. Finally, RANSAC has been utilized as the classifier. Experimental result clearly indicates the utility of the proposed feature and the classification scheme.

**Acknowledgment.** The work is partially supported by the facilities created under DST-PURSE program in Computer Science and Engineering Department of Jadavpur University, India.

## References

1. Dowling, W.J., Harwood, D.L.: Music Cognition. Academic Press, Inc. (1986)
2. Solatu, H., Schultz, T., Westphal, M., Waibel, A.: Recognition of music types. In: IEEE Conf. on Acoustics, Speech and Signal Processing, pp. 1137–1140 (1998)
3. Scheirer, E.: Tempo and beat analysis of acoustic musical signals. *Journal of Acoustical Society of America* 103 (1998)
4. Laroche, J.: Estimating tempo, swing and beat locations in audio recordings. In: Workshop on Appln. of Sig. Proc. to Audio and Acoustics, WASPAA (2001)
5. Foote, J.T., Uchihashi, S.: The beat spectrum: A new approach to rhythmic analysis. In: ICME (2001)
6. Tzanetakis, G., Cook, P.: Music genre classification of audio signals. *IEEE Trans. on Speech Audio Processing* 10, 293–302 (2002)
7. Grimaldi, M., Cunningham, P., Kokaram, A.: An evaluation of alternative feature selection strategies and ensemble techniques for classifying music. In: Workshop on Multimedia Discovery and Mining (2003)
8. Jiang, D.N., Lu, L., Zhang, H.J., Tao, J.H., Cai, L.H.: Music type classification by spectral contrast feature. In: ICME (2002)

9. Lee, C.H., Shih, J.L., Yu, K.M., Su, J.M.: Automatic music genre classification using modulation spectral contrast feature. In: ICME, pp. 204–207 (2007)
10. Lidy, T., Rauber, A.: Evaluation of feature extractor and psychoacoustics transformations for music genre classification. In: Int. Conf. ISMIR, pp. 34–41 (2005)
11. Saunders, J.: Real-time discrimination of broadcast speech/music. In: IEEE Conf. on Acoustics, Speech, Signal Processing, pp. 993–996 (1996)
12. Matityaho, B., Furst, M.: Classification of music type by a multilayer neural network. *Journal of the Acoustical Society of America* 95 (1994)
13. Zhang, T., Kuo, C.C.J.: Audio content analysis for online audiovisual data segmentation and classification. *IEEE Trans. on Speech and Audio Processing* 3, 27–36 (2001)
14. Eronen, A.: Comparison of features for musical instrument recognition. In: IEEE Workshop Appl. Signal Process. Audio Acoust., pp. 19–22 (2001)
15. Sadjadi, S.O., Ahadi, S.M., Hazrati, O.: Unsupervised speech/music classification using one-class support vector machines. In: ICICS (2007)
16. Ghosal, A., Chakraborty, R., Dhara, B.C., Saha, S.K.: Instrumental/song classification of music signal using ransac. In: 3rd Intl. Conf. on Electronic Computer Technology. IEEE CS Press (2011)
17. Gonzalez, C.R., Woods, E.R.: *Digital Image Processing*, 3rd edn. Prentice-Hall, NJ (2006)
18. Walker, W., Lamere, P., Kwok, P., Raj, B., Gouvea, E., Wolf, P., Woelfel, J.: Sphinx-4: A flexible open source framework for speech recognition (2004), [cmusphinx.sourceforge.net/sphinx4/doc/Sphinx4Whitepaper.pdf](https://cmusphinx.sourceforge.net/sphinx4/doc/Sphinx4Whitepaper.pdf)
19. Logan, B.: Mel frequency cepstral co-efficients for music modelling. In: Intl. Symposium on Music Information Retrieval (2000)
20. Ghosal, A., Chakraborty, R., Dhara, B.C., Saha, S.K.: Automatic Identification of Instrument Type in Music Signal using Wavelet and MFCC. In: Venugopal, K.R., Patnaik, L.M. (eds.) ICIP 2011. CCIS, vol. 157, pp. 560–565. Springer, Heidelberg (2011)
21. Rabiner, L.R., Juang, B.H.: *Fundamentals of Speech Recognition*. Prentice-Hall (1993)
22. Haralick, R.M., Shapiro, L.G.: *Computer and Robot Vision*, vol. I. Addison-Wesley (1992)
23. Fischler, M.A., Bolles, R.C.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *ACM Communications* 24, 381–395 (1981)

# Gabor Based Gender Classification with Classifier Independent Feature Selection

Aun Irtaza<sup>1</sup>, M. Arfan Jaffar<sup>1,2</sup>, and Tae-Sun Choi<sup>2</sup>

<sup>1</sup> National University of Computer & Emerging Sciences, Islamabad, Pakistan

<sup>2</sup> Gwangju Institute of Science and Technology, Korea

aun.irtaza@gmail.com, {arfanjaffar, tschoi}@gist.ac.kr

**Abstract.** The study presents an efficient gender classification technique. The gender of a facial image is the most prominent feature, and improvement in the existing gender classification methods will result in the high performance of the face retrieval and classification methods for large repositories. The method presented in this paper selects the effective set of Gabor features which are ranked on the base of entropy and are merged with mean Gabor feature values. This forms an effective feature vector to be used by the classifiers. The method presented in this paper enjoys the features like high speed and low space requirements. Proposed technique has been tested on different datasets. It shows good performance as comparative to existing techniques.

**Keywords:** Gabor features, Entropy, Classification, Feature reduction, Feature selection.

## 1 Introduction

Gender classification problem is an active and challenging research area which has attracted a great deal of attention recently. Gender classification techniques aims to determine the gender of a subject from face images. The face images analysis plays an important role in computer vision, e.g. it has been successfully used in many applications i.e. automated security/surveillance systems, demographic studies, safety monitoring systems, biometric and human-robotic interaction.

Most challenging task in gender classification is the representation of the face image in term of a vector. This vector provides input to a trained classifier and classifies the face images in gender classes. The overall process can be summarized in two main points (1) Effective feature extraction. (2) Pattern recognition through trained classifiers.

A number of different techniques based on facial images have been reported in the literature for solving this problem. These techniques include geometrical feature based methods [17], graph matching methods [16], support vector machine [14] and neural network based methods [15]. These approaches are classified into feature based and feature and template based techniques.

Gabor filtering is a feature based method of extracting texture information. The use of Gabor filters is motivated by Gabor filtering being strongly correlated with the

human visual system. [18] Gabor filtering is a way of extracting feature information from an image in the form of a response image. Several filters with varying parameters are applied to an image to acquire the response. Then from those response images features are extracted and a feature vector is developed which is then used for the classification purposes.

In practice, the dimension of a Gabor feature vector is so high that the computation and memory requirements are very large. For this reason, several sampling methods have been proposed to determine the “optimal” subset for extracting Gabor features. This study proposes an effective Gabor feature selection to extract the informative Gabor features representing the facial characteristics for gender classification. Entropy is used as a criterion to measure the importance of the feature, and then selected features are combined with mean Gabor feature values. This approach reduces the feature dimension without losing much information and also decreases computation and storage requirements.

The overall process involves the following steps. Face portion of the image is extracted from the image using Viola & Jones [1], this helps in reducing the complexity of the proposed method. Then Histogram equalization is performed to reduce the effect of varying illumination. Images are then divided into 9x9 non overlapping regions, and for every region Gabor based odd feature components are computed as they do not produce the imaginary values [18]. Then from those Gabor based response images best features are selected using the entropy criteria.[19] These features are then combined with mean Gabor feature values and generated a feature vector. This feature vector is then passed to the classifiers to classify the images on the base of gender. The method is tested on Stanford university medical student’s frontal face images dataset.

Rest of this paper is structured as follow. In section 2 a brief description of the related work is provided. In section 3 the proposed method is discussed. Section 4 describes our experimentation and results followed by discussion on the results obtained and section 5 concludes the paper.

## 2 Related Work

Gender classification is addressed in computer vision from a long period of time, but researchers are still searching for improved gender classification.

Ziyi Xu et al. [8] came with hybrid face coding method for facial feature extraction. They have extracted geometry features and fused them with features extracted by AdaBoost algorithm to form a feature vector and used SVM classifier for classification. Baback et al. [5] applied the SVMs to gender classification with low-resolution thumbnail faces. The SVMs produced much better results then the techniques such as Radial Basis Function (RBF) classifiers and large ensemble-RBF networks. Zhiguang YANG et al. [7] presented an experimental study on automatic face gender classification that focused on the different texture normalization methods to the performances of gender classification by SVM, Linear Discriminant Analysis (LDA) and Real Adaboost.

Michael Lama, et.al [18] proposed a content based image retrieval system for computed tomography nodule images. Their system takes the input image and generates haralik, Gabor, and markov random field features. On the extracted features they perform Euclidian, Manhattan and Chebychev distances to find the relevant images from image database. This paper has made good contribution by effectively retrieving images from image database, so it is an effective guideline for gender classification problem as well. F. Scalzo, et.al [10] came with a new Feature Fusion Hierarchical (FFH) method for gender classification using Genetic algorithm. Their algorithm worked in two phases. In first phase, Gabor and Laplace features are extracted and used as input to feature fusion level. In second phase, classifiers fusion uses output of future fusion level to produce a result. Chien-Cheng Lee, et.al [19] proposed a facial expression recognition technique based on improved radial bases function network. Their technique is significant in a sense that they are using Gabor features for this purpose and they have introduced an efficient feature selection technique which is based on entropy criteria; the technique reduces the processing time and space requirement for the classification purposes.

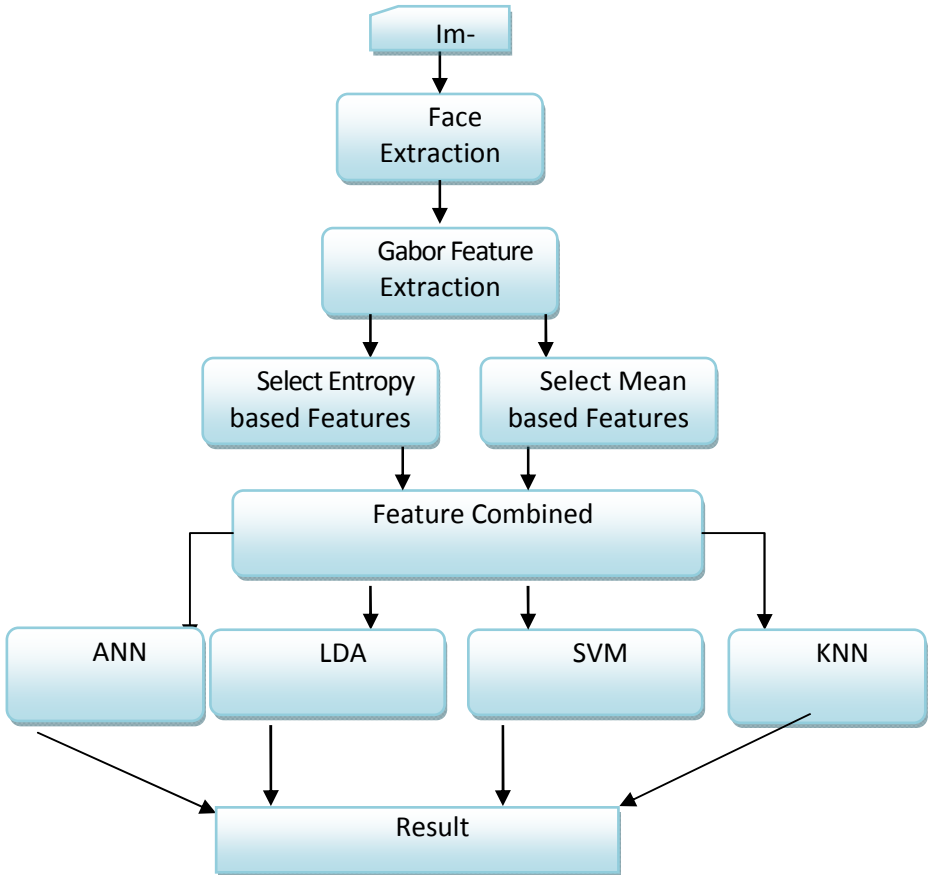


**Fig. 1.** A sample of SUMS face database

### 3 Proposed Method

The gender classification module which we presented here consists of three main steps: face detection, feature extraction/selection, and classification. An input facial image is passed to face detector to extract face from the image, Viola and Jones [1] face detection method is used for this purpose. To overcome illumination variation in images histogram equalization is performed as a next step. It stretches the contrast of the image. In [4] the authors showed that low-resolution images have equal level of classification accuracy, so we can decrease computational cost by reducing the size of the image. After face detection, the image is resized to 36x36. The resized image is divided into 9x9 non-overlapping regions. Then Gabor features are extracted with varying parameters. After the Gabor feature generation, we select the Gabor features

by ranking them on the base of entropy, and also calculated the mean Gabor feature values from Gabor response images against every pixel. These mean Gabor feature values are then combined with ranked feature values calculated on the base of entropy. These features are stored in vector and passed to the classifiers for gender classification. Figure 2 shows the general architecture of the proposed gender classification method.



**Fig. 2.** Architecture of Gender Classification

### 3.1 Face Detection

Viola and Jones [1] presented a well known and robust face detection technique. It works with an amazing speed to detect the frontal face images. This detector extract faces from the image by starting from top left corner and ending at bottom right corner of an image. The technique works in following way: First images are represented in the form of Integral Images, due to which feature computation becomes very fast. Then Adaboost learning algorithm for feature selection is applied, and



finally using a cascade of AdaBoost classifier to remove backgrounds and detect the prominent objects like image regions for detection of the faces. In figure 3 face extraction results are given of some selected images from the dataset.



**Fig. 3.** A sample of extracted faces from the images

### 3.2 Gabor Features

Gabor filtering is a transform based method of extracting texture information. The use of Gabor filters is motivated by Gabor filtering being “strongly correlated with the human visual system.”[18]. in literature Gabor features are frequently used for image retrieval and classification purposes. [10][18][19]. Gabor filter is a sinusoid function modulated by a Gaussian and extracts feature information from an image in the form of a response image by applying varying parameters. [18]. the filters we used are defined by the following equation: [18].

$$G(x, y) = e^{\left( \frac{x_{\theta}^2 - \gamma^2 y_{\theta}^2}{\alpha^2} + \frac{2\pi x \theta^i}{\lambda} \right)} \quad (1)$$

Where

$$x_{\theta} = x \cos \theta + y \sin \theta \quad (2)$$

$$y_{\theta} = -x \sin \theta + y \cos \theta \quad (3)$$

And  $\sigma$  is the standard deviation of the Gaussian function,  $\lambda$  is the wavelength of the harmonic function,  $\theta$  is the orientation, and  $\gamma$  is the spatial aspect ratio which is left constant at 0.5. The spatial frequency bandwidth is the ratio  $\sigma/\lambda$  and is held constant and equal to .56. Thus there are two parameters which changes when forming a Gabor filter  $\theta$  and  $\lambda$

Input image is divided into 9x9 non-overlapping regions. The Gabor filter is then convolved with different parameters; and it will generate the response images. As per the work done by Michael Lama et.al [18] we are using only the odd component of the Gabor filter which does not produce imaginary output:

$$G_o(x, y) = \exp\left(\frac{x_\theta^2 - \gamma^2 y_\theta^2}{\alpha^2}\right) \sin\left(\frac{2\pi x\theta}{\lambda}\right) \quad (4)$$

We convolve the image with 12 Gabor filters tuned to four orientations ( $\theta$ ) and three frequencies ( $1/\lambda$ ). Orientation varied from 0 to  $3\pi/4$  (stepping by  $\pi/4$ ) and frequency varied from .3 to .5 (stepping by .1).

### 3.3 Feature Vector Generation

When Gabor features are generated they can be used for classification purposes. But the main issue which appears in this case is that if we use all features for the classification purposes, it will increase the processing time and also imposes the memory constraints due to the high dimension of Gabor features. Another fact is that, all of the features are not informative and they can become a reason for misclassification; so it is needed that the use of those features must be suppressed. To achieve these two main requirements we have used entropy to discriminate informative features from those features which are not as much informative. Entropy is an effective way to know the importance of the features amongst all of the features. Entropy is the measure of uncertainty associated with the random variable. The less entropy means that the uncertainty is less and vice versa.

Convolve the Gabor filters on input image to have the Gabor responses. Now to select the informative features only we have taken the help of entropy. For this first of all we have calculated the position probability of the feature using following equation.

$$P_{m,n,x,y}(r) = \frac{O_{m,n,x,y}(r)}{nL} \quad (5)$$

Where

$m,n,x,y$  means all Gabor values against each pixel and  $nL$  is the sum of all Gabor values against each pixel. [19]

Now when all the position probabilities have been generated, calculate the entropy of each pixel in every Gabor response image using the following equation:

$$H_{m,n,x,y}(R) = \sum_r P_{m,n,x,y}(R = r) \log(P_{m,n,x,y}(R = r)) \quad (6)$$

Where 'R' is a random variable which is representing the occurrence of Gabor magnitude response against one orientation and wavelength on every pixel. The value of the entropy H indicates the uncertainty of the feature at each pixel position against all response images. A larger value of H means the feature magnitudes vary from different images. Thus, features along this axis of the feature space can improve the discriminating power between gender classes. [19] On the other hand, a smaller value of the entropy H indicates the corresponding features tend toward the same magnitude. That is, features along this axis contribute less to discrimination. To reduce the feature space, these features should not be considered in the classification phase. Now against each Gabor response image use these features only by sorting them in descending order and selecting top 10% of the features for the feature vector.

Our experiments have shown that randomness in the feature vector improves the classification, so for this purpose we have generated one combined image having the mean Gabor feature values amongst twelve Gabor response images. Randomly use 10% of the mean features in feature vector as well.

Now combine the entropy based ranked features with mean Gabor features and make a vector of features. This feature vector will be passed to the classifiers for the classification purposes.

### 3.4 Classification

For classification purposes we have used KNN, SVM, ANN, and LDA classifiers. The classifiers train themselves on different test sizes and use the remaining images for testing purposes.

## 4 Experiment and Results

As a first step we read the image into memory and we applied Viola and Jones face detector to get the face portion of the image. Histogram equalization was applied to the extracted face images to normalize for different lighting conditions. Then we resized the image to the size of 36x36 and passed it to feature extractor. Feature extractor first divides the image to the blocks of size 9x9, then picks each block of face image and applies Gabor filters with odd components. Then features were ranked using the entropy criteria. Mean against each pixel for all the Gabor response images was calculated and randomly merged with ranked features. Then the features were stored in feature vectors and passed to the classifiers for the classification purposes. We have used 40% of the data for training purposes and remaining 60% for the testing purposes. Following experiments were performed to see the classification accuracy. (i) Classification with low un-certainty based features. (ii) Classification with mean based features (iii) Classification with combined feature vectors. The results of these experiments are shown in fig 4. In table 1 classifiers are compared on the base of their accuracy for mean case. Table 2 is showing the results for classification on entropy based ranked features; and Table 3 is showing the results on combined features.

**Table 1.** Comparison of classifiers (Mean based features)

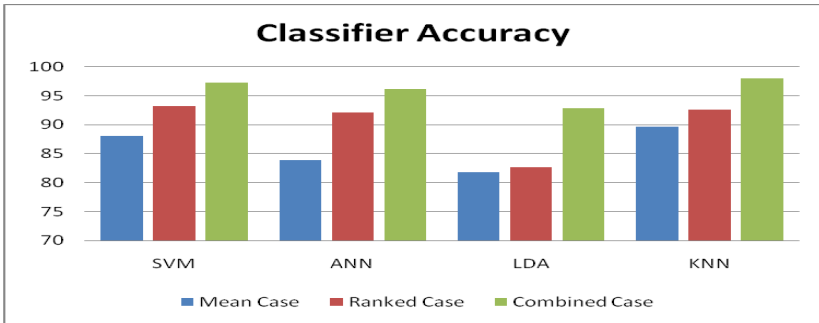
	Method	Accuracy
SVM		88.0%
ANN		83.90%
LDA		81.80%
KNN		89.62%

**Table 2.** Comparison of classifiers (Ranked on the base of entropy)

Method	Accuracy
SVM	93.20%
ANN	92.15%
LDA	82.60%
KNN	92.62%

**Table 3.** Comparison of classifiers (Using combined feature vectors)

Method	Accuracy
SVM	97.20%
ANN	96.15%
LDA	92.80%
KNN	98.00%



**Fig. 4.** Comparison of different Classifiers with different feature sets

## 5 Conclusion

In this paper, problem of gender classification has been addressed. The observations have shown that Gabor features are when ranked and combined with random mean feature values they improve classification accuracy. The proposed method is robust to varying illumination effects and un-even size images. The method is very efficient because high classification accuracies can be obtained with the proposed technique.

**Acknowledgement.** This work was supported by the Bio Imaging Research Center at Gwangju Institute of Science and Technology (GIST), Korea.

## References

1. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2001), vol. 1, p. 511 (2001)
2. Majid, A., Khan, A., Mirza, A.M.: Gender classification using discrete cosine transformation: a comparison of different classifier. In: 7th International Multi Topic Conference, INMIC 2003, December 8-9, pp. 59–64 (2003)
3. Khan, A., Majid, A., Mirza, A.M.: Combination and optimization of classifiers in gender classification using genetic programming. *International Journal of Knowledge-based and Intelligent Engineering Systems* 9(1), 1–11 (2005)
4. Shakhnarovich, G., Viola, P.A., Moghaddam, B.: A unified learning framework for real time face detection and classification. In: Proc. Internat. Conf. on Automatic Face and Gesture Recognition (FGR 2002), pp. 14–21. IEEE (2002)
5. Baback, M., Ming-Hsuan, Y.: Learning Gender with Support Faces. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 24 (2005)
6. Tolba, S.: Invariant gender identification. *Digital Signal Processing* 11(3), 441–470 (2001)
7. Yang, Z., Li, M., Ai, H.: An Experimental Study on Automatic Face Gender Classification. In: Proceedings of 18th International Conference on Pattern Recognition (ICPR), Hong Kong, August 24, vol. 3, pp. 1099–1102 (2006)
8. Xu, Z.Y., Lu, L., Shi, P.F.: A hybrid approach to gender classification from face images. In: ICPR, pp. 1–4 (2008)
9. Pan, Z., Adams, R., Bolouri, H.: Dimensionality Reduction for face Images Using Discrete Cosine Transformation for Recognition. Technical Report, Science and Technology Research Centre (STRC)
10. Scalzo, F., Bebis, G., Nicolescu, M., Loss, L., Tavakkoli, A.: Feature Fusion Hierarchies for gender classification. In: 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1–4 (2008)
11. Majid, A., Khan, A., Mirza, A.M.: Intelligent combination of Kernels information for improved classification. In: International Conference on Machine Learning and its Applications, Los Angeles, California, USA (December 2005)
12. Gabrysa, B., Ruta, D.: Genetic algorithms in classifier fusion. *Application of Soft Computing in Information & Communication Technology (ICT)* 6(4), 337–347 (2006)
13. Costen, N.P., Brown, M., Akamatsu, S.: Sparse models for gender classification. In: Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2004), pp. 201–206 (2004)
14. Sun, Z., Bebis, G., Yuan, X., Louis, S.J.: Genetic feature subset selection for gender classification: a comparison study. In: IEEE Proceedings on Applications of Computer Vision, pp. 165–170 (2002)
15. Sun, Z., Yuan, X., Bebis, G., Louis, S.J.: Neural-network-based gender classification using genetic search foreign-feature selection. In: International Joint Conference on Neural Networks, IJCNN 2002, vol. 3, pp. 2433–2438 (2002)
16. Mäkinen, E., Raisamo, R.: An experimental comparison of gender classification methods. *Pattern Recogn. Lett.* 29(10), 1544–1556 (2008)
17. Han, X., Ugail, H., Palmer, I.: Gender Classification Based on 3D Face Geometry Features Using SVM. In: International Conference on CyberWorlds, pp. 114–118 (2009)
18. Lama, M., Disney, T., et al.: Content based image retrieval for pulmonary computed tomography nodule images. In: SPIE Medical Imaging Conference, San Diego, CA (February 2007)
19. Lee, C.-C., Shih, C.-Y.: Gabor Feature Selection and Improved Radial Basis Function Networks for Facial Expression Recognition. In: International Conference on Information Science and Applications, ICISA (2010)

# Data Hiding in Images Using Some Efficient Steganography Techniques

Chandreyee Maiti\*, Debanjana Baksi, Ipsita Zamider, Pinky Gorai,  
and Dakshina Ranjan Kisku

Department of Computer Science and Information Technology,  
Asansol Engineering College, Kanyapur,  
Asansol – 713305, India

{chandreyee.jannat, debbie\_2k6, er.pink07\_it}@yahoo.co.in,  
ipsita.zamider@rediffmail.com,  
drkisku@ieee.org

**Abstract.** Steganography is the art of hiding data in a seemingly innocuous cover medium. For example – any sensitive data can be hidden inside a digital image. Steganography provides better security than cryptography because cryptography hides the contents of the message but not the existence of the message. So no one apart from the authorized sender and receiver will be aware of the existence of the secret data. Steganographic messages are often first encrypted by some traditional means and then a cover image is modified in some way to contain the encrypted message. The detection of steganographically encoded packages is called steganalysis. In this paper, we propose three efficient Steganography techniques that are used for hiding secret messages. They are LSB based Steganography, Steganography using the last two significant bits and Steganography using diagonal pixels of the image. Symmetric and asymmetric key cryptography has been used to encrypt the message.

**Keywords:** Steganography, Steganalysis, Cryptography, Data Hiding.

## 1 Introduction

Data security [1]–[3] over the networks is an important challenge for researchers and computer engineers for decades. Internet is a great convenience which offers secure data communication of important messages, secret information, variety of images and documents. In order to prevent the unauthorized access of important messages and images from malicious fraudsters, one need to make it more secure by sending the encrypted messages over the networks. To accomplish and build such secure systems, many data hiding and encryption techniques have been proposed in the last few decades. Both the data hiding [3] and encryption techniques [3] are found to be the main mechanisms in data security. However, use of former mechanism has been increasing recently due to some demerits have been found in the later mechanism.

---

\* Corresponding author.

The formal mechanism of data encryption [3], [5] uses the method to convert a message into a ciphertext message by using some encryption algorithm and the ciphertext message is then sent to the recipient who has the authorization to receive and get the original message. To receive the original message which has been sent by the sender, recipient uses a key to obtain the decrypted message. Any malicious user who does not have the key cannot break the security of ciphertext which looks like some meaningless code. Though data encryption is proved to be a secure method to hide data, it has some weaknesses. For example, sometimes appearance of ciphertexts could give a clear impulse to an unauthorized user and this might lead to unauthorized access to the original content by breaking it. As a result the original receiver would not be able to receive the cipher texts sent by the sender. Often unauthorized users may take advantage by destroying the cipher text when it cannot be recovered. Another major drawback to encryption is that the existence of data is not hidden. Data that has been encrypted, although unreadable, still exists as data. If given enough time, someone could eventually decrypt the data. For this reason research on data hiding has been increasing recently.

A solution to this problem is data hiding. Data hiding techniques [3]-[4] could play a major role to embed important data into multimedia files such as images, videos or sounds. Because digital images are insensitive to human visual system, therefore images could be good cover carriers. Data hiding has two major applications [5] – watermarking and steganography. Watermarking merely extends the cover source with extra information. Steganographic techniques are used to store watermarks in data.

Steganography [4]-[5] is an ancient art of hiding messages for making the messages not detectable to malicious users. In this case, no substitution or permutation was used. The hidden message is plain, but unsuspected by the reader. Steganography's intent is to hide the existence of the message, while cryptography scrambles a message so that it cannot be understood. Steganography has been widely used, including in recent historical times and the present day. Possible permutations are endless and known examples include: (i) hidden messages within wax tablets, (ii) hidden messages on messenger's body, (iii) hidden messages on paper written in secret inks, under other messages or on the blank parts of other messages, and (iv) agents used photographically produced microdots to send information back and forth.

Steganography includes the concealment of information within computer files. In digital Steganography, electronic communications may include Steganographic coding inside of a transport layer, such as a document file, image file, program or protocol. Media files are ideal for Steganographic transmission because of their large size. As a simple example, a sender might start with an innocuous image file and adjust the color of every 100th pixel to correspond to a letter in the alphabet, a change so subtle that someone not specifically looking for it is unlikely to notice it. Digital Steganography has three basic components. (a) Obtain the data to be hidden, i.e., secret message, (b) embed the secret message into the cover medium, i.e., images, sounds or videos, etc., and (c) lastly, obtain the stego-carrier to be sent.

In the last decades, many Steganography based data hiding techniques have been proposed in [4]-[6]. There exist plenty of Steganography based data hiding techniques which use LSB based algorithms for hiding the secret messages by embedding it into cover media like images. In [7], authors proposed a data hiding technique which is based

on simple LSB substitution method by selecting optimal numbers of  $k$  LSB substitution method to solve the problem while  $k$  is found to be large. Another work proposed in [8] uses dynamic programming approach to find optimal LSB which later embedded into image. The proposed method reduces the computation time while it is compared the work discussed in [7] uses approximate optimal LSB. Sometimes LSB based substitution method is not sufficient to generate for embedding the secret message into an image. Therefore, the authors in [9] uses genetic algorithm to find an optimal substitution matrix for embedding of the secret message into the mages. Authors also use local pixel adjustment process (LPAP) to improve the quality of the stego-image as carrier.

In this paper, substitution based three different Steganography techniques have been presented and they are LSB based Steganography with the least significant bit, Steganography using last two significant bits and LSB based Steganography using diagonal pixels of the image. However, LSB based techniques are well-known techniques whereas the Steganography using diagonal pixels of the image is the novel technique, which is proposed in this paper along with the well-known techniques.

The rest of the paper is organized as follows. Section 2 has introduced the basic paradigm of LSB based data hiding operation. The next section presents the proposed Steganographic techniques. Results obtained from the proposed techniques are discussed in Section 4 and conclusion is made in the last section.

## 2 Preliminaries

To perform the experiment, gray scale and color images are taken and then Steganography techniques are applied by generating the LSB based substitution matrices. The texts, which are used as the hidden texts are evenly distributed among all the pixels of the  $M \times N$  image matrix. Finally, the resultant stego image is generated.

### 2.1 Basic Paradigm of LSB Based Data Hiding Operation

Since the rightmost bits are used for LSB substitution [6]-[9] in each pixel in the given image, therefore the first operation used rightmost bit and the second operation used rightmost two bits for LSB substitution. The last operation uses diagonal pixels of an image and the rightmost one bit of those diagonal pixels is used for substitution. In these operations, 8-bit grayscale and 24-bit color images are used. In 8-bit grayscale image, rightmost two bits are used in each pixel and rightmost bit is used for diagonal pixels also. The color image uses three color components – red, green and blue which constitute each pixel. The identical phenomenon is used in color image as that of grayscale image. However, for the color image three different matrices are generated and therefore, LSB substitution is used separately for these three matrices.

Let,  $I_{Grayscale}$  be the 8-bit grayscale cover image of size  $P_{I_{gray}} \times Q_{I_{gray}}$  pixels. It can be represented by

$$I_{gray} = \left\{ x_{ij} \mid 0 \leq i \leq P_{I_{gray}}, 0 \leq j \leq Q_{I_{gray}}, x_{ij} \in \{0,1,2,\dots,255\} \right\} \quad (1)$$



Also let,  $I_{color}$  be the 24-bit color cover image of size  $P_{I_{color}} \times Q_{I_{color}} \times 3$  pixels. Therefore, it can be represented for three color components red, green and blue by

$$\begin{aligned}
 I_{color-red} &= \left\{ x_{ij}^{red} \mid 0 \leq i \leq P_{I_{color-red}}, 0 \leq j \leq Q_{I_{color-red}}, x_{ij}^{red} \in \{0,1,2,\dots,255\} \right\} \\
 I_{color-green} &= \left\{ x_{ij}^{green} \mid 0 \leq i \leq P_{I_{color-green}}, 0 \leq j \leq Q_{I_{color-green}}, x_{ij}^{green} \in \{0,1,2,\dots,255\} \right\} \\
 I_{color-blue} &= \left\{ x_{ij}^{blue} \mid 0 \leq i \leq P_{I_{color-blue}}, 0 \leq j \leq Q_{I_{color-blue}}, x_{ij}^{blue} \in \{0,1,2,\dots,255\} \right\}
 \end{aligned} \tag{2}$$

Suppose  $S$  is the  $n$  – bit secret message and it can be defined by

$$S = \left\{ s_i \mid 0 \leq i \leq n, s_i \in \{0,1\} \right\} \tag{3}$$

The secret message  $S$  of  $n$  – bits is to be embedded into the 8-bit grayscale as well as 24-bit color image with three color components. The secret message  $S$  is rearranged to form a  $k$  – bit virtual image  $S'$  which can be described as

$$S' = \left\{ s'_i \mid 0 \leq i \leq n', s'_i \in \{0,1,\dots,2^{k-1}\} \right\}; \tag{4}$$

where  $n' = P_{I_{gray}} \times Q_{I_{gray}}$  and  $n' = P_{I_{color}} \times Q_{I_{color}}$ . Now a mapping is defined between the secret messages  $S = \{s_i\}$  and the embedded message  $S' = \{s'_i\}$ . Further this can be described by the following mathematical formulation

$$s'_i = \sum_{j=0}^{k-1} s_i \times k + j \times 2^{k-1-j} \tag{5}$$

At this stage, all the pixels are chosen from the cover image where the rightmost one bit and rightmost two bits are chosen for the proposed first and second methods and rightmost one bit is selected for the third method in which a subset of pixels are selected containing diagonal pixels only of the image matrix. Hence, the embedding process is completed by replacing the  $k$  ( $k=1,2$ ) LSBs of each pixel by  $s'_i$ . Mathematically, each pixel is storing the  $k$  – bit message to form the stego-pixel as follows.

$$x'_i = x_i \bmod 2^k + s'_i \tag{6}$$

Embedding process for a subset of pixels which contain diagonal pixels only is completed by replacing the  $k$  LSBs of each pixel in the subset by  $s'_i$ . Mathematically, it can be represented by

$$x'_i = x_i - x_i \bmod 2^k + s'_i \quad (7)$$

In Equations (6) and (7),  $x_i$  and  $x'_i$  be the original pixel in cover image and stego-pixel in stego-image respectively.

The embedded message extraction process is accomplished from stego-image by without referring to original cover image. Therefore,  $k$  LSBs of all pixels and subset of pixels are extracted and reconstruct the secret message bits. The embedded message can be extracted from stego-image by the following mathematical formulation

$$s'_i = x'_i \bmod 2^k \quad (8)$$

### 3 LSB Substitution Based Steganographic Techniques

#### 3.1 LSB Substitution in Grayscale Image

A grey scale digital image is an image in which the value of each pixel carries only intensity information. They are also known as black and white images and are composed of shades of grey varying from black at the weakest intensity to white at the strongest. The proposed Steganographic implementation chooses rightmost LSBs ( $k = 1, 2$ ) of each pixel to replace with the secret message bits. The secret message is evenly distributed among all the pixels of the image matrix for the first and second method. However, for the last method a subset of diagonal pixels of the image matrix are used and the secret message is evenly distributed among the diagonal pixels only. The message is encoded in the least significant bit of each pixel in the cover image. This produces no visible change in the original image. The process of LSB substitution in grayscale image is given below.

- An image is read. In case of a gray scale image, a 2-dimensional matrix of unsigned integers with values between 0 and 255 is obtained.
- The pixels are extracted accordingly and converted to binary.
- The secret message can be encrypted using symmetric key or RSA cryptographic techniques.
- The text is encoded in the least significant bits of the pixels. The pixel values of the matrix are changed with a value of (+1) or (-1).
- The pixels are re-inserted into the image.
- Save the image using any lossless compression technique.

#### 3.2 LSB Substitution in Color Image

Each pixel in RGB image is specified by three values, one each for red, blue and green color components. The RGB image is represented by row×column×3 array of class uint8/uint16 or double. In this section, LSB substitution based Steganography is

presented where RGB color image is used. The secret message or plaintext is evenly distributed among the three color components red, green and blue. A subset of pixels of the  $n^{\text{th}}$  column or diagonal elements of each dimension of an image is used. The secret message has been encoded in the least significant bits of these pixels. The process of LSB substitution in color image is given below.

- A RGB image of 3-D matrix is read and the pixel corresponding to the  $n^{\text{th}}$  column and diagonal elements of each dimension is extracted and converted into binary. The last significant bits are extracted from binary matrix.
- A secret message entered and which is encrypted using symmetric key or RSA cryptography techniques. The encrypted message is then converted to binary sequence.
- The message has been encoded in the bits of the  $n^{\text{th}}$  column or diagonal pixels and the secret message is evenly distributed among the three color components - red, green and blue.
- The extracted bits are changed according to the text bits and inserted into the binary matrix. Thus each bit is changed with a value of 1.

### 3.3 Steganalysis

Steganalysis [10] is the process of decoding the secret message from the stego-image. The appropriate pixels of the image, in which the text is stored, are extracted. The pixels are then converted into binary form. Eight bits are extracted at a time and converted into a string. The extracted string can be decrypted using the decryption key. The original message is obtained after string manipulation. In Figure 1, the block diagram of Steganalysis is illustrated. After obtaining the Steganographic or stego-image Steganalysis approach is applied while decryption key is available and finally original message is obtained.

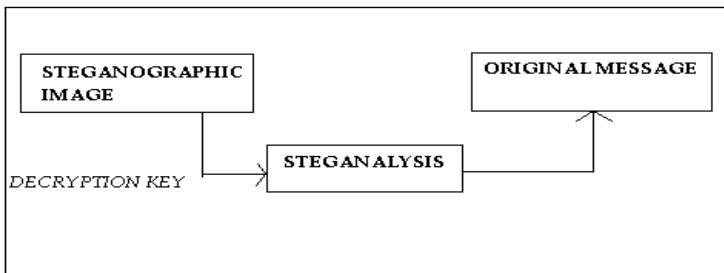


Fig. 1. Block Diagram of Steganalysis

### 3.4 Message Encryption Using RSA Algorithm

To increase the security of the message hidden inside the Stego-image, the message can be encrypted using RSA [11]. The encrypted message can then be hidden using any of Steganographic techniques. Firstly, plaintext message is converted into cipher text using encryption key determined from initial constraints and this cipher text is

then embedded into an image using the proposed Steganographic techniques. Finally, recipient uses the decryption key or private key to decrypt the cipher text message and this decryption key is determined from the preliminary considerations of prime numbers and its associated operations. The RSA algorithm is described below.

Step 1: Choose two large prime numbers  $p$  and  $q$ .

Step 2: Calculate  $n = p \times q$ .

Step 3: Calculate  $\phi = (p-1) \times (q-1)$ .

Step 4: Choose the public key (encryption key)  $e$  such that it is not a factor of  $\phi$  and also calculate private key (decryption key)  $d$  so that  $d \times e = 1 \text{ mod } \phi$ .

Step 5: Make  $e$  and  $n$  public and keep  $\phi$  and  $d$  secret.

Step 6: Calculate cipher text  $ct$ , such that  $ct = pt^e \text{ mod } (n)$ , where  $pt$  is the plain text.

Step 7: The message is decrypted using the formula  $pt = ct^d \text{ mod } (n)$ .

## 4 Results and Discussions

None of the Steganographic methods used in this work produce any visible change in the color or appearance of the image. The size of the image does not change. The proposed work provides two levels of security. It hides the existence of secret message from malicious users. Since the message is further encrypted using RSA encryption algorithm, an intruder will be unable to decipher the image. The proposed three LSB substitution based steganographic techniques have been tested with grayscale and color images. In this section, results of first two methods are not shown. Since these two techniques are found to be simple when compared with the third method in which LSB substitution is made with the diagonal pixels. Figure 2 and Figure 3 show the results obtained by applying the third method which uses  $n^{\text{th}}$  column of diagonal pixels for LSB substitution. In Figure 2, original grayscale image is taken for steganography application and a resultant image stego-image is obtained. In Experiment with grayscale image, only one matrix is generated and considered for LSB substitution. In contrast, RGB image contains three color components – red, green and blue and due to that 3D matrix is generated. These three matrices are treated separately for LSB substitution for each color component. In Figure 3, left RGB image shows cover image whereas the right image depicts stego-image in which the secret message is hidden. In this experiment, 24-bit true color image is used. In Figure 4, original matrix which is generated from the original image is shown in left and the stego matrix which is generated from stego-image is shown in right. In this stego matrix the secret message is hidden and evenly distributed among all diagonal pixels.

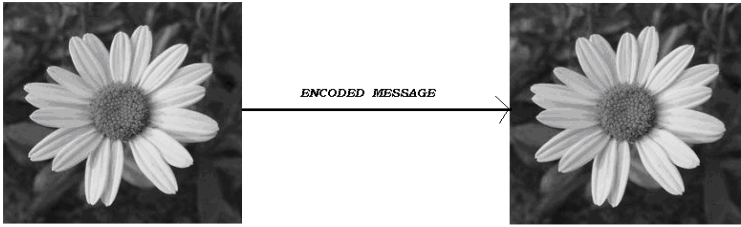


Fig. 2. Original Image (left) and Stego-Image (left) are shown

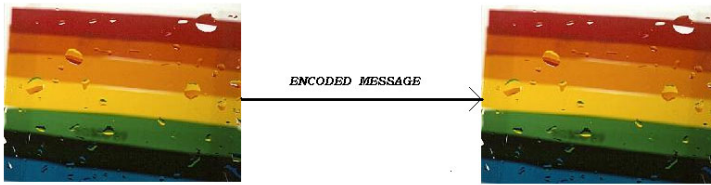


Fig. 3. Original (left) and Stego (right) Images are shown

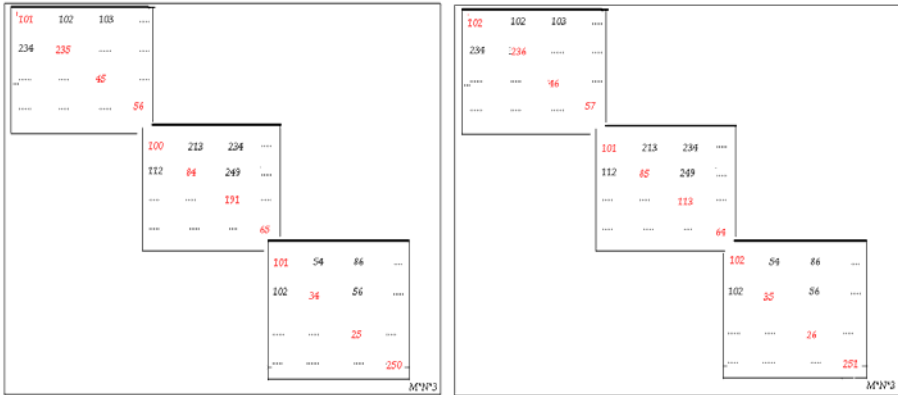


Fig. 4. Original Matrix (left) and Matrix with Encoded Texts (right) are shown

### 4.1 Limitations of the Proposed Techniques

A few demerits have been determined from the proposed methods. Since, the amount of data hidden inside an image depends upon its size. Therefore, a fairly large cover image has to be selected. Secondly, a lossy compression technique cannot be used in compressing an image concealing a secret message. The hidden message will not survive this operation and lost after the transformation. Lastly, any distortion in the image during transmission may lead to loss of the secret message.

## 5 Conclusion

This paper proposes three simple and efficient steganographic techniques which can be used to allow the users to securely transmit a confidential message through images without any detection by an intruder or malicious users. The methods presented do not produce any visible change in the cover image. Among these three methods the last steganographic method can hide maximum information by selecting the diagonal pixels only from the image matrix for LSB substitution. The proposed methods show remarkable performance in terms of accuracy and less distortions of extracted secret message from stego-image while these Steganographic techniques are used.

## References

1. Johnson, N.F., Jajodia, S.: Exploring Steganography: Seeing the Unseen. *Computer* 31(2), 26–34 (1998)
2. Artz, D.: Digital Steganography: Hiding Data within Data. In: *IEEE Internet Computing*, pp. 75–80 (2001)
3. Li, X., Wang, J.: A Steganographic Method based Upon JPEG and Particle Swarm Optimization Algorithm. *Information Sciences* 177(15), 3099–3109 (2007)
4. Chandramouli, R., Memon, N.D.: Analysis of LSB based image steganography techniques. In: *IEEE International Conference on Image Processing*, vol. 3, pp. 1019–1022 (2001)
5. Kutter, M., Hartung, F.: Introduction to Watermarking Techniques in Information Techniques for Steganography and Digital Watermarking. In: Katzenbeisser, S.C. (ed.), pp. 97–119. Artec House (1999)
6. Mohamed, M., Al-Afari, F., Bamatraf, M.: Data Hiding by LSB Substitution Using Genetic Optimal Key-Permutation. *International Arab Journal of e-Technology* 2(1), 11–17 (2011)
7. Wang, R.Z., Lin, C.F., Lin, J.C.: Image Hiding by Optimal LSB Substitution and Genetic Algorithm. *Pattern Recognition* 34(3), 671–683 (2001)
8. Chang, C.C., Hsiaob, J.Y., Chan, C.S.: Finding Optimal Least-Significant-bit Substitution in Image Hiding by Dynamic Programming Strategy. *Pattern Recognition* 36, 1583–1595 (2003)
9. Chan, C.K., Cheng, L.M.: Hiding Data in Images by Simple LSB Substitution. *Pattern Recognition* 37(3), 469–474 (2004)
10. Krenn, J.R.: Steganography and Steganalysis. XIDC.NL (2004)
11. Rivest, R., Shamir, A., Adleman, L.: A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications of the ACM* 21(2), 120–126 (1978)

# Context Based Speech Analysis of Bengali Language as a Part of TTS Conversion

Nabanita Mukherjee<sup>1</sup>, Imon Mukherjee<sup>2</sup>, Debnath Bhattacharyya<sup>2</sup>,  
and Tai-hoon Kim<sup>3,\*</sup>

<sup>1</sup> Department of Information Technology,  
Pailan College of Management and Technology,  
Kolkata, India

nabanita\_ganguly@yahoo.co.in

<sup>2</sup> Computer Science and Engineering Department,  
Institute of Technology and Marine Engineering,  
24-Parganas (South), West Bengal, India

{mukherjee.imon, debnathb}@gmail.com

<sup>3</sup> GVSA and School of Information Science,  
University of Tasmania, Australia  
taihoonk@utas.edu.au

**Abstract.** Speech emotion recognition, as a vital part of affective human computer interaction, has become a new challenge to speech processing. The area of context based speech analysis for a meaningful Bengali Text to Speech conversion has not been worked upon to the extent desired. Emotion plays a vital role in communication as the same sentences can be conveyed on pronounced using different tones with different emotion. Therefore, context based speech analysis of Bengali language as apart of Text to Speech conversion is considered for this work.

**Keywords:** Text-to-Speech, Emotion, Grapheme, Phoneme, Vector Space Model.

## 1 Introduction

The speakers of the sweet language of Bengali ranks fifth in number in the world. It is the official language in the state of West Bengal and Tripura in the neighboring Bangladesh. A large number of Bengali speaking persons are also settled in different parts of the world. Many of these Bengali speakers are illiterate and/or visually handicapped. A text to speech (TTS) system can be of immense help to this segment of population as TTS can empower and emancipate them by removing constraints with the potent tool of Human Computer Interaction (HCI). HCI integrates speech and a language technology to make spoken dialogue in Bengali with computer is possible. In recent years, a range of Bengali language products were developed. This includes word process, spell checker and also few works on text to speech system.

As for Bengali TTS conversion, an area not adequately worked on context based speech analysis. In this work it has been tried to develop how a same sentence is

---

\* Corresponding author.

pronounced with different tones with different emotions. For this purpose, an attempt is therefore, made to develop a TTS system for Bengali language using Festival [1] with Human Computer Interaction (HCI) integrates speech and language technologies to cater to the need of illiterates people with impaired vision and others from providing information and knowledge, this TTS system may be useful for Bengali speaking children.

## 1.1 TTS System

A TTS (Text to Speech) system converts text of normal language into speech. That means it converts grapheme to phoneme.

## 1.2 Speech Synthesis

Speech synthesis is the artificial production of human speech. A computer system used for this purpose is called speech synthesizer and can be implemented in software or hardware.

Text to speech (TTS) synthesis system for Bengali language has been developed using Festival [1]. In this case, different phases of the synthesis task are performed, there are

**Text analysis:** All non standard words converts to standard ones. For example, non standard word like numbers (year, time, floating point), abbreviations was represented into normalized form.

**Phonetic analysis:** This is a Grapheme to Phoneme converter, converting the written text into a sequence of phonemic symbols.

For example shown in Fig. 1,

Letter	Translation
A	a
B	aa

**Fig. 1.** Sample letter and its transform

**Prosodic analysis:** This module takes the phoneme sequence and assigns to each phoneme the required pitch and duration.

**Speech synthesis:** Synthesizing of speech is concatenating into speech unit.

## 2 Background Study

There is a technique of Bengali TTS system has been developed using Festvox. The methodology takes into account the new optimal text selection algorithm, which is used to select appropriate text from the database. It also using grapheme to phoneme converter and for automatic segmentation process, a FestVox frame work was used [2].

The algorithm selects sentences recursively, selecting the sentences with the largest unit first. All the units covered are then removed from the list of units and the next sentence covering the largest number of the remaining units is selected next. This process is continued until all units are covered. The units to be covered can be assigned weights, which may be related to frequency of their occurrence in the corpus.



The basic building blocks of a synthesizer and their implementation strategy are included for developing Bengali TTS (Text to Speech) system [3]. The synthesizer is composed of three parts:

## 2.1 Preprocessing Module

This module related with pitch detection, amplitude normalization, segmentation etc.

## 2.2 Text Analyzer

Words have been detected. Check whether the input word is in the database for exceptional words, if it matches; take the corresponding phoneme to the synthesizer for future processing. If it is not matches, take the corresponding phoneme to the synthesizer for further processing.

## 2.3 Synthesis Module

The concatenation is done by joining the ends of the two phoneme or partnemes. The position of the two segments is determined by the position going zero crossing at the concatenation point.

It has been tried to develop appropriate Bengali pronunciation (or letter to sound) rules. For implementation, these rules are classified into consonant and vowel rules [4].

For example,

CyA, A is often modified to ae,  
e.g biskhyAta (famous)-> bikhkhaeto.

There are other rules for Bengali pronunciation has been developed using rule based automated pronunciation generator. It was said that Bengali script is not completely phonetic in view of the fact that not every word is pronounced in accordance with its spelling [5].

For example, the words ‘/onek’ & ‘/oti’ both start with ‘A /ɔ’ but their pronunciation are [onek] & [oti] respectively.

Rules for that

$$\begin{aligned} A + c + - (H L | j) &> A \\ A + C / c + \phi (C L | j) &> J \quad \text{Where } c \text{ is consonant.} \end{aligned}$$

There is an implementation methodology of international rule for text reading in Standard Colloquial Bengali (SCB) along with the generation of such rules through study of intonation patterns. Syllabic pitches are represented by tree types of syllabic intonations namely rise, fall and flat. Pitch movements of a word can be considering as the sequence of syllabic pitch movements. Intonation pattern of a sentence can be generated by considering both word level as well as the sentence level intonation. Extract of speech information is done by state-phase approach and performing stylization of the pitch contours generates the normalized patterns. Aspects of syllabic stylization and pitch modification techniques have been discussed. Intonation rule was tested on twenty five numbers of sentences. The model has

produced a neutral intonation of standard colloquial Bengali. It has been observed that the output of the synthesized speech is adequately intonated, a step to bring the naturals into the system [6].

F. Arciniegas and M. J. Embrechts suggested artificial neural networks (ANNs) for phoneme recognition for text -to-speech applications. This engine supports the mapping of missing words in the database and it can also reduce contradictions related to different pronunciations for the same word [7].

### 3 Proposed Work

#### 3.1 System Architecture

Fig. 2, shows a block diagram of the context based Bengali Text to Speech System. For this purpose is following some steps. There are

- Taking sort sentences as input.
- Converts it grapheme to phoneme.
- Choose different context for same sentence.
- Getting different context based speech output through the speech database.

**Step 1:** Select text T.

Here basically short sentences (like two words) are selected as a text which will be used in different context.

Sentences like-

fl-a hpz  
i|a c|Jz  
L|S L|z  
-Mm-a k|hz  
iu LI-Rz

**Step 2:** For any text T, let

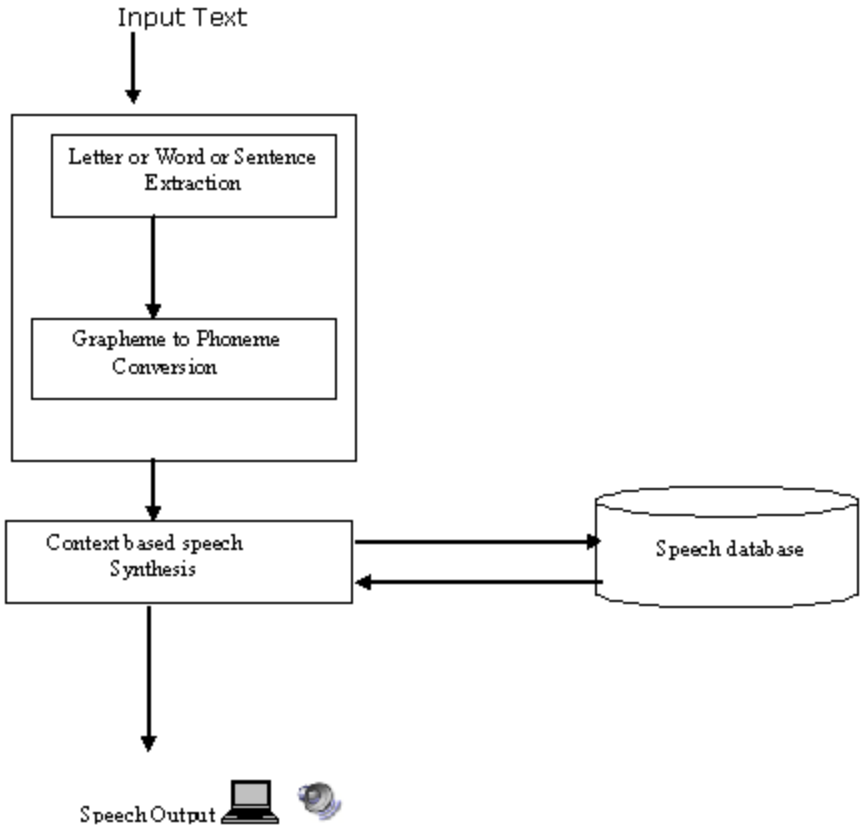
$$L_k := \{ \{A, B, \dots, K\}, \{L, M, \dots, y\}, \{i, -, \dots, -i\}, \dots, \{\dots, \dots, '\} \}$$

$$P_{ij} := \{ \text{'space'}, \text{' '}, \text{' ;'}, \dots, \text{'z'} \}$$

$W_{ij}$  := Words at  $i^{\text{th}}$  row and  $j^{\text{th}}$  column of T is formed by taking n number of  $L_k$  at a time maintaining the Bengali semantics

$S_{ij}$  := A collection of  $W_{ij}$  and  $P_{ij}$  maintaining the Bengali semantics

From  $S_{ij}$  determine  $L_k$  and convert it to its corresponding transliteration given in table 1 and table 2.



**Fig. 2.** Block diagram of the System Architecture of Context based Bengali text to speech conversion

**Table 1.** Vowels (in Bengali language) and its corresponding Transliteration

Letter	Transliteration	Letter	Transliteration	Letter	Transliteration
A	A	B	aa	C	Oi
D	Oii	E	u	F	Uu
G	Ri	H	ea	I	Eai
J	O	K	ouu		

**Step 3:** Grapheme to phoneme conversion according to letter to sound rule.

As Bengali script is basically phonetic, each character usually has default pronunciation. A major task in Bengali text analysis is to modify default pronunciation as per context by appropriate pronunciation (or letter to sound) rules. Some examples of Bengali words and their lexicons are shown in Fig. 3. Table 3 shows the Bengali words and their IPA Transcriptions. Example is shown in (fig. 3).

**Step 4: Emotion Detection.**

In order to detect the emotion the Vector Space Model (VSM) [8] is widely used in information retrieval where each document is represented as a vector, and each dimension corresponds to a separate term. If a term occurs in the document then its value in the vector is non-zero. We have  $n$  distinct terms in our lexicon. Then, the  $k^{\text{th}}$  lexicon,  $L_k$ , can be defined as a set of ordered terms. In other word,

$$L_k = \{t_1, t_2, t_3, \dots, t_n\} \quad (1)$$

An arbitrary document vector, is defined as follows:

$$\vec{d}_i = \{w_{1i}, w_{2i}, w_{3i}, \dots, w_{ni}\} \quad (2)$$

where  $w_{ki}$  represents the weight of  $k^{\text{th}}$  term in document  $i$ .

In literature, there several different ways of computing these weight values have been developed. One of the best known schemes is *tf-idf* weighting. In this scheme, an arbitrary normalized  $w_{ki}$  is defined as follows:

$$w_{ki} = C(t_k, d_i) = \frac{tf_{ik} [\log(\frac{N}{n_k})]}{\sqrt{\sum_{k=1}^n (tf_{ik}^2) [\log(\frac{N}{n_k})]^2}} \quad (3)$$

where, 6

$t_k = k^{\text{th}}$  term in document  $d_i$

$tf_{ik}$  = frequency of the word  $t_k$  in document  $d_i$

$idf_k = \log(N/n_k)$  inverse document frequency of word  $t_k$  in entire dataset

$n_k$  = number of documents containing the word  $t_k$ ,

$N$  = total number of document in the dataset.

Each emotion class,  $M_j$ , is represented by a set of documents,  $M_j = \{d_1, d_2, \dots, d_c\}$ . Then an arbitrary emotion,  $\vec{E}_j$  for a model vector  $V_l$  can be represented by taking the mean of  $d_j$  vectors for an arbitrary emotion class. More formally, each  $\vec{E}_j$ , is computed is follows:

$$\vec{E}_j = \sum_{\vec{d}_j \in M_j} \frac{1}{|M_j|} \vec{d}_j \quad (4)$$

Where,

$|M_j|$  represents the number of documents in  $M_j$ . After preparing model vectors for each emotion class, the whole system is represented with a set of model vectors,  $D = \{E_1, E_2, \dots, E_s\}$  where  $s$  represents the number of distinct emotional classes to be recognized.

In Vector Space Model, documents and queries are represented as vectors, and cosine angle between the two vectors used as similarity of them. Then normalized similarity between a given query text,  $Q$ , and emotional class,  $E_j$ , is defined as follows:

$$\text{sim}(Q, E_j) = \sum_{k=1}^n w_{ki} * E_{ki} \tag{5}$$

In order to measure the similarity between a query text and the  $D$  matrix of size  $s \times n$ , first we convert the query text into another matrix  $n \times 1$  similar to  $D$  where  $n$  is the size of the lexicon and  $s$  is the number of emotions. Then for each emotion (each row of  $D$  matrix), we make multiplication between the query matrix  $Q$  and one row of  $D$  matrix. After these multiplications we have  $m$  scalar values representing the cosine similarity. The index of the maximum of these values is selected as the final emotional class. More formally:

The classification result is then,

$$\text{VSM}(Q) = \arg \max_j (\text{sim}(Q, E_j)) \tag{6}$$

The basic hypothesis in using the VSM for classification is the contiguity hypothesis where documents in the same class form a contiguous region, and regions of different classes do not overlap.

**Table 2.** Consonant (in Bengali language) and its corresponding Transliteration

Letter	Transliteration	Letter	Transliteration	Letter	Transliteration
L	K	M	kh	N	G
O	Gh	P	ng	Q	Ch
R	Chh	S	j	T	Jh
U	Nio	V	t	W	Th
X	D	Y	dh	Z	N
A	Ta	B	thh	c	D
D	Dhh	E	nn	f	P
G	Ph	H	B	i	Bh
J	M	K	z	l	R
M	L	H	bw	n	Sh
O	Shh	P	S	q	H
U	Y	S	Rh	t	Rhh
v	Ott	W	Omn	x	Uh
Y	Onm				

## 4 Experimental Result

In our experiment we have considered a large database carrying 8 types of emotions like normal, anger, request, Disgust, order, Joy, fear, sad. In other word we have considered 8 different classes of emotion. We have used WEKA tool [9] for the classification of the different emotions of voices. Table 4 shows the context based Bengali Translation and Table 5 provides the different pitch vibration in different contexts of same sentence(s).

**Table 3.** Sample Bengali words and their IPA Transcriptions

WORD CATEGORY	WRITTEN FORMAT	PRONOUNCIATION	IPA TRANSCRIPTION
CARDINAL NUMBER	198	HL eu BV	Eak noy aat
ORINAL NUMBER	3u	a«a£u	Tritiyo
TIME	12:10	h <sub>j</sub> -l <sub>j</sub> V <sub>j</sub> cn çjçøV	Barota dos minit
DATE	6/5/2011	Ru -j cœ q <sub>i</sub> S <sub>j</sub> l HN <sub>j</sub> -l <sub>j</sub>	Choy may du hajar agaro
RATIO	6:10	Ru Aeœf <sub>j</sub> a cn	Choy anupat dos
ACRONYM	çp.H.çh	çœ²-LV AÉ <sub>j</sub> -p <sub>j</sub> çp-une Ag -h%œm	Cricket Association of Bengal
ABBREVIATION	çj:	çjø <sub>j</sub> l	Mister

<u>Word</u>	<u>lexicon</u>
i <sub>j</sub> a (Rice)	bhat
c <sub>j</sub> J (Give)	dao
L <sub>j</sub> S (Work)	kaj
L <sub>j</sub> ( Do)	karo
-Mm-a (Play)	khelte
k <sub>j</sub> h (Go)	jabo
fl-a (Read)	porte
hp (Sit)	boso
iu (Fear)	bhoy

**Fig. 3.** Sample words and there Lexicons

**Table 4.** Context based Bengali Translation

Context	Normal	Anger	Request	Disgust	Order	...	Fear
Bengali Sentence	L <sub>i</sub> S Llz	L <sub>i</sub> .. S L..l Z	L <sub>i</sub> S L-l <sub>i</sub> .. Z	L <sub>i</sub> S L-l <sub>i</sub> <sub>x</sub> Z	L <sub>i</sub> S LIÚ Z		iú LI-R Z
Translation	kaj karo.	Kaaj karro	Kaj karo...	Kajj karroh.	Kaj karr.		Bhoy korchhe

**Table 5.** For same sentence gets different pitch vibration in different context


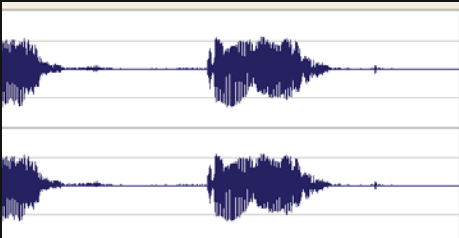
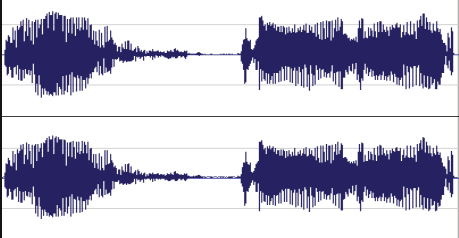

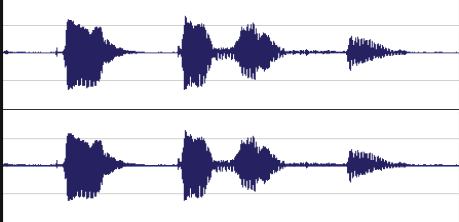
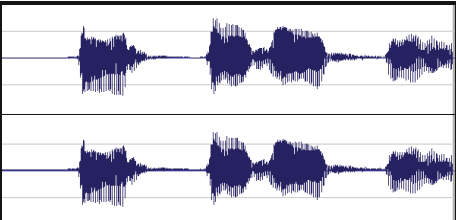
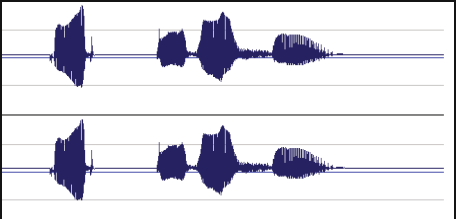

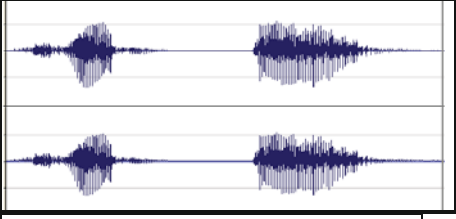
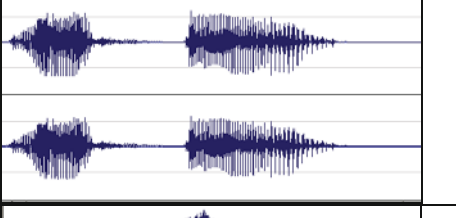
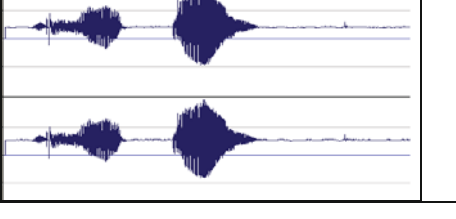

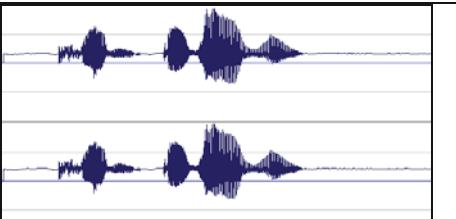
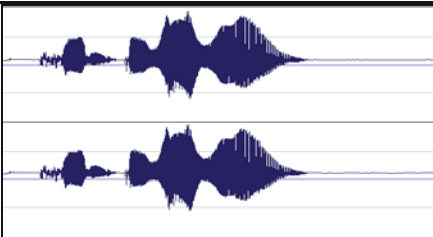

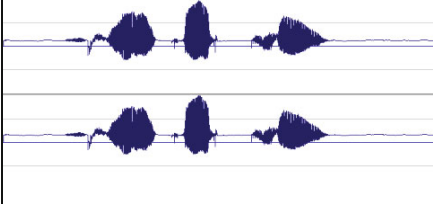
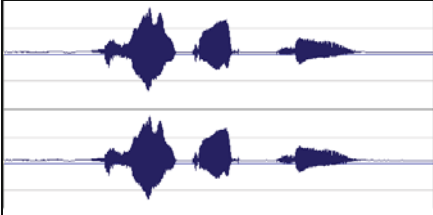
Sentence	Speech output	Mode	Pitch vibration
L <sub>i</sub> S Llz	 L <sub>i</sub> S LI	Normal	
		Order	
fl-a hp	 fl-a hp	Normal	

Table 5. (continued)

		Order		
		Angry		
ija cJ	ija cJ 	Normal		
		Angry		
		Request		
-Mm-a kjh	-Mm-a kjh 	Normal		
		Joy		



**Table 5.** (continued)

				
iu LI-R	iu LI-R 	Normal		
		Fear		

**Acknowledgement.** This work was supported by the Security Engineering Research Center, granted by the Korea Ministry of Knowledge Economy.

**References**

1. Alam, F., Nath, P.K., Khan, M.: Text to Speech for Bangla Language using Festival. BRAC University (2007)
2. Sarkar, T., Keri, V., Santhosh, M., Prahallad, K.: Building Bengali Voice Using FestVox. In: ICLSI (2005)
3. Bandyopadhyay, A.: Some Important Aspect of Bengali speech synthesis system. In: IEMCT, pp. 95–100. Tata McGraw-Hill (June 2002)
4. Sen, A.: Bangla Pronunciation Rules and A Text to speech system. In: Symposium on Indian Morphology, Phonology & Language Engineering, p. 39 (2004)
5. Mosaddeque, A.B., Uzzaman, N., Khan, M.: Rule based Automated Pronunciation Generator. Center for Research on Bangla Language Processing (2005)
6. Bandyopadhyay, A., Mandal, S.K.D., Pal, B., Mitra, M.: Implementation of intonation pattern in Bengali text to Speech synthesis. Speech & Signal Processing Group, ER & DCI (2003)
7. Manimegalai, T.: Enhancing Text To Speech using Spectral Smoothing. In: National Conference on Recent Trends in Intelligent Computing, KGEC, West Bengal, November 17-19 (2006)
8. Danisman, T., Alpkocak, A.: Feeler: Emotion Classification of Text Using Vector Space Model. In: AISB 2008 Convention, Communication, Interaction and Social Intelligence, Aberdeen, UK. Affective Language in Human and Machine, vol. 2, pp. 53–59 (April 2008)
9. <http://weka.en.malavida.com/>

# Image Searching with Eigenfaces and Facial Characteristics

Ayesha Kurukulasooriya and Anuja T. Dharmarathne

University of Colombo School of Computing, 35,  
Reid Avenue, Colombo 07, Sri Lanka

**Abstract.** Automated face recognition is an application domain for automatically identifying or verifying a person in a digital image or a video frame from a video source. This is becoming an increasingly important function in many areas ranging from static matching of controlled photographs and credit card verification to surveillance video images.

However for many years face recognition technology has been focused on one face: frontal, well illuminated and looking directly at the capture device. In reality the situation is different. Human face is being viewed in variety of poses, under a range of illumination conditions and changing expression in real world. In recent years, a considerable number of researches have been carried out to deal with each of these forms of variations individually. However combining techniques to deal with multiple forms of variation at once has received less scrutiny. In this research we utilize two established techniques which cater for these forms of variations individually and unify them into a single novel algorithm for better recognition of human faces.

## 1 Introduction

The nature of the concepts regarding images in many domains are imprecise, and the interpretation of finding similar images is also ambiguous and subjective on the level of human perception. However because of the steady growth in computer power, rapidly declining cost of storage and ever increasing access to the internet, need of an efficient image search system has been increased tremendously in many application areas.

Face recognition is one of the most relevant applications of image analysis. It is a true challenge to build an automated system which equals human ability to recognize faces. Although humans are quite good identifying known faces, they are not very skilled when they must deal with a large amount of unknown faces. So how the computers, with an almost limited memory and computational speed, should overcome humans' limitations. That is why it has remained a difficult problem in the area of computer vision, where some 20 years of research is just beginning to yield useful technological solutions.

The problem of face recognition can be stated as identifying an individual from images of the face and encompasses a number of variations other than the most familiar application of mug shot identification. That is because variations

between images of the same face can be greater than those from different faces such as illusion, pose (angle), accessories and expression.

Face recognition remains as an unsolved problem and a demanded technology. There are many different industry areas which can be benefit from what it offers. Some examples include video surveillance, human-machine interaction, photo cameras, virtual reality or law enforcement. So building a good face recognition system is more advantageous.

## 2 Related Work

Face recognition is a demanded technology. There are many different industry areas interested in what it could offer. This multidisciplinary interest pushes the research and attracts interest from diverse disciplines. Therefore, it is not a problem restricted to computer vision research. Face recognition is a relevant subject in pattern recognition, neural networks, computer graphics, image processing and psychology [3]. In fact, the earliest works on this subject were made in the 1950's in psychology [4]. They came attached to other issues like face expression, interpretation of emotion or perception of gestures.

Engineering started to show interest towards face recognition in the 1960's. In 1960, Bledsoe, along other researches, started Panoramic Research, Inc., in Palo Alto, California. The majority of the work done by this company involved AI-related contracts from the U.S. Department of Defense and various intelligence agencies [5]. During 1964 and 1965, Bledsoe, along with Helen Chan and Charles Bisson, worked on using computers to recognize human faces [10-13]. But only a little work of this was published. He continued later his researches at Stanford Research Institute. Bledsoe designed and implemented a semi-automatic system [10]. Some face coordinates were selected by a human operator, and then computers used this information for recognition. In 1973, Fischler and Elschanger tried to measure similar features automatically [11]. Their algorithm used local template matching and a global measure of fit to find and measure facial features.

The first mention to eigenfaces in image processing, was made by L.Sirovich and M. Kirby in 1986. Their methods were based on the Principal Component Analysis. Their goal was to represent an image in an oval dimension without losing much information, and then reconstructing it. Their work would be later the foundation of the proposal of many new face recognition algorithms. The 1990's saw the broad recognition of the mentioned eigenfaces approach as the basis for the state of the art and the first industrial applications. In 1992 Mathew Turk and Alex Pentland of the MIT presented a work which used eigenfaces for recognition. Their algorithm was able to locate, track and classify a subject's head. Since the 1990's, face recognition area has received a lot of attention, with a noticeable increase in the number of publications. Many approaches have been taken which has lead to different algorithms. Some of the most relevant are PCA, ICA, LDA and their derivatives.

The technologies using face recognition techniques have also evolved through the years. It will allow a new way to interact with the machine. In recent years, a considerable number of researches have been carried out to deal with each of these forms of variations individually. All these methods have their own advantages and limitations. Global feature based methods such as Eigenfaces have been proven to very successful under some limitations. Even though these methods are distinctive, robust to occlusion and do not require to segment the component parts from face, variations of face affect the performance. On the other hand feature-based methods overcome some limitations of Eigenfaces whereas Eigenface advantages become limitations of feature-based matching methods. However combining techniques to deal with multiple forms of variations at once has received less scrutiny. In this research, we utilize Eigenfaces and feature-based matching which cater for these forms of variations individually and unify them into a single novel algorithm for better recognition of human faces.

### 3 System Design and Implementation

The main goal of this project is identifying human faces from a photo and finding similar faces for the identified faces. This research consists of three phases: face detection, preprocessing and face recognition. Given a still or video images of a scene the face detection module first identifies human faces in that image and lists them out. The detected faces are then preprocessed in preprocessing module and then they are input to the face recognition module to identify who are in the input image.

#### 3.1 Face Detection and Classification

Face detection is the first step for any face recognition system. It is involved in finding the spatial location within the scene where human faces are located. This is quite challenging due numerous issues, e.g. pose, presence or absence of structural components, facial expression occlusion, orientation, imaging conditions.

In our approach, to detect the faces in the input image Haar Cascade Classifier of the openCV has been used. OpenCV's face detector uses a method known as Viola-Jones method that Paul Viola and Michael Jones published in 2001. The features that Viola and Jones used are based on Haar wavelets. Haar wavelets are single wavelength square waves. In two dimension space, a square wave is a pair of adjacent rectangles - one light and one dark. The presence of Haar feature is determined by subtracting the average dark-region pixel value from the average light-region pixel value.

In this method a machine-learning method called AdaBoost is used to select specific haar features and to set threshold levels. AdaBoost combines many "weak" classifiers to create one "strong" classifier. A series of AdaBoost classifiers are combined as a filter chain. Each filter is a separate AdaBoost classifier with a fairly small number of weak classifiers.

The acceptance threshold at each level is set low enough to pass all, or nearly all, face examples in the training set. The filters at each level are trained to classify training images that passed all previous stages. During use, if any one of these filters fails to pass an image region, that region is immediately classified as "Not Face." When a filter passes an image region, it goes to the next filter in the chain. Image regions that pass through all filters in the chain are classified as "Face." This filtering chain is dubbed as a cascade. This classification assumes a fixed scale for the face. Since faces in an image might be smaller or larger than this, the classifier runs over the image several times, to search for faces across a range of scales.

### 3.2 Image Preprocessing

Image pre-processing is an extremely important function that should be carried out before supplying detected faces to the face recognition phase. Most face recognition algorithms are extremely sensitive to lighting conditions and many other issues. In some algorithms the face should also be in a very consistent position within the images such as the eyes being in the same pixel coordinates. Other conditions are consistent size, rotation angle, hair and makeup, emotion (smiling, angry, etc), position of lights (to the left or above, etc). This is why it is so important to use good image preprocessing filters before applying face recognition. And also removing the pixels around the face that are not used, to only show the inner face region, not the hair and image background, since they change more than the face does.

So in this research using appropriate filters color images are converted to gray scale, and then Histogram Equalization is applied as a method of automatically standardizing the brightness and contrast of the facial images. Also, the images are resized to a standard size, while keeping its aspect ratio the same.

### 3.3 Face Recognition

Many paradigms are available for implementing the recognition/classification phase. Some of the most important are geometric feature based matching which describes the overall configuration of the face by a vector of numerical data representing the relative position and size of the main facial features: eyes and eyebrows, nose and mouth, global feature based methods such as eigenfaces where a set of orthonormal basis vectors computed from a collection of training face images which provide a basis of low dimensional representation of the facial images and are optimal in the minimum least square error sense. Among those methods Eigenfaces have been proven to very successful under some limitations. In this research, we utilize Eigenfaces and feature-based matching which cater for facial variations individually and unify them into a single novel algorithm for better recognition of human faces.

**Recognition Using Eigenfaces.** Automatic face recognition systems try to find the identity of a given face according to their memory; the memory of a face recognizer is generally simulated by a training set. Therefore something should be carried out to train the system. In this project face recognition phase can be divided into two parts: training and recognition.

In the training phase, eigen vectors for each image in the training set are extracted. A set of M training images  $I_1, I_2 \dots I_M$  are obtained. Let  $X_A$  be a training image of person called A which has pixel resolution of  $N * N$ . In order to extract PCA features of  $X_A$ , image is first converted to a pixel vector  $\Gamma_A$  by concatenating each of M rows into a single vector. The length of the vector PA will be  $N * 1$ . This is carried out with regard to each image in the training set.

Then for the vectors  $\Gamma_1, \Gamma_2, \Gamma_3 \dots \Gamma_M$  face is calculated using the equation 1.

$$\Psi = (1/M)\sum_{i=1}^M \Gamma_i \tag{1}$$

Using the equation 2, the difference between each image vector  $\Gamma_i$  and the average image  $\Psi$  is calculated.

$$\Phi_i = \Gamma_i - \Psi \tag{2}$$

Thus the covariance matrix C is calculated as follows.

$$C = AA^T \text{ where } A = [\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_M] \tag{3}$$

Then the eigenvectors and eigenvalues of this covariance matrix are calculated using standard linear methods. Each eigenvector,  $u_i$  describes a set of axes within the image space along which there is the most variance in the face images and the corresponding eigenvalues represent the degree of variance along these axes. The M eigenvectors are sorted in order of descending eigenvalues and the K greatest eigenvectors are chosen to represent face space.

Then each face in the training set (minus the mean),  $\Phi_i$  is represented as a linear combination of these Eigenvectors  $u_i$ .

$$\Phi_i = \sum_{j=1}^k W_j u_j \tag{4}$$

Where  $u_j$ 's are eigen faces

Each normalized vector is then represented as a weight vector in this basis.

$$\Omega_i^T = [w_1, w_2, \dots, w_k] \tag{5}$$

where  $i = 1, 2 \dots M$ . That means corresponding to every image in the training set we have to calculate such a weight vector and store them as templates.

In recognition phase if an unknown probe face  $\Gamma$  is to be recognized then the incoming probe  $\Gamma$  is normalized as  $\Phi = \Gamma - \Psi$ .

Then this normalized probe is projected onto the Eigenspace (the collection of Eigenvectors/faces) and find out the weights.

$$w_i = u_i^T \Phi \tag{6}$$

The normalized probe  $\Phi$  can then simply be represented as

$$\Omega^T = [w_1, w_2, \dots, w_k] \quad (7)$$

To determine the face which shows the best match for the given input image, the minimum Euclidean distance should be calculated and it can be obtained by equation 8.

$$\xi_k^2 = \|\Omega - \Omega_i\|^2 \quad (8)$$

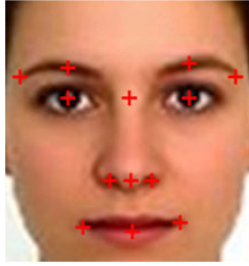
An acceptance (the two face images match) or rejection (the two images do not match) is determined by applying a threshold. Any comparison producing a distance,  $\xi_k$  below the threshold is a match.

**Recognition Using Facial Features.** To Enhance the recognition, the output obtained from eigenface method is then input to another feature based matching module. That is the similar images obtained to a given image using eigenface are then further processed to extract facial features.

First the desired facial region is located for further extracting facial features. To achieve this feature invariant methods are being used. The face segmentation algorithms usually include analysis of skin color, skin density regularization, and geometric filtering. The face segmentation algorithm that we have implemented uses the color information to segment the facial regions. The pixels of the input image can be classified into skin color and non-skin color. The facial region has a special color distribution, which differs significantly from those of the background objects.

*Facial Feature Locating.* After face segmentation, facial features are extracted from the segmented image. Facial features are located based on geometrical relationships between facial components such as nose, eyes, mouth and so on. For these facial features, the degree of darkness of eyebrow depends strongly on the density or color, whereas that of the nostril relies on how much they are occluded. As to the shape of mouth, it is frequently displayed in various appearances and size variances due to facial expressions. In addition, the eyes can be considered as a salient and quite stable feature of a face. Based on these observations, eyes, lip, and eyebrows as well as nose are located sequentially. Fig. 1 illustrates the extracted facial features.

Using these feature points a feature vector for each image is constructed. Then the difference between the feature vector of the face to be matched and the feature vectors of other images filtered out from eigenface is calculated. Then acceptance or rejection is determined by applying a threshold. If the difference is above the defined threshold value then those faces are rejected and the other faces are presented as the matching faces.



**Fig. 1.** Extracted facial feature points

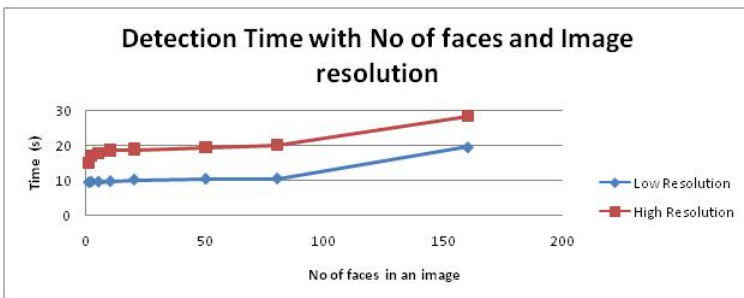
## 4 Results and Discussions

This section presents a quantitative analysis using the final face recognition system. The discussion includes the details on the results obtained on a real world training set.

### 4.1 Face Detection

The image data set used to test the face detection algorithm consisted of 100 images with each image having one or more faces. Those images are chosen in different illumination conditions, different head poses, different age groups, different facial expressions and different image quality (different resolutions).

In the testing phase two image sets were given to the detection phase. Both have images that contain same number of faces. But the resolution of the images of one set is about twice the resolution of the other set. Then the time that each set takes to detect faces when the number of faces in an image increases were measured. Fig. 2 shows how image resolution affects the face detection algorithm.



**Fig. 2.** Detection Time Variation with the Image Resolution

When the image resolution increases it takes much time to process the proposed detection algorithm.



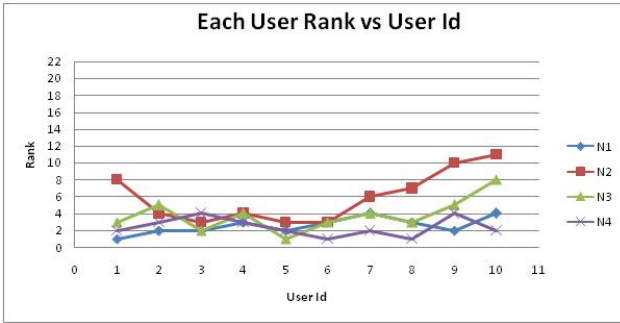
### 4.2 Face Recognition

We compare the performance of the eigenface recognition alone and then evaluate the performance of the combined algorithm.

**Testing the Performance of Eigenface Recognition.** First one image from each person total of 10 people was trained. Then by inputting an image that was not trained, similar images are obtained.

From the results obtained it could be seen that training one image of each person gives the correct result rarely and when the trained image is input to recognize it always gives the correct result.

Then training process was carried out by training 2 images, 3 image . . . ,10 images of each person. Fig. 3 shows how the output varies with the number of training images. Here N1, N2, N3 and N4 are number of trained images from each face. There are 1, 2, 3 and 4 respectively.



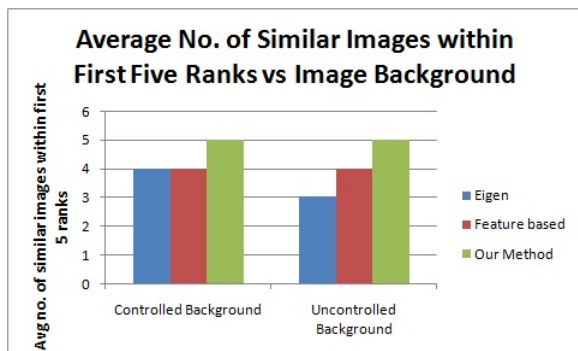
**Fig. 3.** Changes in rank of each user when number of training images change

From this figure it is clear that when the number of trained images increases the recognition rate becomes accurate.

**Enhancing the Output by Combining Eigenface and Feature Based Matching Method.** The results obtained from eigenface recognition were then input to the feature point extraction method. The output became more accurate. To evaluate our approach two experiments were conducted. The first experiment was carried out using 50 individuals each having 5 trained images (250 trained images in the database) in controlled background but with different facial expressions and head poses. Then measured the average number of similar images of each individual within first five ranks using eigenfaces, facial feature matching method and our method.

In the second experiment images of 50 individuals were used each individual having 5 trained images in uncontrolled background (different appearances, different facial expressions, different lighting conditions, different image qualities

etc). Then the number of similar images of each individual within first five ranks using eigenfaces, facial feature matching method and our method. Fig. 4 shows the experiment result.



**Fig. 4.** Number of similar images within first five ranks under background variation

It could be clearly seen that we can distinguish between the individuals accurately using our method. Hence it is clear that the combined method we implemented has a great improvement in recognition.

## 5 Conclusions and Future Work

In this research, two major approaches to the face recognition problem have been studied and a combination of those two methods has been implemented.

The computational models and techniques used by this project were chosen after extensive study of the literature. The successful results conclude that the choices are reliable.

In this research we have implemented a fully automated face detection and recognition application using *c#.net*.

Analysis of the eigenface recognition technique using multiple face eigenspace gives evidence that it proves accuracy to a considerable extent. The combination of it with feature based matching methods achieves accuracy nearly 100%. The images taken to train the application are chosen where they express different variations. That is they are chosen in a way that we can address the problems of illumination, pose and facial expression.

As face recognition is a promising area in surveillance applications and law enforcement, implementing a fully automated face detection and recognition system is significant. In this research we have not taken into consideration about the age. As humans we have the capability of recognizing our grandmother from her wedding photo by just looking at that. But our system does not have that capability i.e. given a face of a person at his/her childhood; recognize all the images of him stored in the database. This is remained as additional work in this project.

## References

1. Maio, D., Maltoni, D.: Real-Time Face Location on Gray-Scale Static Images. *Pattern Recognition* 33(9), 1525–1539 (2000)
2. Rowley, H.A., Baluja, S., Kanade, T.: Neural Network-Based Face Detection. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 23–38 (1998)
3. Rowley, H.A., Baluja, S., Kanade, T.: Rotation Invariant Neural Network-Based Face Detection. In: *IEEE Conf. Computer Vision and Pattern Recognition*, pp. 38–44 (1998)
4. Yang, M.H., Ahuja, N.: Detecting Human Faces in Color Images. In: *Proc. IEEE Int'l Conf. Image Processing*, pp. 127–139 (1998)
5. : Conventional versus neural networks techniques for facial recognition. In: *Proc. of the Symposium of Electronics and Telecommunications, Etc 2000*, pp. 61–64. Timisoara (2000)
6. Sahoo, P.K., Soltani, S., Wong, A.K.C., Chen, Y.C.: A survey of thresholding techniques. *Computer Vision, Graphics, and Image Processing* 41, 233–260 (1988)
7. Zhao, W., Chellappa, R., Rosenfeld, A., Phillips, P.: Face recognition: A literature survey. In: *ACM Computing Surveys*, pp. 399–458 (2003)
8. Bruner, J.S., Tagiuri, R.: The percepton of people. *Handbook of Social Psychology* 17(1) (1954)
9. Ballantyne, M., Boyer, R.S., Hines, L.: Woody bledsoe: His life and legacy. *AI Magazine* 17(1), 7–20 (1996)
10. Bledsoe, W.W.: The model method in facial recognition: a Technical report pri 15. Panoramic Research, Inc., Palo Alto, California (1964)
11. Bledsoe, W.W.: Man-machine facial recognition: a Report on a large scale experiment, Technical report pri 22, Panoramic Research, Inc., Palo Alto, California (1966)
12. Bledsoe, W.W.: Some results on multi category pattern recognition. *Journal of the Association for Computing Machinery* 13(2), 304–316 (1966)
13. Bledsoe, W.W.: Semiautomatic facial recognition: a Technical report sri project 6693, Stanford Research Institute, Menlo Park, California (1968)
14. Kanade, T.: Picture Processing System by Computer Complex and Recognition of Human Faces: a PhD thesis, Kyoto University (1973)
15. Marques, I.: Face Recognition Algorithms: a literature review on the face detection and recognition

# Design of DT-CNN for Imputing Data at Unobserved Location of Geostatistics Image Dataset

Sathit Prasomphan<sup>1,\*</sup>, Hisashi Aomori<sup>2</sup>, and Mamoru Tanaka<sup>2</sup>

<sup>1</sup> Department of Computer and Information Science, Faculty of Applied Science  
King Mongkut University of Technology North Bangkok, Thailand

<sup>2</sup> Department of Information and Communication Sciences  
Faculty of Science Technology, Sophia University, Japan  
sathitp@kmutnb.ac.th, aomori@ieee.org, Mamoru.Tanaka@gmail.com

**Abstract.** The presence of missing values in a geostatistics dataset can affect the performance of using those dataset as generic purposed. In this paper, we have developed a novel method to estimate missing observation in geostatistics by using sigma-delta modulation type of Discrete-Time Cellular Neural Networks(DT-CNN). The nearest neighboring pixels of missing values in an image are used. The interpolation process is done by using B-template with Gaussian filter. The DT-CNN is used for reconstructing the imputed values from analog image value to digital image value. We have evaluated this approach through the experiments on geostatistics image which has different characteristics of missing pixels such as Landsat 7 ETM+ SLC-off and standard geostatistics image. The experimental results show that by using sigma-delta modulation type of Discrete-Time Cellular Neural Networks, we can achieve a high PSNR for various image datasets and at different characteristics of missing image.

**Keywords:** Discrete-Time Cellular Neural Networks, Gaussian filter, Geostatistics images, Missing pixels.

## 1 Introduction

The problem of incomplete geostatistics image or the occurrence of missing pixels in an image is one of the most interesting topics that the researcher wishes to solve. However, this problem is still difficult and challenge task because there are too many different missing pattern in an image. The missing geostatistics image interpolation is a process of estimation of unknown values of the image pixels in a spatial domain. Some previous works for spatial interpolation problem were introduced to improve the accuracy of interpolation. There are several methods which are proposed for this problem. In [10] proposed the spatial interpolation algorithms based on Kriging algorithms. The

---

\* Corresponding author.

Kriging predictor is the minimum variance unbiased predictor of the value at an unobserved location which is a weighted linear sum of data. The weights are chosen so that the prediction error variance for a given spatial covariance functions of the underlying random field [10]. However, one major problem faced with this method is the selection of suitable corresponding covariance or variogram function models. The usage of the wrong model will affect to the correctness of the interpolation process. Another solution for interpolating the geostatistics dataset is artificial neural network. Sitharam et al.[11] introduced a combination of three models: geostatistical, neural network, and support vector machine to predict a reduced level of rock. We found that their method used the global training and testing to interpolate the unobserved data. The problem will occur if in that dataset compose of several distribution. Therefore, the main objective of our research is to propose a new interpolation technique for incomplete geostatistics image dataset based on DT-CNN by applying sigma-delta modulation type of DT-CNN to the interpolation problem in geostatistics domain. This paper is motivated by the occurring of malfunction of the scan line corrector (SLC) in Landsat 7 ETM+ which caused many missing pixels of image during the sensor's scanning process. In addition, the occurring of some damaged area in an image caused by the malfunction or noise signal that degraded by some degrading mechanism is other problem. Hence, the method which has the ability to perform the interpolation which can reduce noises should be focused. Accordingly, with the ability of sigma-delta modulation type of DT-CNN which is working with two-dimensional grid as same as human retina and have a noise shaping property, therefore, we will apply this technique for the geostatistics interpolation problem. In the remainder of this paper, we first describe some related work of DT-CNN in Section 2. The proposed method is presented in section 3. In section 4 and 5, the experimental setups and results are described, respectively. Conclusions are presented in section 6.

## 2 Related Works

### 2.1 Cellular Neural Network

Cellular Neural Networks is an analogic which is applied to many fields of datasets such as image processing, signal reconstruction, noise shaping proportion, converting signal into original sequence[1][2][3][6]. The first contribution of this technique was proposed by Leon O. Chua to be a paradigm for solving nonlinear spatial-temporal wave equations [1][2]. The architecture of CNNs composed of three major components: an input, a state, and an output. It is a two or three dimensional array which cells are locally connected and interacts directly only with the cell within its radius of  $r$ - neighborhood. The main characteristic of CNN which is different from other method is using nearest values with array of size  $5 \times 5$  to perform the calculation with A-template and B-template.

## 2.2 Discrete-Time Cellular Neural Network

Discrete-Time Cellular Neural Networks (DT-CNN) was derived from the Continuous-Time Cellular Neural Networks (CT-CNN) as proposed by Harrer and Nossek[5][6][9]. The state equation of DT-CNN is,

$$\begin{aligned} x_{ij}(t+1) &= \sum C(i, j; k, l)x_{kl}(t) \\ &+ \sum A(i, j; k, l)y_{kl}(t) \\ &+ \sum B(i, j; k, l)u_{kl} + T_h \end{aligned} \quad (1)$$

$$y(t) = f(x(t)) \quad (2)$$

$$f(x) = \begin{cases} 1 & , x \geq 0 \\ -1 & , x \leq 0 \end{cases} \quad (3)$$

where  $C(i, j; k, l)$  is the neighborhood connection coefficients between state variable where  $(i, j)$  denoted the current position,  $(k, l)$  denoted the nearest neighbors of  $(i, j)$  position. The distance between  $(i, j)$  position and  $(j, k)$  position is less than or equal the defined distance. The variable:  $x_{ij}$ ,  $u_{kl}(t)$ ,  $y_{kl}(t)$ ,  $T_h$ ,  $f(\cdot)$ ,  $A(i, j; k, l)$ ,  $B(i, j; k, l)$ ,  $C(i, j; k, l)$  are the internal state, the input of cell, the output, the threshold, the quantization function, the feed-forward template, the feedback template, and the output template, respectively. The  $r$ -neighborhood ( $N_r(i, j)$ ) given by,

$$N_r(i, j) = C(k, l) | \max\{|k - i|, |l - j|\} \leq r \quad (4)$$

The A-template, B-template and the parameter T of DT-CNN can be determined as

$$\begin{aligned} A &= A(i, j; k, l), C(k, l) \in N_r(i, j) \\ &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2}\right) \end{aligned} \quad (5)$$

$$\begin{aligned} B &= B(i, j; k, l), C(k, l) \in N_r(i, j) \\ &= \begin{cases} 1 & , \text{if } k=i \text{ and } l=j \\ 0 & , \text{otherwise} \end{cases} \end{aligned} \quad (6)$$

$$T = 0 \quad (7)$$

The Lyapunov energy function[3] of DT-CNN is defined as

$$E(t) = -\frac{1}{2}y^t(A - \delta I)y - y^tBu - T^ty \quad (8)$$

where  $\delta$  is the possible constant value to determine the quantizing regions. The new application of DT-CNN is studied as an sigma-delta modulation which is proposed by Aomori et al.[3]. They used DT-CNN as sigma-delta modulator for image reconstruction from analog signal into digital signal by using A-template and B-template to execute as a modulator. In this process, bi-level digital output are generated by a low resolution quantized which is incorporate within a feedback loop template and the input analog signal is reconstructed using the low-pass filter. The objective of this application is obtaining a halftone image and reconstructing image by summation of bi-level pulse digital sequence [3].

### 3 The New Model of DT-CNN for Reconstructing a Damaged Geostatistics Dataset

In this section, we propose new interpolation algorithms based on sigma-delta type of DT-CNN for interpolating missing pixels in geostatistics image datasets. The assumption of our proposed method based on the usage of nearest pixels inside the template for interpolating the missing pixels. This assumption is lead to the usage of DT-CNN which cell is locally connected and the state of the cell will be influenced by the cell located outside the neighborhood. This influence is due to the propagation effect determined through the cell dynamics[4]. Hence, to interpolate each missing pixels inside this image, the two-step calculation is performed. The first step is using B-template with Gaussian filter to interpolate all missing pixels. The second step is reconstructing this interpolated analog image into digital image. We use B-template with Gaussian filter for primary imputing missing values. The state equation of this template is given by

$$\begin{aligned} \hat{B} &= \hat{B}(i, j; k, l), C(k, l) \in N_r(i, j) \\ &= \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2}\right) \end{aligned} \quad (9)$$

Suppose  $u_{ij}$  is a missing value at position  $i, j$  on an image. At first, we search for nearest neighborhood surrounding  $u_{ij}$  with an array of size 5x5. The average value is calculated with this equation,

$$\bar{y}_{kl} = \frac{1}{(mn-1)} \sum u_{kl}, u_{kl} \in N_r(i, j), k \neq i, l \neq j \quad (10)$$

where  $\bar{y}_{kl}$  is an average value,  $m$  and  $n$  are size of template which is an array of size 5x5,  $u_{kl}$  is nearest neighborhood. Next, calculate the interpolated value of a missing pixel  $u_{ij}$  as

$$u_{ij} = \sum \hat{B}(i, j; k, l) \bar{y}_{kl} \quad (11)$$

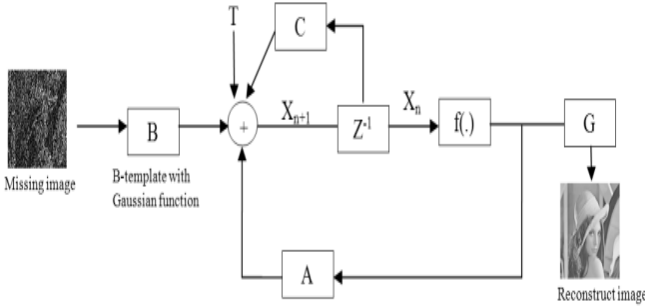
where  $\hat{B}$  is B-template,  $\bar{y}_{kl}$  is an average values of neighborhood surrounding  $u_{ij}$ . All of missing pixels in an image are repeatedly interpolated by using B-template as shown in equation (11). We use sigma-delta type of DT-CNN to reconstruct the analog signal into digital signal after all missing values are already interpolated. In this step, the pulse digital signal is generated with a value range of  $[-1, 1]$ . The number of pulses which is generated depends on the amount of iteration which is used. Next, the reconstructed analog image is obtained from the summation of bi-level pulse sequence digital image. Finally, denoise the reconstructed analog image by using Gaussian filter with this equation.

$$\hat{G} = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(k-i)^2 + (l-j)^2}{2\sigma^2}\right) \quad (12)$$

Details of the algorithms are shown as follows.

1. Let  $u_{ij}$  be missing pixels.
2. Find an arrays size  $5 \times 5$  of nearest values surrounding  $u_{ij}$ .
3. Calculate average value which is taken from step 2 with equation (10).
4. Interpolate missing pixels with B-template using equation (11).
5. Repeat to interpolate all missing values.
6. Use DT-CNN to reconstruct the analog image signal to digital image signal.
7. Calculate summation values of digital image signal to reconstruct the analog image signal.
8. Denoise analog signal by using Gaussian filter with equation (12).

The whole processes of our proposed algorithms are shown as Figure 2.



**Fig. 1.** Missing Image Interpolation Using Sigma-Delta Modulation Type of DT-CNN

## 4 Experimental Setup

### 4.1 Datasets

We used three geostatistics datasets for our experiments. There are different in data density and characteristics. We can classify these three datasets into two groups of data. Details of these two groups of image can be described as follows.



**Landsat 7 ETM+ with SLC-off Imagery.** In our experiments the Bangkok imagery was used. Two similar images acquired before and after the malfunction of Landsat 7 ETM+ imagery sensor (SLC-off) occurred were collected. We used these two images which have the same position and resolution for experiments. The radiometric values from Landsat 7 ETM+ imagery was used for testing the accuracy of our proposed algorithms. In fact, each scene of Landsat 7 ETM+ imagery was composed of 8 bands separately. However, in this research, for testing our method we used only band 1. The old images which do not have missing values were acquired at path 129, row 50 and the acquired date was 02/11/2000. The missing Landsat 7 ETM+ imagery caused by SLC-off were acquired at the same position of the missing image. These two images were downloaded for experiments on 27/10/2010. The size of these two images is 500x500 pixels and the image was taken at [11].

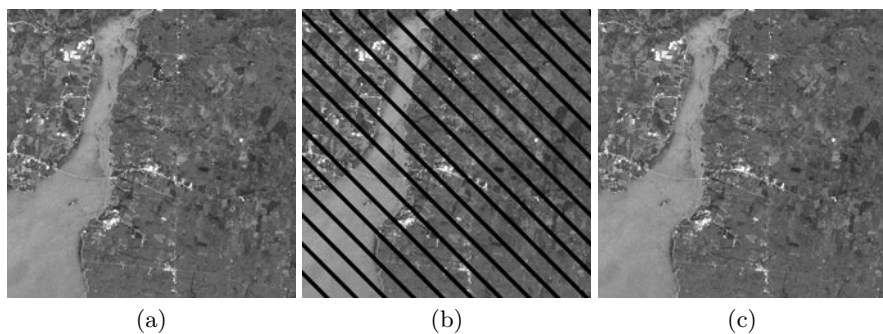
**Standard Geostatistics Gray Scale Image.** The second group of image to be used in the experiment is standard geostatistics gray scale image: aerial and airplane image with 512x512 pixels in which 8 bit gray scale image. For testing the accuracy of our proposed method, a damaged image was generated with various percentage of missing pixels. We generated number of missing pixels in 10%, 20%, 30%, 40%, and 50%. The position of missing pixels is fixed during the pre-process of interpolation.

## 4.2 Accuracy Evaluation

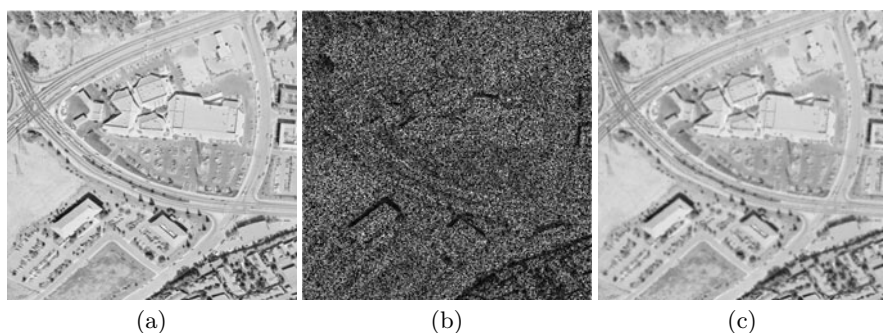
To evaluate the performance of our proposed interpolation algorithms, we used PSNR (peak signal-to-noise ratio) as an indicator to measure. The peak signal-to-noise ratio, PSNR, is a ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. PSNR is usually expressed in terms of the logarithmic decibel scale. The PSNR is most commonly used as a measure of quality of reconstruction of lossy compression codecs. The signal in this case is the original data, and the noise is the error introduced by interpolation. A higher PSNR would normally indicate that the reconstruction is of higher quality. The performance of the method is evaluated in terms of both visual quality and the PSNR value of the restored images. Recall that the peak signal to noise ratio (PSNR) is defined by

$$PSNR = 10 \log_{10} \frac{255^2}{\frac{1}{mn} \sum_{i=1}^m \sum_{j=1}^n (\hat{x}_{ij} - x_{ij})^2} \quad (13)$$

where  $m$  and  $n$  are the width and height of the image,  $x_{ij}$  and  $\hat{x}_{ij}$  are the intensity values of the original image and of the restored image respectively. Typical values for the PSNR in lossy image and video compression are between 30 and 50 dB, where higher is better.



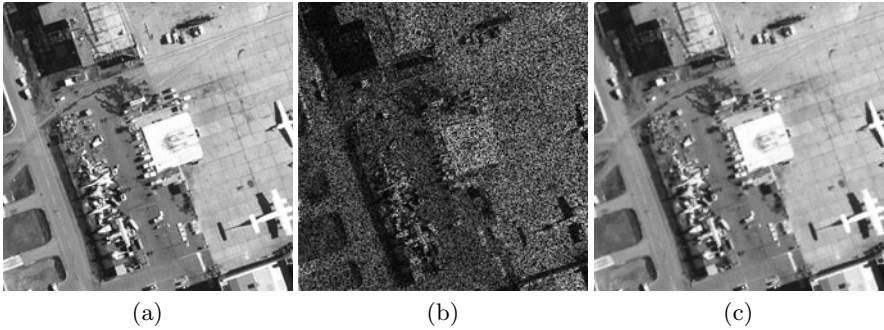
**Fig. 2.** Show the interpolation of missing image using DT-CNN. (a) Original analog image (b) Damaged analog image (c) Reconstructed analog image using sigma-delta type of DT-CNN.



**Fig. 3.** Show the interpolation of missing image using DT-CNN. (a) Original analog image (b) Damaged analog image with 50% of missing values (c) Reconstructed analog image using sigma-delta type of DT-CNN.

## 5 Experimental Results and Discussion

Figure 2. shows image interpolation results by using DT-CNN with Landsat 7 ETM+ with SLC-off image. The original analog image is shown in Figure 2(a).The damage analog image with missing values is shown in Figure 2(b). The reconstructed analog image using sigma-delta type of DT-CNN is shown in Figure 2(c). Figure 3. shows image interpolation results by using DT-CNN with aerial image. The original analog image is shown in Figure 2(a).The damage analog image with 50% of missing values is shown in Figure 2(b). The reconstructed analog image using sigma-delta type of DT-CNN is shown in Figure 2(c). Figure 4. shows image interpolation results by using DT-CNN with Airfield image. The original analog image is shown in Figure 4(a).The damage analog image with 50% of missing values is shown in Figure 4(b). The reconstructed analog image using sigma-delta type of DT-CNN is shown in Figure 4(c). The values of PSNR



**Fig. 4.** Show the interpolation of missing image using DT-CNN. (a) Original analog image (b) Damaged analog image with 50% of missing values (c) Reconstructed analog image using sigma-delta type of DT-CNN.

**Table 1.** The value of PSNR in each images

Dataset	$\sigma$	The proposed method	Gacsadi's method [4]
Landsat 7 ETM+ with SLC-off	0.895	26.16	25.28
Aerial (% of missingness)	0.705		
10%		29.76	30.45
20%		28.34	29.97
30%		26.67	25.38
40%		25.74	25.19
50%		25.13	24.65
Airfield (% of missingness)	0.805		
10%		28.61	31.25
20%		27.37	27.13
30%		25.94	26.86
40%		24.97	23.48
50%		24.37	21.64

after interpolating missing pixels in each image are details in Table 1. From the experimental results we found that, by using sigma-delta modulation type of DT-CNN with A-template and B-template can interpolate missing pixels and can reconstruct this image from analog to digital image. The performance of reconstructed image gives high value of PSNR as shown in Table 1.

An apparent problem on our method as well as any method is the edge effect problem. If missing pixels locates in the edge of regions of available data, it may give a wrong imputed value. This problem may cause to the low PSNR as shown with the PSNR image which have many edges on that picture.

## 6 Conclusions

The new framework for image interpolation based on sigma-delta modulation type of DT-CNN has been proposed. We used B-template with Gaussian filter for primary interpolation of analog image which has missing values. Discrete-Time Cellular Neural Networks with A-template was used again for image modulation from analog to digital signal. The experimental results show that our proposed method for image interpolation gives the high values of PSNR in various percentages of missing values and at different image dataset.

**Acknowledgment.** This research was partially supported by the STEC-2011, Sophia Lecturing Research grants of Faculty of Science and Technology, Sophia University, Japan.

## References

1. Chua, L., Yang, L.: Cellular Neural Networks: Theory. *IEEE Trans. on Circuits and Systems* 35(10), 1257–1272 (1988)
2. Chua, L., Yang, L.: Cellular Neural Networks: Applications. *IEEE Trans. on Circuits and Systems* 35(10), 1273–1290 (1988)
3. Aomori, H., Otaket, T., Takahashi, N., Tanaka, M.: A Spatial Domain Sigma Delta Modulator Using Discrete Time Cellular Neural Networks. In: *Int'l Workshop on Cellular Neural Networks and Their Applications* (2006)
4. Gacsadi, A., Szolgay, P.: Image Inpainting Methods by Using Cellular Neural Networks. In: *Int'l Workshop on Cellular Neural Networks and Their Applications* (2005)
5. Harrer, H., Nossek, J.A.: Multiple layer discrete-time cellular neural networks using time-variant templates. *IEEE Trans. Circuits Syst.* 40(3), 191–199 (1993)
6. Harrer, H., Nossek, J.A.: Some examples of preprocessing analog images with discrete-time cellular neural networks. In: *The 3rd IEEE Intl.*, pp. 201–206 (1994)
7. Aomori, H., Otaket, T., Takahashi, N., Tanaka, M.: Sigma-delta cellular neural network for 2D modulation. *Neural Networks* 21(2-3), 349–357 (2008)
8. Hirano, M., Aomori, H., Otake, T., Tanaka, M.: A Second Order Sigma-Delta Modulation by Cascaded Sigma-Delta CNNs. In: *12th WSEAS International Conference on CIRCUITS*, pp. 86–90 (2008)
9. Harrer, H., Nossek, J.A.: Discrete-time cellular neural networks. *International Journal of Circuit Theory and Applications* 20(5), 453–467 (1992)
10. Zhanga, C., Lia, W., Travis, D.: Gaps-fill of SLC-off Landsat 7 ETM+ satellite image using a geostatistical approach. *International Journal of Remote Sensing* 28(22), 5103–5122 (2007)
11. Sitharam, T.G., Samui, P., Anbazhagan, P.: Spatial Variability of Rock Depth in Bangalore Using Geostatistical, Neural Network and Support Vector. *Machine Models. Geotech. Geol. Eng.* 26(5), 503–517 (2008)
12. Landsat Information, [http://eros.usgs.gov//Find.Data/Products\\_and\\_Data\\_Available/ETM](http://eros.usgs.gov//Find.Data/Products_and_Data_Available/ETM)

# Enhanced Edge Localization and Gradient Directional Masking for Moving Object Detection

Pranab K. Dhar<sup>1</sup>, Mohammad I. Khan<sup>1</sup>, D.M.H. Hasan<sup>1</sup>, and Jong-Myon Kim<sup>2,\*</sup>

<sup>1</sup> Chittagong University of Engineering and Technology, Bangladesh

<sup>2</sup> University of Ulsan, Ulsan, Korea

{pranab\_cse, muhammad\_ikhancuet, simon\_cuet}@yahoo.com,  
jongmyon.kim@gmail.com

**Abstract.** Moving object detection has been widely used in intelligent video surveillance system. This paper proposes a new moving object detection method based on enhanced edge localization mechanism and gradient directional masking. In our proposed method, initially gradient map images are generated from the input image and the background image using gradient operator. The gradient difference map is then calculated from gradient map images. Finally, the moving object is extracted by using appropriate directional masking and thresholding. Simulation results indicate that the proposed method outperforms conventional edge based methods under different illumination conditions including indoor, outdoor, and foggy cases to detect moving object. In addition, it is computationally faster and applicable for real-time processing.

**Keywords:** Video surveillance, gradient directional masking, edge localization, and object detection.

## 1 Introduction

Moving object detection is a challenging task in visual surveillance. It has many applications in diverse discipline such as intelligent transportation system, airport security system, video monitoring system, and so on. Moving object detection often acts as an initial step for further processing such as classification of the detected moving object. In order to perform more sophisticated operations such as classification, we need to first develop an efficient as well as an accurate method for detecting moving object. One of the simplest and popular method for moving object detection is background subtraction method which often uses background modeling and takes long time to detect moving objects [1]. Temporal difference method is very simple and it can detect objects in real time but it does not provide robustness against illumination change. Texture based boundary evaluation methods are not reliable for real time moving object detection.

In this paper, we propose a new method for detecting moving objects from a stationary background based on improved edge localization mechanism and gradient directional masking. In our proposed method, gradient map images are generated

---

\* Corresponding author.

from the input image and the background image using gradient operator. Then the gradient difference map is calculated from gradient map images. Finally, moving object is detected by using appropriate directional masking and thresholding. Simulation results indicate that our proposed method provides better result than well known edge based methods under different illumination conditions, including indoor, outdoor, sunny, and foggy cases to detect moving object.

The rest of this paper is organized as follows. Section 2 provides a brief description of related research. Section 3 introduces our proposed moving object detection method. Section 4 compares the performance of our proposed method with well known edge based methods. Finally, section 5 concludes this paper.

## 2 Related Research

A significant number of moving object detection methods have been reported in recent years. Mahbub *et al.* [2] proposed a method using statistical background modeling. This method detects moving object by matching every edge segment of current frame with every edge segment of background. However, this method fails to detect a moving edge segment which falls into a background edge segment. Li and Leung [3] proposed a method that combines intensity and texture differences and it is based on the argument that texture is less sensitive to illumination changes, whereas intensity is more representative of homogeneous regions. Unfortunately, an exception occurs under weak illumination when intensity is strongly affected by noise and texture is poorly defined. Liu *et al.* [4] introduced a change-detection scheme that compares circular shift moments (CSMs), which represent the reflectance component of the image intensity, regardless of illumination. The shading model (SM) [5] exploits the ratio of intensities in the corresponding regions of two images to adapt with illumination changes. However, both the SM and CSM methods poorly perform over dark regions, as the denominator of the ratio becomes insignificant. Hossain *et al.* [6] presented a method that utilizes the edge segment and it requires a number of initial training frames for generating background image. Kim and Hwang [7] presented a scheme that adapts edge difference method to compute current moving edges and temporal moving edges for detecting moving objects. This method fails to handle the changes of dynamic background, resulting in higher false alarm. Dewan *et al.* [8] proposed an edge segment based approach that utilizes three consecutive frames to compute difference images between each two consecutive frames. This method extracts edges from two difference images and finally detects moving objects using an edge segment matching algorithm. However, it fails to detect slow moving object because it does not utilize any background modeling. Lu *et al.* [9] introduced a method for extracting moving objects from their background based on multiple adaptive thresholds and boundary evaluation. Dailey *et al.* [10] devised a method that generates two edge maps from the edge difference images of three consecutive frames. The moving edges are extracted by applying logical AND operation between the two edge maps. Due to random noise or small camera fluctuation, pixel position of edges may change in consecutive frames. Thus, edge matching is not sufficient for detecting moving objects. This method also fails to detect moving objects with slow motion which indicate that it is not useful for real-time applications.

### 3 Proposed Method

Our proposed method aims to extract moving objects in an input image from their background. As depicted in Fig. 1, the proposed method consists of four steps: (a) Gradient map and edge map generation; (b) Gradient difference image calculation; (c) Gradient Masking and threshold selection; (d) Update background. Details of these steps are described in the following sections.

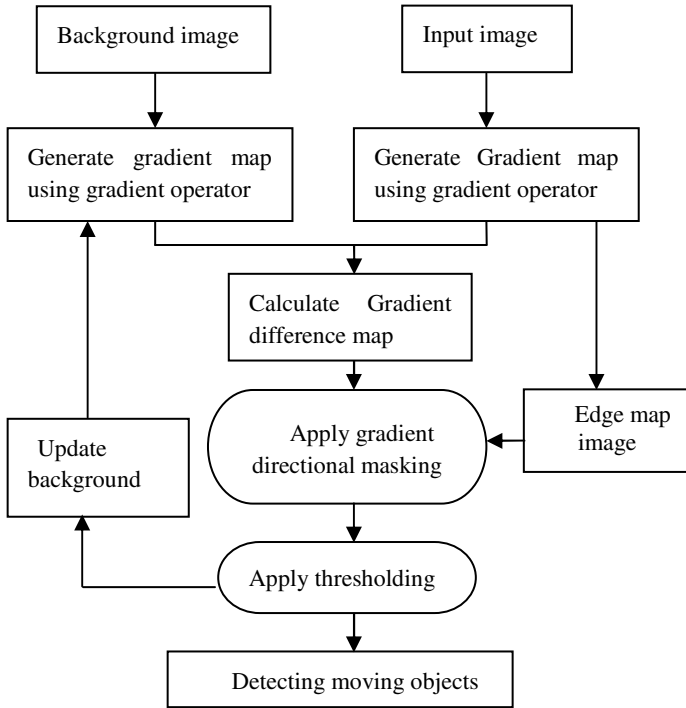


Fig. 1. Overview of the proposed method

#### 3.1 Gradient Map and Edge Map Generation

Edges are local changes in the image and are important features for analyzing images. Most of the edge based methods used conventional Canny edge detector to generate edge map. In this paper, we propose a new method that detects edges directly from RGB images i.e there is no grayscale conversion which reduces computational cost. It computes gradient magnitude and direction by using Euclidian distance and vector angle concept. Fig. 2 shows two 3x3 orthogonal and diagonal mask that are used for gradient computation. For each pixel let  $\underline{V}_1$  be the vector from the origin of the RGB color space to that pixel and  $\underline{V}_2$  be the vector along the main diagonal which represents gray scale line. Then, the angular distance of that pixel from the gray scale line through the origin can be determined as follows:

$$\gamma = \underline{V}_1 \cdot \underline{V}_2 / (|\underline{V}_1| * |\underline{V}_2|) \quad (1)$$

1	0	1
1	0	1
1	0	1

(a)

1	1	1
0	0	0
1	1	1

(b)

**Fig. 2.** (a) Orthogonal mask (b) Diagonal mask

The 3x3 orthogonal mask is applied to each pixel of both input image and background image to calculate the Euclidian distance between each two adjacent neighbors in horizontal and vertical direction. Let the distance in horizontal direction is  $E_H$  and the distance in vertical direction is  $E_V$ . Similarly, 3x3 diagonal mask is applied to calculate Euclidian distance between each two adjacent neighbors in two diagonal directions. Let, the distance along one diagonal direction is  $E_{D1}$  and the distance along another direction is  $E_{D2}$ . Finally, the total distance between each two adjacent neighbors along four directions can be approximated in the gray scale line using the following equation:

$$\delta M_{(x,y)} = \sum E * \gamma, \quad (2)$$

where  $\sum E$  is the total distance between each pair of adjacent neighbors in each four directions and  $\delta M_{(x,y)}$  is the gradient map image. The gradient direction is perpendicular to edge direction. This direction specifies the maximum changes occur in each direction. In our proposed method, the gradient direction is taken as the direction in which the Euclidian distance between two adjacent neighbors is maximum i.e.

$$\theta_{(x,y)} = (H, V, D1, D2) | \text{MAX}(E_H, E_V, E_{D1}, E_{D2}), \quad (3)$$

where  $H, V, D1, D2$  represent horizontal direction, vertical direction, 1<sup>st</sup> diagonal direction, and 2<sup>nd</sup> diagonal direction respectively.

The edge map image  $EM$  is extracted from the gradient map image by applying non-maxima suppression and thresholding. If  $\delta M_{(x,y)}$  is greater than certain threshold value then,

$$EM_{(x,y)} = \text{True};$$

Otherwise,

$$EM_{(x,y)} = \text{False};$$

### 3.2 Gradient Difference Image Calculation

Gradient difference image shows the structural difference between two images. Gradient difference image is calculated from the input gradient map and the background gradient map. This gradient difference is significantly large when changes occur between the images. In our proposed method, the gradient difference image  $\Delta M$  is computed by the following equation:



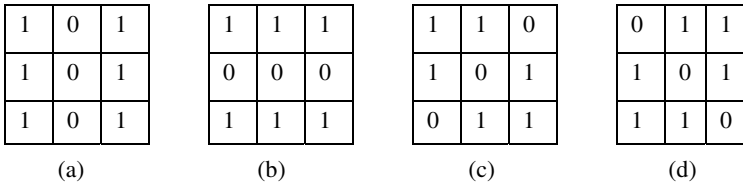
$$\Delta M = | \delta M_{I(x,y)} - \delta M_{B(x,y)} | \tag{4}$$

where  $\delta M_{I(x,y)}$  is the input gradient map and  $\delta M_{B(x,y)}$  is the background gradient map.

### 3.3 Masking and Threshold Selection

Gradient difference map discriminates the structure of the input image and the background image. This gradient structure provides robustness against illumination change. Most existing edge based methods utilize proper edge segment matching criteria to differentiate between moving edges and background edges. However, when a moving edge falls just over a background edge, edge segment matching algorithm detects moving edge as background edge. To overcome from this problem, in our proposed method we use gradient map which holds the position of edges as well as the information of edges. This information can be used to differentiate between overlapped moving edge and background edge.

Our proposed method introduces four directional masks such as Horizontal mask ( $M_H$ ), Vertical mask ( $M_V$ ), 1<sup>st</sup> Diagonal mask ( $M_{D1}$ ), and 2<sup>nd</sup> Diagonal mask ( $M_{D2}$ ) for four gradient directions as shown in Fig. 3. Gradient direction is perpendicular to edge direction. The gradient structure of an edge of both background and input gradient map exhibits almost similar pattern. Hence, the similarity between two gradient maps at certain pixel is calculated by using directional masking on gradient difference map. If this similarity is less than a certain threshold value then the pixel is considered as a background pixel. For simplicity, the number of identical gradient directions on each mask having non-zero value can also be considered.



**Fig. 3.** (a) Horizontal mask  $M_H$  (b) Vertical mask  $M_V$  (c) 1<sup>st</sup> Diagonal mask  $M_{D1}$  (d) 2<sup>nd</sup> Diagonal mask  $M_{D2}$

For each edge pixel of input edge map if  $EM_{(x,y)}$  is true, then the respective mask is selected according to gradient direction for that pixel. If the direction  $\theta_{(x,y)} = H$ , then the horizontal mask  $M_H$  is selected. If the direction  $\theta_{(x,y)} = V$ , then the vertical mask  $M_V$  is selected. The 1<sup>st</sup> diagonal mask ( $M_{D1}$ ) is selected when the direction  $\theta_{(x,y)} = D1$ . If the direction  $\theta_{(x,y)} = D2$ , then the 2<sup>nd</sup> diagonal mask ( $M_{D2}$ ) is selected. The selected mask is applied to the gradient difference map. The value,  $\Delta\delta_{(x,y)}$  at a position  $(x,y)$ , is calculated by the following equation:

$$\Delta\delta_{(x,y)} = \sum \Delta M_{(x+m,y+n)} * Mask_{(m,n)}, \tag{5}$$

where  $m$  and  $n$  varies from -1 to 1 and  $Mask_{(m,n)}$  is the mask value at position  $(m,n)$ .

The threshold value for the mask at any pixel can be determined as follows:

$$T_{(x,y)} = (\delta M_{I(x,y)} + \mu * S) / (S - M), \quad (6)$$

where  $\delta M_{I(x,y)}$  is the input gradient map,  $M \times M$  is the size of the mask.

Here,  $S = M * (M - 1)$  and  $\mu = \sum \delta M_{I(x,y)} / N$  where  $N$  is the number of pixels for which  $EM_{(x,y)}$  is True. If  $\Delta \delta_{(x,y)} < T_{(x,y)}$  then the pixel is considered as a background pixel otherwise it is a moving pixel.

### 3.4 Update Background

Updating background is very important to adapt the changes in background. Random noise may affect any pixel in the background but masking based detection alleviates this affect to detect moving object. In our proposed method we only update the background pixels. As the gradient structure is very consistent, each background pixel  $(x, y)$  is updated as follows:

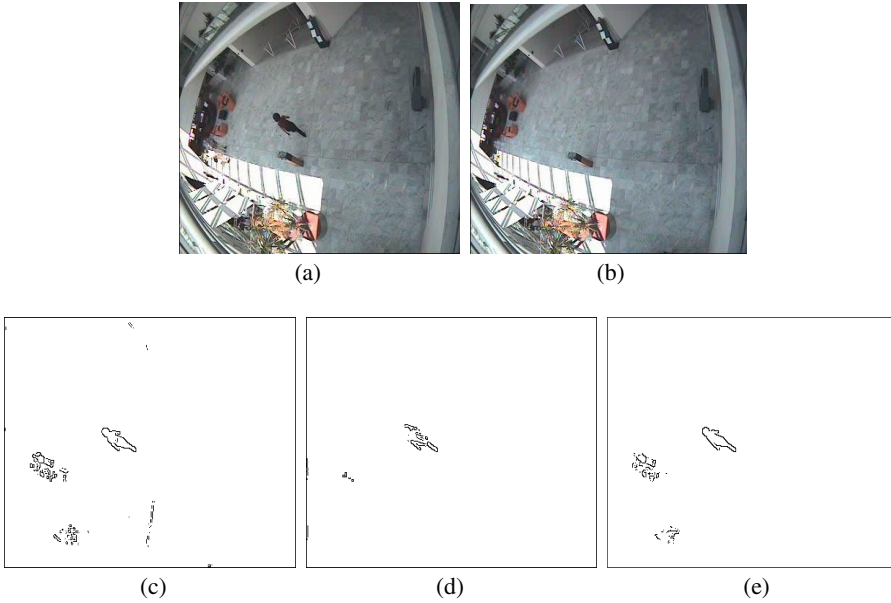
$$\delta M_{B(x,y)} = \beta * \delta M_{I(x,y)} + (1 - \beta) * \delta M_{B(x,y)} \quad (7)$$

where  $\delta M_{B(x,y)}$  is the background gradient map,  $\delta M_{I(x,y)}$  is the input gradient map,  $\beta$  is the learning rate.

## 4 Simulation Results and Discussion

In this section, we have compared the performance of our proposed method with and Kim and Hwang method [7] and Dewan and Chae method [8]. Experiments have been carried out with several video sequences captured from indoor, outdoor sunny, foggy video sequences to verify the effectiveness of our proposed method. We have used the video format of size 640x520 and used Intel Pentium IV 3.06 GHz processor and 256 MB of RAM. Visual C++ 6.0 has been used as of our working tools for implementation.

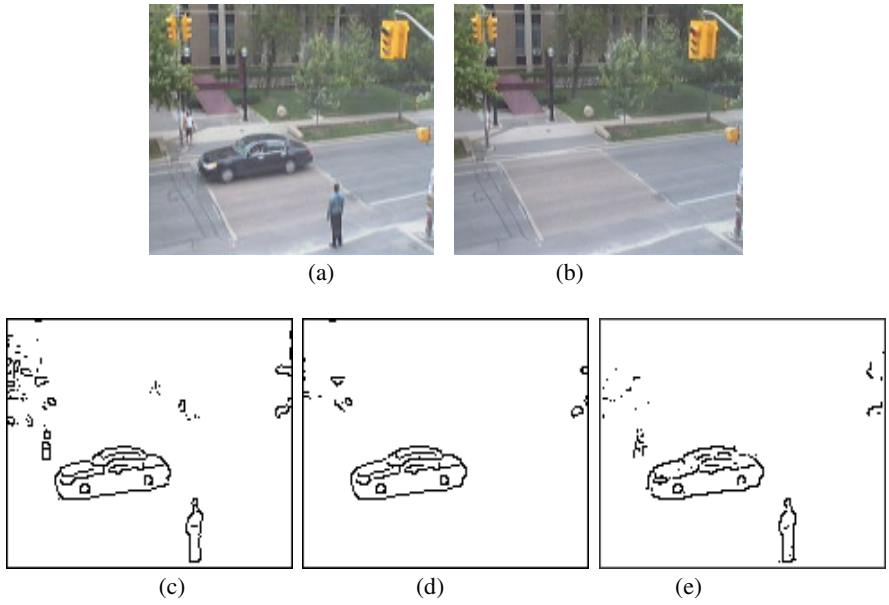
Fig. 4 shows the experimental results for an indoor video sequence. Fig. 4(a) and Fig. 4(b) show the 339<sup>th</sup> input frame and the background image respectively. Figures 4(c), 4(d), and 4(e) show the detection results obtained by Kim and Hwang method [7], Dewan and Chae method [8] and the proposed method. Fig. 4(c) shows that due to illumination variation Kim and Hwang method detects a lot of scattered edge pixels. The area illuminated by sunlight changes their position with time. As this method does not utilize background updating mechanism, this change detects many false edges. Fig. 4(d) shows that some moving objects with slow motions are absent in the detected region. Moreover, this method fails to detect properly the overlapped objects from consecutive frames. On the other hand, our proposed method effectively detected moving objects from the selected frame because of using edge localization mechanism and gradient directional masking.



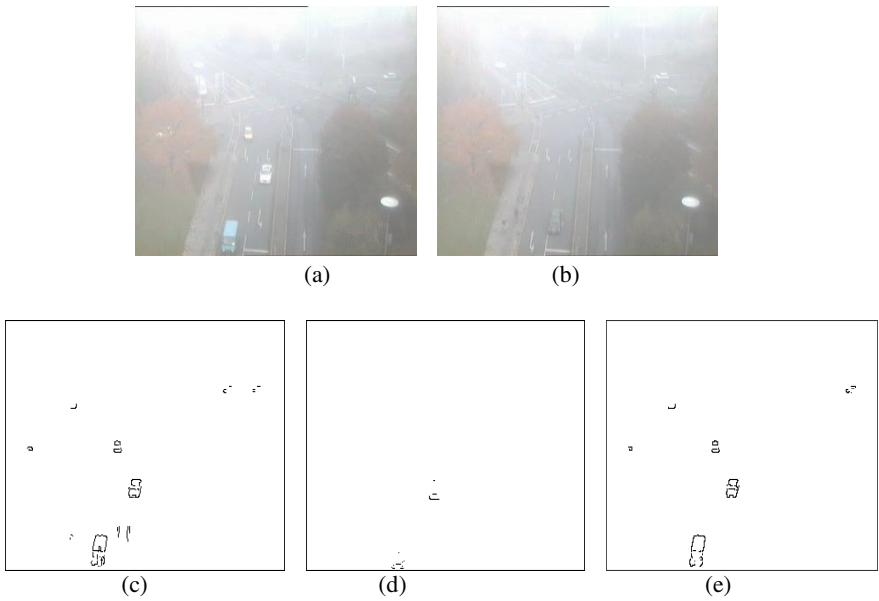
**Fig. 4.** (a) the 339<sup>th</sup> input frame (b) Background image; Detected moving object by (c) Kim and Hwang method; (d) Dewan and Chae method (e) proposed method

Fig. 5 shows the experimental results for an outdoor video sequence. Fig. 5(a) and Fig. 5(b) show the 952<sup>th</sup> input frame and the background image respectively. Figures 5(c), 5(d), and 5(e) show the detection results obtained by Kim and Hwang method [7], Dewan and Chae method [8] and the proposed method. Fig. 5(c) shows that the leaves of the blowing tree with small vibrations have strong effects on detecting objects. If the leaves of the tree change their position, a significant difference in each pixel at that location causes false detection. Dewan and Chae method [8] failed to detect moving objects which are suddenly stopped because it uses two consecutive difference images among three consecutive frames as shown in Fig. 5(d). Moving object in first two consecutive frames causes significant difference in the first difference image. If any object suddenly stopped at third frame, the second difference image contains no significant information, hence no edges are found in that image. Fig. 5(e) shows that our proposed method detected moving objects except the leaves of the blowing tree with small vibrations because of using gradient directional masking.

Fig. 6 shows the experimental results for a foggy video sequence. Fig. 6(a) and Fig. 6(b) show the 126<sup>th</sup> input frame and the background image respectively. Figures 6(c), 6(d), and 6(e) show the detection results obtained by Kim and Hwang method [7], Dewan and Chae method [8] and the proposed method. Fig. 6(c) shows that Kim and Hwang method failed to adapt the background changes because it does not use background updating. In addition, this method also failed to detect slow motion object effectively. Our proposed method overcomes those problems and it properly detected the moving objects as shown in Fig. 6(e).



**Fig. 5.** (a) the 952<sup>th</sup> input frame (b) Background image; Detected moving object by (c) Kim and Hwang method; (d) Dewan and Chae method (e) proposed method



**Fig. 6.** (a) the 126<sup>th</sup> input frame; (b) Background image; Detected moving object by (c) Kim and Hwang method; (d) Dewan and Chae method (e) proposed method

Table 1 shows the approximate time required to execute different modules of the proposed method. It shows that the total time required to process an image of size 640x520 is about 64 ms. Therefore, our method can process about 15 frames per second which is relatively good for real time detection.

**Table 1.** Mean processing time (in ms) for a 640x520 size image

Processing steps	Mean time (ms)
Gradient map & edge map extraction	30
Gradient difference map calculation	4
Masking & thresholding	26
Background updating	4
Total time required	64

Overall, our proposed method outperforms well known methods under different illumination conditions, including indoor, outdoor, and foggy cases for detecting moving object in real time.

## 5 Conclusion

In this paper, a new moving object detection method based on improved edge localization and gradient directional masking was introduced. Simulation results indicate that our proposed method provides better result than well known edge based methods. This is because it works on most recent successive frames and utilizes edge localization for detecting moving object. In addition, it is robust against different illumination changes. These results demonstrate that our proposed method can be a suitable candidate for moving object detection in video surveillance system. In future, we will extend our proposed method for tracking, recognition and classification of moving object.

**Acknowledgement.** This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MEST) (No. 2011-0017941).

## References

1. Sappa, A.D., Dornaika, F.: An Edge-Based Approach to Motion Detection. In: Alexandrov, V.N., van Albada, G.D., Sloot, P.M.A., Dongarra, J. (eds.) ICCS 2006. LNCS, vol. 3991, pp. 563–570. Springer, Heidelberg (2006)
2. Murshed, M., Ramirez, A., Chae, O.: Statistical Background Modeling: An Edge Segment Based Moving Object Detection Approach. In: Proc. of IEEE International Conf. on Advanced Video and Signal Based Surveillance, pp. 300–305 (2010)
3. Li, L., Leung, M.K.H.: Integrating Intensity and Texture Differences for Robust Change Detection. IEEE Trans. Image Process. 11(2), 105–112 (2002)

4. Liu, S., Fu, C., Chang, S.: Statistical Change Detection with Moments under Time-varying Illumination. *IEEE Trans. Image Process.* 7(9), 1258–1268 (1998)
5. Skifstad, K., Jain, R.: Illumination Independent Change Detection for Real World Image Sequences. *Comput. Vis., Graph. Image Process.* 46(3), 387–399 (1989)
6. Hossain, M.J., Dewan, M.A.A., Chae, O.: Moving Object Detection for Real Time Video Surveillance: An Edge Based Approach. *IEICE Transaction on Communication E90-B(12)*, 3654–3664 (2007)
7. Kim, C., Hwang, J.-N.: Fast and Automatic Video Object Segmentation and Tracking for Content-based Applications. *IEEE Transactions on Circuits and Systems for Video Technology* 12(2) (2002)
8. Dewan, M.A.A., Hossain, M.J., Chae, O.: Reference Independent Moving Object Detection: An Edge Segment Based Approach. In: Apolloni, B., Howlett, R.J., Jain, L. (eds.) *KES 2007, Part I. LNCS (LNAI)*, vol. 4692, pp. 501–509. Springer, Heidelberg (2007)
9. Wang, L., Yung, N.H.C.: Extraction of Moving Objects from Their Background Based on Multiple Adaptive Thresholds and Boundary Evaluation. *IEEE Transactions on Intelligent Transportation Systems* 11(1), 40–51 (2010)
10. Dailey, D.J., Cathey, F.W., Pumrin, S.: An Algorithm to Estimate Mean Traffic Speed using Uncalibrated cameras. *IEEE Transactions on Intelligent Transportation Systems* 1(2), 98–107 (2000)

# Automatic Detection of Face and Facial Landmarks for Face Recognition

Hajra Momin<sup>1</sup> and Jules-Raymond Tapamo<sup>2</sup>

<sup>1</sup> School of Computer Science,  
University of KwaZulu-Natal Durban, 4000, South Africa  
210556823@ukzn.ac.za

<sup>2</sup> School of Electrical, Electronic, and Computer Engineering,  
University of KwaZulu-Natal Durban, 4041, South Africa  
tapamoj@ukzn.ac.za

**Abstract.** Automatic face and landmarks detection on images is very important for face recognition. In this paper, we present an approach for detecting face and facial features such as eyes, nose and mouth in gray scale images. We make use of thresholding and connected component labelling algorithm to detect a face and extract features that characterize this face. This approach has the advantage that no manual interaction is required for choosing and extracting components. Experiments show promising results for face images having different orientation and facial expression.

**Keywords:** Binarization, Connected Component Labelling, Face Recognition, Landmarks Detection, Thresholding.

## 1 Introduction

Face and facial landmark detection has been a topic of extensive research for several decades due to important applications such as face recognition and identification and facial expression analysis. The first and very important step of face recognition is face detection in an image. There are several algorithms for face detection that have been proposed in the literature; more information about the subject could be found in [1, 2, 3, 4].

Color based approach for face detection is important due to its invariance to face scaling, poses, and facial expression. A face detection algorithm based on complexional segmentation and eyes location is presented in [5]. The algorithm establishes a proper color model in the concrete color space. Then uses facial complexional segmentation to distinguish the facial regions and the non-facial regions. After detecting a face in an image, the next phase is to extract facial features. These features generally include eyes, nose, mouth and chin. Methods for landmarks detection on face can be divided into three categories:

- (i) The first category is based on shape model. Cootes et al [6] proposed a shape based method that uses Active Appearance Model (AAM) and (Active Shape Model) ASM. Their algorithm uses the texture residual between the target and the estimated images to iteratively update the model's parameters. The ASM employs only shape parameters and it is guided by a local search around each point of the shape.

- (ii) The second category is based on face geometry. Majumder et al. [7] used facial geometry to locate the mouth, nose and eyes positions in an image. In [8] facial features are extracted using the geometric relation between the face and the eyes.
- (iii) In the third category, face skin color segmentation is used to detect faces in a digital image. In [9] skin color information is used to extract the face and features from an image. Ben Jemaa and Khanfir [10] also used color information to detect faces in images. They make use of chrominance components to detect the features in the image. Mauricio et al. [11] proposed a methodology for face detection using edge detection, region clustering, and shape analysis and landmark detection using surface curvature information and depth relief curves.

Component based automatic face recognition has been of interest to a large number of researchers in the past fifteen years. Rakhmadi et al. [12] proposed a connected component labelling algorithm for detecting face in digital images. Huang et al. proposed an Support Vector Machines based face recognition system which decomposes the face into a set of components that are interconnected by a flexible geometrical model [13]. They used 3D morphable model to generate 3D face models from only two input images. In [14] Heisele et al presented a component-based face recognition method and compared it to two global face methods, in respect to the robustness against the pose changes. Components were first extracted from the face, and features were then extracted from each component and combined in one single features vector. Support Vector Machines were then used in the component based and global methods to classify faces. Component based method outperformed the global methods. However, component-based classifier was still performing poorly with the pose changes. Fonou Dombeu and Tapamo [15] proposed an algorithm, based image thresholding and connected component labelling, that extracts the face components from the gray scale image. They used convex hull algorithm to validate the detected components.

In this paper, we present a method for detecting face and facial features from a gray scale image. Section 3 gives the detail description of the proposed face detection model. The detection of face Landmarks is done by splitting the face into two parts and masking them into the upper part which contains the eyes and the lower part which contains mouse and nose. The relevant components are then extracted using a method based on segmentation and connected component extraction.

The remainder of this paper is organized as follows. In section 2, we present materials and methods that we used for face detection and facial feature extraction. The methodology for face detection in an image is discussed in section 3. Section 4 is devoted to landmarks localization from the extracted face. Section 5 presents experimental results followed by a conclusion in section 6.

## 2 Materials and Methods

### 2.1 Binarization and Connected Component Labelling

Several binarization algorithms have been proposed in the literature. The one we used is based on the standard deviation and the mean of the input image as described in [15, 16]. An input image  $I$  with  $m$  rows and  $n$  columns is defined as



$$I = \{(i, j, x_{ij}) | 0 \leq i \leq n-1, 0 \leq j \leq m-1, 0 \leq x_{ij} \leq g-1\} \quad (1)$$

where  $n$ ,  $m$  and  $g$  are positive integers. The threshold  $T$  for an input image can be computed as follows.

$$t = k_1\sigma + k_2\mu \quad (2)$$

The value of  $k_1$  and  $k_2$  should be chosen from 0 to 2. In our work, we have established that  $k_1 = 0.65$  and  $k_2 = 0.9$  give the optimal solution.  $\mu$  and  $\sigma$  are the mean and standard deviation respectively. Formally, binarization transforms a gray level image  $I$  into a binary image  $I_B$  defined as

$$I_B(i, j) = \begin{cases} 1 & \text{if } I(i, j) \geq T \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

After the binarization process, the face will appear on the foreground of the binary image as shown in Fig. 2(b). Connected component labelling will then be applied on the resulting image to single out the main face components.

Given the Binary image,  $I_B$ , connected component labelling consists of partitioning  $I_B$  as follows

$$I_B = B \cup (\cup_{i=1}^n F_i) \quad (4)$$

Where  $B$  is the background and  $F_i$  for  $i = 1, 2, \dots, n$  are the connected components belonging to the foreground. All pixels in each component have the same label. Several algorithms for connected component labelling have been proposed in the literature [17], [18]. Fig. 3(a) and (b) show some examples of connected components of binary images.

### 3 Face Detection Method

In this section we present the segmentation approach used to extract the face region from the image. There are two constraints imposed on the input image. The image should

1. have dark background.
2. not have more than one face.

The proposed segmentation algorithm can be divided into two main modules. The first module will convert the gray scale input image into a binary image. The second module involves locating the face in the binary image thereafter and put bounding box around that face. The flow diagram of the face detection is shown in Fig. 1.

Once we get the binary image as shown in Fig. 2(b). We then find the 8-connected components in the binary image, a post-processing, consisting of hole-filling inside the faces, is then performed as shown in Fig. 2(c). There may be more than one blob in an image after applying hole filling algorithm. A blob with the biggest surface area is then selected. The final result consists of a bounding box put around the biggest blob, which is most probably the face detected as shown in 2(d). Fig. 2 shows step by step the results of face detection using the proposed method.

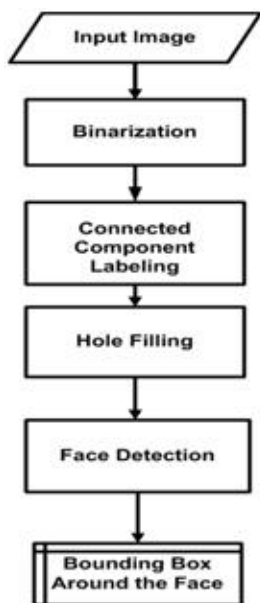


Fig. 1. Face detection process

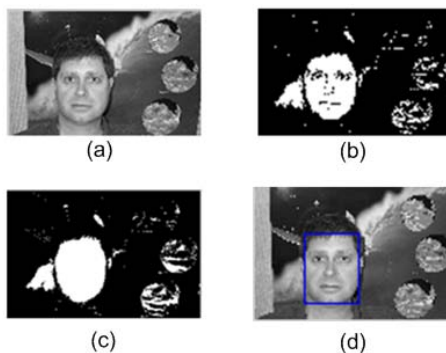
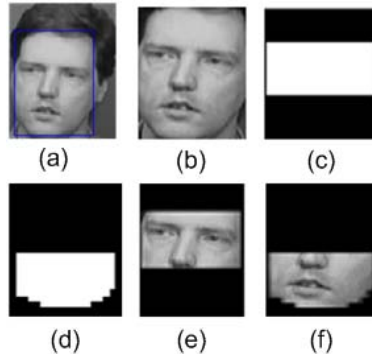


Fig. 2. Different phases of face detection. (a) Original image (b) Binarized Image (c) Resulting image after hole filling process. (d) Detected face with rectangular boundaries.

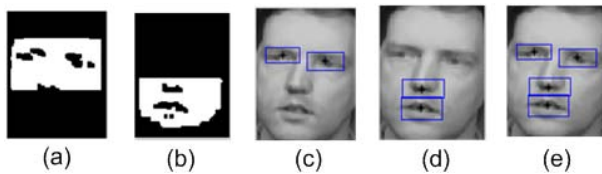
## 4 Face Landmark Localization

After detecting a face in an image, the next phase will involve finding the landmarks to prepare face recognition. Most common landmarks in a face are eyes, nose and mouth. The face is then divided into two parts: the upper and lower parts as shown in the Fig. 3(e) and (f). This is done by using masking. The mask used is shown in Fig. 3(c) and (d) for eyes and nose, and mouth respectively. The result of the application of those masks on the original image is shown in Fig. 3(e) to (f). Square mask was used for the eyes and oval mask for nose and mouth.



**Fig. 3.** Region of interest detection. (a) Detected face in input image (b) cropped face region (c) mask for upper part (eyes) (d) mask for lower part (nose and mouth) (e) masked image for upper part of image (f) masked image for lower part of image.

Our algorithm firstly searches and removes all the components with pixels touching the outer border of the image. Then, the remaining components situated in the inner face space are most probably face components. Thereafter, the size of each remaining component is computed as its total number of pixels. Based on the fact that, in the interior face space, the biggest components are most probable the eyes, the nose and the mouth, the number of remaining components is tested. We successively select the two biggest components from each masked image. As shown in Fig. 4(a) and (b) the biggest component will be eyes, nose and mouth. After the detection, bounding boxes are placed around eyes, nose and mouth on the original image as shown in Fig 4(e).



**Fig. 4.** (a) and (b) labeled image (c) and (d) detected bounding box for the component (e) detected components on single image

### 4.1 Proposed Algorithm

Given an Image  $I$ , the region of interest containing the face previously detected, the proposed Method to detect the landmarks is is summarised in *Algorithm 1*.

In Algorithm 1 *ExtractBoundingBox(CC)* extracts the bounding box around the connected component  $CC$ . The output of the proposed algorithm is a bounding box around each of the following face components: the two eyes, the nose and the mouth and centroids of face components. Fig 5(d) are examples of detected face components.

**Algorithm 1.** Automatic Detection of Face Components

---

**Require:**  $I$  ▷ Image Region of Interest containing the face

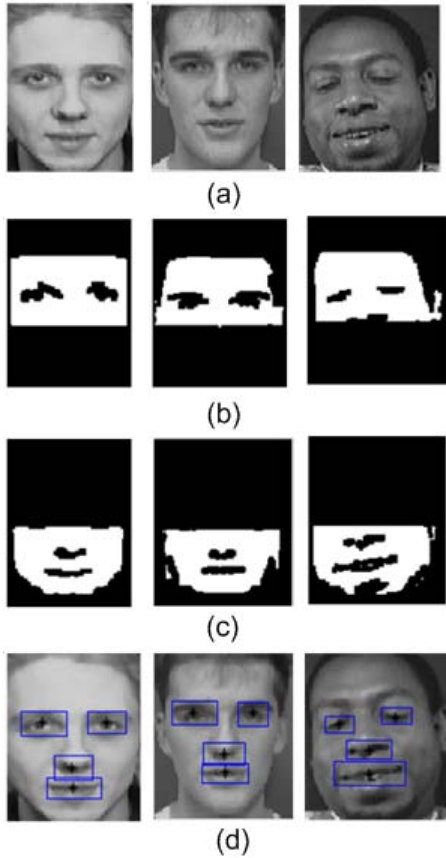
**Ensure:**  $LEye, REye, Nose, Mouth, C_{le}, C_{re}, C_n, C_m$  ▷ Bounding boxes representing the left eye, right eye, nose and mouth and their centroids

- 1: **Split**  $I$  into two regions, upper ( $Part_1$ ) and lower( $Part_2$ ) parts.
- 2: **Binarize**  $Part_1$  and  $Part_2$  using suitable threshold values of  $k_1$  and  $k_2$ . ▷ We have used  $k_1 = 0.65$  and  $k_2 = 0.9$  for both  $Part_1$  and  $Part_2$
- 3: **Calculate**, using 8-connectivity, the connected components, remove the components which are touching the outer border of the image, to obtain  $CPart_1 = (CPart_{1i})_{i=1,2,\dots,p}$  and  $CPart_2 = (CPart_{2i})_{i=1,2,\dots,q}$  of relevant connected components of  $Part_1$  and  $Part_2$  respectively
- 4: **Sort**  $CPart_1$  and  $CPart_2$  in increasing order of sizes of components.
- 5: **if**  $p < 2$  **then** ▷ The two eyes have not been detected; this may be due to hairs covering eye, or that the eye got connected to the border during the binarization process.
  - Extract** the second eye component based on human facial geometry.
- 6: **else**  $LEye = ExtractBoundingBox(CPart_{1,p-1})$ ,  $REye = ExtractBoundingBox(CPart_{1,p})$
- 7: **end if**
- 8: **if**  $q < 2$  **then** If only one component is detected, the second one is computed depending on the component already detected and if we don't find any valid component in this part then we can extract the mouth and nose of the person depending on the component detected in the upper part of the face. ▷ In most cases it is due to beard
- 9: **else**  $Nose = ExtractBoundingBox(CPart_{2,q-1})$ ,  $Mouth = ExtractBoundingBox(CPart_{2,q})$
- 10: **end if**
- 11:  $C_{le} = Centroid(CPart_{1,p-1})$  ▷ Left eye centroid
- 12:  $C_{re} = Centroid(CPart_{1,p})$  ▷ Right eye centroid
- 13:  $C_n = Centroid(CPart_{2,q-1})$ , ▷ Nose centroid
- 14:  $C_m = Centroid(CPart_{2,q})$  ▷ Mouth centroid

---

## 5 Experimental Results

Our method was tested on images from *The ORL face database* [19]. This database contains 400 images of 40 different subject having 10 images per person in different scale and orientation and different facial expressions. The images are gray scale images of size 112 x 90. Table 1 shows the detection rate of face components in frontal, left, and right view. In frontal view, the detection rate of left eye and right eye are nearly 100%, the nose 90%, and the mouth over 86%. In left view, the ratio of detection of the nose, the mouth, and the right eye is over 80% and the left eye over 70%. Some results of detected components of faces in different orientations are shown in Fig. 6. Table 2 shows the detection rate of face components in different facial expression. The detection rate of left eye and right eye for smiling facial expression is 90%, and nose and mouth is 75%. The detection rate for sad facial expression is, 90% for left and right



**Fig. 5.** (a) input image (b) connected component in upper half part (c) connected component in lower half part (d) detected component with bounding box

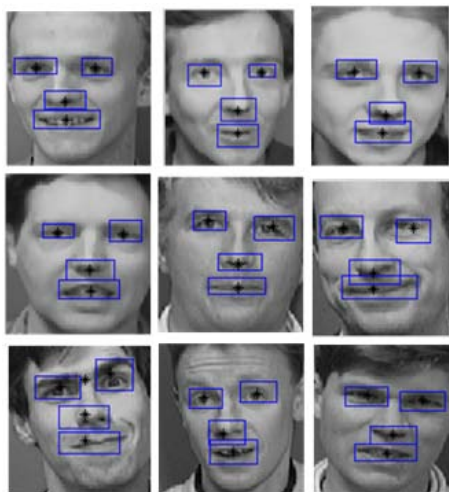
**Table 1.** Detection rate (%) of facial component in the frontal, left, right view

Face Orientation	Left Eye	Right Eye	Nose	Mouth
Front View	96	96	92	88
Left View	75	96	90	93
Right View	92	82	93	87

eye, 80% for the mouth and 75% for mouth. The detection rate for closed eyes has no effect on detection rate that is if eyes are closed still it will be detected using our method. The detection rate for mouth open is 90% for left and right eye, 85% for nose and 80% for mouth. Fig. 5 shows examples of detected facial components in different facial expressions.

**Table 2.** Detection rate (%) of facial component for different facial expression

Facial Expression	Left Eye	Right Eye	Nose	Mouth
Smiling	95	95	75	75
Sad	90	90	75	80
Closed eyes	96	96	87	85
Mouth Open	92	92	85	80

**Fig. 6.** Faces in different orientations with bounding boxes around the detected components**Fig. 7.** Faces with different facial expressions and bounding boxes around the detected components

## 6 Conclusion

A fully automatic method was designed. This method detects a face and facial landmarks in frontal, left and right face view and also the expressive images of high complexity. The complexity of the expression was presented by closed/semiclosed eyes, variety of mouth appearances including open and tight mouth, Smiling and sad. Our segmentation approach extracts the entire face region by combining thresholding and connected component labeling. The system locates facial features, which are eyes, nose, and mouth in an image. It also does not require any manual intervention by users, like manually creating a feature location model.

Extensive experiments demonstrate that our model has achieved very good results with face images with frontal, left and right view and changes in facial expression. With the proposed method we also achieve very good detection rate on faces taken in different orientations and facial expression. The presented approach is simple and fast for detecting facial components. However, there are some challenging cases, such as faces with specs and beard, for which our method has a poor performance. We expect to extend the method to handle faces with those problems.

## References

- [1] Heisele, B., Serre, T., Pontil, M., Poggio, T.: Component-based face detection. In: CVPR 2001, vol. I, pp. 657–662 (2001)
- [2] Viola, P., Jones, M.: Rapid object detection using boosted cascade of simple features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2001), pp. 511–518 (2001)
- [3] Yang, M.-H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1), 34–57 (2002)
- [4] Marius, D., Pennathur, S., Rose, K. Face detection using color thresholding and eigenimage template matching, [http://www.stanford.edu/class/ee368/project\\_03/project/reports](http://www.stanford.edu/class/ee368/project_03/project/reports)
- [5] Zhang, Q., Liu, Z.J.: Face detection based on complexional segmentation feature extraction. In: International Conference on Intelligent Information Hiding and Multimedia Signal Processing (2006)
- [6] Cootes, T.F., Edwards, G.J., Taylor, C.J.: Active appearance models. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23(6), 681–685 (2001)
- [7] Majumder, A., Behera, L., Subramanian, V.K.: Automatic and robust detection of facial features in frontal face images. In: UKSim 13th International Conference on Modeling and Simulation, pp. 331–336 (2011)
- [8] Chaudhari, S., Kale, A., Kinage, K.S., Bhirud, S.G.: Face feature detection and normalization based on eyeball center and recognition. In: 2nd International Conference on Future Computer and Communication, vol. 3, pp. 503–507 (2010)
- [9] Zhi-fang, L., Zhi-sheng, Y., Jain, A.K., Yun-Qiong, W.: Face detection and facial feature extraction in color image. In: 5th International Conference on Computational Intelligence and Multimedia Applications (2003)
- [10] Jemaa, Y.B., Khanfir, S.: Automatic local gabor features extraction for face recognition. *International Journal of Computer Science and Information Security* 3(1) (2009)
- [11] Segundo, M.P., Silva, L., Bellon, I.R.P.: Automatic face segmentation and facial landmark detection in range images. *IEEE Transactions on Systems, Man, and Cybernetics-part B: Cybernetics* 40(5) (2010)

- [12] Rakhmadi, A., Rahim, M.S.M., Bade, A., Haron, H., Amin, I.M.: Loop back connected component labelling algorithm and its implementation in detecting face. *World Academy of Science, Engineering and Technology* (64) (2010)
- [13] Huang, J., Blanz, V., Heisele, B.: Face Recognition using Component-Based SVM Classification and Morphable Models. In: Lee, S.-W., Verri, A. (eds.) *SVM 2002*. LNCS, vol. 2388, pp. 334–341. Springer, Heidelberg (2002)
- [14] Heisele, B., Ho, P., Wu, J., Poggio, T.: Face recognition: component-based versus global approaches. *Computer Vision and Image Understanding* 91(1-2), 6–21 (2003)
- [15] Fonou-Dombeu, J.V., Tapamo, J.R.: Validation of detected facial components for an accurate face recognition. In: 18th Annual Symposium of Pattern Recognition Association of South Africa (PRASA), pp. 141–146 (November 2007)
- [16] Ritter, G.X., Wilson, J.N.: *Handbook of Computer Vision Algorithms in Image Algebra*, 2nd edn. CRC Press (2001)
- [17] Yang, Y., Zhang, D.: A novel line scan clustering algorithm for identifying connected components in digital images. *Image and Vision Computing* 21(5), 459–472 (2003)
- [18] Stefano, L.D., Bulgarelli, A.: A simple and efficient connected component labelling algorithm. In: *International Conference on Image Analysis and Processing*, pp. 322–327 (1999)
- [19] AT&T: The orl database of faces, [www.uk.research.att.com/facedatabase.html](http://www.uk.research.att.com/facedatabase.html)



# A Tool for Ranking and Enhancing Aesthetic Quality of Paintings

W.A.P. Wickramasinghe<sup>1</sup>, Anuja T. Dharmaratne<sup>2</sup>, and N.D. Kodikara<sup>2</sup>

<sup>1</sup> Faculty of Visual Arts, University of the Visual & Performing Arts,  
No. 46, Horton Place, Colombo 07, Sri Lanka

<sup>2</sup> University of Colombo School of Computing,

No: 35, Reid Avenue, Colombo 07, Sri Lanka

ajith.w@gmail.com

{atd,ndk}@ucsc.cmb.ac.lk

**Abstract.** Measuring aesthetic value of an artwork is a significant task in the field of visual & performing arts. Artists follow several techniques manually using traditional methods to balance the visual aesthetic value of different aesthetic products such as a film, a drama, a painting etc. Today, artists are enthusiastic on emerging information technology techniques for judgment and enhancement of designed product aesthetically and efficiently while applying traditional concepts to design initial form of the artwork. Computational aesthetics is the research of computational methods that do make applicable aesthetic decisions in a similar fashion as human can. This paper introduces a new tool that can be used to rank a given digital image of paintings based on a common parameter set with their weighting factors which are supposed to be adjusted for changing the aesthetic level of a particular painting in the area of computational aesthetics.

**Keywords:** computational aesthetics, aesthetics quality, artwork (painting) evaluation, analytic hierarchy process (AHP).

## 1 Introduction

The theory of aesthetics introduced by George David Birkhoff (1933) in his book *Aesthetic Measure* involves a lot of computational methods. Aesthetic quality assessment is a challenging task as it is a subjective matter and there are not clear standard procedures which can be applied directly to measure it. Therefore, particular area is needed to study well in terms of evaluation and enhancement of a given artwork in different field of visual & performing arts. Practically, it depends on the quantitative and qualitative parameters which lead to change the aesthetic feelings of human beings regarding the particular product that we are concerning. In case of comparing a given set of aesthetic products, evaluators may follow particular criteria and rank them as very good, good, bad etc. Basically, it is a qualitative measurement and a team of judges will end up with a result like first place, second place etc. Therefore, it is required to proceed with the observations made by them using a particular hierarchical methodology and to find an approach to propose a model which is applicable to rank them quantitatively.

In this paper, it will propose a new tool using a mathematical approach, called Analytic Hierarchy Process (AHP) [1] for ranking and enhancing digital image of paintings. Further, this paper discusses pros and cons of the experimental results of that approach to justify the suitability of the proposed tool. The paper is organized as follows. The first section contains a discussion on what aesthetics quality is and the importance of computing aesthetics using a model. The second section discusses about related work while emphasizing the challenges of aesthetic computing and enhancement of the particular aesthetic product. Next section discusses about the tool for ranking aesthetic quality of paintings. Then experimental results have also been presented there followed by evaluation and discussion. After that, conclusion and future work have been appended ending with references.

## 2 Related Work

Only a few research works have been published as aesthetic visual quality assessment is still a new research area. Especially for assessing paintings, we were able to find two previous works to our best knowledge [1], [2]. It is required a multi-criteria decision making method for computing aesthetic value, since it depends on several parameters of an artwork. Thus, a new approach called AHP was applied as the core theory of this tool implementation. According to Wickramasinghe et al., AHP is a multi-criteria decision-making method originally developed by Thomas L. Saaty (1980). It can be used to select the best painting with relative weights of its parameters that contribute to the final goal: *highest visual aesthetic value*. When dealing with aesthetic decisions in artworks, normally, evaluator or a group of evaluators will have to express satisfaction feelings based on the aesthetic quality of the artwork. In the current tool, a set of different image groups (paintings) that can be analyzed based on their form (physical attributes, shape, composition, etc.) were given as the input paintings selected by referring some reputed painting archives publicly available[3],[4].

## 3 Proposed Tool for Ranking Given Set of Paintings

According to AHP theory, a hierarchy of parameters was set up having main goal in the top level with level 1 parameters and level 2 parameters in next two levels by consulting experts and conducting a preliminary survey. Parameters of Level 1 were selected as the most significant criteria (factors) to achieve the final goal. Level 2 parameters are sub criteria of selected level 1 parameters. Then, it was done a survey by selecting 50 participants and data were collected for three different paintings under five painting themes for the above mentioned hierarchy. After formulating pair-wise comparison matrixes for each criteria and sub criteria, relative weights (priority vectors) should be calculated. There are several methods for calculating the eigenvector (priority vector). Multiplying together the entries in each row of the matrix and then taking the  $n^{\text{th}}$  root of that product gives a very good approximation to the correct answer. Comparison matrix data were directly entered to an online AHP calculation tool [8] and relative local weights were calculated separately for each and

every member in the survey. After getting relative weights (priority vectors) for all members, mean relative weights were calculated using a Microsoft Excel data sheet. Accordingly, global weights were calculated using AHP theory. A painting which was taken maximum total (global) weight is the one which is having the highest visual aesthetic value. Therefore, this approach is specific where comparison results have to be used only for the cases similar to given data set, neglecting the variety inbuilt with the available digital image of paintings to be compared. So, authors plan was to go beyond this limitation of the AHP technique providing more rooms to accommodate comparisons to the diverse category of paintings. One of the possible solutions is to develop a tool (system) to get online response from the user/users and do the comparisons for the given set of paintings. Then, AHP theory can be utilized as the main technique to do all the required calculations to rank the given set of paintings. To proof the concept, comparisons for three digital images of paintings were selected in the current proposed tool. The architecture of this tool is given below (figure 1).

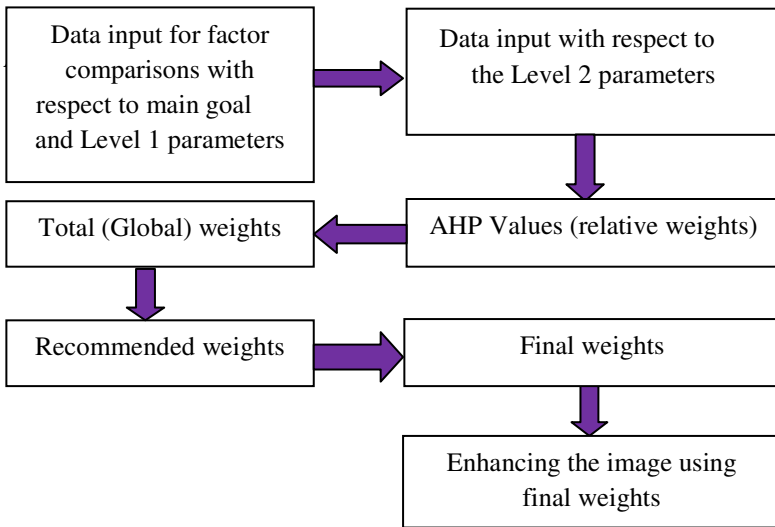


Fig. 1. The architecture of the tool

#### 4 Implementation and Experimental Results

To generalize the above concept, tool was developed having three main windows. Net beans [9] software was used as the main IDE and some other open source software like ImageJ[10] developed by National Institute of Health, USA was integrated to the developed tool specially for the enhancement of the digital image. User will have to follow the scale of measurement from -9 to +9.

In this tool, second window will appear by clicking start comparisons button on first window. There, user has to enter responses within the given scale (-9 to +9) by comparing level 1 parameters two by two with respect to the main goal. And also, all

level 2 parameters will have to be compared with respect to the level 1 parameters. In third window, responses should be entered by comparing three paintings (painting 1, painting 2 & painting 3) two by two with respect to the level 2 parameters.

#### 4.1 Example Test Case: Three Animal Theme Paintings

There were three input painting sets used for the comparisons by different users. Following figures (figure 2, figure 3 and figure 4) will show one of the example test cases for the purpose of explaining the output results.



**Fig. 2.** Painting 1



**Fig. 3.** Painting 2



**Fig. 4.** Painting 3



**Fig. 5.** Adjusted painting 1



**Fig. 6.** Adjusted painting 3

According to the responses given by a particular user for three paintings under animal theme, tool will calculate AHP (relative weights) values and the total weights of the parameter contributions (table 1).

**Table 1.** Total weights of parameter contributions

Painting	Total weights
Painting 1	33.6
Painting 2	34.6
Painting 3	31.6

According to AHP theory, painting 2 (table 1) was selected as the best aesthetic quality painting comparing to other two paintings. Painting 2, Painting 1, Painting 3 is the ranking order. Now an extension work can also be done based on the findings. Now, painting 1 and painting 3 can be enhanced to the aesthetic level of painting 2 by changing their individual parameter weights to the same values as the painting 2. That enhancement can

be done using another tool called ImageJ[10] which is integrated to the developed tool. In this enhancement, Authors decided to exclude three parameters called texture, straight lines and curve lines as those are directly connected with the initial design of the painting. So, remaining 6 parameters are considered and weights are adjusted accordingly and displayed as recommended weights in the proposed tool (table 2).

**Table 2.** Recommended weights of parameter contributions

Painting	Dark Color	Light Color	Low Contrast	High Contrast	Low Brightness	High Brightness
Painting 1	32.5	10.3	13.9	12.9	14.8	15.6
Painting 2	21.6	17.2	12.6	15.3	16.2	17.2
Painting 3	23.8	15.1	15.8	13.6	18.7	13.4

It is assumed that dark color or light color will dominate the other. And also, same concept can be applied to the contrast and brightness. So, ultimately, following three parameters are most critical parameters that should be adjusted in painting 1 & 3(table 3).

**Table 3.** Adjusted weights of parameter contributions

Dark Color	High Contrast	High Brightness
21.6	15.3	17.2

As the throughput of dark color and high brightness is dark color or low brightness and weight should be the difference of those two values, final weights will be displayed as in table 4.

**Table 4.** Final weights of parameter contributions

Dark Color or Low brightness	High Contrast
4.4	15.3

After adjusting these weights of those painting 1 & 3 by using ImageJ, final out put can be seen as in figures 5and 6.

## 5 Evaluation and Discussion

In the evaluation of the results set, feedback was taken from 3 experts in the subject area. They agreed to the final output, as the tool is straightforward to get the output within short period of time. And also, they were able to compare the dynamic views of more than one user, even though, it limits for only three given painting comparisons. Rather than the appearance of the final output, they agreed to the flexibility to get the dynamic views of several participants. Specially, they commented on the behavior of the tool which is considered any type of paintings for the comparisons. One of them commented on the limitation of the system only for three painting comparisons.

## 6 Conclusion and Future Enhancement

According to the evolution and comments of the experts, we can conclude that tool will do a satisfactory job in the area of computational aesthetics. Currently, there is an online tool available only for computing AHP values [8]. It is limited for comparing a particular local parameter set only. But our tool facilitates to compare a hierarchy of factors related to a set up goal to be achieved. And also, there isn't any system or tool available for comparison of paintings using a computer. Therefore, this will play an important role for the artist who is really looking at emerging IT techniques for their aesthetic decisions. This was developed as a proof of concept and there is a chance to enhance the tool for more than three paintings as a future enhancement of this work.

## References

1. Wickramasinghe, W.A.P., Dharmaratne, A.T., Kodikara, N.D.: A mathematical model for computational aesthetics. In: Proceedings of the International Conference on Computational Vision & Robotics (ICCV 2010), Bhubaneswar, India, pp. 130–137 (2010)
2. Li, C., Chen, T.: Aesthetic Visual Quality Assessment of Paintings. *IEEE Journal of Selected Topics in Signal Processing* 3(2) (2009)
3. Kalwick, W.J.: Painting archives, <http://www.kalwick.com>
4. Potter, J.: Painting archives, <http://www.johnpotter.com>
5. Saaty, T.L.: *The analytic Hierarchy Process*. McGraw-Hill, New York (1980)
6. Birkhoff, G.D.: *Aesthetic Measure*. Harvard University Press, Cambridge (1933)
7. Goldman: *Aesthetic value*, Westview Press, Colorado (1995)
8. Tools, Canadian Conservation Institute, [http://www.cci-icc.gc.ca/tools/ahp/index\\_e.asp](http://www.cci-icc.gc.ca/tools/ahp/index_e.asp)
9. NetBeans IDE 7.0, <http://netbeans.org/>
10. ImageJ: Image Processing and analysis in java, <http://rsbweb.nih.gov/ij/download.html>

# Aging Progression of Elderly People Using Image Morphing

L.L. Gayani Kumari and Anuja T. Dharmaratne

University of Colombo School of Computing (UCSC),  
Colombo, Sri Lanka  
gayanillgk@gmail.com, atd@ucsc.lk

**Abstract.** Aging is an inevitable process and its effects cause major variations in the appearance of human faces. Human face identification has a significant amount of information depend on his age, gender, ethnicity and etc. In addition, facial expression and facial gestures often reveal the emotional state of an individual. Consequently human facial analysis has received considerable attention and has led to the development of novel approaches to perform face recognition, facial expression characterization, face modeling, etc. Facial aging is attributed by changes in facial features, shape and texture and other biological factors like weight loss/gain, facial hair, etc. Age seems to be the main cause of the facial change and it has become forefront. Human life cycle can be classified in to four main stages with the age. Those are babies, children, young adults and elderly. Significant amount of facial changes can be identified in each of these stages. This paper introduces a methodology for elderly facial shape changes, hairlines recede and hair colour change using image morphing.

**Keywords:** Image Warping, Image Morphing, Age Progression Techniques, Wrinkle Transformation, Texture Enhancement, Compositional and Dynamic Model, Graph based Image Representation, Automatic Age Estimation.

## 1 Introduction

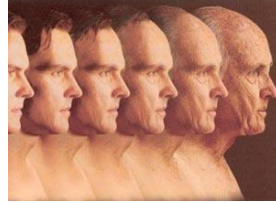
The last decade has brought several important advances in face recognition research. Face recognition is a wide area in research, and there are varieties of ways that it could be carried out such as Graph matching, anthropometric studies age estimation, facial feature analysis, image warping, image morphing, etc. Human Facial Aging can be categorized into two main phases as formative phase and adult phase. A methodology for the formative phase (age from 0 to 18) aging is being proposed in [1]. Adult phase can be further classified as adult primary phase (age from 19 to 50) and adult secondary phase (age 50 upwards). A methodology for the adult primary phase aging is being proposed in [2]. This paper focus on shape changes of faces in elderly aging progression using image morphing techniques. Our previous papers [3] and [4] discussed a collection of aging progression techniques for elderly people. Most researchers used FG-NET [5] and MORPH [6] databases for model construction and testing purposes. We could observe a variety of key facial feature changes in



elderly stage. Those are fine lines and wrinkles on the face, lips thin out, hair colour changes, hairlines recede, balding, facial texture changes, facial shape changes, and muscles drop. However, the individual diversity is broad. Factors such as race, gender, genetics and lifestyle come into play during the aging process. Fig. 1 and Fig. 2 show the facial feature diversity of female and male.



**Fig. 1.** Female facial feature diversity



**Fig. 2.** Male facial feature diversity

## 2 Related Work

The basis of Shan et al.'s approach, Image Based Surface Detail Transfer (IBSDT) [7], is that if the bumps of an old person are transferred to a young person, then the young person's face will look old. This was further discussed by Maulin R. Gandhi [8]. Karl Ricanek Jr and Tamirat Tesafaye use the IBSDT method to age a novel image to a target age that is represented by one of the age prototypes and then use the aging function to determine the exact amount of wrinkle information to transfer from the age prototypes to the novel facial image to simulate accurate and realistic aging [6]. Maulin R. Gandhi & Martin D. Levine proposed a method for Automatic Synthesis of Aged Human Facial Images [9]. They explore and demonstrate new results use of Support Vector Machines (SVM's) [10] for predicting the age of a given facial image. Tiddeman et al proposed a methodology for texture enhancement and age transformation [11]. This methodology produced very convincing results using a new, wavelet-based method for prototyping facial textures and for artificially transforming the age of facial images. This was discussed in detail [7, 8]. Maulin R. Gandhi & Martin D. Levine proposed an improvement [9] to the wavelet-based method for prototyping facial textures done by Tiddeman et al [11]. J. Suo, S. C. Zhu, S. Shan and X. Chen presented a compositional and dynamic model for face aging [12]. The compositional model represents faces in each age group by a hierarchical And-Or graph, in which And nodes decompose a face into parts to describe details (e.g. hair, wrinkles, eyes, mouth, etc.) crucial for age perception and Or nodes represent large diversity of faces by alternative selections and then a face instance is a transverse of the And-Or graph—parse graph. Gayathri Mahalingam and Chandra Kambhamettu presented a graph based image representation and an aging model constructed using Gaussian Mixture Model (GMM) [13] for each individual to model their age variations mainly in shape and texture [14]. A modified Local Feature Analysis that uses Fisher score to extract the feature points and uniform Local Binary Pattern (LBP) [15] operator is applied to each feature point to compute a feature descriptor for each feature point, and is used in the graph representation. X. Geng, Z.-H. Zhou and Kate

S-M proposed an automatic age estimation method named AGES (AGing pattErN Subspace). [16], which improves their previous research [17]. AGES relies on many landmark points in the face images, eventually these landmarks should be determined by applying automatic land marking algorithms like [18]. Beier and Neely in [19] have introduced a warping algorithm called feature based warping which can be divided into two mechanisms: single line pair warping and multiple line pair warping. In multiple line pair warping, the pixels between two lines can be transformed into the gap between another two given lines. So this is the best warping mechanism we have applied for proposed solution because here, a given set of features are going to be transformed along with the aging details to change those features accordingly.

### 3 Methodology

Researchers strongly discussed about wrinkle transformation, texture changes in elderly people aging. Face shape change is also a key factor in elderly people aging. Face shape change in elderly people, is a topic that got less attention in the research pool. Novelty of this project began with the intension of proving the face shape with aging progression. The main objective of this research is to change the adult face shape, hairlines recedes and hair colour change. Thus the facial features such as forehead, chin, cheeks, lips and hairline thickness are considered in the research. With aging the forehead should grow, chin should sharp and broad, cheeks should shrink, and lips should thin out, hairlines should recede and hair colour should gradually become white. These facial features analyze and synthesize with the age. With the intention of analyzing this scenario, we have implemented a proof of concept which is a software application. The key methodology, primary process and flow are described below. The application requests user to upload the face image of age 40. Then user needs to mark ten feature points as per given in the fig 3. After that, facial image is morphed using a warping algorithm [19] and output images are generated for ages 50, 60 and 70. The face shape change and hairlines recede are purely achieved using warping algorithm [19]. Hair colour change was a result of predefined hair affected image set which is similar to original image of age 40. It could be a limitation of this application. To have a fine result at the end, we have used 3 sets of hair colour change images for each uploaded image and morphed at the runtime with the resultant image to get the hair effect in each age 50, 60 and 70. All important steps of the process are described in below sections 3.1, 3.2, 3.3 and 3.4.

#### 3.1 Feature Extraction from the Original Source Image

As shown in the Fig. 3 we need to take ten feature points of the original image which is the source image. Thickness of the front hair lines (3 feature lines), Thickness of the side hair lines (2 feature lines), Height of the forehead, Height of the chin, Width of the left cheek, Width of the right cheek and width of the lip line. This methodology mostly successful for the face images which are faced to the fore front, mainly because of easiness of the localizing the feature points. Otherwise due to some of the face poses, it would be extremely difficult to capture the feature points as they are. Thus there should be a rotation mechanism to set the face horizontally.

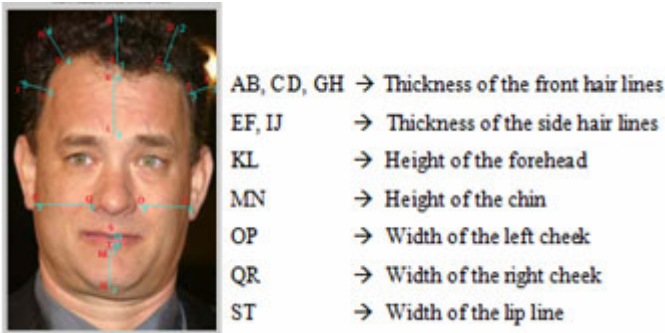


Fig. 3. Facial feature points taken for the analysis

### 3.2 Calculate Feature Difference Mean Value

As shown in Fig. 4 we have measured (manually) 10 features of each image in the sample data set. For each age range the feature mean value differences are calculated and saved in a data file where the application can read it at the runtime.

Feature point	AB	CD	EF	GH	IJ	KL	MN	OP	QR	ST
X Difference	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
Y Difference	Y1	Y2	Y3	Y4	Y5	Y6	Y7	Y8	Y9	Y10

Fig. 4. Facial feature points value example matrix

For each image the sample database measure for the above features are listed as in Figure 4. There after for each age it takes the mean difference value of each feature as shown in Fig. 5.

Age 50										
Feature point	AB	CD	EF	GH	IJ	KL	MN	OP	QR	ST
X Difference	X'1	X'2	X'3	X'4	X'5	X'6	X'7	X'8	X'9	X'10
Y Difference	Y'1	Y'2	Y'3	Y'4	Y'5	Y'6	Y'7	Y'8	Y'9	Y'10
Age 60										
Feature point	AB	CD	EF	GH	IJ	KL	MN	OP	QR	ST
X Difference	X''1	X''2	X''3	X''4	X''5	X''6	X''7	X''8	X''9	X''10
Y Difference	Y''1	Y''2	Y''3	Y''4	Y''5	Y''6	Y''7	Y''8	Y''9	Y''10
Age 70										
Feature point	AB	CD	EF	GH	IJ	KL	MN	OP	QR	ST
X Difference	X'''1	X'''2	X'''3	X'''4	X'''5	X'''6	X'''7	X'''8	X'''9	X'''10
Y Difference	Y'''1	Y'''2	Y'''3	Y'''4	Y'''5	Y'''6	Y'''7	Y'''8	Y'''9	Y'''10

Fig. 5. Facial feature points mean value example matrix for age ranges

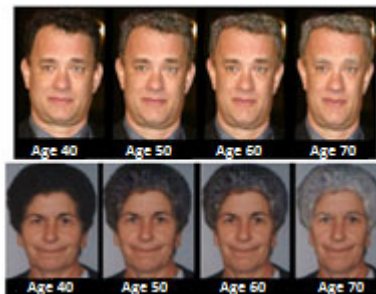
### 3.3 Calculate Destination Feature Values

From the feature points the 'x', 'y' values captured from the original image can be derived from the marked horizontal and vertical lines as per shown in the Fig 3. Then for the age groups (40-50), (50-60) and (60-70) the mean feature values for the above

mentioned distance differences are read from the data file (data file is similar to Fig. 5) from the application. Then new lengths for the facial features according to the expected age-range have been calculated. From that we can derive the destination image feature values.

### 3.4 Simulate the Final Image from Source and Destination Images

Morphing is a transformation of images. It is used in many animations and movies to warp a first image into a second image. The morph process transforms two images to the resulting image. Due to that the resulting image does not deviate from the base "shape". The feature based warping mechanism introduced by Beier and Neely [17] is used here as the warping technique. The basic idea is that feature lines are interactively selected by the "animator" in the two images. These feature lines create a reverse mapping, by calculating pixel by pixel the destination image by sampling the correct pixel from the source image. A morph operation blends between two images source and destination. A set of corresponding lines are defined in both images. A new set of lines are calculated for each intermediate frame of the transformation. This is accomplished by interpolating endpoints of the lines in two images and then they are simply cross-dissolved. Digging into more details with the same, this mechanism uses a technique which is nothing but 'warping with multiple line pairs'. It uses a weighted combination of points defined by each pair of corresponding lines. Based on the features, a mesh is put onto the image to be warped. Then, the warped image can be obtained by distorting the mesh according to the new feature points. The application generated 2 sets of images are shown below in Fig. 6. Each of the images is generated with a ten year gap and the original image is age of 40 years. First image series is shown male aging from 40 – 70 and the second image series shown female aging from 40 - 70.



**Fig. 6.** Male and female face changes with aging

To elaborate the methodology further more, following image sets in Fig.7, Fig.8 and Fig.9 are provided with graphical explanation.

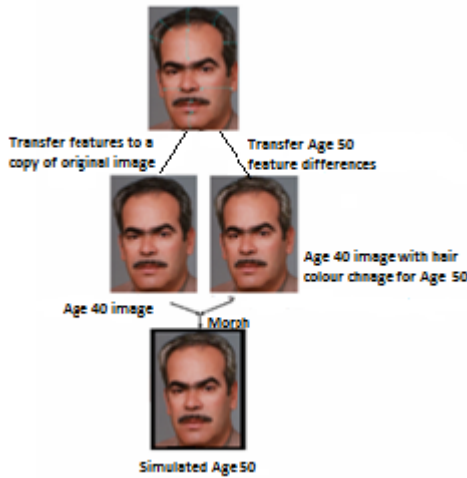


Fig. 7. Simulation of age 50 image

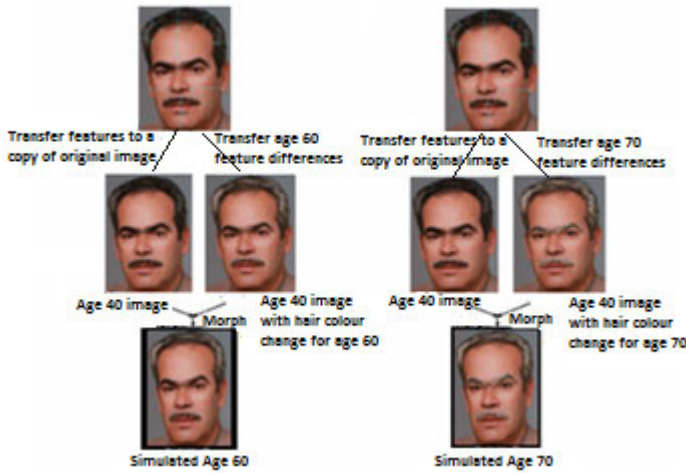


Fig. 8. Simulation of age 60 image



Fig. 9. Simulation of age 70 image

### 4 Testing and Evaluation

In order to test this application, a set of images for a given facial image is generated for age 50, 60 and 70. Then comparison is carried out with actual images. Test data set had images of 20 people. Each individual in the dataset has got many photographs of various age levels. There are about 120 images in the dataset. Finally the simulated images are compared with the original images to see the accuracy of the application. It can be observed that the accuracy of the application would be in a level of 60% - 65% by the comparison carried out between actual age image and the simulated image

as shown in Fig. 10. Application used by 20 users and gave their opinion as per in figure 11. Some users agreed that the simulated faces have similar facial features with compared to the original face images. But some users disagreed with the simulated faces. But none of the users strongly agreed or strongly disagreed with simulated faces. Simulated faces have remained the quality of the image and did not destroy the primary appearance of the image. That’s a key advantage of image morphing techniques.



Fig. 10. Accuracy of the application



Fig. 11. Application user feedback

Fig. 12 shows the original and the simulated images for age 50, 60 and 70.

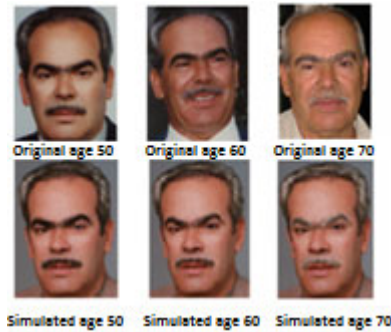


Fig. 12. Original images and simulated images for age 50, 60 and 70

## 5 Conclusion

The many interesting studies on facial aging have provided a good understanding of the versatile problem and have highlighted the many challenges associated with the problem. Main challenges in face recognition are illumination variations, pose variations, facial expressions etc. and the main challenge in temporal nature of the facial aging problem is creating a formal dataset for the problem. The success of the research is totally based on how thoroughly the facial aging datasets represent the problem with facial aging invariance. In other words, with multiple factors such as ethnicity, gender, age group etc. being identified as factors that affect facial aging effects [20], it is critical that face datasets assembled for this problem span the complete spectrum of the problem. In addition, the faces that comprise the aging dataset should ideally be devoid of variations due to other factors such as illumination, head pose, facial expressions, occlusions etc.

As we all know, the human faces are subject to change mainly due to age, gender and ethnic group. Among all the age is the main cause of the facial changes. When compared with these 3 reasons, the changes of the facial appearance due to the aging show some unique characteristics. Facial changes are prominent in man's infancy, childhood, youth hood and old age. During infancy and childhood, the facial growth happens due to their cranium. During youth hood and old age, only the facial features are changing, but the cranium does not change. However, after the age of 40, the wrinkles are appearing, hair colour changes, eye line colour changes, hairline recede and lips thin out happens on the faces. Hence after 40 years, it is not enough to consider only the changes of the facial features. In addition to that the wrinkle information, hair colour changes, eye line colour changes, hairline recede and lips thin out also have to be taken into consideration. This research is conducted to synthesize the facial images of people whose age is beyond 40 years. This methodology cannot be applied for the children as their faces grow according to the cranium.

The main goal of this research is to find out how the appearance of a face changes due to the aging in elderly age. So in order to accomplish this target, this research used an approach of calculating the mean distance differences of the identified ten facial features. Then synthesized facial images are matched with a given set of images to recognize the person's identity. By considering the rate of correct identifications of the generated image with the images of the database where they are stored to be matched with that image, and by considering the evaluation given by the above taken sample, it is proven that the simulated image would be at a satisfactory level.

## References

1. Ariyaratne, K.S., Dharmaratne, A.T.: Age Related Morphing Progression of Young Faces. In: Int'l Conf. in Machine Vision, Hong Kong (2010)
2. Jayasinghe, U., Dharmaratne, A.: Matching Facial Images Using Aging Related Morphing Changes. *World Academy of Science, Eng. & Technology* 60 (2009)
3. Kumari, L.L.G., Dharmaratne, A.: Image Processing Techniques in Human Aging Progression Benefited For Society & Welfare. In: Int'l Conf. Science Technology & Society, Department of Sociology, India, March 12-13 (2011)
4. Kumari, L.L.G., Dharmaratne, A.: A survey on Age Progression Techniques for Elderly People. In: 11th Int'l Conf. on Pattern Recognition & Information Processing (PRIP 2011), Minsk, Belarus (2011)
5. FG-NET Database, <http://www.fgnet.rsunit.com/>
6. Ricanek Jr., K., Tesafaye, T.: MORPH: A Longitudinal Image Database of Normal Adult Age-Progression. In: IEEE 7th Int'l Conf. on Automatic Face and Gesture Recognition, Southampton, UK, pp. 341-345 (April 2006)
7. Shan, Y., Liu, Z., Zhang, Z.: Image-Based Surface Detail Transfer. In: CVPR 2001, Hawaii, vol. II, pp. 794-799 (December 2001)
8. Gandhi, M.R.: A Method for Automatic Synthesis of Aged Human Facial Images. Dep. of Electrical & Computer Engineering McGill University (2004)
9. Gandhi, M.R., Levine, M.D.: A Method for Automatic Synthesis of Aged Human Facial Images. McGill University, Canada H3A 2A7
10. Vapnik, V.: *The Nature of Statistical Learning Theory*. Springer, New York (1995)

11. Tiddeman, B., Burt, M., Perrett, D.: Prototyping and transforming facial textures for perception research. *Computer Graphics and App's*. IEEE (2001)
12. Suo, J., Zhu, S.C., Shan, S., Chen, X.: A Compositional and Dynamic Model for Face Aging. *Journal of Latex Class Files* (January 2009)
13. McLachlan, J., Peel, D.: *Finite mixture models* (2000)
14. Mahalingam, G., Kambhamettu, C.: Age Invariant Face Recognition Using Graph Matching. IEEE (2010)
15. Ojala, T., Pietikainen, M., Maenpaa, T.: A generalized local binary pattern operator for multi-resolution gray scale and rotation invariant texture classification. In: *2nd Int'l Conf. on Advances in Pattern Recognition, Brazil* (2001)
16. Geng, X., Zhou, Z.-H., Kate, S.-M.: Automatic Age Estimation Based on Facial Aging Patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*
17. Geng, X., Zhou, Z.-H., Zhang, Y., Li, G., Dai, H.: Learning from facial aging patterns for automatic age estimation. In: *Proc. the ACM Int'l Conf., CA* (2006)
18. Duta, N., Jain, A.K., Dubuisson-Jolly, M.-P.: Automatic construction of 2D shape models. *IEEE Trans. Pattern Anal. Mach. Intel.* (2001)
19. Beier, T., Neely, S.: Feature-based image metamorphosis. *SIGGRAPH Computer. Graphics* 26(2), 35–42 (1992)
20. Patterson, E., Sethuram, A., Albert, M., Ricanek, K., King, M.: Aspects of age variation in facial morphology affecting biometrics. In: *IEEE Int'l Conf. on Biometrics: Theory, Applications and Systems, Crystal City* (2007)



# Off-line Signature Verification Based on Combination of Modified Direction and Microstructure Features

Danfeng Yang<sup>1</sup>, Yuzhu Qin<sup>2</sup>, Zhimin Huang<sup>3</sup>, and Yue Lu<sup>1</sup>

<sup>1</sup> Department of Computer Science and Technology

East China Normal University, Shanghai 200241, China

<sup>2</sup> School of Information and Communication Engineering

Beijing University of Posts and Telecommunications, Beijing 100088, China

<sup>3</sup> The Third Research Institute of the Ministry of Public Security  
Shanghai 200031, China

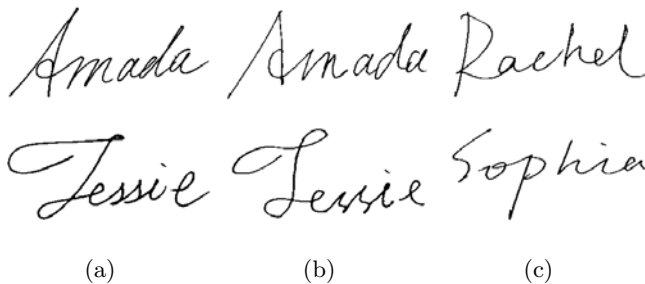
**Abstract.** Off-line signature verification is an important form of behavioral biometric identification. We present a method utilizing Modified Direction Feature(MDF) and Microstructure Feature(MSF) to tackle the problem. MDF and MSF belong to geometric structure features, but these two features are different from each other in each emphasis. In our study, global information in signatures' boundaries is represented by MDF, while local information is represented by MSF. In order to get features with lower dimensions, principal component analysis is employed to reduce redundant dimensions. In addition, we adopt support vector machine as classifier for verification process. The proposed strategy is evaluated on the GPDS and MCYT corpora. Experimental results have demonstrated that the proposed method is effective to improve off-line signature verification accuracy.

**Keywords:** Off-line Signature Verification, Modified Direction Feature, Microstructure Feature, Combination, Support Vector Machine.

## 1 Introduction

Automatic signature verification, as an important form of behavioral biometric identification, can be categorized into two kinds: One is on-line signature verification, and the other one is off-line signature verification [1,2]. Off-line signature verification is more difficult than on-line one. This is because on-line signature verification can get dynamic features compared to off-line signatures. In off-line signature verification systems, usually three kinds of forgeries are considered: random forgery, simple forgery, and skilled forgery [2]. Random forgeries are sampled from other signer's signature. Simple forgeries are written by other signers who just know the signer's real name. Skilled forgeries are written by other signers who have practiced writing with other signer's real signatures. Skilled forgeries are more difficult to distinguish from genuine signatures than simple and random forgeries. Because of limitation of signature database, usually two

kinds of forgeries may be considered: random forgery, skilled forgery(see Fig.1). Up to now, off-line skilled forgeries detection is still an open research problem for many researchers who make efforts to address it.



**Fig. 1.** Signature examples (a) genuine signatures, (b) skilled forgeries, and (c) random forgeries

The same as other pattern recognition problems, feature extraction takes an important place in off-line signature verification. Geometric structure features are widely used for off-line signature verification, because the method not only describes the signature's global information but also the local information. Many feature extraction methods have been proposed by researchers. Huang and Yan [3] used structural features for off-line signature verification. Wen et al. [4] proposed a method using the edge orientation distance histogram and directional gradient density features. Nguyen and Blumenstein [5] presented a method based on modified direction feature, in which modified direction feature recorded the location of the transition and the corresponding direction values. Nguyen et al. [6] used enhanced modified direction features for off-line signature verification.

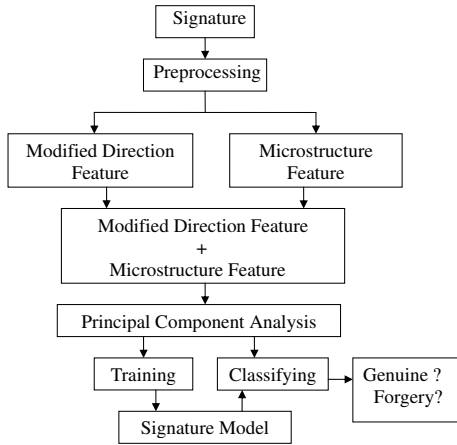
Fig. 2 illustrates the main steps of the approach presented in this paper. The rest of the paper is organized as follows. Section 2 presents the feature extraction. SVM is introduced in Section 3. Details about experiments and results of are introduced in Section 4. Section 5 gives the comparison with other methods. Finally, conclusions are discussed in Section 6.

## 2 Feature Extraction

Binarization of signature images is the first step of preprocessing. We use mathematical morphology method to obtain de-noising effects. Then we get the boundary of the binary image.

### 2.1 Modified Direction Feature

The MDF has been successfully used in character recognition [7]. Previous work [5,6] has shown that the MDF generated encouraging results in signature verification. MDF not only extracts direction information but also detects transitions between background and foreground pixels. The following steps depict the detail to obtain direction features and transition features.



**Fig. 2.** Main steps involved in building a signature verification system

**Labeling Foreground Pixels.** The first unlabeled foreground pixel in the left-bottom side of the image is defined as starting point of the signature, and the starting point is labeled with 8. The algorithm below is used to label other foreground pixels.

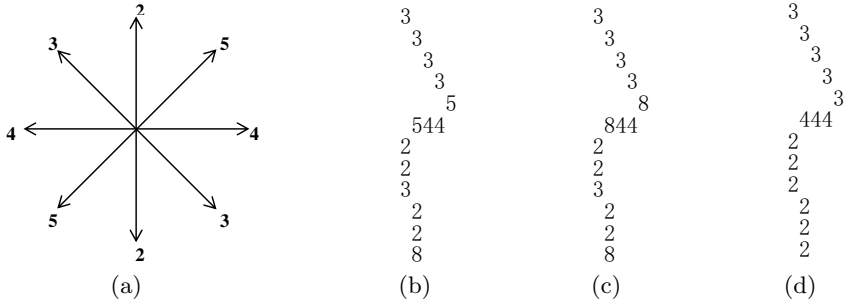
- (a) Take the starting point as  $p_1$ .
- (b) Find the unlabeled foreground pixel  $p_2$  around  $p_1$ . If  $p_2$  can not be found, the algorithm is terminated.
- (c)  $p_2$ 's direction value is determined by the direction of the vector  $\overrightarrow{p_1p_2}$ . Fig. 3(a) defines the direction values.
- (d) Take  $p_2$  as  $p_1$ , then go to (b).

Fig. 3(b) gives an example, in which foreground pixels are assigned direction values according to above process.

**Distinguishing Line Segments.** Each line segment's starting direction value is defined as 8. Fig. 3(c) shows this strategy. The rules used for distinguishing line segments are:

- (a) Direction values 3 and 5 can not be contained in the same line segment.
- (b) The kind of direction values in a line segment is less than 3.
- (c) The length of the previous direction type is greater than three pixels.

**Direction Value Normalization.** After the connected domain is split into several line segments, we use the most occurring value to replace the values of the other pixels in the current line segment. Fig. 3(d) presents the result of the final direction values.



**Fig. 3.** Dealing with direction values: (a) Direction values of MDF, (b) labeling with direction values, (c) distinguishing line segments, (d) direction value normalization

**Features Extraction.** The features are based on the calculation of transition from background to foreground pixels in the scanning direction.

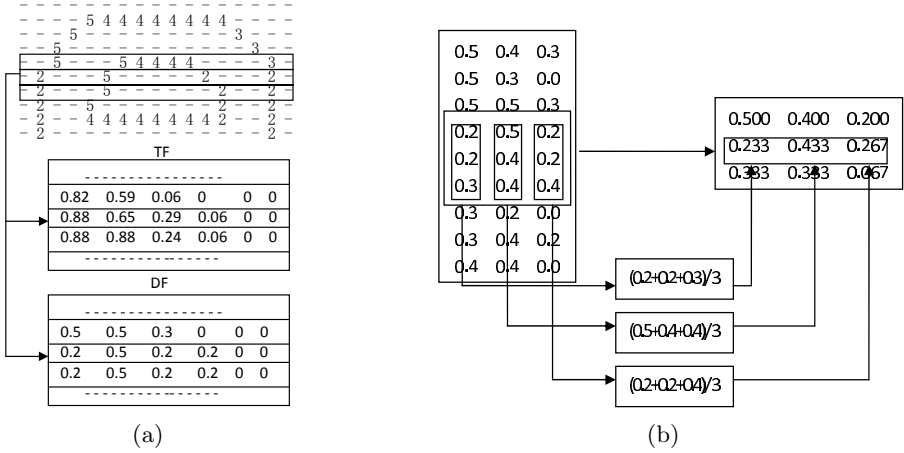
Two values ( direction value  $dv$  and transition value  $tv$  ) are recorded while transition occurs in the scanning direction.  $dv$  is calculated by dividing  $p$ 's direction value by 10.  $p$  stands for the pixel in the transition position. Pair values  $(x, y)$  mean that  $p$  is at the  $x$ th row and the  $y$ th column.  $w$  stands for the width of the image while  $h$  stands for the height of the image. If we scan by rows,  $pos$  is equal to  $y$  and  $len$  is equal to  $w$ . Otherwise,  $pos$  is equal to  $x$  and  $len$  is equal to  $h$ .  $tv$  is computed by:

$$tv = 1 - \frac{pos}{len} \tag{1}$$

Obviously, there are four ways to scan pixels in the image. Two ways are scanning each row from left to right and from right to left. The other two ways are scanning each column from top to bottom and from bottom to top. A maximum value  $MAXT$  is defined to be the largest number of transitions. If the number of transitions  $n$  is less than  $MAXT$ , the remaining  $MAXT - n$  transitions would be assigned values of 0. In our experiments, we set  $MAXT$  to be 10.

We will get two matrixes when scanning each direction. One records direction values while the other one records transition values. Finally, we get 8 matrixes, four matrixes record direction information while the other four matrixes record transition information. Four matrixes have  $w \times MAXT$  dimensions, and the other four matrixes have  $h \times MAXT$  dimensions. Fig. 4(a) illustrates this process.

**Size Normalization.** We get eight matrixes after extracting direction features and transition features. Since the size of each image is different, the matrix we get is different from each other. We re-sample these matrixes into  $n \times MAXT$  matrixes. In our experiments, we set  $n = 30$ . After extracting features from an image, we will get a feature vector in  $8 \times 30 \times 10$  dimensions. Fig. 4(b) gives an example for re-sampling a  $9 \times 3$  matrix to a  $3 \times 3$  matrix.



**Fig. 4.** (a) Example of extracting transition feature and direction feature in the left-to-right direction while  $MAXT = 6$ . (b) Re-sample a  $9 \times 3$  matrix to a  $3 \times 3$  matrix.

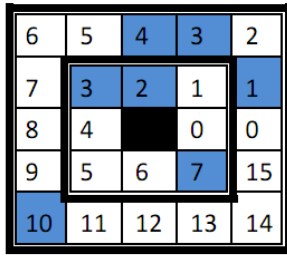
**2.2 Microstructure Feature**

The grid microstructure feature is extracted from the boundary of signature. These features recording the positions of some special contour pixel pairs in every local grid are used to calculate the appearance probability of different position pairs and express the writing style by the probability density distribution [8].

For a pixel  $p$  in an image, pair values  $(x, y)$  mean that  $p$  is location at the  $x$ th row and the  $y$ th column.  $C_d^p$  stands for the pixels whose chessboard distance to  $p$  is  $d$ . The chessboard distance of two pixels  $m, n$  is described as below.

$$D(m, n) = \max(\text{abs}(x_m - x_n), \text{abs}(y_m - y_n)) \tag{2}$$

For each foreground pixel  $p$  in an image's boundary, pixels in  $C_d^p$  are labeled in anticlockwise with values  $r$  where  $r = 0, 1, 2, 3, \dots, 8d - 1$ .  $C_d^p(r)$  stands for the  $r$ th pixel in  $C_d^p$ . For instance,  $C_1^p(2)$  means the second pixel in  $C_1^p$ .



**Fig. 5.** Example of extracting microstructure feature of a foreground pixel

MSF uses a matrix  $M_d^p$  to record the microstructure information. If two conditions are satisfied, we get  $M_d^p(i,j)=1$ . Otherwise,  $M_d^p(i,j)=0$ . The first condition is that  $C_d^p(i)$  and  $C_d^p(j)$  are foreground pixels, and the second condition is  $0 \leq i < j \leq 8d - 1$ . Take Fig. 5 as an example, the black cell means the foreground pixel which we want to extract microstructure information, and the blue cells mean the other foreground pixels. The distribution we get is:  $M_1^p(2,3) = 1$ ,  $M_1^p(2,7) = 1$ ,  $M_1^p(3,7) = 1$ ,  $M_2^p(1,3) = 1$ ,  $M_2^p(1,4) = 1$ ,  $M_2^p(1,10) = 1$ ,  $M_2^p(3,4) = 1$ ,  $M_2^p(3,10) = 1$ ,  $M_2^p(4,10) = 1$ . We do not record the pairs like  $M_1^p(3,2)$ , because they do not satisfy the second condition.

Then MSF get matrix  $M_d$  to record features extracted from all the foreground pixels. We just use  $M_d(i,j)$  when  $0 \leq i < j \leq 8d - 1$  and  $1 \leq d \leq MAXD$ . Therefore, the total microstructure feature of an image is  $\sum_{d=1}^{MAXD} C_{8d}^2$ . In our experiment, we set  $MAXD$  as 4. After we extract MSF from an image, we will get a feature vector in 920 dimensions.

### 2.3 Principal Components Analysis

Because of high dimensional vectors created by MDF (2400 dimensions) and MSF (920 dimensions), reducing the dimensions of feature vectors is an essential part of signature verification. Principal Components Analysis (PCA) [9] is a method that reduces data dimensionality by performing a covariance analysis between factors. Therefore, we use PCA to reduce redundant dimensions for MDF and MSF separately. The number of features extracted by MDF is reduced to 103 while the number of features extracted by MSF is reduced to 106. Finally, we combine MDF and MSF together as geometric structure features.

## 3 Support Vector Machine

The support vector machine algorithm was invented by Vladimir Vapnik [10]. It is a very popular machine learning method for classification, regression, and other learning tasks [11]. Given training instances  $x_i \in R^n, i \in \{1, 2, \dots, l\}$  and  $y_i \in \{-1, 1\}$  which means there are only two classes. SVM solves the primal optimization problem as

$$\begin{aligned} \min_{w,b,\varepsilon} \quad & \frac{1}{2}w^T w + C \sum_{i=1}^l \varepsilon_i \\ \text{subject to} \quad & y_i (w^T x_i + b) \geq 1 - \varepsilon_i \\ & \varepsilon_i \geq 0, i = 1, 2, \dots, l \end{aligned} \tag{3}$$

where  $C \geq 0$  is the regularization parameter. Due to the computational complexity of (3), it is a better choice to solve (3)'s dual problem as

$$\begin{aligned} \min_{\alpha} \quad & \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i y_i \alpha_j y_j K(x_i, x_j) - \sum_{i=1}^l \alpha_i \\ \text{subject to} \quad & \sum_{i=1}^l \alpha_i y_i = 0 \\ & 0 \leq \alpha_i \leq C, i = 1, 2, \dots, l \end{aligned} \tag{4}$$

where  $\alpha$  stands for dual parameters. The decision function is

$$f(x) = \text{sgn} \left( \sum_{i=1}^l \alpha_i y_i K(x_i, x) + b \right) \quad (5)$$

$K$  stands for kernel method which is widely used in pattern recognition problems. We choose linear kernel  $K(x_i, x_j) = \langle x_i, x_j \rangle$  in our experiments.

## 4 Experiments and Results

Experiments are conducted on the publicly available signature database GPDS [12] and MCYT [13] corpora respectively. For evaluation of the proposed method, we adopt ER (Error Rate), FRR (False Rejection Rate) and FAR (False Acceptance Rate) to evaluate the verification performance. Random forgeries and skilled forgeries are considered in our experiments, so there are two types of FAR. FAR1 stands for the random forgeries acceptance rate and FAR2 stands for the skilled forgeries acceptance rate. In our experiments, we employ support vector machine as our classifier.

### 4.1 Experiments on GPDS Corpus

GPDS-300 signature database consists of 300 individuals. Each contains 24 genuine signatures and 30 forgeries. Therefore,  $24 \times 300$  genuine signatures and  $30 \times 300$  forgeries in the signature database. The forgers simulate the genuine signatures static image, and they are allowed to practice writing the signatures as long as them wish.

We choose 300 persons' signatures for testing. Each person has 24 genuine signatures, 30 skilled forgeries, and random forgeries are taken from the remaining 299 persons' genuine signatures. Then we use 4-fold cross validation to get an average evaluation. We get the corpus's average ER, FRR, FAR1 and FAR2 by averaging all the persons' results.

**Table 1.** Comparison on GPDS corpus

Feature	FAR			
	ER(%)	FRR(%)	FAR1 (%)	FAR2(%)
MDF	0.439	26.764	0.039	12.391
MSF	0.706	34.167	0.238	<b>4.548</b>
MDF+MSF	<b>0.208</b>	<b>14.736</b>	<b>0.007</b>	5.722

From Table 1, we can find that the performance of the combination of MDF and MSF outperforms either MDF or MSF. ER, FRR and FAR1 are all the best. Although FAR2 of the combination method is greater than MSF, its ER of 0.208% is the smallest of all and FRR is reduced to 14.736%. This is a significant improvement on signature verification of FRR.

## 4.2 Experiments on MCYT Corpus

MCYT off-line signature database consists of 2250 signature images of 75 individuals. 15 genuine signatures and 15 skilled forgeries for each individual. The forgeries are provided with the signature images to be forged, they train to imitate the signature images several times, and then they are asked to practice the shape with natural dynamics such as without breaks or slowdowns.

We choose 75 persons' signatures in the test. Each person has 15 genuine signatures, 15 skilled forgeries, and random forgeries are taken from the remaining 74 persons' genuine signatures. Then we use 4-fold cross validation to get average evaluation parameters. We get corpus's average ER, FRR, FAR1 and FAR2 by averaging all the persons' results.

**Table 2.** Comparison on MCYT corpus

Feature	FAR			
	ER(%)	FRR(%)	FAR1 (%)	FAR2 (%)
MDF	1.016	19.694	0.013	6.889
MSF	3.243	37.306	0.780	8.583
MDF+MSF	<b>0.799</b>	<b>13.611</b>	<b>0.007</b>	<b>4.722</b>

We can see, from Table 2, the performance of the combination of MDF and MSF is the best. Its ER of 0.799% is the smallest of all while its FRR is reduced to 13.611%.

It is obvious that the geometric structure features combined by microstructure features and modified direction features can get a significant improvement on signature verification's FRR.

## 5 Comparison with Other Methods

We do two comparisons with the latest published work based on MDF and SVM.

### 5.1 Comparison with [5]

We compare the proposed method with [5], which was based on global features and SVM.

In [5], Nguyen et al. adopted 160 persons' signature data sets in the GPDS database. In each training, we also use 12 genuine samples and 400 random forgeries just as they did. The random forgeries were chosen from 100 randomly selected writers, four genuine signatures from each writer. The remaining 12 genuine signatures, 59 random and 15 skilled forgery signatures were used in the testing phases. This experiment is evaluated by AER, which means the average of FAR and FRR2. Table 3 shows the comparison. Our method has achieved the average of 16.15%, which is comparable to the best result 17.25% reported by [5].



**Table 3.** Performance comparison of different features

Method	Main features	Classifier	Result (AER)
Nguyen et al. [5]	Global feature	SVM	17.25%
Our method	MDF+MSF	SVM	16.15%

## 5.2 Comparison with [6]

We also do some comparison with Nguyen et al.'s study in [6]. They proposed a method using enhanced modified direction features in conjunction with neural classifiers and support vector machines. In their experiments, they also adopted 160 persons' signature data sets in the GPDS database. 20 genuine samples, 25 skilled and 400 random forgeries were used for training. 4 genuine samples, 5 skilled forgery samples and 59 random samples were used for testing.

Using the same setting, the comparisons are shown in Table 4. They use Distinguishing Error Rate(DER) which is the average of FRR and FAR2 to evaluate their performance. And their best result in the experiments with SVM was 10.73%. Our method achieved 9.44%. When we used 300 sets, the result obtained can be as low as 8.90%.

**Table 4.** Performance comparison of different features

Method	Main features	Classifier	Result (DER)
Nguyen et al. [6]	Enhanced MDF	SVM	10.73%
Our method	MDF+MSF	SVM	9.44%

## 6 Conclusion

In this paper, we proposed a new method by extracting geometric structure features composed of MDF and MSF. The combination of these two features has a great ability to recognize forgeries and tolerate intrapersonal variances. After features are extracted from each signature's boundary, we use PCA to reduce redundant dimensions, and build our classification model on support vector machine. Experiments on GPDS and MCYT corpora demonstrated that the combination of MDF and MSF is effective to improve signature verification's performance.

**Acknowledgements.** This work is partially supported by the Science and Technology Commission of Shanghai Municipality under research grant 09510701800.

## References

1. Impedovo, D., Pirlo, G.: Automatic signature verification: the state of the art. *IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews* 38(5), 609–635 (2008)

2. Wen, J., Fang, B., Tang, Y., Zhang, T.P.: Model-based Signature Verification with Rotation Invariant Features. *Pattern Recognition* 42(7), 1458–1466 (2009)
3. Huang, K., Yan, H.: Off-line Signature Verification Using Structural Feature Correspondence. *Pattern Recognition* 35(11), 2467–2477 (2002)
4. Wen, J., Fang, B., Tang, Y.Y., Wang, P.S.P., Cheng, M., Zhang, T.P.: Combining EODH and directional gradient density for offline signature verification. *International Journal of Pattern Recognition and Artificial Intelligence*, 1161–1177 (2009)
5. Nguyen, V., Blumenstein, M., Leedham, G.: Global Features for the Off-Line Signature Verification Problem. In: 10th International Conference on Document Analysis and Recognition, pp. 1300–1304 (2009)
6. Nguyen, V., Blumenstein, M., Muthukkumarasamy, V., Leedham, G.: Off-line Signature Verification Using Enhanced Modified Direction Features in Conjunction with Neural Classifiers and Support Vector Machines. In: 9th International Conference on Document Analysis and Recognition., vol. 2, pp. 734–738 (2007)
7. Blumenstein, M., Liu, X.Y., Verma, B.: An Investigation of the Modified Direction Feature for Cursive Character Recognition. *Pattern Recognition* 40(2), 376–388 (2007)
8. Li, X., Ding, X.Q.: Writer identification of chinese handwriting using grid microstructure feature. In: International Conference on Advances in Biometrics, pp. 1230–1239 (2009)
9. Esbensen, K., Geladi, P., Wold, S.: Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems* 2, 37–52 (1987)
10. Vapnik, V.: *Statistical Learning Theory* (1998)
11. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines (2001)
12. Vargas, J.F., Ferrer, M.A., Travieso, C.M., Alonso, J.B.: Off-line Handwritten Signature GPDS-960 Corpus. In: International Conference on Document Analysis and Recognition, vol. 2, pp. 764–768. IEEE (2007)
13. Ortega, G.J., Fierrez, A.J., Simon, D., Gonzalez, J., Faundez, Z.M., Espinosa, V., Satue, A., Hernaez, I., Igarza, J.J., Vivaracho, C., et al.: MCYT Baseline Corpus: a Bimodal Biometric Database. In: *Vision, Image and Signal Processing, IET*, vol. 150, pp. 395–401 (2004)

# Heart Sound Feature Reduction Approach for Improving the Heart Valve Diseases Identification

Mostafa A. Salama<sup>1</sup>, Aboul Ella Hassanien<sup>2</sup>, Aly A. Fahmy<sup>2</sup>,  
and Tai-hoon Kim<sup>3</sup>

<sup>1</sup> Department of Computer Science, British University in Egypt, Cairo, Egypt  
mostafa.salama@gmail.com

<sup>2</sup> Cairo University, Faculty of Computers and Information, Cairo, Egypt  
{aboitcairo,aly.fahmy}@gmail.com

<sup>3</sup> Hannam University, Korea  
taihoonn@hannam.ac.kr

**Abstract.** Recently, heart sound signals have been used in the detection of the heart valve status and the identification of the heart valve disease. Due to these characteristics, therefore, two feature reduction techniques have been proposed prior applying data classifications in this paper. The first technique is the chi-Square which measures the lack of independence between each heart sound feature and the target class, while the second technique is the deep believe network that uses to generate a new data set of a reduced number of features according the partition of the heart signals. The importance of feature reduction prior applying data classification is not only to improve the classification accuracy and to enhance the training and testing performance, but also it is important to detect which of the stages of heart sound is important for the detection of sick people among normal set of people, and which period important for the classification of heart murmur. Different classification algorithms including naive bayesian tree classifier and sequential minimal optimization was applied on three different data sets of 100 extracted features of the heart sound. The extensive experimental results on the heat sound signals data set demonstrate that the proposed approach outperforms other classifiers and providing the highest classification accuracy with minimized number of features.

**Keywords:** Hear sound, heart valve diseases, feature selection.

## 1 Introduction

The diagnosis of diseases like heart valve diseases using data mining tools is an essential requirement in daily life. Most of heart valve diseases have an effect on the heart sound of patients [1]. Classification can be applied to detect whether the patient's heart sound signal is patient or not, and also can detect the type of the heart disease in sick patients [2]. Such an approach could be

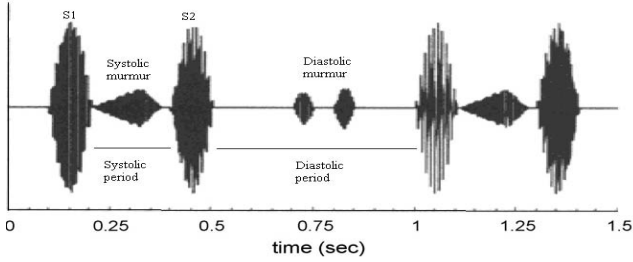
useful in the diagnosis of heart disease remotely, just by sending a record of the heart signals to a medical back end system that replies automatically by the problem in heart. Also it is considered as a low cost approach rather than the high cost medical examinations. Also a computerized system could provide physicians with suggestions about the diagnostic suggestions about the diseases. Due to the sensitivity of heart diagnosis results, a high classification accuracy and performance are required with the least error percentage. After extracting features from heart sound signals, preprocessing is applied on these features [3]. The most important preprocessing step is the feature reduction of the input data set. The data set contain features that are considered as noisy or irrelevant features, these features could have a negative impact on the classification accuracy of the instances, patients. Feature reduction methods are either feature extraction or feature selection method. Feature extraction method apply operation on the original features and extract a lower number of features that carries the same characteristics. Feature selection methods rank and select the most important features, where if only a subset of features with the highest rank are used in classification, a high classification accuracy could be achieved.

The extracted heart sound data are three different data sets, each of 100 features where they are slitted into six different parts. the first data set is required to classify whether the heart of the patients are normal or not. The second and third data set is required for the detection of the heart valve disease. The heart valve diseases under investigation in this paper are aortic stenosis *AS*, aortic regurgitation *AR*, mitral stenosis *MS* and mitral regurgitation *MR*. This disease classification is performed in two steps where the first step is applied on the second data set for determining the type of the systolic murmur which means *AS* or *MR*, and the second step is applied on the third data set of a diastolic murmur diseases which means *AR* or *MS*. The second importance of feature selection method is to determine which stage of the heart sound could have the greatest indication to heart valve disease in the case of each murmur type. The four stages of a heart sound are the first heart signal *S1*, the systolic period, the second heart signal and the diastolic period [4].

The rest of this paper is organized as follows: Section 2 gives a brief introduction of the heart signals data collection and its characteristics. Section 3 shows an overview of the two feature selection and reduction methods. The proposed feature reduction and classification approaches are given in section 4. The experimental results and conclusions are presented in Section 5 and 6 respectively.

## 2 The Heart Sound Signals Data Set

A lot of research have been applied on heart sound for the detection of heart valve disease. Features are extracted from the heart sound signal into a data set that is composed of 100 features. Then, a classification algorithm is applied on such data set for detection of heart valve disease. Features are extracted in three phases, segmentation [5] [6], transformation and extraction.



**Fig. 1.** Heart signal: systolic period and diastolic period [4]

These extracted feature represent the four stages of a heart signal which are  $S1$  signal, systolic period,  $S2$  signal and diastolic period as shown in figure 1. These features are divided into six groups as follows:

- F1:F4 are the standard deviation of all heart sounds,  $S1$ ,  $S2$  and average heart rate.
- F5:F12 represents signal  $S1$ .
- F13:F36 represents the systolic period.
- F37:F44 represents signal  $S2$ .
- F45:F92 represents the diastolic period.
- F93:F100 the four stage of a heart signal are passed from four band-pass frequency filters. The energy of each output is calculated to form these last 8 features.

### 3 The ChiMerge and Deep Believe Networks: A Brief Introduction

Nearly most of the feature selection method discretize the continuous feature which could leads to the distortion of data and loose of its characteristics and hence the decrease feature classification. In this paper we introduced two reducation approaches, the first is ChiMerge approach which determines the Chi-Square value while perform the discrimination of features which leads to more accurate results. The second method is a deep belief network [7], [8]. The deep belief network is applied in a novel method in order to generate a new data set of a reduced number of features according the partition of the heart signal discussed in the previous section. This section provides a brief introduction of the ChiMerge and the deep belief network approaches.

#### 3.1 ChiMerge Feature Evaluation

One of the most popular feature selection methods is the chi-Square  $\chi^2$  method, which measures the lack of independence between each feature  $A$  and the target class  $c$ . ChiMerge or Chi2-Square [9] is a  $\chi^2$ -based discretization method. It uses the  $\chi^2$  statistic to discretize numeric features repeatedly until some inconsistencies are found in the data that achieves feature selection via discretization.

The  $\chi^2$  value is calculated for each continuous feature as follows: Initially, each distinct value of a numeric feature  $A$  is considered to be one interval. The values of feature  $A$  are sorted and the  $\chi^2$  is applied for every pair of adjacent intervals as follows:

$$\chi^2 = \sum_{i:1..2} \sum_{j:1..k} \frac{(A_{ij} - E_{ij})^2}{E_{ij}} \quad (1)$$

Where:

- $A_{ij}$  is the number of values in the  $i$ th interval and  $j$ th class,
- $R_{ij}$  is the number of values in the  $j$ th class =  $\sum_{j:1..k} A_{ij}$ ,
- $C_{ij}$  is the number of values in the  $i$ th interval =  $\sum_{i:1..2} A_{ij}$ ,
- $N$  is the total number of values =  $\sum_{i:1..2} R_{ij}$ ,
- $E_{ij}$  is the expected frequency of  $A_{ij} = \frac{R_{ij} * C_{ij}}{N}$

Adjacent intervals with the least  $\chi^2$  values are merged together, because low  $\chi^2$  values for a pair indicates similar class distributions. This merging process proceeds recursively until all  $\chi^2$  values of all pairs exceeds a parameter *signlevel* (initially 0.5). Then repeat the previous steps with a decreasing *signlevel* until an inconsistency rate is exceeded, where two patterns are the same but classified into different categories.

### 3.2 Deep Believe Networks

Deep Belief Networks (DBNs) as shown in figure 2 is a deep architecture that consists of a stack of Restricted Boltzmann Machines (RBM). RBM is an energy-based undirected generative model that uses a layer of hidden variables to model a distribution over visible variables [10]. The undirected model for the interactions between the hidden and visible variables is used to ensure that the contribution of the likelihood term to the posterior over the hidden variables is approximately factorial which greatly facilitates inference. In DBN architecture the hidden activity vectors of the first RBM layer will behave as the training data (visible layer) for next RBM layer. The deep architecture has the benefit that each layer learns more complex features than layers before it. Each layer in DBN (RBM) depends on Contrastive Divergence method for input reconstruction which increases the performance of the network. In case of not using class labels and back-propagation in the DBN Architecture (unsupervised training), DBN could be used as a feature extraction method for dimensionality reduction.

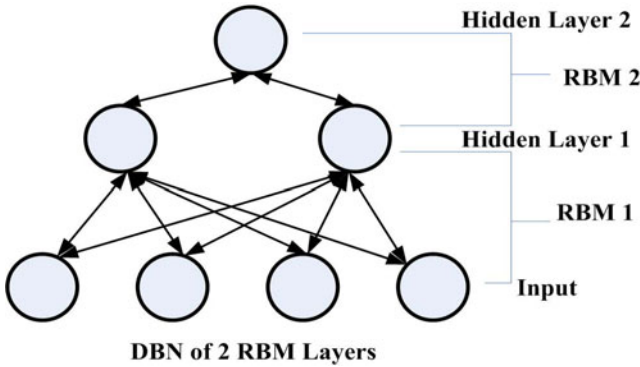


Fig. 2. Deep belief networks architecture

## 4 Heart Sound Feature Selection and Classification

The main objectives of the proposed approaches are (1) to increase the performance of classification accuracy and (2) determine which of the heart signal stages is an indicator to the presence of a heart disease. The first approach were used a ChiMerge feature selection method that satisfy the nature of the input data set and find out the features that leads to the best classification performance. Then, it detects which of the heart signal stage that leads to the best classification accuracy according to these selected features. The second approach applies the DBN on each group of features from the six groups discussed above. Then, it applies the feature classification technique on the output of one each DBN to find which stage can be as the best indicator for heart disease, then on the combination all outputs. So, the first approach selects the most relevant features then determine which stage of the signal is the best identifier while the second approach check each stage directly. This section gives a details description of the two proposed feature selection approaches, and see their impact on the classification accuracy using many classifiers.

### 4.1 Feature Selection Using ChiMerge

The chiMerge Feature selection method calculates the  $\chi^2$  value of each feature [7]. Features are sorted according to the corresponding to the  $\chi^2$  value. A test is applied on a data set containing the features of the highest  $\chi^2$  value, then applied incrementally on data sets of two features until the data set contains  $n$  features, all features of the original data set. The classification accuracy increases as the number of selected features increases until a certain number of features where a specific peak of accuracy then the accuracy [11], also it must be stated that the chart may contain many local extrema besides the global maximum value. The selected features will be those who produces the max classification accuracy. This approach is declared in the following detailed algorithm [1].

---

**Algorithm 1.** Feature selection using ChiMerge

---

```

1: Input:Dataset of 100 features;
2: Extract the rank of each feature using ChiMerge method
3: Sort the features according to their ranks in a descending order
4: maxAccuracy = 0 : maximum accuracy reached;
5: noFeatures = 0: number of features lead to maximum accuracy;
6: for iterator i = 1 to 100 do
7:   Data set contains the first i features in the sorted list
8:   Apply the classifier
9:   State the resulted accuracy of classification
10:  if resulted accuracy > maxAccuracy then
11:    noFeatures = i
12:    maxAccuracy = resulted accuracy
13:  end if
14: end for
15: Return the maximum classification accuracy maxAccuracy
    and the corresponding number of attributes noFeatures

```

---



---

**Algorithm 2.** Feature extraction using Deep Belief network

---

```

1: Input:Dataset of 100 features;
2: Divide the 100 features into six groups of features
3: for each group of features do
4:   Construct an RBM of input nodes equal to number of feature of the group and
     one output node
5:   Train the RBM
6:   Apply the classifier on output from the RBM
7:   Save the RBM weights that shows the best classification accuracy
8: end for
9: The output from the six RBM networks are collected in a single data set
10: Apply the classifier on the resulted data set
11: Return the classification accuracy

```

---

## 4.2 Feature Selection Using Deep Belief Network

In this approach, each group of features of the six heart sound signal passed through a deep belief network. The DBN consists layers of RBMs, such that the first visible layer contains the number of features in the group while the next layer contains half the number of features, and finally the last layer contains only one feature. The generated output from the DBN is passed to a classifier to get the classification accuracy. Then the output from DBN applied on the first group is combined to those from the other groups to form a data set of the reduced number of features which is six features. Finally apply the classifier on this reduced data set. The detailed steps of the approach is described in algorithm 2.

One of the problems that may face this approach is that the weights in the DBN are randomly initialized, and some weights could lead to a decrease the



classification accuracy. In order to overcome this problem, the training of each DBN is applied several times and for each time the classification accuracy is calculated, then the weights that produces the best classification accuracy will be used. This will allow that the over all generated data set will have a better classification results and that the resulted analysis doesn't depend on the randomly initiated weights. Algorithm 2 shows the main steps of the feature extraction using DBN.

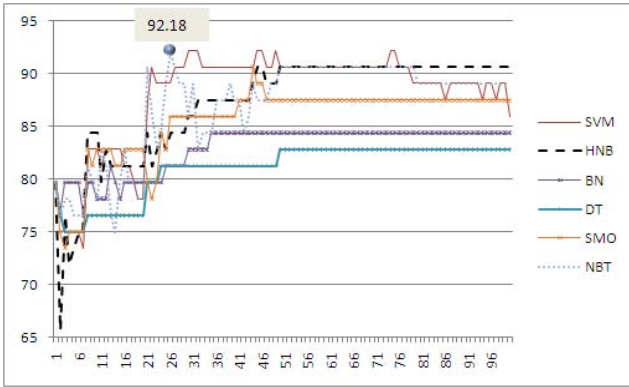
## 5 Experimental Results and Discussions

Three data sets with equal number of instances in every class will be examined by the previously discussed two approaches. The results will be compared to the case when no feature reduction is applied and an analysis on the results will be discussed to show the importance of feature selection in heart sound identification. The first data set “*HS\_N\_S*” contains 64 instance, where 32 instances represent healthy patients and the other 32 represents unhealthy, murmur diseased patients. The second data set “*HS\_AS\_MR*” is about systolic diseases where it contains 37 instances of aortic stenosis *AS* cases and 37 instances of mitral regurgitation *MR* cases. The third data set “*HS\_AR\_MS*” is about diastolic diseases where it contains 37 instances of aortic regurgitation *AR* cases and 37 instances of mitral stenosis *MS* cases. Seven well known classifiers was used in all experiments for comparative analysis including decision trees (DT), (SVM) support vector machine, (NBT) naive bayesian tree, (SMO) sequential minimal optimization, and (HNB) hidden navie bayesian.

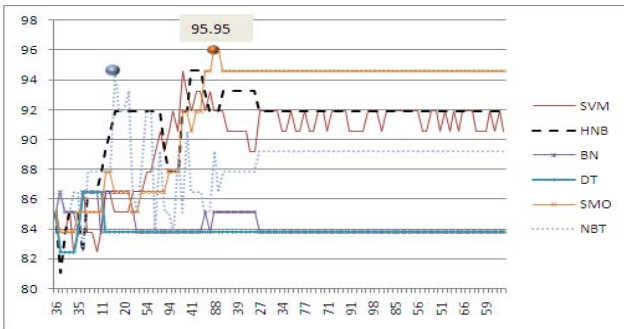
### 5.1 Results of the ChiMerge Approach

Using the ChiMerge approach, seven classifiers have been applied for comparison and the best results appears in the case of using naive bayesian tree classifier *NBT* and Sequential minimal optimization *SMO* for least number of features. The validation procedure developed is stratified 10-fold cross-validation. The data is divided randomly into 10 parts, in which each class is represented in approximately the same proportions as in the full data set. In the case of Normal-Sick data set, the best classification accuracy appears when ‘26’ features are selected. The features used are “54, 58, 53, 65, 64, 67, 59, 97, 61, 70, 96, 98, 55, 60, 66, 51, 52, 49, 62, 63, 23, 22, 45, 50, 2, 56”. The classifiers that shows the best classification accuracy is *SMO* where it achieves 92.18% accuracy as shown in figure 3.

In the case of the systolic “*HS\_AS\_MR*” data set, the best classification accuracy appears when ‘14’ features are selected. The features used are “36, 13, 12, 14, 15, 35, 4, 2, 18, 17, 11, 3, 92, 5”. The classifiers that shows the best classification accuracy is *SMO* where it achieves 95.94% accuracy as shown in Figure 4.



**Fig. 3.** Classifier accuracy (normal-sick data set): Decision trees (DT), (SVM) support vector machine, (NBT) naive bayesian tree, (SMO) sequential minimal optimization and (HNB) hidden naive bayesian

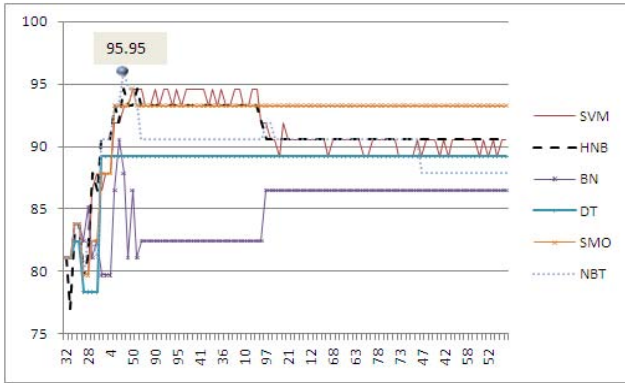


**Fig. 4.** Classifier accuracy (systolic data set): Decision trees (DT), (SVM) support vector machine, (NBT) naive bayesian tree, (SMO) sequential minimal optimization, and (HNB) hidden naive bayesian

In the case of the diastolic “*HS\_AR\_MS*” data set, the best classification accuracy appears when ‘14’ features are selected. The features used are “32, 31, 30, 29, 100, 28, 33, 27, 3, 5, 4, 49, 48, 45”. The classifiers that shows the best classification accuracy is *NBT* where it achieves 94.59% accuracy as shown in Figure 5. And when the *SMO* classifier is used the accuracy achieves 95.94% when 36 attributes are used.

### 5.2 Results of Deep Believe Networks Approach

The second approach have used the DBN in feature reduction of every group of features into a single feautre, then apply the classification on the generated 6



**Fig. 5.** Classifier accuracy (diastolic data set): Decision trees (DT), (SVM) support vector machine, (NBT) naive bayesian tree, (SMO) sequential minimal optimization, and (HNB) hidden navie bayesian

features. As shown in table I, for NBT classifier, the accuracy of classification achieved for the *HS\_AS\_MR* data set 93.24% and highest classification accuracy among the four parts that described the heart sound signal is the second part, the systolic stage.

**Table 1.** Classification results of the *HS\_AS\_MR*

Partition	NBT	RBM-NBT
Part 1	81.08	81.08
Part 2	79.72	45.94
Part 3	78.37	83.78
Part 4	75.67	79.72
Part 5	50.00	45.94
Part 6	85.13	85.13
All	87.83	93.24

### 5.3 General Analysis

It is clear from the results the importance of feature reduction in enhancing the classification accuracy and performance. But on the other hand, its needed to determine the stages of the heart signal that is most important in determining both the existence of the disease and the type of the disease. The results in the first approach shows (1) A high percentage from selected features from the systolic and diastolic data sets contains features from part 3 that represent the systolic period. Also the features of the highest ranks are from this period only, (2) in the case of the normal-sick data set, a high percentage of the selected features are from the the diastolic period.

## 6 Conclusions

The selection of the relevant features to the classifier and remove the irrelevant and misleading features has a great effect in the enhancement of the classification accuracy. This paper proposed two feature reduction techniques prior applying data classifications. The first technique is the chi-Square which measures the lack of independence between each heart sound feature and the target class, while the second technique is the deep believe network that uses to generate a new data set of a reduced number of features according the partition of the heart signals. Different classification algorithms including naive bayesian tree classifier and sequential minimal optimization was applied on three different data sets of 100 extracted features of the heart sound. The extensive experimental results on the heat sound signals data set demonstrate that the proposed approach outperforms other classifiers and providing the highest classification accuracy with minimized number of features. Moreover, we conclude that the diastolic period is a good identifier for the existence of heart disease, while the systolic period is a good identifier for distinguishing between aortic stenosis and mitral regurgitation, and between aortic regurgitation and mitral stenosis.

**Acknowledgment.** We want to thank Ilias Maglogiannisa from University of Central Greece, Department of Computer Science and Biomedical Informatics, Lamia, Greece, for supporting us and providing with heart sound data set.

## References

1. Chen, T., Kuan, K., Celi, L., Clifford, G.: Intelligent heart sound diagnostics on a cellphone using a hands-free kit. In: Proceedings of AAAI Artificial Intelligence for Development (AID 2010). Stanford University, California (2010)
2. Hebden, J.E., Torry, J.N.: Neural network and conventional classifiers to distinguish between first and second heart sounds. *Artificial Intelligence Methods for Biomedical Data Processing IEE Colloquium (Digest) 3*, 1–6 (1996)
3. Kumar, D., Carvalho, P., Antunes, M., Paiva, R.P., Henriques, J.: Heart murmur classification with feature selection. In: Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, vol. 1, pp. 4566–4569. Buenos Aires, Argentina (2010)
4. Maglogiannis, I., Loukis, E., Zafiropoulos, E., Stasis, A.: Support vectors machine-based identification of heart valve diseases using heart sounds. *Journal of Computer Methods and Programs in Biomedicine* 95, 47–61 (2009)
5. Liang, H., Lukkarinen, S., Hartimo, I.: Heart sound segmentation algorithm based on heart sound envelopgram. *Computers in Cardiology*, 105–108 (1997)
6. Liang, H., Lukkarinen, S., Hartimo, I.: A heart sound segmentation algorithm using wavelet decomposition and reconstruction. In: Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, USA, vol. 4, pp. 1630–1633 (1997)
7. Salama, M.A., Hassanien, A.-E., Fahmy, A.: Deep belief network for clustering and classification of a continuous data. In: IEEE International Symposium on Signal Processing and Information Technology, Luxor, Egypt, pp. 473–477 (2010)

8. Platt, J.C.: Sequential minimal optimization: A fast algorithm for training support vector machines. *Advances in Kernel Methods - Support Vector Learning* (1998)
9. Liu, H., Setiono, R.: Chi2: feature selection and discretization of numeric attributes. In: *Proceedings of the Seventh International Conference on Tools with Artificial Intelligence*, Virginia, USA, November 8, p. 388 (1995)
10. Mohamed, A.R., Dahl, G., Hinton, G.E.: Deep belief networks for phone recognition. In: *NIPS 22 Workshop on Deep Learning for Speech Recognition* (2009)
11. Janecek, A.G.K., Gansterer, W.N., Demel, M., Ecker, G.F.: On the relationship between feature selection and classification accuracy. In: *JMLR: Workshop and Conference Proceedings*, Antwerp, Belgium, vol. 4, pp. 90–105 (2008)

# Comparison of Different Ontology-Based Query Expansion Algorithms for Effective Image Retrieval

C.H.C. Leung and Yuanxi Li

Department of Computer Science  
Hong Kong Baptist University  
Hong Kong, China

{clement,yxli}@comp.hkbu.edu.hk

**Abstract.** We study several semantic concept-based query expansion and re-ranking scheme and compare different ontology-based expansion methods in image search and retrieval. In particular, we exploit the two concept similarities of different concept expansion ontology-WordNet Similarity, Wikipedia Similarity. Furthermore, we compare the keywords semantic distance with the precision of image search results with query expansion according to different concept expansion algorithms. We also compare the image retrieval precision of searching with the expanded query and original plain query. Preliminary experiments have been able to demonstrate that the two proposed retrieval mechanism has the potential to outperform unaided approaches.

**Keywords:** Image retrieval, query expansion, ontology, concept distance.

## 1 Introduction

The presence of particular objects in an image often implies the presence of other objects. If term  $U \rightarrow V$ , and if only  $U$  is indexed, then searching for  $V$  will not return the image in the result, even though  $V$  is present in the image. The application of such inferences will allow the index elements  $T_i$  of an image to be automatically expanded according to some probability which will be related to the underlying ontology of the application. There are two types of expansion:

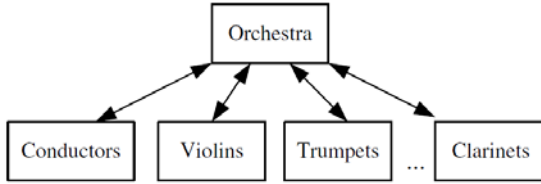
(a) Aggregation hierarchical expansion

This relates to the aggregation hierarchy of sub-objects that constitute an object.

The objects can be classified as:

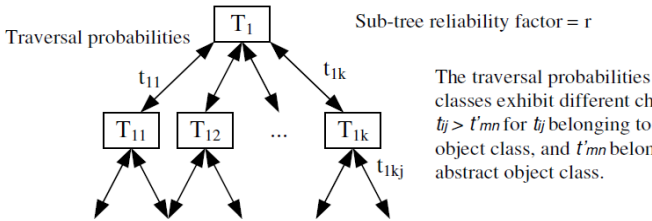
(i) Concrete, where the relevant objects are well-defined (e.g. an orchestra expanded to conductors, violins, trumpets, clarinets etc, and see Fig. 1)

(ii) Abstract, where the objects are not concretely defined (e.g. although ‘conflict’ is not a definite visual object, it does contain certain common characteristics).



**Fig. 1.** The concrete aggregation hierarchical expansion, where the relevant objects are well-defined. In this example, an orchestra is expanded to conductors, violins, trumpets, clarinets etc.

Associated with each branch is a tree traversal probability  $t_{ij}$  (Fig. 2) which signifies the probability of occurrence of the branch index given the existence of the parent index. In general, the traversal probabilities of different object classes exhibit different characteristics, with  $t_{ij} > t'_{mn}$  for  $t_{ij}$  belonging to the concrete object class, and  $t'_{mn}$  belonging to the abstract object class.



**Fig. 2.** A tree traversal probability  $t_{ij}$  which signifies the probability of occurrence of the branch index given the existence of the parent index

(b) Co-occurrence expansion

This relates to the expectation that certain semantic objects tend to occur together. The relevant weighting is expressed as a conditional probability given the presence of other objects. An expansion to associate an image object  $O_j$  given the presence of object  $O_i$  is taken to be indexable when

$$Prob [O_j | O_i] \geq h' \tag{1}$$

where  $h'$  is a preset threshold value that depends on the tradeoff between precision and recall performance of the system. More generally, complex probabilistic rules taking the form

$$Prob [O_j | O_1, \dots, O_n] \geq h' \tag{2}$$

will be applied. The ontology expansion tree is traversed bi-directionally in the course of the expansion. Top-down traversal will lead to an expansion factor  $> 1$ , while bottom-up traversal will have an expansion factor  $< 1$  at each level of expansion.

There are, in general, many sub-trees whose roots are the nodes of the ontology expansion tree. Each sub-tree has been fully expanded, and it has an expansion reliability factor  $0 < r < 1$ , which signifies the dependability and completeness of the

associated expansion. For high precision retrieval ( $\pi \approx 1$ ), only sub-trees, which are having a significant reliability factor, need to be traversed; and nodes with a small value for  $r$  will be pruned. Decision rules linking expandability with  $\pi$  and  $r$  can be determined. [1]

In this paper, we mainly focus on WordNet Similarity and Wikipedia Similarity. Details of the two will be discussed in this paper later.

This paper is organized as follows. Related work will be introduced in section two; our ontology-based query expansion similarity measures will be presented in detail in section three; then the experiments, as well as results will be reported in section four, followed by a conclusion in the last section.

## 2 Related Work

### 2.1 Semantic Knowledge from Ontology Approach

As for redefining image indexing, the most popular way is to simplify the semantic knowledge into the semantic similarity between concepts. Semantic similarity [2] or semantic relatedness is a concept whereby a set of documents or terms within term lists are assigned a metric based on the likeness of their meaning / semantic content.

Concretely, this can be achieved for instance by defining a topological similarity, by using ontologies to define a distance between words (a naive metric for terms arranged as nodes in a directed acyclic graph like a hierarchy would be the minimal distance—in separating edges—between the two term nodes), or using statistical means such as a vector space model to correlate words and textual contexts from a suitable text corpus (co-occurrence).

### 2.2 Semantic Knowledge from Probabilistic Approach

Probabilistic query expansion is usually based on calculating co-occurrences of terms in documents and selecting terms that are most closely related to query terms [3]. In [4], it categorizes existing approaches for visual concept-based retrieval similarly into three categories for text-based approaches – lexical, global statistical, and local statistical approaches. Lexical approaches for visual concept-based query expansion are based on textual descriptions of the visual concepts, which essentially reduce the problem to that of lexical text-based query expansion. Typically, each concept is represented with a brief description or a set of representative terms (e.g., synonyms). Given a textual query, the query words are then compared against the concept descriptions, and any matched concepts are used for refinement, where the matching may be exact or approximate. Global techniques extract their co-occurrence statistics from the whole document collection and might be resource intensive, although some of the calculations can be performed offline. Local techniques extract their statistics from the top- $n$  documents returned by an initial query and might use some corpus wide statistics such as the inverse document frequency, but they must be fast because they delay the response of the system. All calculations for local methods are done online; just after the user supplies the query and before presenting the results to the user [3].



### 3 Concept-Based Similarity Distance

In our previous research [5], web images could be mainly categorized into the following four different types: (i) images without any caption nor annotation (ii) images with caption but no annotation nor tags (iii) images annotated with keywords and tags (iv) images provided with the full MPEG-7 annotation. For those with some basic caption, such primitive level of information may be exploited to carry out inferential reasoning based on domain content. It has been found in our study in [6] that captions may sometimes harm annotation correctness, and QBE techniques can be additionally deployed to attempt to filter out the misleading captions and provided keywords.

#### 3.1 WordNet Similarity

WordNet [7], is one of these applications of semantic lexicon for the English language and is a general knowledge base and commonsense reasoning engine.

The purpose of the work is both to produce a combination dictionary-and-thesaurus that is more intuitively usable, and to support automatic text analysis and artificial intelligence applications.

For example, by using WordNet, ‘downtown’ has been expanded to ‘business district’, ‘commercial district’, ‘city centre’ and ‘city district’, while ‘city district’ has been expanded to ‘road’, ‘building’, ‘architecture’, ‘highway’ and ‘hotel’. The semantic knowledge is hierarchically expansible from the query terms and concepts and knowledge can be expanded extensively. The more extensive and complete such hierarchies, the greater the scope for rich semantic manipulation.

Recent research [8] on the topic in computational linguistics has emphasized the perspective of semantic relatedness of two lexemes in a lexical resource, or its inverse, semantic distance.

The first line of research [9], which brings together ontology and corpus, tries to define the similarity between two concepts  $c_1$  and  $c_2$  lexicalized in WordNet, named WordNet Distance (WD). It indicates by the information content of the concepts that subsume them in the taxonomy. Formally, define:

$$sim(c_1, c_2) = \max(c \in S(c_1, c_2))[-\log p(c)] \quad (3)$$

where  $p(c) = \frac{\sum_{n \in \text{words}(c)} \text{count}(n)}{N}$  and  $N$  is the total number of nouns observed. And  $S(c_1, c_2)$

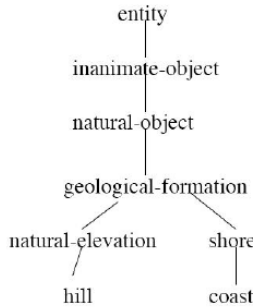
is the set of concepts that subsume both  $c_1$  and  $c_2$ . Moreover, if the taxonomy has a unique top node, then its probability is 1. In practice, we often measure word similarity rather than concept similarity. Using  $s(w)$  to represent the set of concepts in the taxonomy that are senses of word  $w$ , define

$$wsim(c_1, c_2) = \max(c \in S(c_1, c_2)) [sim(c_1, c_2)] \quad (4)$$

where  $c_1$  ranges over  $s(w_1)$  and  $c_2$  ranges over  $s(w_2)$ .

It defines two words as similar if near to one another in the thesaurus hierarchy. For example, refer to Fig. 3, ‘entity’ can expand to ‘inanimate-object’. Then

‘inanimateobject’ can expand to both ‘natural-object’ followed by ‘geological-formation’. It then expands to both ‘natural-elevation’ and ‘shore’. The former can expand to ‘hill’ while the latter expands to ‘coast’.



**Fig. 3.** Example demonstrate the way WordNet Distance defines two words are similar if nearby in thesaurus hierarchy

Beside the original work, researchers [10] propose a similarity measure between arbitrary objects. It uses the same elements but in a different fashion:

$$d(c_1, c_2) = \frac{2 \times \log p(lso(c_1, c_2))}{\log p(c_1) + \log p(c_2)} \tag{5}$$

### 3.2 Wikipedia Similarity

Wikipedia is the world’s largest collaboratively edited source of encyclopedic knowledge. In spite of its utility, its contents are barely machine-interpretable. Each article in Wikipedia describes a single topic; its title is a succinct, well-formed phrase that resembles a term in a conventional thesaurus. Meanwhile, each article must belong to at least one category of Wikipedia. Hyperlinks between articles keep many of the same semantic relations as defined.

WikiRelate [11] was the first to compute measures of semantic relatedness using Wikipedia. Their approach took familiar techniques that had previously been applied to WordNet and modified them to suit Wikipedia. Implementation of WikiRelate follows hierarchical category structure of Wikipedia.

The Wikipedia Link Vector Model (WLVM) [12] uses Wikipedia to provide structured world knowledge about the terms of interest. Their approaches are using the hyperlink structure of Wikipedia rather than its category hierarchy or textual content [13].

Probability of WLVM is defined by the total number of links to the target article over the total number of articles. Thus if  $t$  is the total number of articles within

Wikipedia, then the weighted value  $w$  for the link  $a \rightarrow b$  is:

$$w(a \rightarrow b) = a \rightarrow b \times \log\left(\sum_{x=1}^t \frac{t}{|x \rightarrow b|}\right) \tag{6}$$

where  $a$  and  $b$  denotes the search terms.

For example, calculate the similarity between *Israel* and *Jerusalem*, one would consider only the nation and its capital city. The commonness of a sense is defined by the number of times the term is used to link to it: e.g. 95% of *Israel* anchors link to the nation, 2% to the football team, 1% to the ancient kingdom, and a mere 0.1% to the Ohio township. According to Equation 4.11, WLVM value of both terms *Israel* and *Jerusalem* is 0.994, which is completely reasonable.

### 4 Preliminary Experiments and Results

We compute the relatedness of the searching keywords and their related concept using WordNet Similarity and Wikipedia Similarity, and then search images by the plain keywords as well as the related words expanded from the WordNet Similarity computation and Wikipedia Similarity computation. As shown in Fig.4 and 5, for Example, the word “downtown” can be inferred to “Brooklyn”, “Maryland”, and so on according to WordNet Ontology Similarity; while “downtown” can be inferred to “driving”, “tower”, and so on according to Wikipedia Ontology Similarity. The blue lines (Relatedness line) indicate the performance of relatedness rates between concept ‘downtown’ and other 327 concepts computed by the WordNet distance/Wikipedia distance. The red lines (Precision 1 line) show the image search precision with only the expanded keyword such as search by “Brooklyn”. The green lines (Precision 2 line) represent the image search precision with the expanded keyword combined with the original plain keyword such as search by “downtown + Brooklyn”. We use Flickr as the Web Image Database to test the precision.

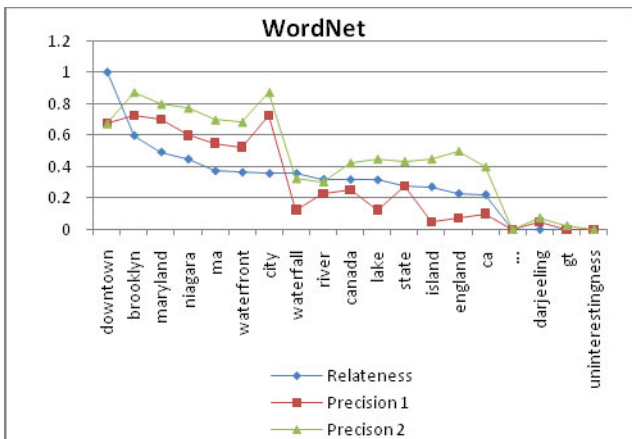


Fig. 4. Performance of relatedness rates between concept ‘downtown’ and other 327 concepts computed by the WordNet distance and image seaching precision

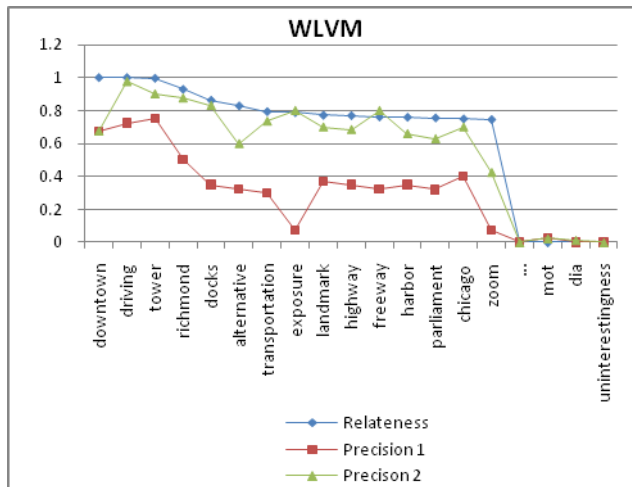


Fig. 5. Performance of relatedness rates between concept ‘downtown’ and other 327 concepts computed by the Wikipedia distance and image searching precision

The precision is calculated as follows:

$$Precision = \frac{|\{relevant\_images\} \cap \{retrieved\_images\}|}{|\{retrieved\_images\}|} \tag{7}$$

We also do the experiment on other 100 concepts, 10,000+ web images. And the results are showing in table 1 and 2:

Table 1. Performance of semantic relatedness measures of WD algorithms with their standard deviations

WD	Relatness	Precision 1	Precision 2
<u>Tagged Term</u>	1	0.630	0.630
word 1	0.687	0.786	0.905
word 2	0.582	0.730	0.827
word 3	0.498	0.629	0.752
word 4	0.467	0.560	0.710
word 5	0.401	0.541	0.633
word 6	0.393	0.429	0.575
word 7	0.352	0.425	0.514
word 8	0.325	0.325	0.405
word 9	0.308	0.270	0.388
word 10	0.301	0.125	0.345
word 11	0.265	0.124	0.235
word 12	0.239	0.05	0.210
...	...	...	...
uninterestingness	0	0	0

**Table 2.** Performance of semantic relatedness measures of WLVM algorithms with their standard deviations. The targeted terms is underlined.

WLVM	Relatedness	Precision 1	Precision 2
<u>Tagged Term</u>	1	0.630	0.630
word 1	0.995	0.790	0.956
word 2	0.980	0.764	0.920
word 3	0.932	0.733	0.885
word 4	0.887	0.634	0.820
word 5	0.813	0.625	0.752
word 6	0.789	0.598	0.630
word 7	0.781	0.521	0.615
word 8	0.774	0.450	0.568
word 9	0.770	0.350	0.485
word 10	0.762	0.315	0.418
word 11	0.750	0.308	0.366
word 12	0.736	0.300	0.354
...	...	...	...
uninterestingness	0	0	0

In these two tables, tagged term is the original keyword, such as “downtown”. Word 1-word 12 are the expanded words/concepts base on WD (WordNet Distance) and WLVM (Wikipedia Link Vector Model). Precision 1 column shows the image search average precision with only the expanded keywords of all tested tags. Precision 2 column represents the image search precision with the expanded keyword combined with the original plain keywords of all tested tags.

As we can see in Table 1 and 2, average precision are all satisfactorily higher compared with the only searching by the original terms. The value of precision increases with the relatedness increases. The performance of semantic relatedness of WD and WLVM could affect the image search performance positively. Meanwhile, the image search precision line goes with the WLVM semantic similarity line a little better than WD semantic similarity line. Furthermore, the precision of searching with original concept combined with the expanded concepts are higher than searching with only expanded keywords. These results indicate that significant improvement in performance may be attained from using the keywords expansion approach base on WordNet or Wikipedia ontology.

## 5 Conclusion

We explore numbers of semantic relatedness measure algorithms and developed an ontology-based, i.e. WD, WLVM. Our system is evaluated quantitatively, and experimental results indicate that this approach is able to deliver highly competent performance. Our approach, both search web image with the additional concepts expanded by WordNet Similarity and Wikipedia Similarity, not only demonstrates the applicability of ontology to the image annotation problem, but also using the sub-objects as surrogate terms for general queries improves the precision in the image sets.

However, there still exist some limitations in our proposed methods. This thesaurus based approach has its limitations: lots of detailed semantic knowledge present has

been lost and it does not provide a quantitative measure of the semantic similarity between concepts. Our future work consist of developing algorithms for increasing image retrieval precision of our model, and working out more accurate and fast re-ranking scheme, to achieve greater user satisfaction.

## References

1. Wong, C.F.: Automatic Semantic Image Annotation and Retrieval. PhD Thesis, Hong Kong Baptist University (August 2010)
2. Wikipedia Article website- Semantic similarity, [http://en.wikipedia.org/wiki/Semantic\\_similarity](http://en.wikipedia.org/wiki/Semantic_similarity)
3. Andreou, A.: Ontologies and Query Expansion (2005)
4. Natsev, A., Haubold, A., Tesic, J., Xie, L., Yan, R.: Semantiv Concept-Based Query Expansion and Re-ranking for Multimedia Retrieval. In: Proceedings of the 15th ACM International Conference on Multimedia, New York, NY, USA, pp. 991–1000 (2007)
5. Li, Y.X., Leung, C.H.C.: Multi-level Semantic Characterisation and Refinement for Web Image Search. In: The 2nd International Conference on Innovative Computing and Communication, pp. 70–73 (2011)
6. Wong, R.C.F., Leung, C.H.C.: Automatic Semantic Annotation of Real-World Web Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30(11), 1933–1944 (2008)
7. Miller, G.A.: Wordnet: a lexical database for english. *Communications of the ACM* 38(11), 39–41 (1995)
8. Budanitsky, A., Hirst, G.: Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures. In: Proceedings of Workshop on WordNet and Other Lexical Resources, Pittsburgh, PA, USA, p. 641. North American Chapter of the Association for Computational Linguistics (2001)
9. Resnik, P.: Using information content to evaluate semantic similarity in a taxonomy. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 448–453 (1995)
10. Lin, D.: An information-theoretic definition of similarity. In: Proceedings of the 15th International Conference on Machine Learning, pp. 296–304. Morgan Kaufmann (1998)
11. Strube, M., Ponzetto, S.P.: WikiRelate! computing semantic relatedness using wikipedia. In: Proceedings of the Twenty-First National Conference on Artificial Intelligence. AAAI Press (July 2006)
12. Milne, D., Witten, I.H.: An effective, low-cost measure of semantic relatedness obtained from wikipedia links. In: WIKIAI 2008: Proceedings of First AAAI Workshop on Wikipedia and Artificial Intelligence, Chicago, IL, USA (2008)
13. Völkel, M., Kröttsch, M., Vrandečić, D., Haller, H., Studer, R.: Semantic wikipedia. In: WWW 2006: Proceedings of the 15th International Conference on World Wide Web, New York, NY, USA, pp. 585–594. ACM (2006)
14. Leung, C.H.C., Li, Y.X.: CYC Based Query Expansion Framework for Effective Image Retrieval. In: The 4th International Congress on Image and Signal Processing (2011)
15. Leung, C.H.C., Chan, W.S., Milani, A., Liu, J., Li, Y.X.: Intelligent Social Media Indexing and Sharing Using an Adaptive Indexing Search Engine. *ACM Transactions on Intelligent Systems and Technology* (to appear, 2011)
16. Tam, A.M., Leung, C.H.C.: Semantic Content Retrieval and Structured Annotation: Beyond Keywords., ISO/IEC JTC1/SC29/WG11 MPEG00/M5738, Noordwijkerhout, Netherlands (March 2000)

# Fast Reconstruction Technique for Medical Images Using Graphics Processing Unit

Mohammad Nazmul Haque<sup>1</sup>, Mohammad Shorif Uddin<sup>2</sup>,  
M. Abdullah-Al-Wadud<sup>3</sup>, and Yoojin Chung<sup>4</sup>

<sup>1</sup> Department of CSE, Daffodil International University, Dhaka, Bangladesh  
nazmul@daffodilvarsity.edu.bd

<sup>2</sup> Department of CSE, Jahangirnagar University, Savar, Dhaka, Bangladesh  
shorifuddin@gmail.com

<sup>3</sup> Department of IME, Hankuk University of Foreign Studies, Yongin, South Korea  
wadud@hufs.ac.kr

<sup>4</sup> Department of CS, Hankuk University of Foreign Studies, Yongin, South Korea  
chungyj@hufs.ac.kr

**Abstract.** In many medical imaging modalities, the Fast Fourier Transform (FFT) is being used for the reconstruction of images from acquired raw data. The objective of the paper is to develop FFT and Inverse FT algorithms to run under GPU for performing in much faster way. The GPU based FFT implementation provides much faster reconstruction of Medical images than CPU based implementation. The GPU based algorithm is developed in MATLAB environment. GPUMat is used to running CUFFT library code in MATLAB. This work exercises the acceleration of MRI reconstruction algorithm on NVIDIA's GPU and Intel's Core2 Duo based CPU. The reconstruction technique shows that GPU based MRI reconstruction achieved significant speedup compared to the CPUs for medical applications at a cheaper cost.

**Keywords:** MRI reconstruction, image reconstruction, medical imaging, GPU based image processing, image processing using CUDA.

## 1 Introduction

The Fourier Transform (FT) is a mathematical operation used widely in many fields. In medical imaging it is used for many applications such as image filtering, image reconstruction and image analysis. When Magnetic Resonance Imaging (MRI) machines take an image of the human body, the output is in the form of raw data. The FT is used to reconstruct the image from this raw data. In medical imaging devices specialized hardware or CPUs are used to reconstruct images from acquired raw data. When the raw data size is relatively small, it takes moderate time to reconstruct an image. But, as the raw data size continues increasing, the time for processing the reconstruction increases as well. Image reconstruction has reached a bottleneck where further speed improvement from the algorithmic perspective is difficult. But requirement for faster FT has not comes in an end. This paper's aim is to increase the performance of FT based medical image reconstruction.

## 2 Literature Review

General-purpose computing on graphics processing units (GPU) supports a broad range of scientific and engineering applications, including physical simulation, signal and image processing, database management, and data mining [1]. There are several excellent reviews of image reconstruction by many other authors. These include: Calvetti, Reichel & Zhang (1999) on iterative methods; Molina et al. (2001) and Starck, Pantin & Murtagh (2002) on image reconstruction in astronomy.

Medical imaging was one of the first GPU computing applications, with computed tomography (CT) reconstruction achieving a speedup of two orders of magnitude on the SGI Reality Engine in 1994 [2]. A wide variety of CT reconstruction algorithms have since been accelerated on graphics processors [2], [3], [4], [5] and the Cell Broadband Engine [1]. In [3] the GPU is used to accelerate Simultaneous Algebraic Reconstruction Technique (SART), an algorithm that increases the quality of image reconstruction relative to the conventional filtered back-projection algorithm under certain conditions. MRI reconstruction on the GPU has not been studied extensively. Research in this area has focused on accelerating the FFT.

## 3 Magnetic Resonance Imaging Basics

Magnetic resonance imaging (MRI) is an imaging technique used primarily in medical settings to produce high quality images of the inside of the human body. MRI is based on the principles of tomographic image technique [7], [8], [9]. It is used primarily in medical fields to produce images of the internal structure of human body. It does not involve the use of ionizing radiation hence, free from associated harmful effects known with other imaging techniques [10].

MRI imaging equation can be express as a two dimensional entity given as

$$S(k_x, k_y) = f[I(x, y)] \quad (1)$$

where  $f$  represent spatial information encoding scheme [7]. If  $f$  is invertible, a data consistent  $I$  can be obtained from the inverse transformation such that

$$I(x, y) = f^{-1}S(k_x, k_y) \quad (2)$$

The desired image intensity function  $I(x, y)$  is a function of: Relaxation times,  $T1$ ,  $T2$  and  $T2^*$  spin density  $\rho$ , Diffusion coefficients  $D$  and so on [11]. Using function notation, this can be written as

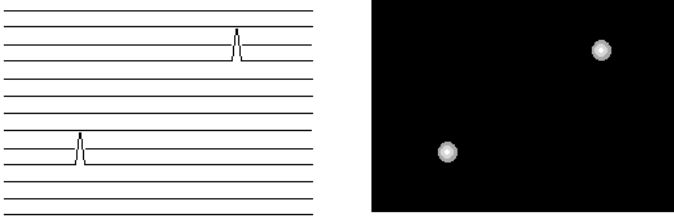
$$I = f[T1, T2, T2, \rho, D] \quad (3)$$

Generally,  $T1$  and  $T2$  are two independent processes and happen simultaneously.  $T1$  is called spin lattice relaxation, because the energy from this process is released to the surrounding tissue (lattice) [7], [8], [11].  $T1$  happens along the z-component axis and its value is always greater than the spin-spin relaxation  $T2$ . The relationship between protons and their immediate surroundings (molecules) is described by the spin-spin relaxation  $T2$  and it happens along x-y plane.



### 3.1 Fourier Transform Based MR-Imaging Principle

FT base tomographic imaging is the most commonly used MRI method utilized today. It uses a type of magnetic field gradient, called a phase encoding gradient, in addition to the slice selection and frequency encoding gradients. The MRI recording process is repeated 128 to 256 times, varying the magnitude of the phase encoding gradient each time, in order to obtain sufficient data, free induction decays or signals, for creating an image [12].



**Fig. 1.** Tomographic image created from Fourier transformed data

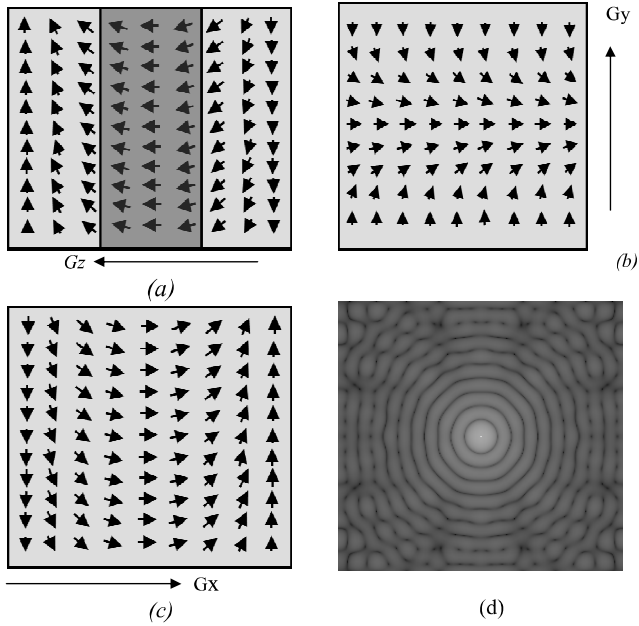
Before actually creating the image, however, the signals must be Fourier transformed. This is done first in the direction in which the spins are located to extract the frequency domain information. Then it is done in the phase encoding direction to obtain information about the spin locations in that direction. The FT data finally becomes an image when the intensities of the data peaks are converted into intensities of pixels. This creates a tomographic image, shown in Fig.1.

### 3.2 MRI Data Acquisition

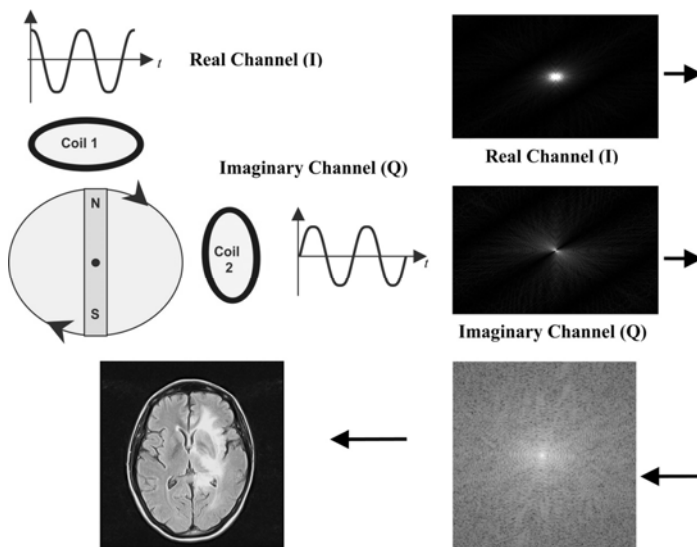
MRI data acquisition involves three major steps shown in Fig. 2, namely:

- Gz, Slice selection by the use of Gz gradient: This select axial slice in the object to be imaged.
- Gy, Phase encoding using the Gy gradient: This creates rows with different phases.
- Gx, Frequency encoding using the Gx gradient: This will create columns with different frequencies.

The data that are captured during MR imaging are called k-space data or, simply, raw data. Typically, the data are collected by using quadrature detection, which provides both real and imaginary k-space data. K-Space data include useful information but can be interpreted only after they are translated into images with the Fourier transform method which is illustrated in Fig 3. The K-space data contains all the necessary information required to reconstruct an image. Also, it gives a comprehensive way for classification and understanding of the imaging properties and method of reconstruction [7], [13], [14].



**Fig. 2.** K-Space Data formation: a)  $G_z$  Gradient data Filling, b)  $G_y$  Gradient Data Filling, c)  $G_x$  Gradient Data Filling and d) K-Space Data



**Fig. 3.** Schematic of the collection and combination of imaginary MR signal data to produce a complex map of k-space

## 4 MR Image Reconstruction

There exist various methods of converting the acquired K-space data to real images in spatial domain. These include: the use of DFT, Radon Transform, Parametric technique, artificial neural network based reconstruction technique and so on. DFT technique involves the application of Fourier series on the linearly or radially sampled k-space data. This work deals with only DTF technique using FFT.

### 4.1 Application of Inverse FFT to K-Space Data

MRI reconstruction using FFT/IFFT is done in two steps; firstly the 2D- IFFT of the data is computed then data are shifted to center for display the image. In practical MRI reconstruction, there are some other pre-processing activities that must be accomplished before the application of IFFT [15], [16], [17]. The reconstruction algorithm is depicted here.

#### *Algorithm : MRI Reconstruction*

*Input: MRI Raw Spectral Domain Data*

*Output: Reconstructed Spatial Domain Image*

Step 1: Read data header information:

Load RAW data contain information about the MRI data file.  
It is a text file containing Information about Offset,  
DATA size, Kx Co-ordinate, Ky Co-ordinate etc.

Step 2: Read in the K-space DATA.

Step 3: IFFT in K(x,y) Direction

Step 4: FFT shift

Step 5: Image Display

K-space is the rawest form of data obtained at MR imaging. An acquisition with a 256 x256 matrix contains 256 lines of data, and each of those lines contains 256 data points. The distance between neighboring points in k-space determines the field of view of the object imaged, and the extent of k-space determines the resolution of the image [18], [19].

## 5 Experimental Setup

The experiment is divided into two sections. First is simulation of performance for FFT-IFT on different sized images. After that the experiment is done on MR data. All experiments are done using both CPU and GPU. The Hardware configurations for them are listed in Table 1 and Table 2. The experiment requires some software and tools for programming and documenting purpose. Following Table 3 lists up all used software and tools.

**Table 1.** CUDA Device Configuration at Experiment

Features	Specification
Name:	GeForce G 103M
CUDA Driver Version:	3.2
Total Global memory:	521601024 bytes
Multiprocessors x Cores:	1 x 8
Clock rate:	1.60 GHz

**Table 2.** Host Machine Configuration at Experiment

Features	Specification
System Model:	Compaq Presario CQ40 Notebook PC
Operating System:	Windows 7 Ultimate 32-bit
Processor:	Intel(R) Core(TM)2 Duo CPU
Clock Speed:	2.00GHz
Memory:	2048MB RAM

**Table 3.** Required Software and Tools

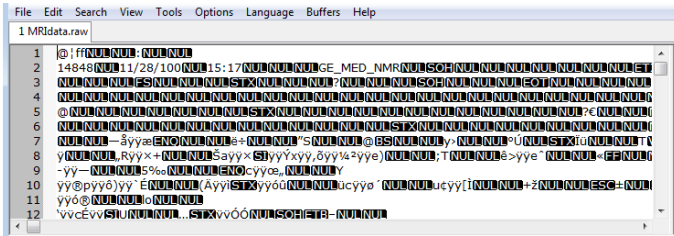
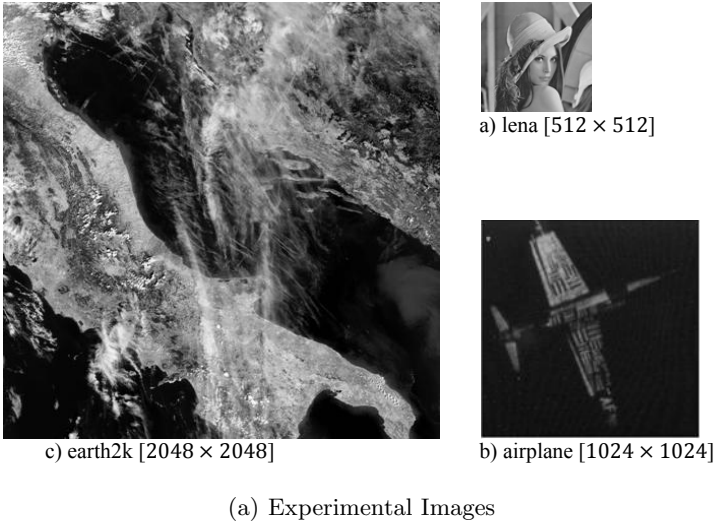
Software	Version	Purpose
NVIDIA GPU Computing SDK	3.2	SDK for NVIDIA's GPU
MATLAB R2010a	7.10	Simulation and Programming
CUFFT	2.3	CUDA capable FFT library
GPUmat	0.27	Wrapper for MATLAB to run CUDA Program

For comparing the runtime of FFT based image Reconstruction using GPU we have used three images shown in Fig 4(a). Images are chosen different resolution for figure out the performance on CPU and GPU based implementation. MRI raw data in Fig 4(b) has been collected from Duke Virtual Imaging Laboratory [6].

## 6 Simulation Results

GPU and CPU performance results are obtained by computing FFT of experimental images in MATLAB and measuring the execution time using commands tic and toc. All CPU iterations are measured and averaged over 100 iterations.

Spatial resolution of lena image is 512x512. Both CPU based and GPU based reconstruction is done 100 times for measuring runtime and plotted in Fig 5(a). The simulation shows an average speed up of GPU vs. CPU by a factor of 3.1 times. The spatial resolution of airplane image is 1024x1024. Both CPU based and GPU based reconstruction is done 100 times for measuring runtime. The simulation output in Fig 5(b) shows an average speed up of GPU by a factor of 4.1 time speedy than CPU. We have used earth2k image having resolution of 2048x2048. For this image, both CPU based and GPU based reconstruction is done 100. The simulation result in Fig 5(c) shows an average speed up of 4.349 times faster in GPU than CPU.



(b) MRI Raw data file

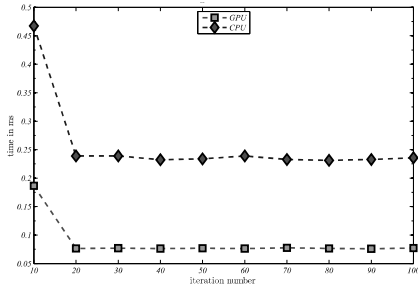
Fig. 4. Experimental (a) Images and (b) Data

FFT and IFT time for 100 iterations are measured and plotted total time requirement in Fig 5(d). From the performance measures of different image size has been observed. The result in Fig 5(d) the reconstruction performance of GPU over CPU was augmented gradually while the image size increased.

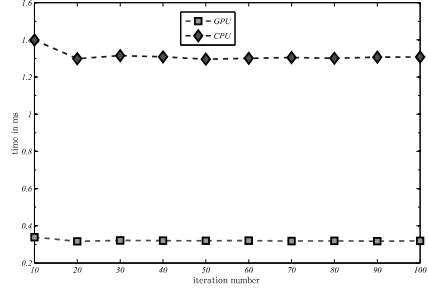
## 7 GPU Based MR-Image Reconstruction

Raw data file is opened and the proposed system separates header info and k-space data. Then raw k-space data is separated into real image and imaginary image data space. After that spatial frequency data is created from both real and imaginary data. Spatial frequency data is resized by reducing the dimension data and matrix transpose operation. The frequency image is shown as in Fig 6. After applying the IFT, the reconstructed MR Phantom Image is acquired. The visual property of CPU and GPU based reconstructions are similar.

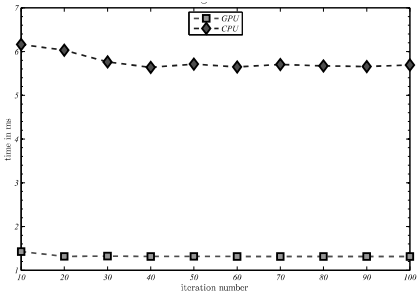
The GPU Based FFT improves the performance of the reconstruction algorithm by a factor of 8.2 which is showed in Fig 7. The resulting image quality is



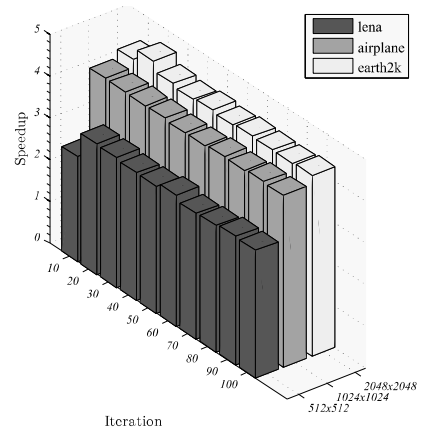
(a) lena image



(b) airplane image

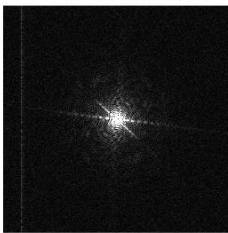


(c) earth2k image

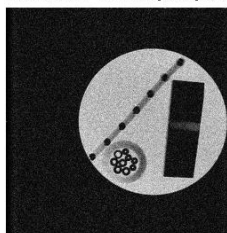


(d) Average Performance

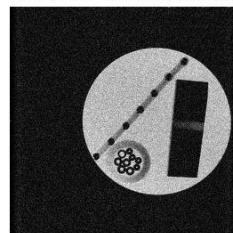
**Fig. 5.** GPU vs. CPU Performance of Reconstruction for (a) Lena Image (b) Airplane Image (c) Earth2k Image and (d) Speedup of Image Reconstruction by Image Size



a) Raw Data



b) MRI Reconstruction by CPU



c) MRI Reconstruction by GPU

**Fig. 6.** MRI Reconstruction Outputs in both CPU and GPU

visually same for both CPU-based and GPU-based implementations. In future, the real-time GPU-based MR image reconstruction will be possible.

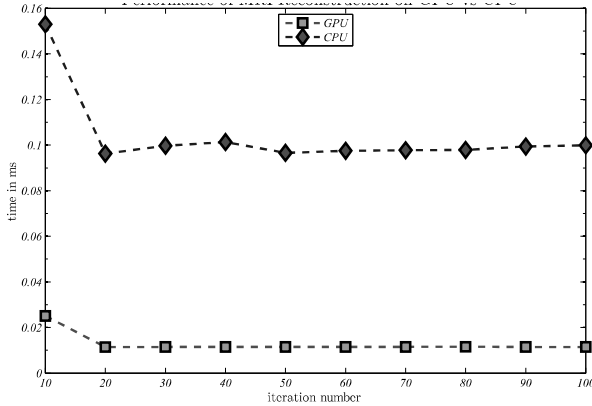


Fig. 7. MR Image Reconstruction Performance of GPU over CPU

## 8 Limitations and Future Work

Currently the work can perform only on data that resides in GPU memory. External memory algorithms based on the hierarchical algorithm can be designed to handle larger data. Computation can also be performed on multiple GPUs. It can also possible to do other processing, such as de-noising and de-blurring on MRI raw data before output the reconstructed image. The work can be ported to real-time visualization of MR image that is very important for envision situation of sensitive internal organs while operating a patient.

## 9 Conclusion

Besides the performance advantage of using a GPU over a CPU for MR-image reconstruction, there are other advantages as well. In MRI device, the CPU can be preoccupied with time-critical tasks such as controlling the data acquisition hardware. In this case, it is beneficial to use the GPU for image reconstruction, leaving the CPU to do data acquisition. Moreover, because of the GPU is free of interrupts from the operating system, it results better performance than interrupt driven CPU. The rate of increase in performance of GPUs is expected to outshine that of CPUs in the next few years, increasing the demand of the GPU as the processor of choice for medical applications.

**Acknowledgment.** This work was supported by Hankuk University of Foreign Studies Research Fund of 2011.

## References

1. Bockenbach, O., Knaup, M., Kachelrie, M.: Implementation of a cone-beam back-projection algorithm on the Cell Broadband Engine processor. In: SPIE Medical Imaging 2007: Physics of Medical Imaging (2007)

2. Mueller, K., Xu, F., Neophytou, N.: Why do commodity graphics hardware boards (GPUs) work so well for acceleration of computed tomography? In: SPIE Electronic Imaging 2007 (2007)
3. Mueller, K., Yagel, R.: Rapid 3-D cone-beam reconstruction with the simultaneous algebraic reconstruction technique (SART) using 2-D texture mapping hardware. *IEEE Transactions on Medical Imaging* 19, 1227–1237 (2000)
4. Chidlow, K., Möller, T.: Rapid emission tomography reconstruction. In: Int'l Workshop on Volume Graphics (2003)
5. Xue, X., Cheryauka, A., Tubbs, D.: Acceleration of uro-CT reconstruction for a mobile C-Arm on GPU and FPGA hardware: A simulation study. In: SPIE Medical Imaging (2006)
6. Virtual Imaging Laboratory, Department of Biomedical Engineering, Duke University, <http://dukemil.bme.duke.edu/MRI/index.html>
7. Liang, Z.P., Lauterbur, P.C.: Principles of Magnetic Resonance Imaging. A signal processing perspective. IEEE Press (2000)
8. Haacle, E.M., Liang, Z.P.: Challenges of Imaging Structure and Function with MRI. *IEEE Transactions on Medicine and Biology* 19, 55–62 (2000)
9. Nishimura, D.G.: Principles of Magnetic Resonance Imaging (April 1996)
10. MRI Basics, <http://www.cis.rit.edu/htbooks/mri/inside.htm>
11. Haacle, E.M., Liang, Z.P.: Challenges of Imaging Structure and Function with MRI. *IEEE Transactions on Medicine and Biology* 19, 55–62 (2000)
12. Uzun, I.S., Bouridane, A.A.A.: FPGA Implementations of Fast Fourier Transforms for Real-Time Signal and Image Processing. In: Proceedings. 2003 IEEE International Conference on Field-Programmable Technology, FPT (2003)
13. Zhuo, J., Gullapalli, R.P.: MR Artifacts, Safety, and Quality. *Physics Tutorial for Residents*, pp. 275–279 (2006)
14. Aibinu, A.M., Salami, M.J.E., Shafie, A.A., Najeeb, A.R.: MRI Reconstruction Using Discrete Fourier Transform: A tutorial. World Academy of Science, Engineering and Technology (2008)
15. Moratal, D., Valls-Luch, A., Mart-Bonmat, L., Brummer, M.: k-Space tutorial: an MRI educational tool for a better understanding of k-space. *Biomedical Imaging and Intervention Journal* (2008)
16. Chaari, L., Pesquet, J.-C., Benazza-Benyahia, A., Ciuciu, P.: Autocalibrated Parallel MRI Reconstruction in the Wavelet Domain. In: IEEE International Symposium on Biomedical Imaging, Paris, France, May 14–17, pp. 756–759 (2008)
17. Schultz, G., et al.: K-Space Based Image Reconstruction of MRI Data Encoded with Ambiguous Gradient Fields. In: Proc. of International Society for Magnetic Resonance in Medicine (2011)
18. Stone, S.S., et al.: Accelerating Advanced MRI Reconstructions on GPUs. In: Proceedings of the 5th International Conference on Computing Frontiers (2008)
19. Kleut, D., Jovanovi, M., Reljin, P.D.B.: 3D Visualisation of MRI images using MATLAB. *Journal of Automatic Control* 16(1-3) (2006)



# Adaptive Image Zooming Based on Bilinear Interpolation and VQ Approximation

Yu-Chen Hu, Wu-Lin Chen, and Jun-Rong Zeng

Dept. of Computer Science & Information Management, Providence University  
200 Chung Chi Rd., Taichung 43301, Republic of China, Taiwan  
{ychu, wlchen, g9871018}@pu.edu.tw

**Abstract.** In this paper, a two-stage image zooming scheme is proposed. In the first stage, the bilinear interpolation technique is employed to recover the unknown pixels in the smooth areas. In the second stage, the vector quantization approximation technique is used to determine the unknown pixels in the edge boundary. From the experimental results, it is shown that the proposed scheme provides better image quality than the linear prediction technique, the local adaptive zooming technique and the vector quantization zooming technique. Besides, the proposed scheme requires little computational cost for image zooming.

**Keywords:** image zooming, bilinear interpolation, vector quantization.

## 1 Introduction

Image zooming [1]-[11] is a basic function of digital image processing. It can be applied to many multimedia applications, such as image database, electronic publishing, digital camera, and medical imaging. Typically, the original image to be enlarged is often called the low-resolution and the enlarged image is called the high-resolution. Basically, the image zooming techniques can be classified into two categories: fixed zooming [2]-[4] and adaptive zooming [5]-[11]. The major difference between these two categories is that the image properties are exploited by the adaptive zooming techniques.

The zooming techniques, such as the pixel copy technique [1], the bilinear interpolation technique [2], and the bi-cubic interpolation technique [3], belong to the fixed zooming category. Among them, the pixel copy technique is the simplest technique for image zooming. The unknown pixels are recovered by its corresponding sampled pixels, and the pixel copy technique is done. However, the blocking effect can be easily found in the pixel copy technique when the scaling factor is high. In addition, the image quality of the enlarged image is poor. The bilinear interpolation technique is a commonly used technique for image zooming. It works well for zooming smooth areas. Nevertheless, the blocking artifact can be observed in the edge boundary of the enlarged image by the bilinear interpolation technique.

From the literature, some adaptive zooming techniques, such as the edge preserving zooming technique [5,6,9,11], the local adaptive zooming (LAZ) technique

[7], the vector quantization zooming technique [8,10], had been proposed. The goal of edge preserving zooming techniques is to remove/reduce artifacts arising in image interpolation. In 1996, Hong et al. proposed the edge-preserving image interpolation system [5]. In this system, the edge direction is determined by exploiting the DCT coefficients. Li et al. proposed the edge-directed interpolation technique [6] based on the local covariance coefficients of the low-resolution image. Then, the interpolation is achieved by using the geometric duality of the variance between the low-resolution covariance and the high-resolution covariance.

In 2002, Battiato et al. proposed the local adaptive zooming technique [7]. The basic idea of LAZ is to perform a gradient-controlled and weighted interpolation. The method speeds up the entire adaptive process without requiring a preliminary gradient computation. Chang et al. proposed the vector quantization zooming technique (VQZ) [8] in 2005. In VQZ, the unknown pixels on the enlarged image are interpolated by using a vector quantization codebook based on their local information. In addition, the interpolative classified vector quantization technique [10] had been proposed in 2008.

The remaining of this paper is organized as follows. In Section 2, the review of the vector quantization (VQ) scheme [12-16] will be given. In Section 3, the proposed scheme for image zooming will be introduced. The experimental results will listed in Section 4. Finally, the conclusions will be given in Section 5.

## 2 The Vector Quantization Scheme

The vector quantization scheme was proposed by Linde et al. [12] in 1980 for grayscale image coding. VQ consists of three procedures: codebook generation, image encoding, and image decoding. The goal of the codebook generation procedure is to design a set of representative codewords for image encoding/decoding. From the literature, the LBG algorithm [12] is the most commonly used algorithm for codebook generation.

To generate a codebook of  $N$  codewords, multiple training images are selected. Each training image is divided into non-overlapping image blocks of  $n \times n$  pixels. Then,  $N$  image blocks from the training image blocks are chosen as the initial codewords. The set of initial codewords is also called the initial codebook. After the initial codebook is generated, an iterative process for several rounds is executed to generate the codebook in the LBG algorithm. In each round, the data clustering process and the centroid updating process are executed.

In the data clustering process, the closest codeword of each training image block in the codebook is to be determined. To find out the closest codeword of each training image block, a total number of  $N$  squared Euclidean distances are calculated. Then, the closest codeword is determined by the minimal squared Euclidean distance from it to the training image block. After the data clustering process is performed, all these training image blocks are classified into  $N$  groups. In the centroid updating process, the mean vector of the training image blocks in each group is computed. These  $N$  calculated image vectors in turn form the new codebook in this round. By repeatedly performing the data clustering process and the centroid updating process in each round, a representative set of codewords can be generated.

In the image encoding procedure, the grayscale image to be compressed is divided into non-overlapping image blocks of  $n \times n$  pixels. Each image block can be treated as a  $k$ -dimensional image vector where  $k = n \times n$ . To encode each image block, we search the codebook of  $N$  codewords to find the closest codeword for the image block based on the minimal squared Euclidean distance criterion. The codeword in the codebook that has the minimal squared Euclidean distance from it to the image block is selected. Then, the  $\log_2 N$ -bit index of the closest codeword is recorded. After each image block is sequentially compressed, the set of the indices, also called the index table, is obtained.

In the image decoding procedure, the same codebook of  $N$  codewords is stored. To recover each image block, the index of  $\log_2 N$  bits of each compressed block is sequentially extracted from the index table. The codeword of each index in the codebook is used to rebuild the compressed image block. In other words, the decoder only needs to perform simple table lookup operations to reconstruct the encoded image.

### 3 The Proposed Scheme

The goal of the proposed scheme is to combine both the advantage of the bilinear interpolation technique and that of the vector quantization scheme for image zooming. We assume that the bilinear interpolation technique works well in the smooth areas. Then, the vector quantization approximation technique is employed to process the complex unknown pixels.

Given the original image  $I$  of  $w \times h$  pixels, the scaling factor  $SF$  is set to 2 in the proposed scheme. The enlarged image  $E$  of  $W \times H$  pixels is to be generated where  $W = w \times 2$  and  $H = h \times 2$ . Three main steps are executed to generate the enlarged image in the proposed scheme. The first step is to copy the corresponding coordinate of each pixel in  $E$  from a pixel in  $I$ . An example of processing the first step is depicted in Fig. 1. The original image of  $4 \times 4$  is listed in Fig. 1(a) and the processed result of the first step is listed in Fig. 1(b).

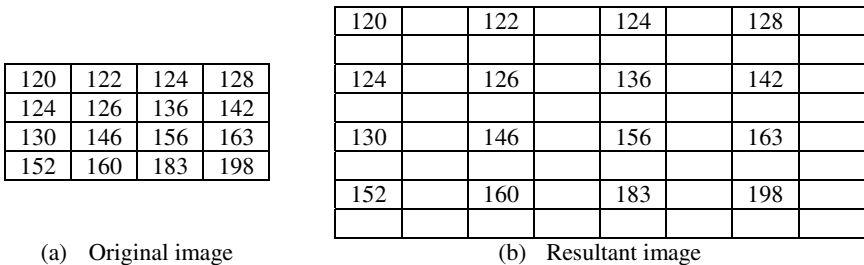
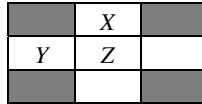


Fig. 1. Example of the first step processing

After the first step is executed, some pixels in  $E$  are still unknown. In fact, only 1/4 of all pixels in  $E$  are processed. In general, the unknown pixels can be classified into three categories. The positional diagram of the unknown pixels is listed in Fig. 2. An unknown pixel is classified as the first category if its adjacent left and right pixels

have already existed; pixel *X* in Fig. 2 belongs to the first category. An unknown pixel is classified as the second category if its adjacent up and down pixels have already been copied; pixel *Y* in Fig. 2 belongs to the second category. One pixel is classified as the third category if its adjacent four diagonal pixels exist. In Fig. 2; pixel *Z* belongs to the third category.



**Fig. 2.** Positional diagram of the unknown pixels to be processed

Before the bilinear interpolation technique is to be performed in the second step of the proposed scheme, the maximal absolute difference (*MAD*) of the adjacent pixels for each unknown pixel will be computed. To compute *MAD* of each unknown pixel, the maximal value (*max*) and the minimal value (*min*) of its adjacent sampled pixels are searched. The *MAD* value of each unknown pixel can be computed as follows:

$$MAD = max - min. \tag{1}$$

After the *MAD* value of each unknown pixel is computed, the rough image generated by the first step is then processed by using the bilinear interpolation technique. Here, a predefined threshold *TH* is used to control the degree of similarity among these neighboring existing pixels. Each unknown pixel is processed by the linear prediction technique if the *MAD* value is less than or equal to *TH*. Otherwise, the unknown pixel is skipped in step 2. When each unknown pixel is sequentially processed by the above mentioned process in the order of left-to-right and then top-to-bottom, the rough image of the second stage is then generated.

An example of processing the second step is depicted in Fig. 3. Here, the rough image as shown in Fig. 1(b) is used. The value of *TH* is set to 20 in this example. The pixels in red are processed by the bilinear interpolation technique. According to the results, it is shown that the unknown pixels in the smooth areas are processed by the bilinear interpolation technique. However, the unknown pixels in the complex edge boundary are still unknown after the bilinear interpolation technique is executed.

120	121	122	123	124	126	128	
122	123	124	127	130	133	135	
124	125	126	131	136	139	142	
127	132	136	141	146			
130	138	146	151	156	160	163	
		156					
152	156	160		183	191	198	

**Fig. 3.** Example of the second step processing when *TH* is set to 20

In the third step, the vector quantization approximation technique is employed to process the remaining unknown pixels. Suppose the codebook of  $N$  codewords had already been generated. The rough image is divided into non-overlapping image blocks of  $n \times n$  pixels. Each image block  $x$  is sequentially processed in the order of left-to-right and then top-to-bottom.

If the number of the unknown pixels in  $x$  is equal to 0,  $x$  is unchanged. Otherwise,  $x$  is processed by the vector quantization approximation technique to determine the values of the unknown pixels. Since there are some unknown pixels in  $x$ , the traditional image encoding procedure of the vector quantization scheme cannot be used here. Instead, the vector quantization scheme with partial dimension search is employed in the proposed vector quantization approximation process. In other words, only the existing pixels in the rough image are used to find out the closest codeword in the codebook. Let  $q$  denote the number of existing pixels in  $x$  where  $0 < q < n \times n$ . These  $q$  pixels in  $x$  are used to determine the closest codeword in the codebook. After the closest codeword of  $x$  is searched by using these  $q$  dimensions, these unknown pixels are recovered by the corresponding values of the searched closest codeword.

Continue the example shown in Fig. 3, and the enlarged image are divided into four image blocks  $B_1$ ,  $B_2$ ,  $B_3$ , and  $B_4$  when the block size is set to  $4 \times 4$ . The four image blocks are listed in the following:

$$\begin{aligned} B_1 &= [120, 121, 122, 123, 122, 123, 124, 127, 124, 125, 126, 131, 127, 132, 136, 141], \\ B_2 &= [124, 126, 128, \_, \_, 130, 133, 135, \_, \_, 136, 139, 142, \_, \_, 146, \_, \_, \_, \_], \\ B_3 &= [130, 138, 146, 151, \_, \_, \_, \_, 156, \_, \_, 152, 156, 160, \_, \_, \_, \_, \_], \\ B_4 &= [156, 160, 163, \_, \_, \_, \_, \_, \_, \_, 183, 191, 198, \_, \_, \_, \_, \_]. \end{aligned}$$

In this example, the numbers of the unknown pixels in these four image blocks are 0, 6, 8, and 10, respectively. The first block  $B_1$  is skipped in the vector quantization approximation technique because all pixels have already existed. In the second block  $B_2$ , 10 dimensions are used to search the closest codeword in the codebook. Similarly, 8 and 6 dimensions are used in  $B_3$  and  $B_4$  to find out their corresponding closest codewords, respectively.

## 4 Experimental Results

The simulations are executed on Microsoft windows XP with an Intel Core Duo 2.2GHz CPU and 512Mbytes RAM. The codes are implemented by using Bloodshed Dev C++. In the simulations, four grayscale images of  $512 \times 512$  pixels "Airplane", "Boat", "Goldhill", and "Toys" are used as the training images for codebook design by the LBG algorithm [13]. The terminated threshold used in the LBG algorithm is set to 0.001. Six testing images of  $512 \times 512$  pixels, "Airplane", "Boat", "Girl", "Goldhill", "Lenna", and "Toys", as shown in Fig. 4 are used to compare the performance of the comparative methods with that of the proposed scheme.

In the simulations, each grayscale image of  $512 \times 512$  pixels is sub-sampled to generate the low-resolution image of  $256 \times 256$  pixels. Then, each low-resolution image is processed to generate the high-resolution (enlarged) image of  $512 \times 512$  pixels. In other words, the scaling factor  $SF$  is set to 2 in the simulations. Then, the similarity between the original image and the enlarged image is compared. To measure the image

quality of the compressed image, the Mean Square Error (*MSE*) between the pixels of the original image and those of the enlarged image is defined as

$$MSE = \frac{1}{W \times H} \sum_{i=1}^W \sum_{j=1}^H (o_{ij} - e_{ij})^2. \quad (2)$$

Here,  $o_{ij}$  and  $e_{ij}$  denote the pixels in the original image and the enlarged image, respectively. The quality of the enlarged image is measured by means of the peak signal-to-noise-ratio (*PSNR*), which is defined as

$$PSNR = 10 \times \log_{10} \frac{255^2}{MSE}. \quad (3)$$

Basically, *PSNR* is considered as an indication of image quality rather than a definitive measurement; however, it is a commonly used measurement for evaluating the image quality.

Experimental results of the image qualities by the pixel copy technique and by the bilinear interpolation technique are listed in Table 1. Trivially, the bilinear interpolation technique achieves better image quality of the enlarged image than the pixel copy technique. Average image qualities of 27.792 dB and 32.842 dB are achieved by using the pixel copy technique and the bilinear interpolation technique, respectively.



**Fig. 4.** Six testing images for performance evaluation

Experimental results of LAZ [7] are listed in Table 2. In LAZ, three thresholds  $TH_1$ ,  $TH_2$ , and  $TH_f$  are needed. Among them,  $TH_1$  and  $TH_2$  are used to classify the pixel activity.  $TH_f$  denotes the threshold value used to determine the edge direction. Here, the  $TH_1$  values are set to 50, 100, 150, 200, and 250. The value of  $TH_2$  is set to the same value of  $TH_1$  in the simulation. Originally, LAZ was proposed to expand the low-resolution image of 256×256 pixels into the enlarged image of 511×511 pixels. To generate the enlarged image of 512×512 pixels, the additional linear extrapolation process is executed in the simulations. It is shown that the image quality of the enlarged image increases with the increase of the value of  $TH_1$ . Compared to the results of the bilinear interpolation technique as shown in Table 1, LAZ provides slightly worse performance.

**Table 1.** Results of the image qualities of the pixel copy technique and the bilinear interpolation technique

Images	Pixel copy	Bilinear interpolation
Airplane	27.208	31.372
Boat	26.536	30.747
Girl	29.300	35.407
Goldhill	28.797	33.001
Lenna	29.285	35.713
Toys	25.625	30.812
<b>Average</b>	<b>27.792</b>	<b>32.842</b>

Experimental results of VQZ [8] are listed in Table 3. In VQZ, four thresholds  $TH_s$ ,  $TH_d$ ,  $TH_r$ , and  $TH_v$  are needed. According to the the authors’ suggestion,  $TH_s$ ,  $TH_d$ , and  $TH_r$  are set to 75, 5, and 5, respectively. Five  $TH_v$  values 50, 100, 150, 200, and 250 are used in the simulations. Similarly, VQZ was proposed to expand the low-resolution image of 256×256 pixels into the enlarged image of 511×511 pixels. To generate the enlarged image of 512×512 pixels, the additional linear extrapolation process is executed in the simulations. According to the results in Tables 2 and 3, VQZ achieves better performance than LAZ. Nevertheless, the average image quality of VQZ is worse than that of the bilinear interpolation technique.

**Table 2.** Average image qualities of the enlarged images by using LAZ

Threshold	$TH_1=50$	$TH_1=100$	$TH_1=150$	$TH_1=200$	$TH_1=250$
$TH_f=50$	29.090	29.218	30.492	31.390	31.634
$TH_f=100$	29.070	29.066	30.405	31.390	31.634
$TH_f=150$	29.071	29.031	30.052	31.389	31.634

**Table 3.** Average image qualities of the enlarged images by using VQZ

Threshold	$N=256$	$N=512$	$N=1024$	$N=2048$
$TH_v=50$	32.187	32.196	32.252	32.273
$TH_v=100$	32.357	32.350	32.379	32.392
$TH_v=150$	32.377	32.359	32.388	<b>32.398</b>
$TH_v=200$	32.362	32.362	32.361	32.362

Experimental results of the proposed scheme with different block sizes are listed in Table 4. In the simulations, three block sizes  $3 \times 3$ ,  $4 \times 4$ , and  $5 \times 5$  are tested. In Table 4, the best average image qualities with different codebook sizes are shown in bold fonts. Average image qualities of 33.197 dB, 33.208 dB, 33.215 dB, and 33.242 dB are achieved in the proposed scheme when the codebook sizes are 256, 512, 1024, and 2048, respectively. In other words, the image quality of the proposed scheme increases as the codebook size increases. From the results, it is shown that the best image qualities are achieved when the block size is set to  $5 \times 5$ .

**Table 4.** Average image qualities of the enlarged images by using the proposed scheme with different block sizes

Threshold		$N = 256$	$N = 512$	$N = 1024$	$N = 2048$
$TH = 50$	$3 \times 3$	31.190	31.212	31.234	31.206
	$4 \times 4$	31.600	31.605	31.691	31.681
	$5 \times 5$	31.692	31.855	31.949	32.035
$TH = 100$	$3 \times 3$	32.678	32.694	32.695	32.696
	$4 \times 4$	32.796	32.751	32.817	32.800
	$5 \times 5$	32.846	32.891	32.926	32.994
$TH = 150$	$3 \times 3$	33.156	33.167	33.147	33.166
	$4 \times 4$	33.145	33.123	33.150	33.144
	$5 \times 5$	<b>33.197</b>	<b>33.208</b>	<b>33.215</b>	<b>33.242</b>
$TH = 200$	$3 \times 3$	32.850	32.849	32.848	32.849
	$4 \times 4$	32.771	32.771	32.773	32.769
	$5 \times 5$	32.856	32.856	32.856	32.854

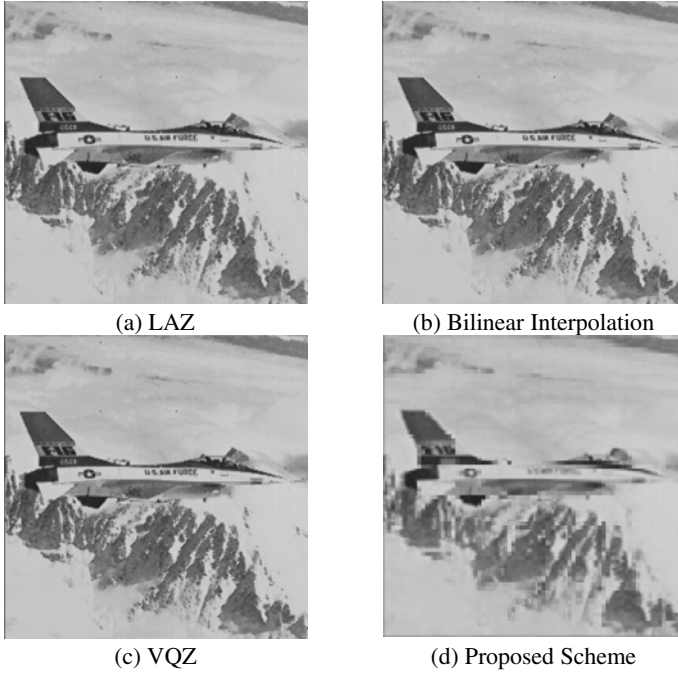
Results of the best image qualities of LAZ, VQZ and the proposed scheme are listed in Table 5. Among these three techniques, LAZ provides the worst image quality for image zooming. Average image qualities of 31.700 dB, 32.362 dB and 33.242 dB are achieved by using LAZ, VQZ and the proposed scheme, respectively. The proposed scheme provides better image quality than LAZ and VQZ. Compared to the results shown in Table 1, it is shown that both LAZ and VQZ provide worse performance than the bilinear interpolation technique. However, LAZ and VQZ achieve better image quality than the pixel copy technique. From the results shown in Tables 1 and 5, the proposed scheme provides the best image quality of the enlarged image. Average image quality gain of 0.4 dB is achieved by using the proposed scheme compared to the bilinear interpolation technique.

**Table 5.** Best image qualities of LAZ, VQZ and the proposed scheme

Images	LAZ	VQZ	Proposed scheme
Airplane	30.434	31.034	31.376
Boat	29.890	30.422	30.748
Girl	33.830	34.735	35.887
Goldhill	32.081	32.666	33.006
Lenna	33.822	34.957	35.631
Toys	30.145	30.575	32.801
<b>Average</b>	<b>31.700</b>	<b>32.398</b>	<b>33.242</b>



Some enlarged images of these comparative methods for the testing image “Lenna” are listed in Fig. 5. Here, the best results of LAZ, VQZ, the bilinear interpolation technique, and the proposed scheme are listed. It is hard to find out the difference between each enlarged image in Fig. 5 and its original image in Fig. 4.



**Fig. 5.** Some enlarged images of “Airplane”

## 5 Conclusions

An adaptive image zooming technique based on the image properties is introduced in this paper. The unknown pixels to be processed are classified into two categories: smooth pixels and the complex pixels. These smooth pixels are recovered by using the bilinear interpolation technique. The remaining complex pixels are rebuilt by using the vector quantization approximation technique.

From the experimental results, it is shown that the proposed scheme provides better image quality of the enlarged image compared to LAZ, VQZ and the bilinear interpolation technique. More than 0.4 dB image quality gain is achieved by using the proposed scheme compared to the bilinear interpolation technique. In addition, the required computational complexity of the proposed scheme is less than that of VQZ because the rough image is divided into non-overlapping image blocks instead of overlapping block division used in VQZ. Furthermore, the proposed scheme is much simpler than both LAZ and VQZ because only one controlling threshold is used. Recall that three and four controlling threshold values are used in LAZ and VQZ, respectively.

**Acknowledgments.** This research was supported by the National Science Council, Taipei, R.O.C. under contract NSC 98-2221-E-126-008 and NSC 99-2221-E-126-004-MY2.

## References

1. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall (2008)
2. Keys, R.G.: Cubic Convolution Interpolation for Digital Image Processing. *IEEE Transactions on Acoustics, Speech, Signal Processing* 29, 1153–1160 (1981)
3. Maeland, E.: On The Comparison of Interpolation Methods. *IEEE Transactions on Medical Imaging* 7(3), 213–217 (1988)
4. Jensen, K., Anastassiou, D.: Subpixel Edge Localization and the Interpolation of Still Images. *IEEE Transactions on Image Processing* 4(3), 285–295 (1995)
5. Hong, K.P., Paik, J.K., Kim, H.J., Lee, C.H.: An Edge-Preserving Image Interpolation System for a Digital Camcorder. *IEEE Transactions on Consumer Electronics* 42(3), 279–284 (1996)
6. Li, X., Orchard, M.: New Edge-Directed Interpolation. *IEEE Transactions on Image Processing* 10(10), 1521–1527 (2001)
7. Battiato, S., Gallo, G., Stanco, F.: A locally Adaptive Zooming Algorithm for Digital Images. *Image and Vision Computing* 20, 805–812 (2002)
8. Chang, C.C., Chou, Y.C., Yu, Y.H., Shih, K.J.: An Image Zooming Technique Based on Vector Quantization Approximation. *Image and Vision Computing* 23(13), 1214–1225 (2005)
9. Chughtai, M.A., Khattak, N.: An Edge Preserving Locally Adaptive Anti-aliasing Zooming Algorithm with Diffused Interpolation. In: 3rd Canadian Conference on Computer and Robot Vision, pp. 49–49 (2006)
10. Hong, S.H., Park, R.H., Yang, S., Kim, J.Y.: Image Interpolation Using Interpolative Classified Vector Quantization. *Image and Vision Computing* 26(2), 228–239 (2008)
11. Wang, G., Huang, X.L., Matiko, J.W.: An Improved Edge Preserving Smoothing Method (IEPS). In: 2010 International Conference on Optoelectronics and Image Processing, pp. 214–216 (2010)
12. Linde, Y., Buzo, A., Gray, R.M.: An Algorithm for Vector Quantizer Design. *IEEE Transactions on Communications* 28, 84–95 (1980)
13. Chang, C.C., Hu, Y.C.: A Fast Codebook Training Algorithm for Vector Quantization. *IEEE Transactions on Consumer Electronics* 44(4), 1201–1208 (1998)
14. Hu, Y.C., Chang, C.C.: An Effective Codebook Search Algorithm for Vector Quantization. *Imaging Science Journal* 51(4), 221–234 (2003)
15. Hu, Y.C., Su, B.H., Tsou, C.C.: Fast VQ Codebook Search Algorithm for Grayscale Image Coding. *Image and Vision Computing* 26(5), 657–666 (2008)
16. Hu, Y.C., Chuang, J.C., Lo, C.C., Lee, C.Y.: Efficient Grayscale Image Compression Technique Based on VQ. *Opto-Electronics Review* 19(1), 104–113 (2011)

# An Enhanced Fuzzy C-Means Clustering (ECFMC) Algorithm for Spot Segmentation

A. Sri Nagesh<sup>1</sup>, G.P. Saradhi Varma<sup>2</sup>, A. Govardhan<sup>3</sup>, and B. Raveendra Babu<sup>4</sup>

<sup>1</sup> CSE Dept., R.V.R. & J.C. College of Engg., Chowdavaram, Guntur, India

<sup>2</sup> Information Technology Department, S.R.K.R. Engg. College, Bhimavaram, India

<sup>3</sup> CSE Dept., JNTUCEH, Jagityal, Hyderabad, India

<sup>4</sup> Delta Technologies, Hyderabad, India

asrinagesh@gmail.com, gpsvarma@yahoo.com,

{govardhan\_cse}@yahoo.co.in, rbhogapathi@yahoo.com

**Abstract.** In this paper, we developed an efficient approach for the automated analysis of microarray images. The proposed approach consists of three steps: pre-processing, gridding and segmentation. Initially, the microarray images are pre-processed by hill climbing algorithm and gridding is applied to accurately identify the location of each spot while extracting spot intensities from the microarray images. In the segmentation stage, we developed an efficient Enhanced Fuzzy C-means (ECFMC) algorithm. The advantages of the proposed clustering algorithm compared with the conventional clustering techniques in segmentation of microarray image are given as: (1) it effectively detects the absent spot since the inclusion of neighborhood pixel information. (2) It is quite independent of the various noises (3) the detection of accurate spot can be done. By taking these advantages, the proposed approach effectively segmented the microarray images with preserving the image details. The proposed method is tested and evaluated using the Microarray image Database and the results confirmed the efficiency of the proposed approach.

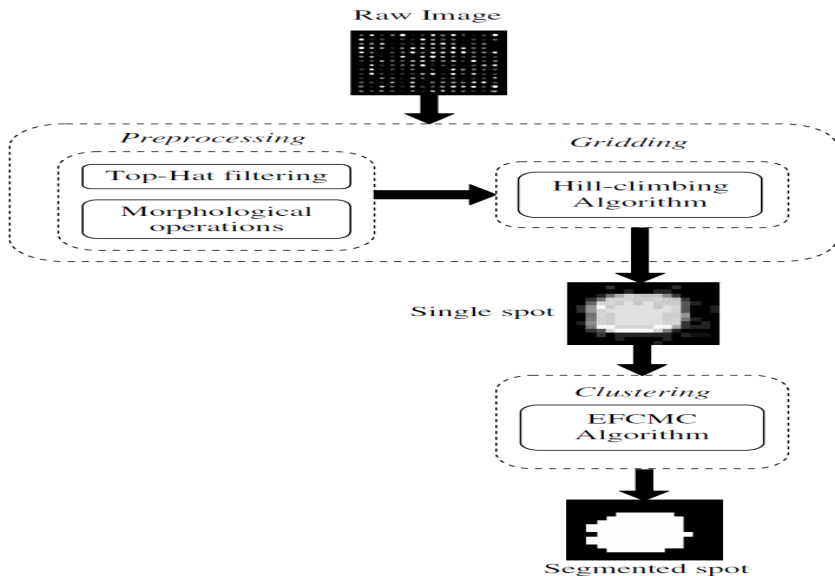
**Keywords:** Bioinformatics, Microarray images, Spot Segmentation, Threshold, Morphological Filtering, Enhanced Fuzzy C-Means Algorithm, (ECFMC), Clustering algorithms.

## 1 Introduction

Microarray image analysis is one of the tools, which are necessary when dealing with vast amounts of biological data [5]. Recently, a handful of researches are presented in the literature for segmentation of microarray images utilizing gridding and clustering methods. By looking into methods, there is an indispensable need to segment spots using proper clustering techniques in the microarray images. With the intention of this, we have developed an efficient clustering algorithm, which is designed based on the fuzzy c-means clustering [5] for spot segmentation. The proposed algorithm is used for the automated analysis of microarray images rather than the k-means [11],

fuzzy c-means [5] and hybrid clustering algorithms. The proposed algorithm integrates the neighborhood pixel information with the gray level information to cluster the microarray images in an efficient fuzzy manner and it can overcome the difficulties of the conventional fuzzy c-means clustering algorithm so that the segmentation performance of the microarray images can be improved significantly.

The advantages of the proposed clustering algorithm compared with the conventional clustering technique in segmentation of microarray image are given as: (1) it effectively detects the absent spot since the inclusion of neighborhood pixel information. (2) It is quite independent of the various noises and, (3) the detection of accurate spot can be highly possible. By taking these advantages, the proposed algorithm can effectively segment the spots presented in the microarray images with preserving the image details and without taking into account the noisy pixels. The proposed automated approach to microarray image segmentation consists of three steps: preprocessing, gridding and spot segmentation. Initially, the microarray images are preprocessed and gridding is applied to accurately identify the location of each spot while extracting spot intensities from the microarray images. In the spot segmentation stage, we have used the enhanced fuzzy c-means clustering algorithm (EFCMC) by taking into the account of neighborhood pixel information. The steps employed in the proposed approach are given in figure 1.



**Fig. 1.** Block diagram of the proposed EFCM Clustering approach

## 2 Review of Related Works

N. Giannakeas and D. Fotiadis [5] have proposed a method for the automated analysis of microarray images. Their proposed method has two stages; they are Gridding and Segmentation. First, template matching has been used to preprocess the micro array

Images, and block and spot identification has been performed. Then, a Voronoi diagram has been used to fit the grid on the image, subsequent to non-expressed spots detection. The segmentation stage has employed K-means and Fuzzy C Means (FCM) clustering. The Stanford Micro array Database (SMD) images have been used in the evaluation of the proposed method. The efficiency of their Fuzzy C means-based work has been shown to be comparable with that of the two already developed K-means-based methods by the obtained results. Their proposed fully automated method could deal with images and artifacts.

The organization of the paper is as follows: Section 2 presents a brief review of some recent significant researches in Microarray image segmentation. The proposed methodology for microarray image segmentation utilizing the enhanced fuzzy c-means clustering algorithm is explained in section 3. Experimental results and analysis of the proposed methodology are discussed in Section 4. Finally, concluding remarks are provided in Section 5.

### **3 Spot Segmentation Using the Enhanced Fuzzy C-Means Clustering Algorithm**

After applying the gridding process using hill-climbing algorithm, an area (cell) around each spot is marked and the marked cell region is extracted to apply the segmentation technique. Here, for each spot identified from the gridding, we have applied an enhanced fuzzy c-means clustering algorithm (EFCMC) to spot segmentation. The proposed algorithm is more appropriate for spot segmentation because it takes into consideration the neighborhood pixel in identifying membership values. Due to the fact, it can clearly identify the boundary values in between the clusters even if they are corrupted by the noises. The detailed discussion of the proposed clustering algorithm is explained in sub-section 3.3.1.

#### **3.1 (a). Enhanced Fuzzy c-Means Clustering Algorithm (EFCMC)**

Dunn first introduced the fuzzy c-means (FCM) clustering algorithm [5] and Bezdek extended it later [6]. Because traditional FCM algorithm does not take into account any information about spatial context, it is very sensitive to noise and other imaging artifacts, though it works properly on most noise-free images. In order to tackle these challenges, several clustering algorithms have been proposed by extending the fuzzy c-means clustering such as, FCM\_S1, FCM\_S2, EnFCM and its variations [7], [8] and [9]. By making use of the concept described in the fuzzy c-means clustering to microarray image segmentation, we have proposed an efficient fuzzy c-means clustering by additionally including the spatial information and the neighborhood dependent fuzzy factor. It can effectively handle the noisy pixels as well as the boundary pixels in between the two clusters for promoting the micro array image segmentation performance.

#### **3.2 (b). Introducing the Neighborhood-Dependent Fuzzy Factor**

In general, clustering can be done by the intensity values of images without taking into consideration of spatial information. It is necessary to include the spatial

information whenever the membership values computed for each pixel associated with cluster centers. The inclusion of spatial information to clustering process will lead to the better image segmentation even though the images are abrupt by the noises. By analyzing the behaviors of the traditional clustering algorithms, it looks more preferable if the decision should be based on the neighborhood pixels. In order to adapt this information into the clustering problem, we have introduced neighborhood dependent fuzzy factor which is included in the membership computation. Also, the dependent fuzzy factor makes the influence of the pixels within its surrounding to change flexibly along with their distance from the central pixel. The proposed clustering algorithm is based on the minimization of the objective function containing neighborhood dependent fuzzy factor given below.

$$G_m = \sum_{i=1}^M \sum_{k=1}^n \left[ w_{ki}^m * |d(y_i, c_k)|^2 + NF_{ki} \right] \tag{1}$$

Where,  $Y = \{y_1, y_2, \dots, y_M\} \subseteq \mathbf{R}^h$  is the data set in the  $h$ -dimensional vector space,  $M$  is the number of data items,  $n$  is the number of clusters,  $w_{ji}$  is the degree of membership of  $y_i$  in the  $k$ th cluster,  $m$  is the weighting exponent on each fuzzy membership varying between any real number greater than 1,  $c_k$  is the center of cluster  $k$ ,  $d(y_i, c_k)$  is a distance measure between data item  $y_i$  and cluster center  $c_k$ ,  $NF_{ki}$  is neighborhood fuzzy dependent fuzzy factor.

$$NF_{ki} = \sum_{\substack{j \in N_i \\ i \neq j}} \frac{1}{D_{ij}^2 + 1} (1 - w_{kj})^m |d(y_j, c_k)|^2 \tag{2}$$

Where,  $i$ th pixel is referred to the center pixel,  $k$  is the reference cluster,  $j$ th pixel is referred to the neighbors corresponding to the  $i$ th pixel,  $N_i$  is the total number of neighbors belongs to  $i$ th pixel.  $D_{i,j}$  is the spatial Euclidean distance between pixels  $i$  and  $j$ ,  $w_{kj}$  is the degree of membership of the  $j$ th pixel in the  $k$ th cluster,  $m$  is the weighting exponent on each fuzzy membership, and  $c_k$  is the center of cluster  $k$ .

**Steps proposed in the EFCMC algorithm:**

**Step 1.** Give a number of input cluster  $c$ , fuzzification parameter  $m$  and the termination threshold  $\beta$ .

**Step 2.** Randomly, initialize the fuzzy partition matrix  $w_{ki}$  and loop counter  $l = 0$ .

**Step 3.** Compute the cluster centers using the following equation.  $c_k = \frac{\sum_{i=1}^M w_{ki}^m y_i}{\sum_{i=1}^M w_{ki}^m}$  (3)

**Step 4.** Update the fuzzy partition matrix  $w_{ki}$  using the following equation.

$$w_{ki} = \frac{1}{\sum_{j=1}^n \left( \frac{|d(y_i, c_k)|^2 + NF_{ki}}{|d(y_i, c_j)|^2 + NF_{ji}} \right)^{1/m-1}} \tag{4}$$

**Step 5.** If  $\max\{W^{(l)} - W^{(l+1)}\} < \beta$ , stop the iteration; otherwise, set  $l = l + 1$  and go to step 3.

**Definition 1: (Distance Measure)**

The Euclidean distance ( $E$ ) between data point  $y$  and cluster centre  $c$  is the length of the line segment  $\overline{yc}$ . In Cartesian coordinates, if  $y = (y_1, y_2, \dots, y_l)$  and  $c = (c_1, c_2, \dots, c_l)$  are two points in Euclidean  $l$ -space, afterwards the distance from  $y$  to  $c$  is described by:

$$d(y, c) = \sqrt{(y_1 - c_1)^2 + (y_2 - c_2)^2 + \dots + (y_l - c_l)^2} = \sqrt{\sum_{i=1}^l (y_i - c_i)^2} \tag{5}$$

## 4 Results and Comparative Analysis

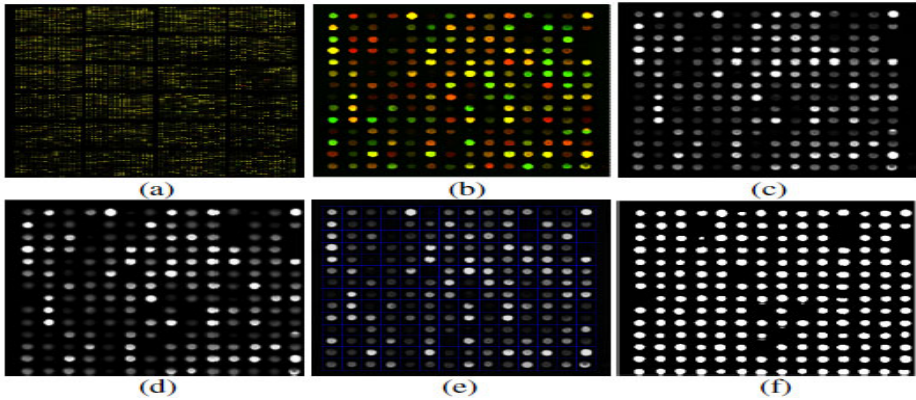
The proposed approach for microarray image segmentation has been implemented using MATLAB 7.10. The experimental results of the proposed approach for microarray image segmentation are explained in this section. Here, the performance of the proposed approach is compared with the traditional algorithms to evaluate the quality of the algorithm in microarray image segmentation.

### 4.1 (a). Experimental Dataset

The performance of the proposed approach is carried out in a set of real microarray images obtained from the publically available database [10]. The image taken from the database contains 24 blocks and each block contains 196 spots, i.e. 14x14 rows and columns of spots. Here, we have taken one block containing 196 spots from the real images and the experimentation is carried out on the extracted block.

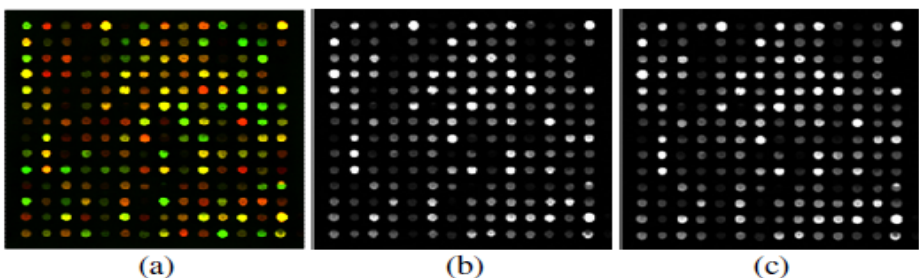
#### 4.2 (b). Segmentation Results

This section describes the segmentation results of the proposed approach, which is then compared with the results obtained by the k-means clustering and fuzzy c-means clustering algorithms described in [10, 1]. The overall segmentation results of the proposed approach are given in figure 2.



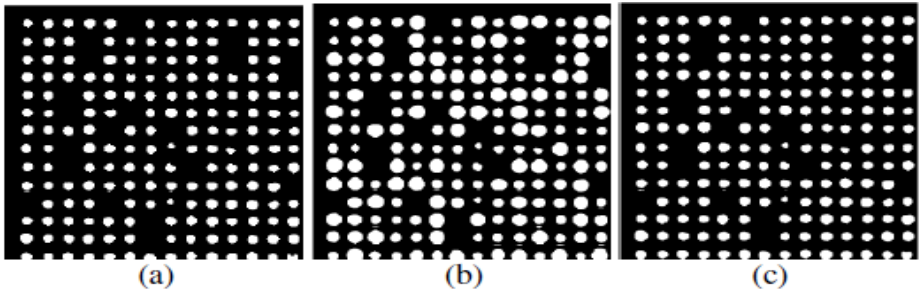
**Fig. 2.** (a) Microarray image from [10] (b) Example of microarray image used in the proposed approach (c) Gray scale image (d) preprocessed image (e) image obtained after gridding (f) Segmentation results obtained by the proposed approach

The input microarray image shown in fig. 3.a is given to the proposed approach which converts it into gray scale image of red channel as well as green channel. Then, the entire steps are executed on this image and finally we obtained the segmentation results as shown in fig. 4.c and fig. 5.c. The same input image is given to the k-means clustering as well as fuzzy c-means clustering and the results obtained are shown in the following figures. By analyzing three algorithms, the efficiency of detecting low intensity spots as well as identifying the absent spots have been significantly improved for our proposed algorithm as per figures shown below.

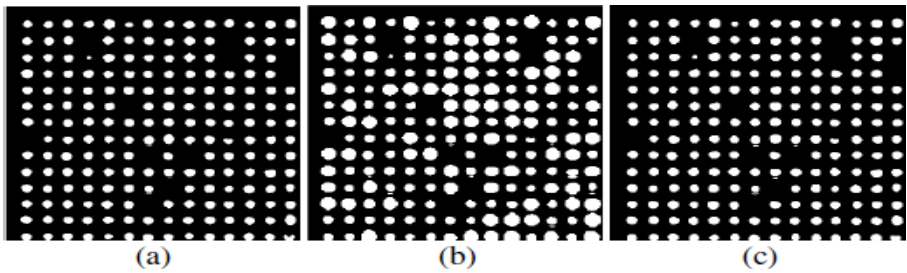


**Fig. 3.** (a) Color microarray image (b) Gray scale image- red channel (c) Gray scale image- green channel





**Fig. 4.** Segmentation results of green channel input (a) k-means clustering (b) FCM clustering (c) EFCMC



**Fig. 5.** Segmentation results of red channel input (a) k-means clustering (b) FCM clustering (c) EFCMC

## 5 Conclusion

We have developed an efficient approach to automated analysis of microarray images by making use of the proposed clustering algorithm (EFCMC). The proposed approach consists of three important stages: preprocessing, gridding and segmentation. The preprocessed microarray image was directly applied to hill-climbing algorithm, which is adapted to fit a spot on the image. Then, the segmentation was carried out utilizing the proposed EFCM clustering algorithm, which is an efficient algorithm for clustering. Therefore, the experimentation was carried out using the real microarray images to find the efficiency of the segmentation

**Acknowledgments.** The authors would thank all the authors and people who have directly or indirectly contributed to this paper and we are thankful to them for the help rendered for implementing this paper.

## References

- [1] Wu, H., Yan, H.: Microarray Image processing Based on Clustering and Morphological Analysis. In: Proc. of the First Asia Pacific Bioinformatics Conference, Adelaide, Australia, pp. 111–118 (2003)

- [2] Microarray Images, <http://llmpp.nih.gov/lymphoma/data/rawdata/>
- [3] Sri Nagesh, A., Varma, G.P.S., Govardhan, A.: An Improved Iterative Watershed and Morphological Transformation Techniques for Segmentation of Microarray Images. In: IJCA Special Issue on Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications, CASCT (2010)
- [4] Unsharp Filter, <http://homepages.inf.ed.ac.uk/rbf/HIPR2/unsharp.htm>
- [5] Giannakeas, N., Fotiadis, D.: An automated method for gridding and clustering-based segmentation of cDNA microarray images. *Computerized Medical Imaging and Graphics* 33(1), 40–49 (2009)
- [6] Bezdek, J.: *Pattern Recognition with Fuzzy Objective Function Algorithms*. Plenum, New York (1981)
- [7] Ahmed, M., Yamany, S., Mohamed, N., Farag, A., Moriarty, T.: A modified fuzzy C-means algorithm for bias field estimation and segmentation of MRI data. *IEEE Transaction on Medical Imaging* 21(3), 193–199 (2002)
- [8] Chen, S., Zhang, D.: Robust image segmentation using FCM with spatial constraints based on new kernel-induced distance measure. *IEEE Transactions on Systems, Man, and Cybernetics* 34(4), 1907–1916 (2004)
- [9] Wang, Y., Ma, M.Q., Zhang, K., Shih, F.Y.: A hierarchical refinement algorithm for fully automatic gridding in spotted DNA microarray image processing. *Information Sciences: an International Journal* 177(4), 1123–1135 (2007)
- [10] Heyer, L.: *MicroArray Genome Imaging & Clustering (MAGIC) Tool*, Davidson College, <http://www.bio.davidson.edu/projects/magic/magic.html>
- [11] Ergüt, E., Yardimci, Y., Mumcuoglu, E., Konu, O.: Analysis of microarray images using FCM and K-means clustering algorithm. In: *Proceedings of International Conference on Signal Processing*, pp. 116–121 (2003)
- [12] Sri Nagesh, A., Govardhan, A., Varma, G.P.S., Prasad, G.S.: An Automated Histogram Equalized Fuzzy Clustering based Approach for the Segmentation of Microarray images. *ANU Journal of Engineering and Technology* 2(2), 42–48 (2010) ISSN: 0976-3414
- [13] Suresh, K., Kundu, D., Ghosh, S., Das, S., Abraham, A., Han, S.Y.: Multi-Objective Differential Evolution for Automatic Clustering with Application to Micro-Array Data Analysis. *Sensors* 9, 3981–4004 (2009)
- [14] Wang, Y., Ma, M.Q., Zhang, K., Shih, F.Y.: A hierarchical refinement algorithm for fully automatic gridding in spotted DNA microarray image processing. *Information Sciences: an International Journal* 177(4), 1123–1135 (2007)

# On Intuitionistic Fuzzy T-ideals in TM-Algebra

Megalai Kandasamy<sup>1</sup> and Tamilarasi Angamuthu<sup>2</sup>

<sup>1</sup> Department of Mathematics, Bannari Amman Institute of Technology,  
Sathyamangalam-638 401, Tamil Nadu, India.

megaguna@rediffmail.com

<sup>2</sup> Department of Computer Science, Kongu Engineering College,  
Perundurai, Tamil Nadu, India.

angamuthu\_tamilarasi@yahoo.co.in

**Abstract.** In this paper, first we define the notions of TM-algebra, T-ideals, fuzzy sets, intuitionistic fuzzy T-ideals and intuitionistic fuzzy closed T-ideals. Using the concept of level subsets, we prove some theorems which show that there are some relationships between these notions. Finally we define the homomorphism of TM-algebras and then we give related theorem about the relationship between their images and intuitionistic fuzzy T-ideals.

**Keywords:** TM-algebra, T-ideal, Fuzzy subset, Intuitionistic fuzzy ideals, Intuitionistic fuzzy T-ideals, Intuitionistic fuzzy closed T- ideals.

## 1 Introduction

The concept of fuzzy set was introduced [1] in 1965. Several researchers explored on the generalization of the notion of fuzzy sets. The concept of intuitionistic fuzzy set was introduced [2], as a generalization of the notion of fuzzy sets in 1986. Since then the literature on these concepts has been growing rapidly.

In 1996, two classes of abstract algebras, BCK-algebras and BCI-algebras have been introduced. We introduce a notion of TM-algebra[3], which is a generalization of Q/BCK/BCI/ BCH algebras. In this paper, we introduce the notion of Intuitionistic Fuzzy T-ideals in TM-algebra and investigate some properties.

## 2 Preliminaries

### 2.1 Definition

A TM-algebra  $(X, *, 0)$  is a non-empty set  $X$  with a constant “0” and a binary operation “\* ” satisfying the following axioms:

- (i)  $x * 0 = x$
- (ii)  $(x * y) * (x * z) = z * y$ , for any  $x, y, z \in X$ .

In  $X$  we can define a binary relation  $\leq$  by  $x \leq y$  if and only if  $x * y = 0$ .

**2.2 Definition**

Let  $(X, *, 0)$  be a TM-algebra. A non-empty set  $I$  of  $X$  is called an ideal of  $X$  if it satisfies

- (i)  $0 \in I$
- (ii)  $x * y \in I$  and  $y \in I$  imply  $x \in I$ , for all  $x, y \in X$ .

**2.3 Definition**

An ideal  $A$  of a TM-algebra  $X$  is said to be closed if  $0 * x \in A$  for all  $x \in A$ .

**2.4 Definition**

Let  $X$  be a non-empty set. A mapping  $\mu : X \rightarrow [0,1]$  is called a fuzzy set in  $X$ . The complement of  $\mu$ , denoted by  $\bar{\mu}(x) = 1 - \mu(x)$ , for all  $x \in X$ .

**2.5 Definition**

An intuitionistic fuzzy set  $A$  in a non empty set  $X$  is an object having the form  $A = \{ (X, \mu_A(x), \lambda_A(x) / x \in X \}$ , where the function  $\mu_A : X \rightarrow [0,1]$  and  $\lambda_A : X \rightarrow [0,1]$  denote the degree of membership (namely  $\mu_A(x)$ ) and the degree of non-membership (namely  $\lambda_A(x)$ ) of each element  $x \in X$  to the set  $A$  respectively, and  $0 \leq \mu_A(x) + \lambda_A(x) \leq 1$ , for all  $x \in X$ .

For the sake of simplicity, we use the symbol  $A = (X, \mu_A, \lambda_A)$  for the intuitionistic fuzzy set  $A = \{ (X, \mu_A(x), \lambda_A(x) / x \in X \}$

**2.6 Definition**

Let  $A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy set in  $X$ . Then (i)  $\square A = (X, \mu_A, \bar{\mu}_A)$  and (ii)  $\diamond A = (X, \bar{\lambda}_A, \lambda_A)$ .

**3 Intuitionistic Fuzzy T-ideal**

**3.1 Definition**

A non-empty set  $I$  of a TM-algebra  $X$  is called T-ideal of  $X$ , if

- (i)  $0 \in I$
- (ii)  $(x * y) * z \in I$  and  $y \in I$  imply that  $x * z \in I$ , for all  $x, y, z \in X$ .

**3.2 Definition**

A fuzzy subset  $\mu$  in a TM-algebra  $X$  is called a fuzzy T-ideal of  $X$ , if

- (i)  $\mu(0) \geq \mu(x)$
- (ii)  $\mu(x * z) \geq \min\{\mu((x * y) * z), \mu(y)\}$ , for all  $x, y, z \in X$ .

**3.3 Definition**

An intuitionistic fuzzy set  $A = (X, \mu_A, \lambda_A)$  in a TM-algebra  $X$  is called an intuitionistic fuzzy T-ideal of  $X$  if,

- (i)  $\mu_A(0) \geq \mu_A(x)$  and  $\lambda_A(0) \leq \lambda_A(x)$
- (ii)  $\mu_A(x * z) \geq \min\{\mu_A((x * y) * z), \mu_A(y)\}$
- (iii)  $\lambda_A(x * z) \leq \max\{\lambda_A((x * y) * z), \lambda_A(y)\}$ , for all  $x, y, z \in X$ .

**3.4 Definition**

An intuitionistic fuzzy set  $A = (X, \mu_A, \lambda_A)$  in a TM-algebra  $X$  is called an intuitionistic fuzzy closed T-ideal of  $X$ , if it satisfies

- (i)  $\mu_A(0 * x) \geq \mu_A(x)$  and  $\lambda_A(0 * x) \leq \lambda_A(x)$
- (ii)  $\mu_A(x * z) \geq \min\{\mu_A((x * y) * z), \mu_A(y)\}$
- (iii)  $\lambda_A(x * z) \leq \max\{\lambda_A((x * y) * z), \lambda_A(y)\}$ , for all  $x, y, z \in X$ .

**3.5 Definition**

Let  $A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy set in a TM-algebra  $X$ . The set

$$U(\mu_A; s) = \{x \in X / \mu_A(x) \geq s\}$$

is called an upper  $s$ -level of  $\mu_A$  and the set

$$L(\lambda_A; t) = \{x \in X / \lambda_A(x) \leq t\}$$

is called lower  $t$ -level of  $\lambda_A$ .

**3.6 Theorem**

Let  $A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy T-ideal (IF T-ideal) of a TM-algebra  $X$ . Then so is  $\square A = (X, \mu_A, \overline{\mu_A})$ .

*Proof*

Since  $A$  is an IF T-ideal of  $X$ ,

$$\mu_A(0) \geq \mu_A(x) \text{ and } \mu_A(x * z) \geq \min\{\mu_A((x * y) * z), \mu_A(y)\}.$$

Now,  $\mu_A(0) \geq \mu_A(x)$

$$\begin{aligned} \Rightarrow 1 - \overline{\mu}_A(0) &\geq 1 - \overline{\mu}_A(x) \\ \Rightarrow \overline{\mu}_A(0) &\leq \overline{\mu}_A(x), \text{ for any } x \in X \text{ and} \\ \mu_A(x * z) &\geq \min\{\mu_A((x * y) * z), \mu_A(y)\}. \\ \Rightarrow 1 - \overline{\mu}_A(x * z) &\geq \min\{1 - \overline{\mu}_A((x * y) * z), 1 - \overline{\mu}_A(y)\} \\ \Rightarrow \overline{\mu}_A(x * z) &\leq 1 - \min\{1 - \overline{\mu}_A((x * y) * z), 1 - \overline{\mu}_A(y)\} \\ \Rightarrow &\leq \max\{\overline{\mu}_A((x * y) * z), \overline{\mu}_A(y)\} \end{aligned}$$

Hence  $\square A = (X, \mu_A, \overline{\mu}_A)$  is an IF T-ideal of  $X$ .

**3.7 Theorem**

Let  $A = (X, \mu_A, \lambda_A)$  be an IF T-ideal of a TM-algebra  $X$ . Then so is  $\diamond A = (X, \overline{\lambda}_A, \lambda_A)$ .

*Proof*

$$\begin{aligned} \text{We have } \lambda_A(0) \leq \lambda_A(x) &\Rightarrow 1 - \overline{\lambda}_A(0) \leq 1 - \overline{\lambda}_A(x) \\ \Rightarrow \overline{\lambda}_A(0) &\geq \overline{\lambda}_A(x), \text{ for any } x \in X. \\ \text{Now, for any } x, y, z \in X & \\ \lambda_A(x * z) \leq \max\{\lambda_A((x * y) * z), \lambda_A(y)\} & \\ \Rightarrow 1 - \overline{\lambda}_A(x * z) \leq \max\{1 - \overline{\lambda}_A((x * y) * z), 1 - \overline{\lambda}_A(y)\} & \\ \Rightarrow \overline{\lambda}_A(x * z) \geq 1 - \max\{1 - \overline{\lambda}_A((x * y) * z), 1 - \overline{\lambda}_A(y)\} & \\ \Rightarrow &= \min\{\overline{\lambda}_A((x * y) * z), \overline{\lambda}_A(y)\} \\ \Rightarrow \overline{\lambda}_A(x * z) \geq \min\{\overline{\lambda}_A((x * y) * z), \overline{\lambda}_A(y)\} & \end{aligned}$$

Hence,  $\diamond A = (X, \overline{\lambda}_A, \lambda_A)$  is an IF T-ideal of  $X$ .

**3.8 Theorem**

If  $A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy closed T-ideal of a TM-algebra  $X$ , then so is  $\square A = (X, \mu_A, \overline{\mu}_A)$ .

*Proof*

$$\begin{aligned} \text{For any } x \in X, \mu_A(0 * x) &\geq \mu_A(x) \\ \Rightarrow 1 - \overline{\mu}_A(0 * x) &\geq 1 - \overline{\mu}_A(x) \\ \Rightarrow \overline{\mu}_A(0 * x) &\leq \overline{\mu}_A(x), \text{ for any } x \in X. \end{aligned}$$

Hence  $\square A = (X, \mu_A, \overline{\mu}_A)$  is an IF closed T-ideal of  $X$ .

**3.9 Theorem**

If  $A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy closed T-ideal of a TM-algebra X, then so is  $\diamond A = (X, \bar{\lambda}_A, \lambda_A)$ .

*Proof*

We have  $\lambda_A(0 * x) \leq \lambda_A(x) \Rightarrow 1 - \bar{\lambda}_A(0 * x) \leq 1 - \bar{\lambda}_A(x)$   
 $\Rightarrow \bar{\lambda}_A(0 * x) \geq \bar{\lambda}_A(x)$ , for any  $x \in X$ .

Hence,  $\diamond A = (X, \bar{\lambda}_A, \lambda_A)$  is an IF closed T-ideal of X.

**3.10 Theorem**

$A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy closed T-ideal of a TM-algebra X if and only if  $\square A, \diamond A$  are IF closed T-ideals of a TM-algebra X.

**3.11 Theorem**

$A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy T-ideal of a TM-algebra X if and only if the non-empty upper s-level cut  $U(\mu_A; s)$  and the non-empty lower t-level cut  $L(\lambda_A, t)$  are T-ideals of X, for any  $s, t \in [0, 1]$ .

*Proof*

Suppose  $A = (X, \mu_A, \lambda_A)$  is an intuitionistic fuzzy T-ideal of a TM-algebra X.

For any  $s, t \in [0, 1]$  define the sets  $U(\mu_A; s) = \{ x \in X / \mu_A(x) \geq s \}$

and  $L(\lambda_A, t) = \{ x \in X / \lambda_A(x) \leq t \}$ .

Since  $U(\mu_A; s) \neq \emptyset$ , for  $x \in U(\mu_A; s) \Rightarrow \mu_A(x) \geq s$

$$\Rightarrow \mu_A(0) \geq \mu_A(x) \geq s$$

$$\Rightarrow \mu_A(0) \geq s$$

$$\Rightarrow 0 \in U(\mu_A; s)$$

Let  $(x * y) * z \in U(\lambda_A, s)$  and  $y \in U(\lambda_A, s)$

$$\Rightarrow \mu_A((x * y) * z) \geq s \text{ and } \mu_A(y) \geq s$$

Since  $\mu_A(x * z) \geq \min \{ \mu_A((x * y) * z), \mu_A(y) \} \geq \min \{ s, s \} = s$ .

Thus  $\mu_A(x * z) \geq s$

$$\Rightarrow x * z \in U(\lambda_A; s)$$

Hence  $U(\mu_A; s)$  is a T-ideal of X.

Similarly we can prove that  $L(\lambda_A, t)$  is a T-ideal of X.

Conversely,

Suppose that for any  $s, t \in [0, 1]$ ,  $U(\mu_A; s)$  and  $L(\lambda_A, t)$  are T-ideals of X,

If possible, assume  $x_0, y_0 \in X$  such that  $\mu_A(0) < \mu_A(x_0)$

and  $\lambda_A(0) > \lambda_A(y_0)$ .

Put  $s_0 = \frac{1}{2} [\mu_A(0) + \mu_A(x_0)]$

$\Rightarrow s_0 < \mu_A(x_0)$  and  $0 \leq \mu_A(0) < s_0 < 1$

$\Rightarrow x_0 \in U(\mu_A; s_0)$  and  $0 \notin U(\mu_A; s_0)$ .

Since  $U(\mu_A; s_0)$  is a T-ideal of  $X$ , we have  $0 \in U(\mu_A; s_0)$  and  $U(\mu_A; s_0) \geq s_0$

Therefore our assumption is wrong.

Hence,  $\mu_A(0) \geq \mu_A(x)$ , for all  $x \in X$ .

Similarly, by taking  $t_0 = \frac{1}{2} [\lambda_A(0) + \lambda_A(y)]$  we can show that

$\lambda_A(0) \leq \lambda_A(y)$  for all  $y \in X$ .

If possible, assume that  $x_0, y_0, z_0 \in X$  such that

$\mu_A(x_0 * z_0) < \min\{\mu_A((x_0 * y_0) * z_0), \mu_A(y_0)\}$

Take  $s_0 = \frac{1}{2} [\mu_A(x_0 * z_0) + \min\{\mu_A((x_0 * y_0) * z_0), \mu_A(y_0)\}]$

$\Rightarrow s_0 > \mu_A(x_0 * z_0)$  and  $s_0 < \min\{\mu_A((x_0 * y_0) * z_0), \mu_A(y_0)\}$

$\Rightarrow s_0 > \mu_A(x_0 * z_0), s_0 < \mu_A((x_0 * y_0) * z_0)$  and  $s_0 < \mu_A(y_0)$

$\Rightarrow (x_0 * z_0) \notin U(\mu_A; s_0)$ .

Since  $U(\mu_A; s_0)$  is a T-ideal of  $X$ ,  $((x_0 * y_0) * z_0) \in U(\mu_A; s_0)$  and  $y_0 \in U(\mu_A; s_0)$  imply that  $x_0 * z \in U(\mu_A; s_0)$ , a contradiction to the assumption.

Therefore  $\mu_A(x * z) \geq \min\{\mu_A((x * y) * z), \mu_A(y)\}$ , for any  $x, y, z \in X$ .

Similarly we can prove,  $\lambda_A(x * z) \leq \max\{\lambda_A((x * y) * z), \lambda_A(y)\}$  for all  $x, y, z \in X$ .

Hence  $A$  is an IF T-ideals of a TM-algebra  $X$ .

**3.12 Theorem**

$A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy closed T-ideal of a TM-algebra  $X$  if and only if the non-empty upper  $s$ -level cut  $U(\mu_A; s)$  and the non-empty lower  $t$ -level cut  $L(\lambda_A, t)$  are closed T-ideals of  $X$ , for any  $s, t \in [0, 1]$ .

*Proof*

Suppose  $A = (X, \mu_A, \lambda_A)$  be an intuitionistic fuzzy closed T-ideal of a TM-algebra  $X$ .

$\Rightarrow \mu_A(0 * x) \geq \mu_A(x)$  and  $\lambda_A(0 * x) \leq \lambda_A(x)$  for any  $x \in X$ .

For any  $x \in U(\mu_A; s)$  implies that  $\mu_A(x) \geq s$ .

$\Rightarrow \mu_A(0 * x) \geq s$

$\Rightarrow 0 * x \in U(\mu_A; s)$  and

$x \in L(\lambda_A, t)$  implies that  $\lambda_A(x) \leq t$

$\Rightarrow 0 * x \in L(\lambda_A, t)$ .



Hence  $U(\mu_A; s)$  and  $L(\lambda_A, t)$  are closed T-ideals of  $X$ , for any  $s, t \in [0, 1]$ .

Conversely,  $U(\mu_A; s)$  and  $L(\lambda_A, t)$  are closed T-ideals of  $X$ .

To show that  $A = (X, \mu_A, \lambda_A)$  is an IF closed T-ideal of  $X$ , it is enough to show that  $\mu_A(0 * x) \geq \mu_A(x)$  and  $\lambda_A(0 * x) \leq \lambda_A(x)$  for any  $x \in X$ .

If possible assume that there exist some  $x_0 \in X$  such that  $\mu_A(0 * x_0) < \mu_A(x_0)$ .

$$\text{Take } s_0 = \frac{1}{2} [\mu_A(0 * x_0) + \mu_A(x_0)]$$

$$\Rightarrow \mu_A(0 * x_0) < s_0 < \mu_A(x_0)$$

$\Rightarrow x_0 \in U(\mu_A; s_0)$ , but  $0 * x \notin U(\mu_A; s_0)$ , a contradiction to the definition of T-ideal.

Hence  $\mu_A(0 * x) \geq \mu_A(x)$  for any  $x \in X$ .

Similarly we can prove that  $\lambda_A(0 * x) \leq \lambda_A(x)$  for any  $x \in X$ .

**3.13 Definition**

Let  $f$  be a mapping on a set  $X$  and  $A = (X, \mu_A, \lambda_A)$  an IF set in  $X$ . Then the fuzzy sets  $u$  and  $v$  on  $f(X)$  defined by  $u(y) = \sup_{x \in f^{-1}(y)} \mu_A(x)$  and  $v(y) = \inf_{x \in f^{-1}(y)} \lambda_A(x)$  for all  $y \in f(X)$ , is called the image of  $A$  under  $f$ . If  $u, v$  are fuzzy sets in  $f(X)$  then the fuzzy sets  $\mu_A = u \circ f$  and  $\lambda_A = v \circ f$  are called the pre-images of  $u$  and  $v$  respectively under  $f$ .

**3.14 Definition**

A function  $f : X \rightarrow X$  is said to be a homomorphism of TM-algebras if  $f(x * y) = f(x) * f(y)$ .

**3.15 Theorem**

Let  $f : X \rightarrow X'$  be an on to homomorphism of TM-algebras. If  $A' = (X', u, v)$  is an IF T-ideal of  $X'$  then the pre-image of  $A'$  under  $f$  is an IF T-ideal of  $X$ .

*Proof*

Let  $A = (X, \mu_A, \lambda_A)$  where  $\mu_A = u \circ f$  and  $\lambda_A = v \circ f$  is the pre-image of  $A' = (X', u, v)$  under  $f$ .

Since  $A' = (X', u, v)$  is an IF T-ideal of  $X'$ , we have

$$u(0') \geq u(f(x)) = u \circ f(x) = \mu_A(x) \text{ and } v(0') \leq v(f(x)) = v \circ f(x) = \lambda_A(x).$$

$$\text{On the other hand } u(0') = u(f(0)) = u \circ f(0) = \mu_A(0) \text{ and } v(0') = v(f(0)) = v \circ f(0) = \lambda_A(0)$$

Therefore  $\mu_A(0) = u(0') \geq \mu_A(x)$  and  $\lambda_A(0) = v(0') \leq \lambda_A(x)$  for all  $x$ .

Now, we show that

$$\begin{aligned} \mu_A(x * z) &\geq \min \{ \mu_A((x * y) * z), \mu_A(y) \} \text{ and} \\ \lambda_A(x * z) &\leq \max \{ \lambda_A((x * y) * z), \lambda_A(y) \}, \text{ for all } x, y, z \in X. \\ \text{We have } \mu_A(x * z) &= u \circ f(x * z) = u(f(x * z)) = u(f(x) * f(z)) \\ &\geq \min\{ u((f(x) * f(y)) * f(z)), u(f(y)) \} \\ &= \min\{ u(f(x * y) * f(z)), u(f(y)) \} \\ &= \min \{ u(f((x * y) * z)), u(f(y)) \} \\ &= \min\{ u \circ f((x * y) * z), u \circ f(y) \} \\ &= \min\{ \mu_A((x * y) * z), \mu_A(y), \text{ for all } x, y, z \in X. \end{aligned}$$

Hence, the result  $\mu_A(x * z) \geq \min\{ \mu_A((x * y) * z), \mu_A(y) \}$  is true for all  $x, y, z \in X$ .

Similarly, we can prove  $\lambda_A(x * z) \leq \max \{ \lambda_A((x * y) * z), \lambda_A(y) \}$ , for all  $x, y, z \in X$ .

Hence the pre image  $A = (X, \mu_A, \lambda_A)$  of  $A'$  is an IF T-ideal of  $X$ .

### References

1. Zadeh, L.A.: Fuzzy Sets. Information Control 8, 338–353 (1965)
2. Atanassov, K.T.: Intuitionistic fuzzy sets. Fuzzy sets and Systems 20, 87–96 (1986)
3. Megalai, K., Tamilarasi, A.: TM-algebra-An Introduction. International Journal of Computer Application, Special Issue on Computer Aided soft Computing Techniques for imaging and Biomedical Application 1, 17–23 (2010)
4. Atanassov, K.T.: New operations defined over the intuitionistic fuzzy sets. Fuzzy Sets and Systems 61, 137–142 (1994)
5. Dudek, W.A., Young Bae, J.: Fuzzification of ideals in BCC-algebras. Glasnik Matematicki 36(56), 127–138 (2001)
6. Isaki, K., Shotoro: An Introduction to the theory of BCK-algebras. Math. Japon 23, 1–26 (1978)
7. Jun, Y.B., Kim, K.H.: Intuitionistic fuzzy ideals of BCK-algebras. International Journal of Mathematics and Mathematical Sciences 24, 839–849 (2000)
8. Satyanarayana, B., Bindu Madavi, U., Durga Prasad, R.: On Intuitionistic fuzzy H-ideals in BCK-algebras. International Journal of Algebra 4(15), 743–749 (2010)

# Mitigating Congestion and Improving the Performance of Wireless Sensor Networks

S. Raj Barath<sup>1,\*</sup>, C. Kezi Selva Vijila<sup>1</sup>, and A. Jaya Prakash<sup>2</sup>

<sup>1</sup> Department of ECE

<sup>2</sup> Department of IT,

Christian College of Engineering and Technology,  
Oddanchatram, Tamil Nadu 624619, India  
rajbarathece@gmail.com

**Abstract.** In wireless sensor networks (WSNs), congestion will cause packet loss which in turn wastes energy and reduces the lifetime of WSNs, Successful event detection in Wireless Sensor Networks (WSN) requires reliability and timeliness. When an event occurs, the base station (BS) is particularly interested about reliable and timely collection of data sent by the nodes close to the event, and the data sent by other nodes have little importance. Data generated in wireless sensor networks may not all be alike: some data may be more important than others and hence may have different delivery requirements. In this paper, we address differentiated data delivery in the presence of congestion in wireless sensor networks. We propose a class of algorithms that enforce differentiated routing based on the congested areas of a network and data priority. The basic protocol, called Congestion-Aware Routing (CAR), discovers the congested zone of the network that exists between high- priority data sources and the data sink and, using simple forwarding rules, dedicates this portion of the network to forwarding primarily high-priority traffic.

**Keywords:** Congestion control, Wireless Sensor Networks, Congestion Aware Routing(CAR), Conzone.

## 1 Introduction

WSNs are undoubtedly one of the largest growing types of networks today. They are fast becoming one of the largest growing types of networks today and, as such, have attracted quite a bit of research interest. They are used in many aspects of our lives including environmental analysis and monitoring, battlefield surveillance and management, emergency response, medical monitoring and inventory management. These networks also play a significant role in areas like agriculture and industries as well. Their reliability, cost effectiveness, ease of deployment and ability to operate in an unattended environment, among other positive characteristics, make sensor networks the leading choice of networks for these applications.

---

\* Corresponding author.

A sensor network normally constitutes a wireless ad-hoc network, meaning that each sensor supports a multihop routing algorithm. Congestion in network will lead to the following problems:

It can lead to indiscriminate dropping of data. Some packets of high priority might be dropped while others of less priority are delivered.

Congestion can cause an increase in energy consumption as links become saturated.

## **2 Existing Protocols**

### **2.1 AODV Protocol**

The AODV routing protocol is a reactive routing protocol; therefore, routes are determined only when needed. Hello messages may be used to detect and monitor links to neighbours. If Hello messages are used, each active node periodically broadcasts a Hello message that all its neighbours receive. Because nodes periodically send Hello messages, if a node fails to receive several Hello messages from a neighbour, a link break is detected. When a source has data to transmit to an unknown destination, it broadcasts a Route Request (RREQ) for that destination. At each intermediate node, when a RREQ is received a route to the source is created. If the receiving node has not received this RREQ before, is not the destination and does not have a current route to the destination, it rebroadcasts the RREQ. If the receiving node is the destination or has a current route to the destination, it generates a Route Reply (RREP). The RREP is unicast in a hop-by hop fashion to the source. Control messages are route request route reply and Hello message.

### **2.2 Dynamic Source Routing (DSR)**

Dynamic Source Routing (DSR) also belongs to the class of reactive protocols and allows nodes to dynamically discover a route across multiple network hops to any destination. Source routing means that each packet in its header carries the complete ordered list of nodes through which the packet must pass. DSR uses no periodic routing messages (e.g. no router advertisements), thereby reducing network bandwidth overhead, conserving battery power and avoiding large routing updates throughout the ad-hoc network. Instead DSR relies on support from the MAC layer (the MAC layer should inform the routing protocol about link failures). The two basic modes of operation in DSR are route discovery and route maintenance.

### **2.3 Performance Comparison of Routing Protocols**

First, by virtue of source routing, DSR has access to a significantly greater amount of routing information than AODV. Second, to make use of route caching aggressively, DSR replies to all requests reaching a destination from a single request cycle. The main difference between both protocols is that in DSR a source routing option is used;

i.e. when a node wants to send something to a destination it sets the whole route for that packet, indicating the addresses of the terminals it has to pass through. In this sense all packets have a DSR header included, and it is needed that all nodes within the ad hoc network know the whole network topology. On the other hand, AODV does not perform source routing at all; when a terminal wants to send something to a destination, it checks its routing table, looking for the next hop towards that destination, and sends the packet to it, and so on[1]. In this sense, data packets "travel" through the ad hoc network without any AODV specific information. AODV, however, outperforms DSR in more stressful situations, with widening performance gaps with increasing stress .

### 3 Mannasim

Mannasim is a Wireless Sensor Networks simulation environment. The Mannasim Framework is a module for WSN simulation based on the Network Simulator (NS-2). Mannasim extends NS-2 introducing new modules for design, development and analysis of different WSN applications. The Script Generator Tool (SGT) is a front-end for TCL simulation scripts easy creation. SGT comes blunded with Mannasim Framework and it's written in pure Java making it plataform independent.

Mannasim goal is to develop a detailed simulation framework, which can accurately model different sensor nodes and applications while providing a versatile testbed for algorithms and protocols. Numerous challenges make the study of real deployed sensor networks very difficult and financially infeasible. At the current stage of the technology, a practical way to study WSNs is through simulations that can provide a meaningful perspective of the behavior and performance of various algorithms.

#### 3.1 Program Code

In our Paper we have considered a temperature data generator tool which enables us to create virtual sensor nodes which act as temperature detectors. The temperature data generator used is based on the following coding .,

```
proc create_temp_data_generator
{sensing_interval sensing_type avg_measure std_deviation}

{
    set temp_gen_ [new DataGenerator/Temperature Gene-
rator ]

    $temp_gen_ set sensing_interval_ $sensing_interval

    $temp_gen_ set sensing_type_ $sensing_type
    $temp_gen_ set avg_measure $avg_measure
```

```

$temp_gen_ set std_deviation $std_deviation

return $temp_gen_
}

```

**Table 1.** Nodes sensing temperature data

Node	Temperature Data	Sensed Time
27	23.163195	12.354383
26	25.263736	12.402511
35	26.454916	12.468805
21	23.207405	12.475237
37	22.9250542	12.490212
36	24.004726	12.522567
41	22.173151	12.525178
24	22.173151	12.525665
34	22.907966	12.527280
40	24.004726	12.545992
22	22.173151	12.556836
30	22.907966	12.599142
12	25.905568	12.615997
20	24.484277	12.622640
09	24.187366	12.692209

## 4 Congestion Aware Routing (CAR)

CAR protocol is designed to provide differentiated routing for packets based on their priority. High priority packets are routed inside a zone known as- conzone using CAR protocol and the low priority packets are routed using the existing AODV protocol.

The CAR protocol design has two phases

Design of conzone and

Routing packets inside the conzone.

The present design comprises of the following steps:

- Access the routing table
- Include priority to the critical nodes
- Access the neighbour table
- Build Conzone using the RREQ broadcast

CAR provides differentiated routing based on priority of the nodes. High priority packets that are generated from the critical nodes are routed inside the conzone

while the low priority packets are routed outside the conzone using the AODV protocol [10]. After the conzone is formed the routes from the critical nodes to the sink are updated in the routing table and the high priority packets will use this new route to the sink. The intermediate nodes which are on the conzone also updates its route to the sink. The nodes are routed and handled using a spanning tree approach.

## 5 Implementation Details

The present routing algorithm is implemented using NS2. The proposed routing approach is implemented with changes in the scripts as follows:

### (i) Routing Table

Every node has its own routing table which stores the information about the various routes. For each destination in the table the corresponding next hop, total hops, and the expiry time are specified. As the expiry time elapses the route gets deleted or updated. In ns2 the routing table can be accessed at any time from the tcl script and the routing updation can be studied.

(ii) A new parameter called “priority” has been added to the routing table in `aodv_rtable` file to specify the priority to critical nodes. The critical nodes are assigned a priority of 1 while the other nodes are assigned 0.

### (iii) Neighbour Table

Every node has its own neighbour table which stores all its neighbour ids. The neighbour table is updated using the HELLO broadcast.

### (iv) Conzone formation

The conzone is built in `recvRequest(..)` function in `aodv.cc` file. The critical node ids are checked and the function to build conzone is called. When all parents for a level is entered, the value in `id_no` array is changed to the new set of values. When the level is 3, the parents of current level nodes are statically assigned the id of the sink node.

### (v) Conzone Routing

The high priority packets are routed inside the conzone. First, the `sendReply(..)` function in `aodv.cc` file is modified inorder to block the reply to the critical nodes. After the formation of conzone the routing table of each node is accessed from tcl file and checked if its on conzone node.

(vi) If the node is on conzone then another function is called in `rtable.cc` which has `index`, `corresponding parent` and `level` as its parameters. This function updates the route entries for the conzone nodes.

## 6 Results and Discussions

The CAR protocol was analyzed in ns-2 and simulated in Network Animator (NAM) for an 8 \* 15 grid with the sink as Node 121. Here the sink is node 121 and critical nodes are node 5, node 6, node 7, node 8 and node 9.

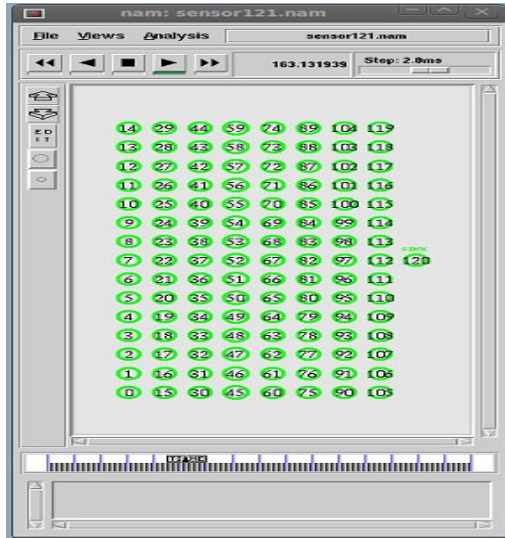


Fig. 1. Network Scenario

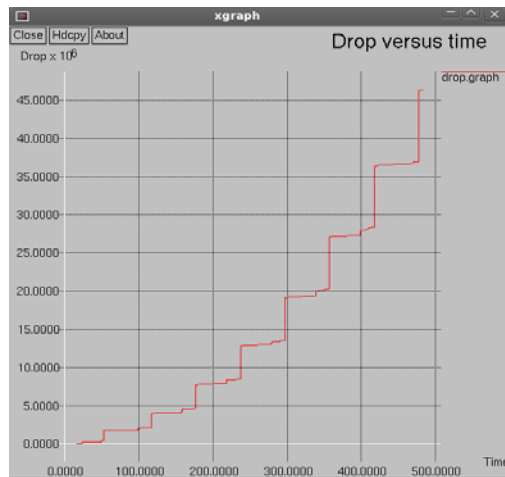


Fig. 2. Packet Drop Versus Time in AODV



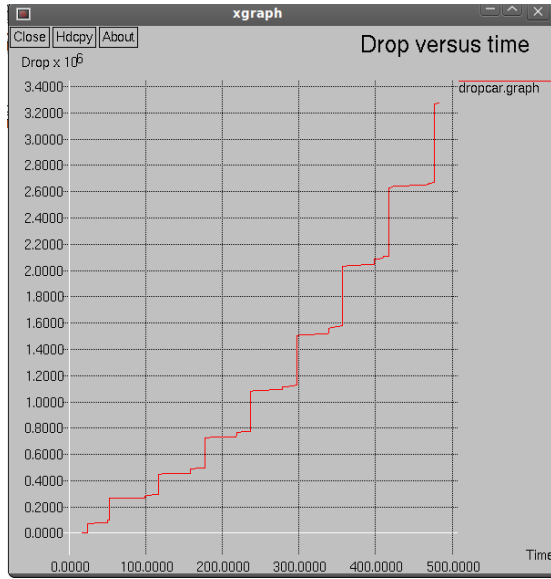


Fig. 3. Packet Drop Versus Time in CAR

From these Graphs we can see that the Packet Drop has been very much reduced with the use of Congestion Aware Routing (CAR) . In CAR the dropped Packets are those which are of the low priority where as the high priority data are routed through the congested zone and they are delivered successfully.

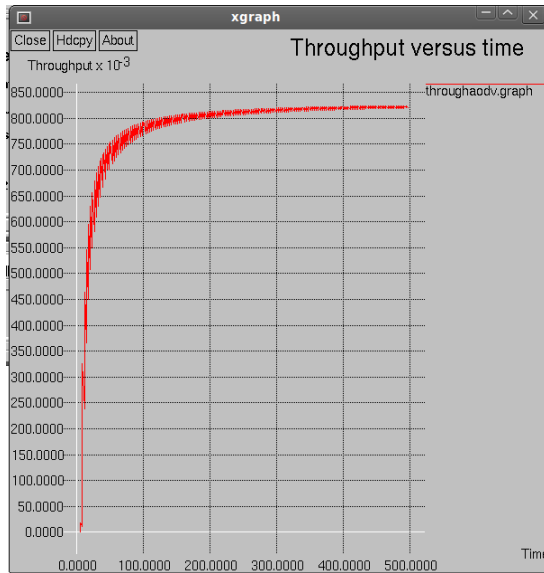
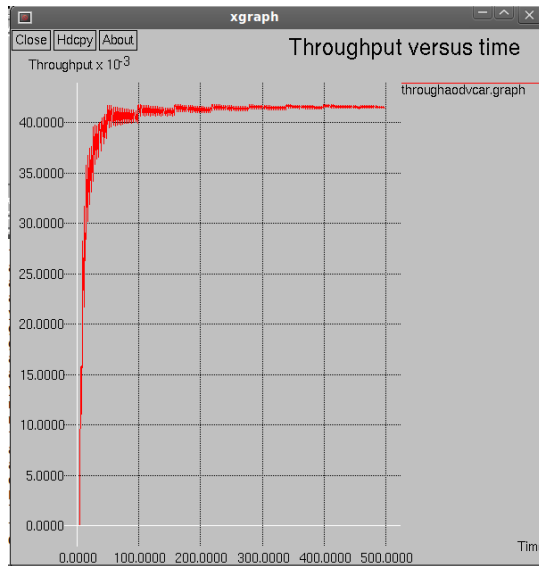
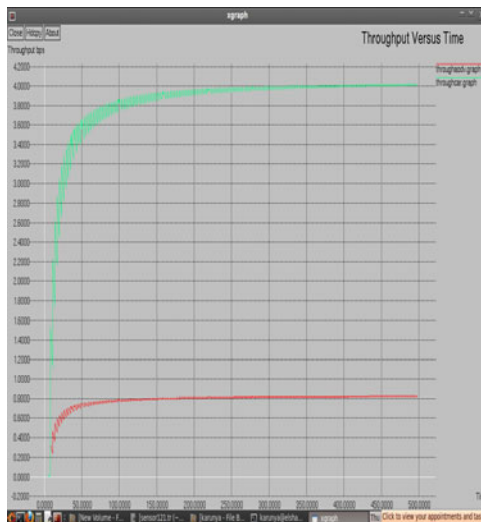


Fig. 4. Throughput Versus Time in AODV



**Fig. 5.** Throughput Versus Time in AODV

In communication networks, throughput or network throughput is the average rate of successful message delivery over a communication channel. This data may be delivered over a physical or logical link, or pass through a certain network node. The throughput is usually measured in bits per second (bit/s or bps), and sometimes in data packets per second or data packets per time slot.



**Fig. 6.** Comparison of Network Throughputs

Figure 6 shows that the throughput under congestion aware routing is much higher when compared to the normal AODV routing. Here in CAR, the nodes requests are handled as they arrive and leave the zone. The node relationships are connected in form of a spanning tree. The path length for reaching the destination node using the spanning tree structure. The performance is improved in proposed design .

## 7 Conclusion

In the present work data delivery issues in the presence of congestion in wireless sensor networks is addressed. Congestion Aware Routing (CAR) is a routing protocol that uses data prioritization and treats packets according to their priorities. We defined a conzone as the set of sensors that will be required to route high priority packets from the data sources to the sink. Our solution does not require active queue management, maintenance of multiple queues or scheduling algorithms, or the use of specialized MAC protocols. Our extensive simulations show that as compared to AODV, CAR increases the fraction of high priority data delivery, decreases delay and jitter for such delivery while using energy uniformly in the deployment. By discovering the required conzone and using differentiated routing we can free the conzone from most of the low priority traffic traveling through the network. This will help nodes on the conzone to provide better service to high priority data.

## References

1. Yin, X., Zhou, X., Huang, R.: A Fairness-Aware Congestion Control Scheme in Wireless Sensor Networks. *IEEE Transactions on Vehicular Technology* (November 2009)
2. Sheu, J.-P., Hu, W.-K.: Hybrid Congestion Control Protocol in Wireless Sensor Networks. In: *IEEE Vehicular Technology Conference(VTC 2008 – Spring)* (2008)
3. Li, Z., Liu, P.X.: Priority-based Congestion Control in Multi-path and Multi-hop Wireless Sensor Networks. In: *IEEE International Conference on Robotics and Biomimetics, ROBIO 2007* (2007)
4. Bhuiyan, M.M., Gondal, I., Kamruzzaman, J.: CAM - Congestion Avoidance and Mitigation in Wireless Sensor Networks. In: *IEEE Vehicular Technology Conference (VTC 2010 – Spring)* (May 2010)
5. Tao, L., Yuecoda, F.: Enhanced Congestion Detection and Avoidance for Multiple Class of Traffic in Sensor Networks. In: *15th IEEE Asia Pacific Conference on Communications – APCC* (October 2009)
6. Ee, C.T., Bajcsy, R.: Congestion Control and Fairness for Many-to-One Routing in Sensor Networks. In: *Proc. Second ACM Conf. Embedded Networked Sensor Systems (SenSys 2004)*, pp. 148–161 (2004)
7. Felemban, E., Lee, C.-G., Ekici, E.: MMSPEED: Multipath Multi-SPEED Protocol for QoS Guarantee of Reliability and Timeliness in Wireless Sensor Networks. *IEEE Trans. Mobile Computing* 6, 738–754 (2006)
8. He, T., Stankovic, J.A., Lu, C., Abdelzaher, T.: Speed: A Stateless Protocol for Real-Time Communication in Sensor Networks. In: *Proc. 23rd IEEE Int'l Conf. Distributed Computing Systems, ICDCS* (2003)

9. Hill, J., Szewczyk, R., Woo, A., Hollar, S., Culler, D., Pister, K.: System Architecture Directions for Network Sensors. In: Proc. Ninth Int'l Conf. Architectural Support for Programming Languages and Operating Systems (ASPLOS 2000) (November 2000)
10. Hull, B., Jamieson, K., Balakrishnan, H.: Mitigating Congestion in Wireless Sensor Networks. In: Proc. Second ACM Conf. Embedded Networked Sensor Systems, SenSys (2004)
11. Madden, S., Franklin, M., Hellerstein, J., Hong, W.: Tag: A Tiny Aggregation Service for Ad-Hoc Sensor Networks. In: Proc. Fifth Symp. Operating System Design and Implementation, OSDI (2002)
12. Intanagonwiwat, C., Govindan, R., Estrin, D.: Directed Diffusion: A Scalable and Robust Communication Paradigm for Sensor Networks. In: Proc. ACM MobiCom 2000 (August 2000)
13. Johnson, D.B., Maltz, D.A.: Dynamic Source Routing in Ad Hoc Wireless Networks, pp. 153–181. Kluwer Academic Publishers (February 1996)
14. Karp, B., Kung, H.: GPSR: Greedy Perimeter Stateless Routing for Wireless Networks. In: Proc. ACM MobiCom (2000)
15. Lu, C., Blum, B., Abdelzaher, T., Stankovic, J., He, T.: RAP: A Real-Time Communication Architecture for Large-Scale Wireless Sensor Networks. In: Proc. Eighth IEEE Real-Time and Embedded Technology and Applications Symp (RTAS 2002), pp. 55–66 (2002)

# Robust Key Points Matching by Ordinal Measure

S. Lakshmi<sup>1</sup> and V. Sankaranarayanan<sup>2</sup>

<sup>1</sup> Jeppiaar Engineering College, Chennai, India

<sup>2</sup> Crescent Engineering College, Director (Rtd.), RCC, Anna University, Chennai, India

**Abstract.** Key point matching is based on finding reliable corresponding in the images. We propose ordinal based measure of image matching. The normalized cross correlation is one of the most accepted techniques for matching key points to facilitate the comparison of images. In the literature a number of fast implementation algorithms are available, there still are ways to improve expedite for many applications. We propose a general frame work for ordinal base image correspondence. In this paper, we introduce a new and effective method for matching the key points before applying cross correlation between neighbor key points that has been adapted to scale and rotation. Several examples are presented and the measure is evaluated on a set of test images.

**Keywords:** Interest point matching Normalized cross correlation, Ordinal measure, Rank correlation, Spearman's Rho, Kendall's tau, similarity measure of images.

## 1 Introduction

There are several methods used in the field of image matching which is a part of prevailing area of image processing in which the given image is matching with the same key points to several group images acquired from the same scene. Matching points between images is very important for object recognition, tracking objects and change detection. Normalized cross correlation is the widely used technique to acquire distance measurements [1]. Correlation is basically used to find the similarity between the image points' neighborhood. A combined corner edge detector developed by Harris make the correspondence between images using various types of filters[2]. The RANSAC was introduced by Fischler and Bolles in 1981. It is commonly used to estimate the fundamental matrix from a set of feature point pairs determined by a correlation [3][4]. It results in a high computing.

The KLT Tracker is one of the most popular methods used to track feature points from one image to the other in robotic applications using vision to self localize. In this method, the feature points are extracted with a SUSAN detector only once, the extracted points are tracked by using pyramidal approach.[5]. It reduces the complexity of processing significantly but mismatches will occur. Traditional image matching measures such as correlation or squared Euclidean distance, which are essentially based on pixel intensity value, fail to face the outlier problem.[6].

The ordinal measures operate on the ranks of pixels rather than directly on the pixel values. This measure is based on the ranking of one image with respect to the ranks of the other. Their measure is robust in the presence of outliers.[7]. The Kendall's  $\tau$  and Spearman's  $\rho$  are the two recognized ordinal measures used in psychological measurements.[8][9].

In this paper a correlation measure is employed to match the images. Experiments are conducted in real images and proved the superiority of the Kendall’s tau rank correlation.

The rest of the paper is organized as follows: In section 2 a brief review of rank correlation techniques is given and explained the procedure for calculating the Kendall’s tau and spearman’s row correlation. Section 3 describes the overall system and Section 4 explains the algorithm. Experimental results are presented in Section 5 and the comparison results are given in Section 6. Finally, the conclusions are drawn in Section 7.

## 2 Review of Rank Correlation Techniques

Linear correlation measures are invariant under positive linear conversions of intensity values. Hence, Spearman’s rho and Kendall’s tau rank correlations can be used to solve the similarity measures.

The ordinal measures operate on the ranks of pixels rather than directly on the pixel values. Thus, only relative ordering between data values is of consequence in determining the distance or correlation between two images. We propose a general framework for performing ordinal-based image correspondence. This framework contains well-known Kendall’s tau and Spearman’s rho measures as special cases. Additionally, it allows one to design other correspondence measures that can incorporate region-based spatial information [10].

### 2.1 Strength of Ordinal (Ranked) Measure Relationships

The two most common ordinal measures of association are Spearman’s rho ( $\rho$ ) and Kendall’s rank order coefficient ( $\tau$ ). Both Spearman’s rho ( $\rho$ ) and Kendall’s rank order coefficient ( $\tau$ ) require that the two variables, X and Y, are paired observations, with the variables measured are at the ordinal level. Both the measures can range between -1.0 and +1.0 when a positive correlation indicating that the ranks increase together, while a negative correlation indicates that as the rank of one variable increases the other one decreases.

### 2.2 Procedure for Computing Spearman’s Rho Rank Correlation Coefficient ( $\rho$ )

It is a measure of association in which the intensity values are converted to ranks before calculating the coefficient. A ranking which represents the relative ordering between values of an ordinal variable is simply a permutation of integers. Let  $I_1$  be the given image and  $I_2$  be the input image or group image. Ordinally invariant ie., rank, measures of association for bivariate ( a pair of two variable) populations are discussed.

	110	121	134
$I_1 =$	124	163	129
	117	135	143
	186	155	121
$I_2 =$	177	182	115
	165	149	133

Firstly, rank matrices  $\pi_i$  are calculated considering pixel values.

$I_1$  can be written as (110,121,134,124,163,129,117,135,143)

$I_2 = (186,155,121,177,182,115,165,149,133)$ .

Let  $\pi_1^i$  be the rank of  $I_1^i$  among the  $I_1$  data,  $\pi_2^i$  be the rank of  $I_2^i$  among the  $I_2$  data. Then the differences on rank matrix between the compared windows are calculated.

Secondly, these changes in order are weighted, according to the intensity variation which has caused the difference.

$$\begin{aligned} \pi_1 &= (1, 3, 6, 4, 9, 5, 2, 7, 8) \\ \pi_2 &= (9, 5, 2, 7, 8, 1, 6, 4, 3) \end{aligned}$$

To define correlation between two rankings  $\pi_1, \pi_2$ , we require a measure of closeness- a distance metric between them.

$$\rho = 1 - \frac{6 \sum D^2}{N(N^2 - 1)}$$

where:

D = the difference between the ranks of corresponding values  $\pi_1$  and  $\pi_2$  and  
 N = the number of pairs of values.

In the above example, the difference between the ranks for the 3x3 window ie., 9 ranks are  $\pi_1 - \pi_2$ .

$$D = (-8, -2, 4, -3, 1, 4, -4, 3, 5)$$

Squaring and summing these values:

$$\begin{aligned} \sum D^2 &= 64+4+16+9+1+16+16+9+25 \\ &= 160 \end{aligned}$$

Computing the denominator:

$$\begin{aligned} N(N^2-1) &= 9(9^2-1) \\ &= 9(81-1) \\ &= 720 \end{aligned}$$

And finally the Spearman's rho

$$\begin{aligned} \rho &= 1 - [(6*160)/720] \\ &= 1 - (960/720) \\ &= -0.333. \end{aligned}$$

The Spearman's rho ( $\rho$ ) of -0.333 would be considered a moderately strong positive correlation.

### 2.3 Procedure for Computing Kendall's Tau ( $\tau$ )

The Kendall's tau is an alternative measure of correlation and so measures the strength of the relationship between two variables. Kendall's tau, like Spearman's rank correlation (Crichton, 1999), is carried out on the ranks of the data. That is, for each variable separately the values are put in order and numbered, 1 for the lowest value, 2 for the next lowest and so on. Conover (1980) gives details of how to calculate Kendall's tau. In common with other measures of correlation Kendall's tau will take values between  $\pm 1$  and  $+1$ , with a positive correlation indicating that the ranks of both variables increase together whilst a negative correlation indicates that as the rank of one variable increases the other one decreases. In almost all situations the values of Spearman's rank correlation and Kendall's tau are very close and would invariably lead to the same conclusions. The main advantage of using Kendall's tau over Spearman's rho is that one can interpret its value as a direct measure of the probabilities of observing concordant and discordant pairs.

Let  $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$  be a set of joint observations from two random variables of 2 images for measuring the similarity of the key points such that all the values of  $(x_i)$  and  $(y_i)$  are unique.

Any pair of observations  $(x_i, y_i)$  and  $(x_j, y_j)$  are said to be *concordant* if the ranks for both elements agree: that is, if both  $x_i > x_j$  and  $y_i > y_j$  or if both  $x_i < x_j$  and  $y_i < y_j$ . They are said to be *discordant*, if  $x_i > x_j$  and  $y_i < y_j$  or if  $x_i < x_j$  and  $y_i > y_j$ . If  $x_i = x_j$  or  $y_i = y_j$ , the pair is neither concordant nor discordant.

The Kendall's tau coefficient is defined as

$$\tau = \frac{C - D}{1/2n(n - 1)}$$

where C = Number of concordant pairs

D = Number of discordant pairs and

n = number of elements

Kendall's tau can also be computed using the formula

$$\tau = \frac{2P}{1/2n(n - 1)} - 1$$

where P is the sum, over all the cases, of cases ranked after the given item by both rankings, and n is the number of paired items.

If  $X_i = X_j$ , or  $Y_i = Y_j$  or both, the comparison is called a 'tie'. Ties are not counted as concordant or discordant. If there are tied i.e., same value then  $\tau$  can be calculated using the formula

$$\tau = \frac{C - D}{\sqrt{C + D + T} \sqrt{C + D + U}}$$

where C is the number of concordant samples, D is the number of discordant samples, T is the number of ties in variable  $V_1$  and U is the number of ties in variable  $V_2$ .

Using the same data as we employed to compute Spearman's rho, we note that the paired observations are sorted in order of the intensity values of the image  $I_1$ , so we will compute P based on the intensity values of the image  $I_2$ .

$I_1$  can be written as (110,121,134,124,163,129,117,135,143)

$I_2 = (186,155,121,177,182,115,165,149,133)$ .

Let  $\pi_1^i$  be the rank of  $I_1^i$  among the  $I_1$  data,  $\pi_2^i$  be the rank of  $I_2^i$  among the  $I_2$  data.

$$\pi_1 = (1, 3, 6, 4, 9, 5, 2, 7, 8)$$

$$\pi_2 = (9, 5, 2, 7, 8, 1, 6, 4, 3)$$

A composite permutation S can be defined as follows:

$$S^i = \pi_2^k, k = (\pi_1^{-1})^i$$

where  $\pi_1^{-1}$  denotes the inverse permutation of  $\pi_1$ . The inverse permutation is defined as follows. If  $\pi_1^i = j$ , then  $(\pi_1^{-1})^j = i$ . Informally, s is the ranking of  $I_2$  with respect to  $I_1$ . Under perfect positive correlation, should be identical to the identity permutation given by  $u = (1,2,\dots,n)$ . [6].

$$S = (9 \quad 1 \quad 2 \quad 7 \quad 3 \quad 8 \quad 5 \quad 6 \quad 4)$$

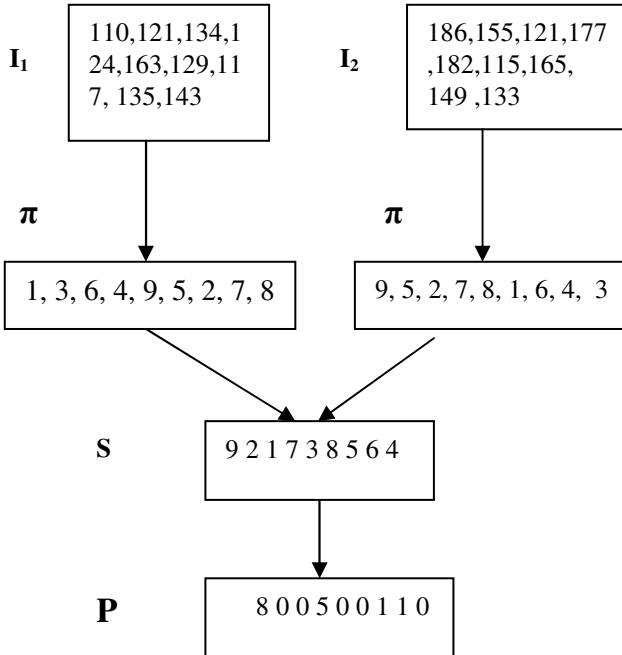


$$P = 8+0+0+5+0+0+1+1+0 = 15$$

$$2P = 2 * 15 = 30$$

$$\begin{aligned} 1/2n(n-1) &= 1/2 * 9(9-1) \\ &= 1/2 * 9 * 8 = 36 \end{aligned}$$

$$\begin{aligned} \tau &= 30/36 - 1 \\ &= 5/6 - 1 \\ &= -1/6 = -0.166 \end{aligned}$$



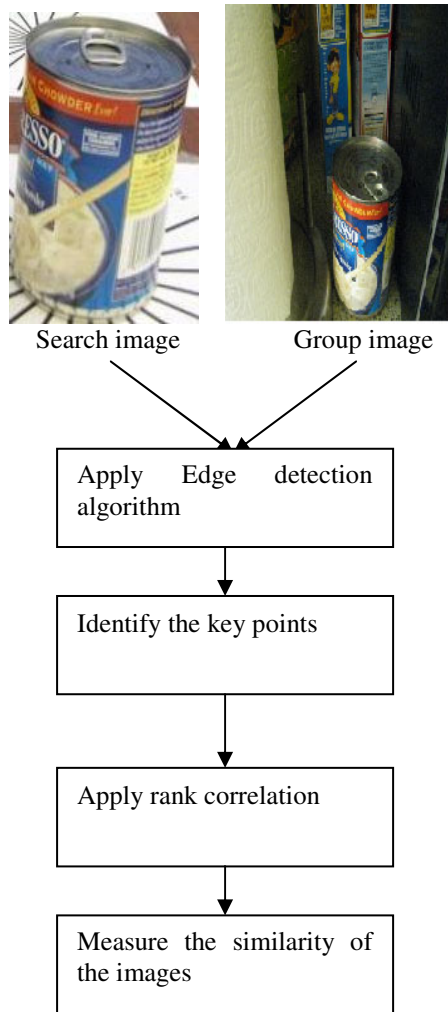
The ordinal measure normally gives a high similarity value which implies a right match. However, obtaining false positives with high correlation value is usual. This is caused by the images homogeneity.

### 3 System Overview

Basic steps involved in the process are given in figure1. As shown, input image and group image is taken from the static camera. For both images, we have to detect the edges and mark the key points. [12].After reducing the number of key points, apply the ordinal measure that is the rank correlation coefficient such as Spearman’s Rho and Kendall’s tau to measure the similarity of the key points.

#### 3.1 The Computation of Rotational Invariant Gradient

It is a highly complicated and also difficult task to match the key points if they are in different orientation and transformation. Traditional similarity measure based on correlation is not invariant to image rotation. We detect the key points K and extract the



**Fig. 1.** Framework of the System

neighborhood points around the key point  $K$  in the given image so that the formation of the neighborhood intensity values will be reliable under various viewing conditions.

Let  $K$  be the key point and  $K_i$  be any one of the sample points around  $K$ . Let us establish the  $x - y$  coordinate system by  $K$  and  $K_i$  for each sample point as in Fig(2), where  $\overrightarrow{KK_i}$  is defined as the positive  $y$ - axis. Such a coordinate system is perceptibly rotational invariant. The gradient is also invariant in this system. So, the invariant gradient can be calculated as follows:

$$D_x(K_i) = I(K_i^1) - I(K_i^3)$$

$$D_y(K_i) = I(K_i^2) - I(K_i^4)$$

Where  $K_i^j$ ,  $j = 1, 2, 3, 4$  are  $K_i$ 's neighboring points along the  $x$ -axis and  $y$ -axis in the local  $x$ - $y$  coordinate system and  $I(K_i^j)$  stands for the intensity at  $K_i^j$ .

Let us assume that  $K_1$  be the center of a key point with scale  $s_1$ , radius  $r_1$  and orientation  $\omega_1$  of the first image  $F_1$ , then the vector  $\vec{be}_{K_1b}$  and  $K_2$  be the center of a key point with scale  $s_2$ , radius  $r_2$  and orientation  $\omega_2$  of the query image  $F_2$ , then the vector  $\vec{be}_{K_2d}$ . Let  $W_1$  be the window size of  $S_1 = \text{round}(\alpha \times r_1)$  with centered  $K_1$  and  $W_2$  be the window size of  $S_2 = \text{round}(\alpha \times r_2)$  with centered  $K_2$ . We have to take the gradient of the sample points randomly from the circle in which the key point centered. Apply rank correlation to those sample points for measuring the similarity by using correlation.

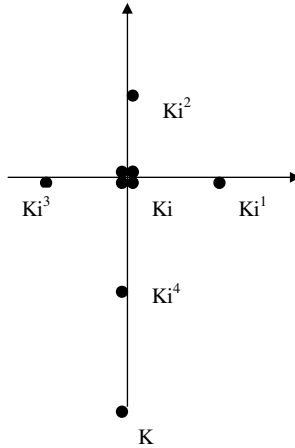


Fig. 2. X-Y co-ordinate system

## 4 Algorithm

Basic algorithm steps for matching the key points are given below.

- 1 Read the detected edge image with the marked key points of the search image as well as group image.
- 2 Divide the group image into 10 blocks.
- 3 Mark the key points
- 4 Extract the intensity values of all extracted key points of size  $3 \times 3$ .
- 5 Now find out the ordinal measure of the extracted key points for the given image ( $I_1$ ) as well as each and every part of the group image ( $I_2$ ).
- 6 Calculate the rank correlation coefficients Spearman's rho and Kendall's tau.
- 7 Based on the correlation value we match the similar key points of the given images.
- 8 Repeat the same procedure for all key points of the given image as well as group image.

## 5 Experimental Results

The algorithm have been tested and implemented on windows XP platform using Java and also the algorithm has been analyzed on a variety of indoor and outdoor environments. The proposed approach has been tested on different images with rotation, scale and view angle changes.

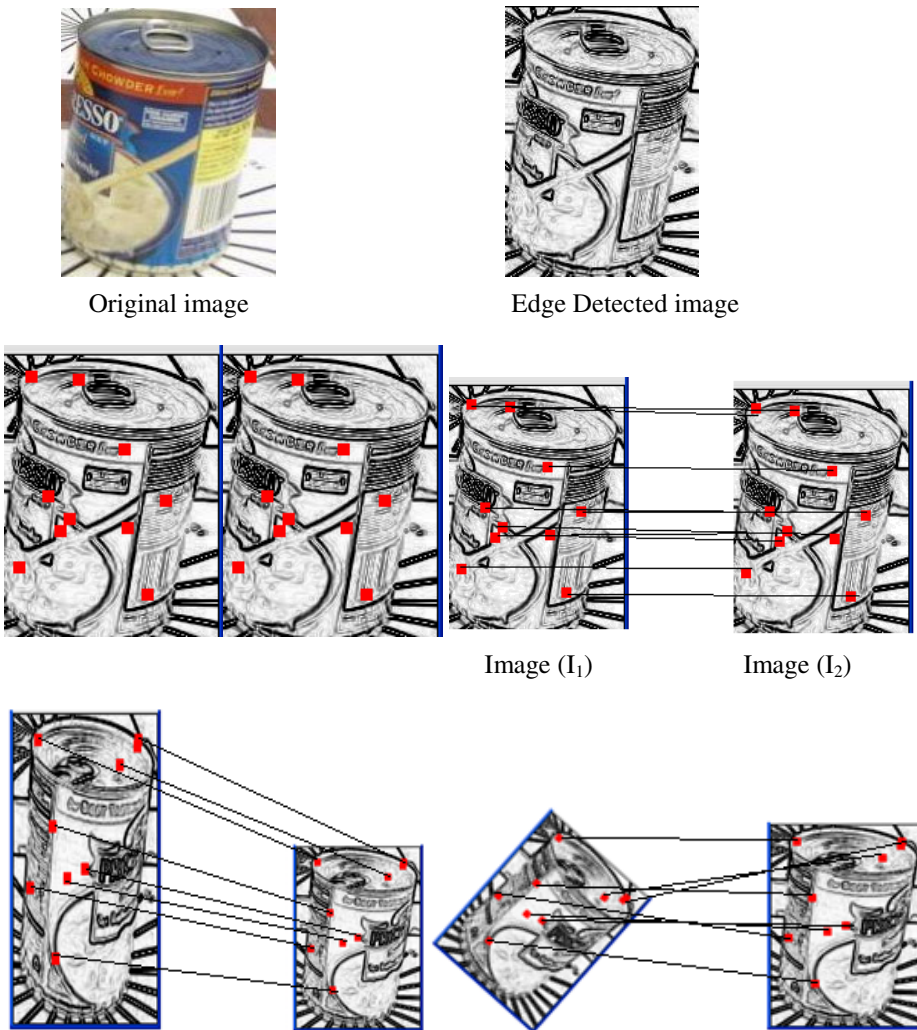


Fig. 3. Key points matching

### 5.1 Performance Evaluation

Key points matching results are shown in figure -3. Performance evaluation for key points matching is given in terms of two metrics for characterizing the Detection Rate(DR) and the False Alarm Rate[FAR] of the system. The key points are matched or mismatched. The key points are matched in image I1 with the image I2.i.e., true positive(TR).Points of Image I1 are not matched with the points of Image I2.i.e., false negative or misses(FN). Points of image I1 are matched with the points are not belonging to the points of image I2. ie., false positive(FP).

$$DR = TP/(TP+FN)$$

$$FAR = FP/(TP+FP)$$

The values of TP, FN and FP for different image blocks are shown in table 1.

**Table 1.** TP,FP and FN of different image blocks

Image Blocks	TP	FP	FN
B1	8	2	1
B2	5   7	2	0   1
B3	15	3	9
B4	33	2	1
B5	16	2	5
B6	18	2	2

From the table -1 we can calculate DR and FAR easily by using the formula.

**Table 2.** DR and FAR for different image blocks

Image Blocks	Detection Rate	False Alarm Rate	
B1	0.89	0.2	
B2	1   0.875	0.286	0.222
B3	0.625	0.1666	
B4	0.970	0.057	
B5	0.762	0.111	
B6	0.9	0.1	

## 6 Comparison with Other Methods

Selected key points are matched by using intensity augmented ordinal measure technique. The ordinal measure gives accurate matches with some false positive.

**Table 3.** Comparison with other methods

Method	Time (ms)
Nearest Neighborhood search (NNS)	25
Approximate Nearest Neighborhood search (using Kd tree structure) (ANN)	20
Best – Bin – First (BBF)	20
Point to Point (P2P) (one to one symmetric (OOS))	35
RANSAC (Outlier rejection) (Random Sample Consensus)	30
K means based tree	26
Proposed Method	20

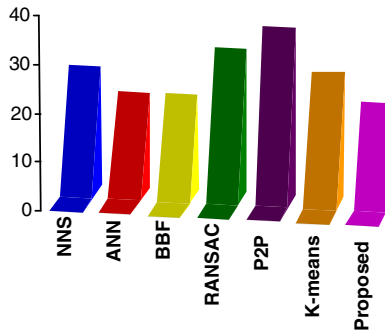


Fig. 4. Comparison chart

## 7 Conclusion

This paper presents a simple but effective method for matching two images. An algorithm is proposed by applying the rank correlation coefficients Kendall's tau and Spearman's rho for matching the key points. In poor lighting conditions also the algorithm is giving better results. The proposed method to detect mismatches is of very low computation time and produces very few mismatches.

## References

- [1] Tsai, V.J.D., Huang, Y.T.: Automated image mosaicing. *Journal of Chinese Institute of Engineers* 28(2), 329–340 (2005)
- [2] Harris, C.J., Stephens, M.: A combined corner and edge detector. In: *Alvey Vision Conference*, pp. 147–151 (March 1988)
- [3] Bolles, F.: Random sample consensus: a paradigm for model fitting with application to image analysis and automated artography. *Communications of the ACM* 24(6), 381–395 (1981)
- [4] Vincent, E.: On feature point matching, in the calibrated and uncalibrated contexts, between widely and narrowly separated images, Ph.D thesis (2004)
- [5] Carlo, T., Kanade, T.: Detection and Tracking of Point Features, Carnegie Mellon University Technical Report (April 1991)
- [6] Bhat, D.N., Nayar, S.K.: Ordinal measures for image correspondence. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20(4), 415–423 (1998)
- [7] Bhat, D.N., Nayar, S.K.: Ordinal measures for visual correspondence. Columbia University, Computer Science, tech. rep. CUCS-009-96 (February 1996)
- [8] Gideon, R.A., Hollister, R.A.: A rank correlation coefficient. *J. Am. Statistical Association* 82(398), 656–666 (1987)
- [9] Kendall, M., Gibbons, J.D.: *Rank correlation Methods*, 5th edn. Edward Arnold, New York (1990)
- [10] <http://personal.systembiology.net/ilya/ORDINAL.html>
- [11] [http://en.wikipedia.org/wiki/Kendall\\_tau\\_rank\\_correlation\\_coefficient](http://en.wikipedia.org/wiki/Kendall_tau_rank_correlation_coefficient)

# Performance Enhanced Hybrid Kohonen-Hopfield Neural Network for Abnormal Brain Image Classification

D. Jude Hemanth<sup>1</sup>, C. Kezi Selva Vijila<sup>3</sup>, A. Immanuel Selvakumar<sup>2</sup>, and J. Anitha<sup>1</sup>

<sup>1</sup> Department of ECE

<sup>2</sup> Department of EEE,

Karunya University, Coimbatore, India

<sup>3</sup> Christian Engineering College, Oddanchatram, India

jude\_hemanth@rediffmail.com

vijila\_2000@yahoo.com

immanuel@karunya.edu

rajiveel@rediffmail.com

**Abstract.** Image classification is one of the pattern recognition techniques which are often used for categorizing the abnormal medical images into different groups. Artificial Neural Networks (ANN) is widely used for this automated application owing to their numerous advantages. Despite the merits, one of the significant drawbacks of the ANN is the convergence problem. Specifically, in Kohonen Neural Networks (KN) & Hopfield Neural Networks (HN), the convergence to the stored patterns is not guaranteed which ultimately leads to misclassification on the input data. In this work, a hybrid approach namely, Kohonen-Hopfield neural network (KHN) is proposed to minimize the convergence problem for medical image classification applications. Experiments are conducted on KHN using abnormal Magnetic Resonance (MR) brain images from four classes. The performance of KHN is analyzed in terms of classification accuracy and convergence rate. Experimental results suggest promising results in terms of accuracy which indirectly indicates the minimization of convergence irregularities. A comparative analysis with other techniques is also performed to show the superior nature of the proposed approach.

**Keywords:** Image classification, Kohonen-Hopfield neural network, Convergence problem and MR brain images.

## 1 Introduction

The application of ANN for image analysis is exponentially increasing in the medical field. Classification is one of the significant applications of medical image analysis which assist in categorizing the abnormal images into different groups. Specifically, abnormal brain tumor images are initially classified into one of the several categories which will aid in treatment planning. Conventional medical image classification is

performed with human observation which is highly prone to error. Hence, automated classification methodologies based on ANN are extremely important for medical imaging applications.

Several automated techniques have been developed for MR abnormal brain image classification. ANN based classification techniques prove to be much superior over other techniques in terms of performance measures. Literature survey reveals the wide usage of ANN techniques for brain image classification applications. Bidirectional associative memory (BAM) neural network based brain image classification is implemented in [1]. Feed forward neural networks are also used for pattern classification among abnormal brain images [2]. The application of evolutionary algorithms for MR brain image classification is reported in [3]. Even though several works have been proposed in the literature, most of them suffer from the drawbacks of inferior classification accuracy which is mainly due to the convergence problem of the neural networks.

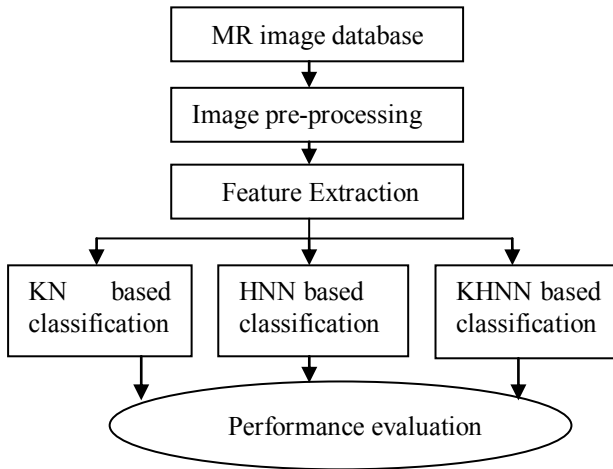
Researchers have also performed some modifications on the conventional neural networks to enhance the performance of the automated system. A modified probabilistic neural network has been developed for brain image analysis by Song et al [4]. A novel training algorithm for back propagation neural networks is proposed in [5]. Training algorithms to improve the performance of the supervised neural networks are tested by Ludwig & Nunes [6]. Successive approximation training algorithms are also used for performance enhancement in feed forward neural networks [7]. These works mainly lay emphasis on supervised neural networks which are mostly used for medical image analysis. But, the convergence problem of the unsupervised neural networks is mostly left unnoticed which limits their usage for practical medical applications.

In this work, a hybrid Kohonen-Hopfield neural network (KHNN) is proposed to tackle the convergence problem of unsupervised neural networks. This modified neural network is a 3 layer network in which the hidden layer employs the “winner take-all” training methodology and the output layer follows the Hopfield neural network procedure. This avoids the convergence problem of Kohonen neural network to a higher extent which ultimately yields accurate results. Being a hybrid approach, KHNN possess the benefit of high accuracy and quick convergence. Experiments are conducted on abnormal images collected from the radiologists. The conventional Kohonen neural network and the Hopfield neural network are also implemented to show the superior nature of the proposed approach. Feature extraction is further performed to assist the training procedure of the ANN. The results of the three networks are analyzed in terms of accuracy and convergence rate. An extensive analysis reveals promising results for the proposed approach in terms of the performance measures.

## 2 Materials and Methods

The proposed automated system for image classification is shown in Figure 1.





**Fig. 1.** Proposed Methodology

The automated classification system is experimented on real-time MR images collected from scan centre at Madurai, India. These are  $256 \times 256$  gray level images from four categories such as meningioma, metastase, glioma & astrocytoma which are available in TIFF format. These images are then filtered using Gabor filters to remove the noises of the raw input image. An extensive set of features are further extracted from these features to train the KHNN. The features used in this work are wavelet transform based textural features.

These features are further used to train the KHNN. Since the extracted features are unique for each category, the training process assists the KHNN to store the information about each category distinctly. After training, the KHNN is tested using the testing images which are not used for the training process. The performance measures of KHNN are further compared with the results of KN and HNN.

### 3 Image Pre-processing and Feature Extraction

Image pre-processing and feature extraction are the mandatory steps of any medical imaging applications. These steps enhance the accuracy of the final results besides reducing the complexity of the entire methodology.

#### 3.1 Filtering

The raw images collected from the radiologists are not suitable for classification since the noises present in the images may change the characteristic features of each class which ultimately lead to inferior results. Hence, a suitable filtering method is initially adopted to generate a better quality image. In this work, Gabor filtering [8] is used since these filters are suitable for medical images than any other filters. The filtering equation of the 2D Gabor filter is given by

$$g_{\lambda,\theta,\gamma,\sigma}(x, y) = \exp\left(-\frac{x'^2 + \gamma^2 y'^2}{2\sigma^2}\right) \cos\left(2\pi \frac{x'}{\lambda}\right) \quad (1)$$

$$x' = x \cos \theta + y \sin \theta$$

$$y' = -x \sin \theta + y \cos \theta$$

where  $\lambda$  corresponds to the wavelength,  $\theta$  corresponds to orientation in degrees,  $\gamma$  corresponds to the aspect ratio and  $\sigma = 0.56\lambda$ . Suitable selection of these parameters yields a better quality image than the original raw image. The mode of selection of these values is discussed in the Implementation section.

### 3.2 Textural Feature Extraction

The purpose of feature extraction is to reduce the original data set by measuring certain properties, or features that distinguish one input pattern from another pattern [9]. The extracted feature should provide the characteristics of the input type to the classifier by considering the description of the relevant properties of the image into a feature space. In this work, wavelet transform is initially used on the pre-processed images. The wavelet transformed image yields more information in a compact form since most of the energy is available in the low frequency coefficients of the input image. This will reduce the misclassification rate of the proposed system. Eight textural features based on the first order histogram of these images are used in this work. The various features such as mean, standard deviation, skewness, kurtosis, energy, entropy, correlation and inverse difference moment are computed using the formulae available in [10,11]. These features work especially well for the MR brain image classification applications.

## 4 Kohonen-Hopfield Neural Network Based Image Classification

In this work, a hybrid Kohonen-Hopfield Neural Network (KHNN) is proposed for abnormal MR brain image classification. For comparative analysis, the conventional Kohonen network (KN) and the Hopfield network (HNN) are also implemented using the same database. Initially, the KHNN is dealt in detail followed by the brief explanation on KN and HNN.

### 4.1 Hybrid Kohonen-Hopfield Neural Network

The conventional Kohonen neural network suffers from the drawback of inaccurate convergence which is mainly due to the lack of standard stabilization criteria for the training algorithm. Hence, the practical application of KN is highly limited even though the convergence rate is high. On the other hand, Hopfield neural networks yield sufficiently accurate results than the KN with the help of the stability function but the convergence rate is quite low. Thus, to avail the facilities of both the neural networks, a hybrid approach such as KHNN is proposed in this work.

The KHNN is a 3-layer network which consists of Kohonen layer (hidden layer) followed by the Hopfield layer (output layer). The output of the Hopfield layer is fed back as input to the input Kohonen layer. Thus, the training algorithm involves the mathematical procedures of both KN and the HNN. This methodology yields the benefits of both the networks in a single module (KHNN) which is the main objective of this work.

**Network Architecture.** The framework of the proposed KHNN is shown in Figure 2.

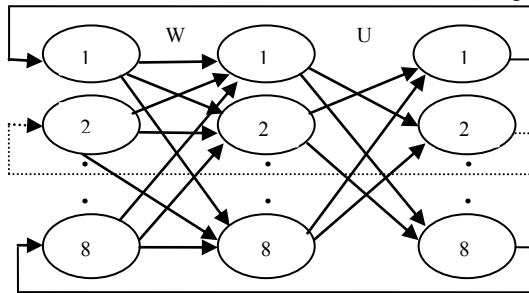


Fig. 2. Architecture of KHNN

The proposed KHNN is a 2-layer network with the input layer, hidden layer and an output layer. The number of neurons in the input layer is equal to the number of input features. The number of neurons used in the hidden and output layer is also eight in order to satisfy two conditions. Since the output layer represents the Hopfield layer, there shall not be any one-to-one connection between the hidden layer and the output layer to ensure stability of the network. Hence, equal number of neurons is used. Secondly, the feedback from the output layer neuron is given to the corresponding input layer neuron. Thus, the number of neurons used in all the layers must be same in the proposed KHNN. Unlike conventional HNN, there is no necessity for the external inputs.

Two sets of weights are used in the architecture with ‘W’ represents the weights of the Kohonen layer and ‘U’ represents the weights of the Hopfield layer. While the weights ‘U’ is determined using the output values of hidden layer neurons, the weights ‘W’ are adjusted through the “winner take-all” training methodology.

**Training Algorithm of KHNN.** The training methodology of KHNN involves both the training procedures of KN and HNN. Unlike conventional neural networks, the training of Kohonen and Hopfield layer takes place simultaneously. Initially, the input database is divided into training dataset and testing dataset. The training inputs are given as input to the Kohonen layer and the hidden layer output is calculated using the “winner take-all algorithm” in a single step. These outputs are given as inputs to the Hopfield layer where the output values of each neuron is estimated with the help of weights ‘U’ & the output of hidden layer neurons ‘D’. These outputs are further fed back to the input layer and the winner neuron’s weights of the previous iteration are changed using the weight adjustment equations. With the new set of weights, the output values are once again calculated and given as input to the Hopfield layer. This procedure continues till the change in energy value becomes zero which yields the stabilized set of weights. The following mathematical steps illustrate the training algorithm more clearly.

Step 1: Initialize weights  $w_{ij}$  and threshold  $T_k$ ;  $i$  = input layer (1,2...8),  $j$  = hidden layer (1,2...8) and  $k$  = output layer (1,2...8).

Step 2: Supply the input features to the input layer and determine the output of the Kohonen layer neurons using the following formula

$$D(J) = \sum_i (w_{ij} - x_i)^2 \quad (2)$$

Step 3: Determine  $J$  for which  $D(J)$  is minimum. This neuron is selected as the winner neuron and stored separately.

Step 4: Determine the weight matrix of the output layer using the following equations

$$u_{jk} = D_j D_k; D = \text{output values of hidden layer neurons} \quad (3)$$

$$u_{jk} = u_{kj}; \forall j \neq k \quad (4)$$

$$u_{jj} = 0; \forall j \quad (5)$$

Step 5: Calculate the output value of the Hopfield layer using the following formulae

$$NET_k = \sum_{j \neq k} u_{jk} D_j \quad (6)$$

$$OUT_k = 1 / (1 + e^{-NET_k}) \quad (7)$$

Step 6: Determine the Liapunov energy function ( $E$ ) using the output values

$$E = (-1/2) \sum_j \sum_k u_{jk} D_j OUT_k + \sum_k T_k OUT_k \quad (8)$$

Step 7: Estimate the change in energy  $E$  by partially differentiating the above equation with respect to  $OUT_k$ . The change in energy is given by

$$\delta E = - \left[ \sum_{i \neq j} (u_{jk} D_j) - T_k \right] \delta OUT_k \quad (9)$$

$$= - [NET_k - T_k] \delta OUT_k \quad (10)$$

Step 8: The values  $OUT_k$  are fed back to the input layer if  $\delta E \neq m$  where 'm' is a constant which must be a minimum value. The ideal value of 'm' is zero.

Step 9: The weights of the winner neuron of the previous iteration is then adjusted using the following equation

$$w_{ij}(t+1) = w_{ij}(t) + \alpha [OUT_k - w_{ij}(t)] \quad (11)$$

Step 10: After weight adjustment, the new values of  $D_j$  is calculated with the new set of weights and Step 3-Step 7 is repeated till the convergence condition ( $\delta E = 0.01$ ) is reached.

The above training methodology is repeated for training images from all categories and the final weight matrix 'W' represents all the four stored patterns. The testing process is then carried out with the stabilized weights and the performance measures are estimated for the testing images. In this method, since the convergence condition

is available, the KHNN weights are more accurate than the weights obtained by conventional KN. The addition of distance calculation equation to the conventional HNN training possesses very less impact on the convergence time period of the automated system. Thus, the benefit of high accuracy at a superior convergence rate is achieved in this system.

#### 4.2 Kohonen Neural Network (KN)

The conventional KN is also implemented in this work to provide a comparative analysis of the classifiers. The architecture of KN consists of 8 neurons in the input layer and 4 neurons in the output layer with a single set of weights between these two layers. The training process is performed using the “winner take-all” algorithm which adjusts the weights of only the winner neuron. The training process is iteration dependent and the weight matrix is observed after a specific number of iterations. Since the convergence condition is not clearly defined for KN, the weights are less accurate which ultimately degrades the performance of the system. A detailed algorithm is available in [12].

#### 4.3 Hopfield Neural Network (HNN)

Hopfield neural networks belong to the associative memory neural network groups. The HNN used in this work is a single layer network with 8 neurons in the input and output layers. An external input is given to the output layer to initiate the training process. The weights are symmetrical and the weights connecting the neurons of the same position are zero. A detailed training algorithm of the HNN is available in [13].

## 5 Implementation

In this study, 374 abnormal images from four different classes are used. 31 images from each class are used for training and the remaining images are used for testing.

In the pre-processing step, the value of  $\theta = 45$ degrees,  $\gamma = 0.5$ ,  $\lambda = 10$  and  $\sigma = 5.6$ . These values are selected on trial and error basis. The values may be changed and the output can be obtained at different parameter values. The value of ‘M’ and ‘N’ used in this work are 9 and 256 respectively. These parameters specify 3×3 neighborhood with the size of the matrix ranging from 0 to 255.

The value of  $\delta E = 0.01$  is used as convergence condition for this work since  $\delta E = 0$  is practically non-feasible. The number of iterations used for training the conventional KN is 2000. The normalization of the input values is performed throughout the implementation for easier mathematical analysis. These experiments are conducted on Intel processor with 1GB RAM. The software used for implementation is MATLAB [14].

## 6 Experimental Results and Discussions

The experiments are conducted on the real-time MR images and the performance measures are analyzed individually for the three classifiers. The performance

measures used in this work are classification accuracy and convergence rate. The classification accuracy is the ratio of the correctly classified images to the total number of images. Convergence rate is the time taken for training and testing the neural network. Initially, the results of the conventional neural networks (KN & HNN) are reported and then the results of the proposed approach (KHNN) are discussed in detail. Table 1 shows the results of Kohonen neural network.

**Table 1.** Classification Accuracy results of the KN classifier

Abnormality type	No. of testing images	No. of correctly classified images	Classification Accuracy (%)
Meningioma	71	52	73.2
Astrocytoma	51	37	72.5
Metastase	65	49	75.3
Glioma	63	45	71.4
Average			73.1

From Table 1, it is evident that the KN classifier fails to yield highly accurate results. The average accuracy is sufficiently low to limit the usage of KN classifier for practical applications. The main reason behind the less accurate results is the lack of standard convergence condition. Table 2 illustrates the results of the HNN classifier.

**Table 2.** Classification Accuracy results of the HNN classifier

Abnormality type	No. of testing images	No. of correctly classified images	Classification Accuracy (%)
Meningioma	71	61	85.9
Astrocytoma	51	43	84.3
Metastase	65	56	86.1
Glioma	63	52	82.5
Average			84.7

The classification accuracy of conventional HNN classifier is better than the KN classifier but still the accuracy is not sufficient for real-time applications such as medical applications. The improvement in the accuracy is due to the stability conditions involved in the training algorithm. Table 3 depicts the results of the proposed KHNN classifier.

**Table 3.** Classification Accuracy results of the KHNN classifier

Abnormality type	No. of testing images	No. of correctly classified images	Classification Accuracy (%)
Meningioma	71	68	95.7
Astrocytoma	51	49	96.0
Metastase	65	62	95.4
Glioma	63	59	93.6
Average			95.2

The KHNN yields highly accurate results than the conventional classifiers which is evident from Table 3. The classification accuracy results also prove the necessity for hybrid approaches such as KHNN. Since the proposed technique involves the benefits of both KN & HNN, the classification accuracy of this hybrid approach is sufficiently high for practical applications. The three classifiers are further analyzed in terms of the convergence rate. Table 4 shows the performance analysis of the classifiers in terms of time period.

**Table 4.** Convergence rate of the classifiers

Classifier	Training time (CPU secs)	Testing time (CPU secs)	Convergence rate (CPU secs)
KN	550	0.215	550.215
HNN	720	0.215	920.215
KHNN	885	0.215	985.215

The KN classifier is better than the other two classifiers in terms of convergence rate. But, this time period is proportional to the number of iterations. Considering the fact that the number of iterations must be high for accurate results, the estimated time period may increase for higher quality results. In any case, the accuracy of the KN classifier is always low irrespective of the high convergence rate. Hence, KN is usually not preferred over the other classifiers

The convergence time period of the HNN classifier is higher than the KN classifier. The stability condition calculations increase the time period to a higher extent. On the other hand, the convergence time period of KHNN is slightly higher than the conventional HNN classifier. But, the KHNN is still preferred for practical applications because of the high quality results. A comparative analysis with other classifiers is shown in Table 5.

**Table 5.** Comparative analysis with other classifiers

Classifier	Accuracy (%)	Convergence rate (CPU secs)
Kohonen network	73	550
Hopfield network	85	920
Back propagation network [16]	95	1210
Linear vector quantization network [17]	89	990
<b>Kohonen-Hopfield network (Proposed approach)</b>	95	985

From Table 5, it is evident that the proposed approach is highly suitable for medical applications in terms of classification accuracy and convergence rate. A good compromise between these two performance measures is achieved in this hybrid neural network. Thus, this work suggests a performance improved neural network which is a suitable alternative for conventional neural networks.

**Acknowledgment.** The authors wish to thank M/s. Devaki Scan centre, Madurai, India for their help regarding database collection and results validation. The authors also thank Foundations of Computer Science (FCS) Inc., for their financial support of this research work.

## References

- [1] Sharma, N., et al.: Segmentation and classification of medical images using texture-primitive features: Application of BAM-type artificial neural network. *J. Med. Phys.* 33, 119–126 (2008)
- [2] Maitra, M., Chatterjee, A.: A Slantlet transform based intelligent system for magnetic resonance brain image classification. *Biomedical Signal Processing and Control* 1, 299–306 (2006)
- [3] Zhang, Y., Wang, Z., Wu, L.: A novel method for magnetic resonance brain image classification based on adaptive chaotic PSO. *Progress in Electromagnetic Research* 109, 325–343 (2010)
- [4] Song, T., Jamshidi, M.M., Lee, R.R., Huang, M.: A Modified Probabilistic Neural Network for Partial Volume Segmentation in Brain MR Image. *IEEE Transactions on Neural Networks* 18, 1424–1432 (2007)
- [5] Altman, C., Zapatin, R.R.: Back propagation training in adaptive quantum networks. *International Journal of Theoretical Physics* 49, 2991–2997 (2010)
- [6] Ludwig, O., Nunes, U.: Novel Maximum-Margin Training Algorithms for Supervised Neural Networks. *IEEE Transactions on Neural Networks* 21, 972–984 (2010)
- [7] Liang, Y.C., et al.: Successive approximation training algorithm for feed forward neural networks. *Neurocomputing* 42, 311–312 (2002)
- [8] Lin, C.T., Huang, C.H.: A complex texture classification algorithm based on Gabor filtering CNN and self-organized fuzzy inference neural networks. In: *IEEE Symposium on Circuits and Systems, Kobe*, pp. 3942–3945 (2005)
- [9] Guyon, I., et al.: *Feature Extraction: Foundations and Applications*. Springer, Heidelberg (2006)
- [10] Sasikala, M., Kumaravel, N.: A wavelet based optimal texture feature set for classification of brain tumors. *Journal of Medical Engineering & Technology* 32, 198–205 (2008)
- [11] Assefa, D., et al.: Robust texture features for response monitoring of glioblastoma multiforme on T1-weighted and T2-flair MR images: a preliminary investigation in terms of identification and segmentation. *Medical Physics* 37, 1722–1736 (2010)
- [12] Chaplot, S., Patnail, L.M., Jaganathan, N.R.: Classification of magnetic resonance brain images using wavelets as input to support vector machine and neural networks. *Biomedical Signal Processing and Control* 1, 86–92 (2006)
- [13] Meyer-Base, A.: *Pattern recognition for medical imaging*. Academic Press, Salt Lake City (2004)
- [14] MATLAB, User's Guide, The MathWorks, Inc., Natick, MA 01760 (2002)



# Middleware for Physical and Logical Context Awareness

Junzhong Gu

Institute of Computer Applications, East China Normal University (ICA-ECNU)  
200062 Shanghai, China  
jzgu@ica.stc.sh.cn

**Abstract.** The human-computer interaction depended computing mode is evolved to an intelligent adaptive context aware computing, involving multiple sensors. The corresponding middleware should be adaptive, reflective, dynamic reconfigurable and physical/virtual context aware. Some key issues, e.g. virtual and meta sensors, reflective context model, multi-agent mechanism, an extension of Web Service-Open Web Service, etc. are analyzed and studied. An experimental system and its design are introduced.

**Keywords:** context awareness, middleware, mobile agent, virtual sensor, meta sensor, reflective system, OpenWebService.

## 1 Introduction

Boundary of the Internet is widely extended. Connected devices are extended from workstations/PCs to handhelds, as well as various sensors. IOT<sup>1</sup> is one of such cases. The fixed network based computing is evolved to a mobile environment. The human-computer interaction depended computing mode is evolved to intelligent adaptive context aware computing. [1, 2, 3] Therefore intelligent agents, especially mobile agents, are in good grace. With characters such as distributed and mobile environment, heterogeneity, context awareness, physical/logical sensor orientation, and so on, developing responsible middleware system is urgently demanded. Different from traditional middleware mechanisms such as message oriented middleware, object-oriented middleware, etc., context aware middleware should be adaptive, intelligent, reflective and dynamic reconfigurable.

Some common factors affect the design of the middleware infrastructure in context aware computing, such as: (Mobile) Devices, Reusable and adaptive, Communication and Privacy.

**Devices:** Things are introduces to Internet. They can be categorized to two classes, computable (devices) and computing enable things. Computing enable things refer that without computing power, such as things (e.g. books, parcels, etc.) are tagged with RFID. They connect to Internet via special gateways (e.g. RFID readers). A gateway here plays a role of computable device. The computable devices are with computing power, such as PCs, PDAs, smart phones, as well as online gateways (e.g. RFID readers) serving enable things.

---

<sup>1</sup> IOT-Internet of Things.

Devices, mainly mobile, vary from one to another in term of resource availability. For example, laptops are different from PDAs and smart phones in CPU computing power, and amount of RAM and disk space. Hence, the corresponding middleware should be designed to flexibly sense context via various sensors and automatically optimize resource utilization. Mobile clients may interact with different types of networks, services, and security policies as they move from one area to another.

Mobile devices (including online gateways serving enable things) often hinder the functionality of their respective systems, making difficult to build more advanced and complex system behaviors, for the limited computational resources on mobile devices, such as battery power, information storage, computing power, and communication:

**Battery Power Constraint-**Context-aware agents often rely on sensors to acquire contexts. Most of the mobile devices do not support sophisticated sensing hardware due to their battery power constraint.

**Information Storage Constraint-**Context-aware agents often need to store acquired contextual knowledge to reduce the cost of performing repetitive context acquisition and to exploit the historical information of contexts. Limited by the disk space on mobile devices, agents usually do not have the privilege to continuously store all knowledge. Agents will be faced with the challenge to be selective about what kind of knowledge should be stored and what kind of knowledge should be deleted. Contextual knowledge acquired from physical environment is potentially inaccurate, agents will be challenged to maintain knowledge consistency and resolve information ambiguity.

**Computing Power Constraint-** Because mobile devices are often lack of enough CPU power and memory, context-aware agents on mobile devices can only process primitive contexts. Even though computation is done at background, it raises the issues in the communication constraint of mobile devices.

Due to these limitations, conventional middleware technologies designed for fixed distributed systems are not suitable.

**Reusable and adaptive** - Context sensing and reasoning are two important steps in context aware applications.[3]

Context sensing is to acquire information via hardware/logical (software) sensors. Facing on a dynamic environment, a great amount of hardware/software sensors are required for agents to acquire a wide divergence of information. Without reusable and adaptive mechanisms, directly attach sensors with context-aware agents also makes difficult for building more advanced and complex systems. Therefore, context aware middleware should be reusable and adaptive.

Context reasoning mechanisms varies from simple intelligent query to complex rule-based programming. With the increasing demand to process more and more advanced contexts, context reasoning mechanisms will become more complex and specialized. Without any reusable and adaptable mechanisms for context reasoning, building context-aware systems from scratch will be difficult and costly.

**Communication Issue** - Previous middleware is always fixed connection oriented, where RPC, object requests, RMI or distributed transactions assume a high-bandwidth

connection of the components, as well as their constant availability. In mobile systems unreachability and low bandwidth are the norm rather than a failure.

Mobile applications should replicate information in order to access them off-line. Replication needs synchronization when a connection is re-established. The commonly used principle of transparency prevents the middleware to exploit knowledge that only the application has, such as which portion of data to replicate and which reconciliation policy to apply.

In a dynamic environment, the communication between agents on mobile devices and the nearby context sources often cannot be pre-defined. Agents may be lack of sufficient knowledge to communicate with context sources (e.g. knowing which sensor can provide what information and how to communicate). And context sources may dynamically join and leave the environment without notifying the agents. Therefore, it will be difficult for context-aware agents on a mobile device to effectively acquire contexts.

**Privacy Issues** - In a context-aware environment, agents collect and share user information, such that it raises great concern for user privacy. Privacy is intrinsically bound up with control - who controls what information. In the existing context-aware systems, users often do not have control over how personal information are acquired. As context sensors are built to be hidden in the physical space, personal information are collected without the explicit consent of the users. And privacy issues also arise when sharing user information.

To rise to the challenges, a suitable middle system is demanded. It's the target of our project-GaCam<sup>2</sup>. It will be detailed in following sections.

## 2 Context Aware Middleware

The role of middleware is to provide an additional layer of abstraction suitable for a specific type of applications.

Conventional middleware technologies have focused on masking out the problems of heterogeneity and distribution to facilitate the development of distributed systems. They allow the application developers to focus on application functionality rather than on dealing explicitly with distribution issues. Different middleware systems such as CORBA, DCOM and Java RMI have proved their suitability for standard client-server applications. However, it is believed that existing traditional middleware systems are not capable of providing adequate support for the mobile computing and dynamic context based environment, as well as IOT. There is a great demand for designing middleware systems that can support new requirements imposed by mobility, as well as context awareness.

GaCam, as a successor of LaMOC [2], is a middleware system to support the efficient development, automatic construction and deployment of context aware applications, launched since 2009, at ICA-ECNU, supported by of the Shanghai Scientific Development Foundation.

---

<sup>2</sup> GaCam-acronym of Generalized Adapt Context Aware Middleware.

LaMOC is a Location Aware System. [2] It involves context sensing, context pre-processing, reasoning and applications, etc., i.e. involves a complex context life cycle. Where, sensors to sense context are various (including logical and physical sensors), contexts sensed are changed dynamically, corresponding pre-processing, as well as reasoning mechanism are also variable and changeable. Challenges appear, such as how to fit in with the variability and dynamically changeable, as well as hardly heterogeneous. A context aware middleware system as GaCam is urgently required, to satisfy the requirements such as physical/virtual sensor orientation, flexible context awareness and inference mechanism.

Assumed computing environment in GaCam is nomadic (i.e. it runs on fixed network and mobile network background at same time), involving various and changeable computing power of nodes, dynamically changed topology of communication network, limited communication bandwidth for mobile nodes, etc.

In order to flexibly support context aware applications, *dynamic reconfiguration*, *Adaptivity*, *context awareness*, *reflection* are as main features of GaCam.

***Dynamic reconfiguration.*** During the system lifetime, the application behavior and context may be altered due to dynamic context changes. Dynamic reconfiguration is thus required. Dynamic changes in system behavior and operating context at runtime can trigger re-evaluation and reallocation of resources. Middleware supporting dynamic reconfiguration needs to detect changes in available resources and either reallocate resources, or notify the application to adapt to the changes.

***Adaptivity*** is also a requirement that allows applications to run efficiently and predictably under a broader range of conditions. The middleware needs to monitor sensors/context such as the resource supply/demand, compute adaptation decisions, and notify applications about changes.

***Context-awareness*** is important to build an effective and efficient adaptive system. For example, the context of a mobile unit is usually determined by its current location which, in turn, defines the environment where the computation associated with the unit is performed. The context may include device characteristics, user's activities, services, as well as other resources of the system.

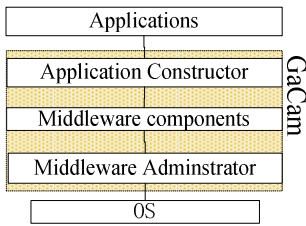
***Reflection.*** A reflective system consists of two levels named as meta-level and base-level. The former performs computation on the objects residing in the lower levels. The latter performs computation on the application domain entities. The reflection approach supports the inspection and adaptation of the underlying implementation (the base-level) at run time. A reflective system provides a meta-object protocol (meta-interface) to define the services available at the meta-level. The meta-level can be accessed via reification, which expose some hidden aspect of the internal representation and hence they can be accessed by the application (the base-level). The implementation openness offers a straightforward mechanism to insert some behavior to monitor or alter the internal behavior of the platform. This enables the application to be in charge of inspecting and adapting the middleware behavior based on its own needs.

Besides, GaCam should also support: multi sensing orientation; flexible application composition; support for heterogeneity; scalability; support for privacy protection; traceability and control; as well as ease of deployment and configuration.

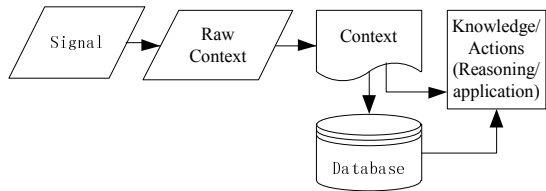
The framework of GaCam can be illustrated as in Fig. 1. GaCam (in shadow) is in the middle, which runs on different operating systems, and supports various context aware applications. It is divided in three layers, i.e. Middleware Administration, Middleware components and Application Constructor.

- **Middleware Administration.** It is used to manage middleware, monitor and audit the running of middleware.
- **Middleware components.** This layer has various context processing components covering the life cycle of context (Fig. 2), such as sensing components, pre-processing components, reasoning components, and application components.
- **Application Constructor.** The Application Constructor is in charge to construct various context aware applications, therefore composition and deployment mechanism as well as software appears here.

GaCam's support covers a life cycle of context as in Fig. 2. The life cycle of context can be illustrated as an evolution from digital signal to knowledge. That is, via sensors signal is sensed to raw context. After pre-processing of context, raw context is converted to context. Continually it is deduced to knowledge after reasoning, as well activation commands. The design of our middleware is around such a life cycle.



**Fig. 1.** Layered GaCam architecture



**Fig. 2.** Context life cycle

### 3 Some Key Issues

GaCam supports flexible and adaptive sensing source and mechanism, compatible with physical and logical sensors. Sensed context is in raw data. For further usage, context pre-processing including context converting, filtering and fusion is required. Because of the changeability of sensors and sensed context, pre-processing mechanism (e.g. converting, filtering and fusion) should adapt the changes, dynamically. Some context should be stored for further uses. The system should satisfy the requirement of the situation, keeping the consistency of stored context, and efficiently store and manage global context database and dynamic context at mobile devices. Reasoning is an important step in context aware computing. A flexible reasoning mechanism is required, to suite the dynamical and adaptive application situation.

### 3.1 Sensor, Virtual Sensor and Meta Sensor

The spectrum of sensor involved here are wide. We category them into two classes, physical sensors and logical sensors, as in Fig. 3.

The physical sensors can be defined as “A *sensor* is a device that measures a physical quantity and converts it into a signal which can be read by an observer or by an instrument. ...” (<http://en.wikipedia.org/wiki/Sensor>).

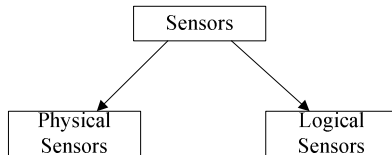


Fig. 3. two classes of sensors

They can be further categorized in acoustic/sound/vibration sensors; automotive/transportation sensors; chemical sensors; ...; and position/angle/displacement/speed/acceleration sensors (Fig. 4).

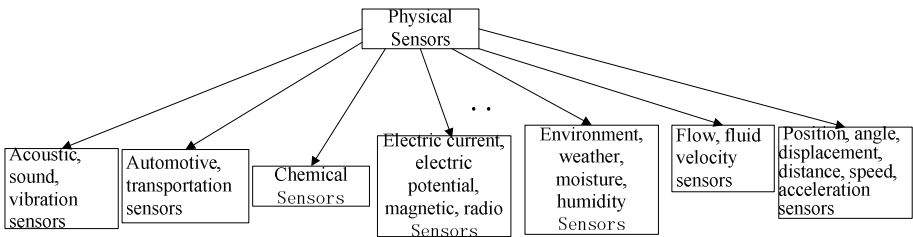


Fig. 4. Physical sensors

Continually, as an example, the Position sensors can be classified in to GPS based, Phone-cell Based, RFID based, etc., as in Fig. 5.

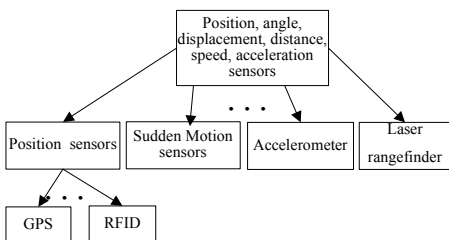


Fig. 5. where is position sensors

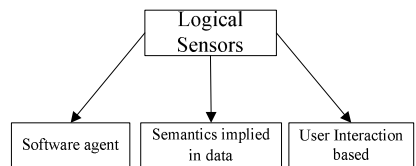


Fig. 6. Logical sensors

The logical sensors are software based, such that they can be classified as software agent based sensors, and semantics implied in data, as well as user interaction based sensors (Fig. 6).

Context sensed by software agent are used more and more popular now. Spiders used in web searching, daemon in operating systems, as well as mail server daemon are such examples of software agent based logical sensors.

Semantics implied in data is another case. It means, user profile, group profile, computer profile, network profile and so on are sensed, and more implied semantics are as context mined.

The third of logical sensing is via user interaction. Based on user interaction, user intentions, usages and operational context are analyzed, and user behaviours are collected and mined as context.

Because of the variety and heterogeneity of sensors, a reflective mechanism is designed in GaCam (Fig. 7).

Meta sensor is an abstraction of sensors, the common properties of sensors are described in meta sensors. The relationship of sensor and meta sensor can be illustrated as in Fig. 7, the mapping (noted as  $\delta$ ) from meta layer to sensor layer is named as reification. (The relationship between sensor to sensor instance is named as *instance-of*.)

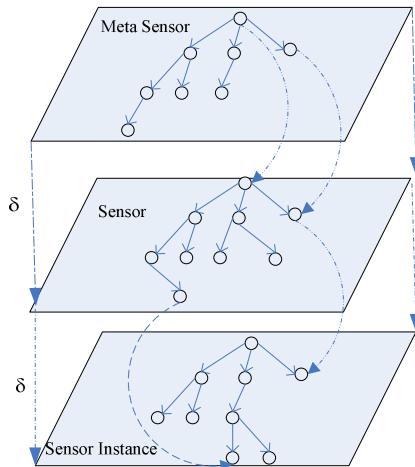


Fig. 7. Meta sensor, sensor and sensor instance

### 3.2 Semantic Context Model

Dey, Christiansen and other scholars proposed different definition of context. [1,2] But they mainly focus on the context sensed by physical sensors. Actually, besides context sensed by physical sensors (named as physical context here), logical context sensed by software is more and more widely used now. Therefore, in GaCam, the concept of context is widely extended as follows.

*Context: any information which can be sensed and used to characterize the situation of any entity, even real and virtual.*

It is emphasized that context is sensible, even by physical sensors or logical sensors, i.e. from real and virtual entities. The real entities are, for example, a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and application themselves. Virtual entities are computation power of devices, profiles, implied semantics in data, etc. Context sensed by physical sensors and virtual sensors is named as physical context and virtual context, respectively.

Therefore, challenges come :

- The model should adapt various sensors;
- Compatible with physical and logical sensors;
- Entail enough semantic information, especially, for later reasoning.

Some context models have been proposed, such as Key-Value pair Models; Markup Scheme Models, Graphical Models, Object Oriented Models, Logic Based Models, and Ontology Based Models. [8]

Key-Value pair Models, Markup Scheme Models and Graphical Models are weak in semantic expression comparing with Object Oriented, Ontology Based and Logic Based Models. The later three models are also different. Object Oriented Model and Ontology Based Model are weaker than logical based model in reasoning. We need a reasonable context model, which is always available and useful during the life cycle of context. The context model should support enough semantic expression, flexible inference, and durable context storage, compatible with physical and virtual context. Therefore, combining with the advantages of last three models, a GaCam context model is proposed by us.

### **GaCam Context Model**

As the description of information, context concept (noted as *ContextConcept* in following) is used to express context. Actual context is an instance of a context concept.

A context concept characterizes the common properties of a group of contexts. The properties are described by corresponding attributes.

#### ***ContextConcept***

*ContextConcept*: = {attribute}

Besides, the dynamic behaviour of *contextConcept* is described as methods. Therefore *ContextConcept* is expressed as:

```
ContextConcept
{
  {attribute1;
  attribute2;
  ...
}
Behaviour
{ getContext();
```



```

    ConvertContext();
    FilterContext();
    FusionContext ();
    ...
  }
}

```

Here, some generic behaviors are presented, such as `getContext()` used to read sensed context from sensor, `ConvertContext()`, `FilterContext` and `FusionContext()` used to pre-process context, i.e. context converting, filtering and fusion.

### **Attribute**

*Attribute*: = <attributeName:type> | < attributeName: ContextConcept> | <attributeName:  $\uparrow$ ContextInstance>

That is, an attribute can be data in simple data type (e.g. real, or integer), an instance of a *ContextConcept*, or a pointer to an instance of a *ContextConcept*.

### **ContextInstance**

*Context Instance*: a context instance is an instance of a *ContextType*, e.g. context sensed by GPS is an instance of *ContextConcept* GPS, i.e.

`MyGPS_Context` is *Context Instance* of GPS.

The pre-defined relationships of context-to-context in GaCam context model are classified into: *is-a*, *has-a* (i.e. *component-of*), and *member-of*. Users can define other relationships.

As example, *ContextConcept* **Position** can be described by its longitude, latitude altitude, and some methods, as follows:

```

ContextConcept Position
{
  {
    Longitude: real;
    Latitude: real;
    Altitude: real
  }
  Behaviour
  {
    ...
  }
}

```

Furthermore, we can define GPS as:

```

ContextConcept GPS
{
  {CurrentPosition: Position;
  CurrentTime: Time
  }
  Behaviour
  {

```

```

...
}
}

```

**GPS** *is-a* relates to **Position**.

Such that, GaCam can be described as:

$\exists s \in \text{sensor}$ , such that

$$s \xrightarrow{\text{getContext}} \text{raw context}$$

Preprocessing

$$\text{raw context} \xrightarrow{\text{convert/filtering}} \text{context}$$

Fusion

$$\text{context} \xrightarrow{\text{fusion}} \text{situation}$$

Reasoning

$$\text{context} \xrightarrow{\text{reasoning}} \text{conclusion}$$

Reaction

$$\text{conclusion} \xrightarrow{\text{driven}} \text{actuator}$$

A context based query language and a corresponding interface, and a context definition tool are designed in GaCam. The detail will be reported in our another papers.

### 3.3 Mobile Agent System

One of the main shortages of traditional software is lack of intelligence. Therefore, intelligent agents, especially mobile agents, are used in GaCam, to satisfy adaptivity and dynamic reconfiguration in physical and virtual context aware environment.

Here, an agent is a software component, which executes specific tasks on behalf of someone (a person) or something (an organization or another agent) with some autonomy. An agent is an active object (i.e., it can decide to execute a method) that can be implemented through one or more threads on one operating system process. Agents have some features, such as autonomous, proactive, self-learning, and sociality.

Agents are categorized to mobile and static agents. *Static agents* are created in the context of a specific application at the user's initiative and become to the user for a long time. *Mobile agents* are created by static agents, by other mobile agents, by other types of objects, or even by humans. They navigate between different places in order to execute some predefined tasks near to desired resources. Mobile agents play main role in GaCam.

A group of agents having common properties is named as an agent type or agent cluster. An agent is an instance of some type.

A *node* in GaCam is a hardware infrastructure on which agents are executed. Typically, a node is a computer, PDA, a smart phone, an iPad or another computing device. Usually, an agent executes in one node. However, during its life cycle, it may navigate through an arbitrary number of nodes. An agent can be duplicated but each agent can only be executed in one node at a time.

GaCam is a mobile agent system. A system responsible for supporting and managing agents is called an *Agent System*, in short AS. One or more AS systems can exist in each node.

The AS system provides a full computational environment to execute an agent, as well as other support features such as agent persistence, security and mobility.

*Users* have agents that execute on their behalf. They can interact with their own agents, or can interact with agents owned by other users if their respective security managers grant that behavior. Users have rights over their agents, to suspend, change their knowledge and goals, or even eliminate them.

Agents are grouped in two groups here, i.e.:

- Context agents: agents responding to context processing during the context life cycle
- System agents: agents for agent monitoring, auditing and so on, named as administrative agent here.

The context agents are further classified in three types, i.e. functional agents, actuator agents, and sensing agents, respectively.

Sensing agents are used to sense context. Actuator agents are for reaction according to the context and context based inference. Functional agents are agents for pre-processing, context storing as well as context based inference.

### 3.4 Open Web Service

Now applications face a more open environment, especially Internet based. Therefore Web Service<sup>3</sup> (WS) seems as more suitable than CORBA, DCOM, as well as RMI. But, Web Service provides loose coupling for distributed software components, rather close coupling as in CORBA, DCOM and RMI. Furthermore, comparing with RPC, IIOP and RMI, Web Service is stateless and lack of synchronized communication support. When a service requestor finds a service and binds the service, there is no working state exchange between the service requestor and service provider during the activation of the service. That is it's a stateless protocol. Accompanying with it, synchronized communication between the service requestor and service provider is also failed in WS. Context aware applications require the synchronized communication and state exchange between requestor and provider. Besides, lack of privacy protection is also its weakness. Therefore, an extended Web Service mechanism is proposed by us, which is named as OpenWebService (OpenWS).

Beside supporting state exchange and synchronized communication, OpenWS is:

1. Enabling distributed context-aware agents (possibly running on resource-limited mobile devices) to contribute to and access a shared model of the context.
2. Allowing users to control the access of their personal information in a context-aware environment.
3. The core of the OpenWS is autonomous agents that manages and controls the context model of a specific domain.

---

<sup>3</sup> <http://www.w3.org/2002/ws/>

4. Maintaining the context model of the domain, including domain contexts from the past and at the present.
5. Resolving inconsistencies and ambiguities of the domain contexts through information fusion.
6. Establishing privacy policies with users before sharing their personal information.
7. Providing knowledge sharing service for context-aware agents through agent communications

Within OpenWS, an agent is looked as a black box, with service as its interface. The Web service is used for such interface. Beside Web service, in some cases, especially for physical sensing, the service as defined in OSGi<sup>4</sup> is also used as its interface.

There are other issues, for example, Inference Engine, Application Composition and Privacy Protection, etc. But they will be shortly described here, restricted by the length of the paper.

### **Inference Engine**

In the life cycle of context, reasoning is an important step. Because facing variable sensors/context, inferential mechanism is adaptive to activate various actuators. Therefore an Inference Engine including a management system of variable methods (algorithms) is designed and developed in GaCam.

### **Application Composition**

An application supporting platform is required for a middleware system. It should support application composition and reconfiguration based on the developed middleware. Applications will be constructed and re-constructed according to the change of context. A workflow based application composition subsystem is designed in GaCam.

### **Privacy Protection**

As described, privacy is an important issue. A role based and multiple layered authority allocation and management mechanism is designed in GaCam. And in OpenWS, SOAP<sup>5</sup> is extended with authority assignment and identification, i.e. using the message with SOAP, the corresponding authority information is exchanged with an extended message envelop.

## **4 Conclusion**

In this paper, context aware middleware is discussed. Here nodes are extended to sensors (even physical or logical). Nodes are various, changeable, and mobile. Therefore, the middleware should be more intelligent, adaptive, reflective and re-configurable. Furthermore the middleware should support the life cycle of context, and application composition automatically.

---

<sup>4</sup> Open Service Gateway Initiative, <http://www.osgi.org>

<sup>5</sup> Simple Object Access Protocol, used in Web Service.

### From Our Practice, It's Found

The middleware should be reasonably layered. If layered granule is very fine, then tasks in each layer are more concrete and detailed. Portability is its strong points. But the system will be more complex. If the system divided in few layers, the system architecture seems as more succinct. But the portability is its weakness.

From our experience, layered according the life cycle of context is reasonable.

Beside the layered architecture, designing a semantic context model and sensor model are also important and fundamental. A reflective mechanism is more suitable.

Limited by the length of the paper, some details will be reported by our other papers.

**Acknowledgement.** This work is supported by the International Cooperation Foundation of Shanghai Science and Technology under Grant No. 09510703000

### References

1. Dey, A.K.: Providing Architectural Support for Building Context-Aware Applications, PhD thesis, Georgia Institute of Technology (November 2000)
2. Baldauf, M.: A survey on context-aware systems. *Int. J. Ad Hoc and Ubiquitous Computing* 2(4) (2007)
3. Gu, J.: Context aware computing. *Journal of East China Normal University (natural science)* (6) (2009) (in Chinese)
4. Gu, J., He, L., Yang, J.: LaMOC –A Location Aware Mobile Cooperative System. In: *Proc. Signal Processing, Image Processing and Pattern Recognition*, Jeju Island, Korea, December 10-12, Springer, Heidelberg (2010)
5. Kjær, K.E.: A Survey of Context-Aare Middleware, [http://www.hydramiddleware.eu/hydra\\_papers/A\\_Survey\\_of\\_Contextaware\\_Middleware.pdf](http://www.hydramiddleware.eu/hydra_papers/A_Survey_of_Contextaware_Middleware.pdf)
6. Chen, H.: An Intelligent Broker Architecture for Context-Aware Systems, PhD Thesis, University of Maryland (2003)
7. Mascolo, C., Capra, L., Emmerich, W.: Mobile Computing Middleware. In: Gregori, E., Anastasi, G., Basagni, S. (eds.) *NETWORKING 2002*. LNCS, vol. 2497, pp. 20–58. Springer, Heidelberg (2002)
8. Strang, T., Linnhoff-Popier, C.: A Context Modelling Survey. In: *Workshop on Advanced Context Modelling Reasoning and Management as Part of UbiComp* (2004)

# The Use of Biorthogonal Wavelet, 2D Polynomial and Quadtree to Compress Color Images

Loay E. Goerge and Bushra A. Sultan

Dept. of Computer Science, Baghdad University, College of Science, Baghdad, Iraq  
{loayedwar57,bushra542001}@yahoo.com

**Abstract.** In this paper, a compression system with high synthetic architect is introduced, it is based on wavelet transform, polynomial representation and quadtree coding. The bio-orthogonal (tap 9/7) wavelet transform is used to decompose the image signal, and 2D polynomial representation is utilized to prune the existing high scale variation of image signal. Quantization with quadtree coding are followed by shift coding are applied to compress the detail band and the residue part of approximation subband. The test results indicate that the introduced system is simple and fast and it leads to better compression gain in comparison with the case of using first order polynomial approximation.

**Keywords:** Image compression, Wavelet transform, Polynomial coding of degree two, Quadtree, Entropy encoder.

## 1 Introduction

Recently, with the rapid proliferation of multimedia contents and the advent of the World Wide Web (WWW) data compression techniques have gained immense significance. Data compression has emerged as significant technology for effectual storage and transmission of digital data. Lately digital images have been extensively made use of in many diverse fields [1].

Image compression reduces the amount of data required to represent an image with closer resemblance to the original image by removing redundant information [2].

The trend in image compression is increasingly wavelet-based due to the fact that it provides high image quality at high compression rate [3]. Kharate and Patil [4] have proposed a new technique which is based on wavelet packet and the best selected tree is based on the Entropy with enhanced run-length encoding which reduces the time complexity of wavelet packets decomposition.

Quadtree coding has been one of the most popular hierarchical segmentation based coding scheme that investigated by researchers [5]. It is recursively divides the image into simple geometrical regions. Hierarchical segmentation based coding schemes (like, quadtree, HV) generally lead to compression performance better than those based on fixed-block partitioning (like, JPEG, subband coding, classical vector quantization) [6].

Generalized Polynomial representation is one of the polynomials type commonly used as computer graphic tool to display surfaces with high degree of smoothness.

Moreover, in the literals, polynomial fitting is utilized in two different ways, either in finding the surface passes through a set of given points (i.e., interpolation) or finding a surface that passes near a set of given point (i.e., approximation) [7]. The polynomial approximation represents the core of our suggested method for reducing the spatial slow variation component associated with image data.

Many research efforts have spent to investigate the use of different types of polynomial to represent the pictographic data, and many of such research work could be found in the literature. Schraringer [8] proposed the use of multi-polynomial (spline) interpolation with wavelet texture coding to build a robust lossy image compression method. EL-Kuhamy et al [9] presented an adaptive algorithm for the implementation of polynomial based interpolation technique (such as Biliner, Bicubic and Cubic Spline); this algorithm is based on the minimization of the average of squared estimation error over all pixels in the interpolated image. Kazinnik et al [10] investigated the compression performance of a technique based on using cubic spline interpolation to reconstruct the image pixels from a set of irregularly sample pixels (i.e., of the Nyquist rate) chosen randomly from the image. Krishnamoorthi and Kannan [11] presented an image coding scheme based on orthogonal polynomial. They proposed the use of different polynomial bases operators (with different width) to represent the image data, also they apply quantization and bit allocation to encode the polynomial coefficients in a manner similar to that used in JPEG.

Finally Goerge and Sultan [12] proposed a simple and hybrid method for compressing color image used wavelet transformation and 2-D polynomial surface representation, the latter is utilized as a technique for reducing the large scale variation (or equivalently the low frequency component) associated with the image signal. Both wavelet and polynomial transform are combined in high synthetic architect followed by quantization and quadtree spatial coding and finally the results further encoded using shift encoder as an efficient entropy encoder.

In this paper the performance of our previous published method, proposed in [12], is improved using higher order of polynomial interpolation (degree two) instead of the 1<sup>st</sup> order polynomial representation. As mentioned before in [12] the main concerns of the proposed system is to get the benefit of energy compaction property of Bio-orthogonal tap 9-7 wavelet transform [13] and to utilize its ability to decompose the image signal into low (i.e., approximation) and high frequency (i.e., detail) subbands. Since the approximation band contains low variation part of the signal, so the use of polynomial transform will be helpful to represent a significant energy part of LL band, and the approximation error is handled as a residue; the later could be represented using quadtree as spatial encoder. The polynomial will prune the smooth part of the image signal. Shift coding is used as fast and simple entropy encoder to minimize the bits required to represent the values of the produced codeword. Shift coding is adopted because it needs a little amount of over head information.

The rest of the paper is structured as follows. Section 2 contains an explanation of the proposed system. The established proposed system is tested using some commonly used images, and the test results are discussed in Section 3, finally the derived conclusions are listed in section 4.

## 2 The Proposed System

The layout of the proposed system is illustrated in Fig. 1. Each step of the proposed system is explained in one of the following subsections:

### A. Color Transform

In this section the color component (R, G, B) are transferred to less correlated color space components (like YUV) in order to reduce the redundancy and getting better compression gain. After that the chrominance component (U, V) are down sampled by 2 (producing dU, dV).

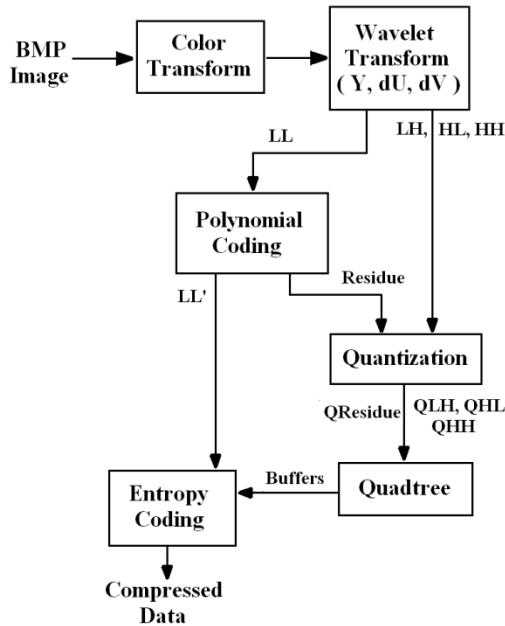


Fig. 1. The Compressed System Layout

### B. Wavelet Transform

In this step, the bio-orthogonal tap 9-7 wavelet filters are applied on Y and the down sampled dU, dV chromatic bands in separate manners. The transform will decompose each color band data to the known subbands (LL, LH, HL, HH). The following steps listed in the below subsections will be applied on each decomposed wavelet subband of (Y, dU, dV) separately.

### C. Polynomial Coding

As a first step, the LL subband is partitioned into blocks each of size (4x4) pixels. Then the coefficients  $\{a_i | i=0,1,..5\}$  of the second order polynomial are computed for each LL block. The following equations are used to determine the polynomial coefficients:



$$a_0 = \frac{1}{s} [T S_g + U(S_{gx} + S_{qy})] \tag{1}$$

$$a_1 = \frac{1}{p} \sum_{y=0}^{L-1} \sum_{x=0}^{L-1} G(x, y)(x - c) \tag{2}$$

$$a_2 = \frac{1}{p} \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} G(x, y)(y - c) \tag{3}$$

$$a_3 = \frac{1}{s} [U S_g + V S_{gx} + Z S_{qy}] \tag{4}$$

$$a_4 = \frac{1}{s} [U S_g + Z S_{gx} + V S_{qy}] \tag{5}$$

$$a_5 = \frac{1}{R} \sum_{y=0}^{L-1} \sum_{x=0}^{L-1} G(x, y)(x - c)(y - c) \tag{6}$$

Where,

$$c = \frac{L-1}{2} \tag{7}$$

$$S = L^2(Q^2 + R^2) + 2P^2(R - Q) \tag{8}$$

$$P = \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} (x - c)^2 \tag{9}$$

$$Q = \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} (x - c)^4 \tag{10}$$

$$T = Q^2 + R^2 \tag{11}$$

$$U = P(R - T) \tag{12}$$

$$V = L^2Q + P^2 \tag{13}$$

$$Z = P^2 - L^2 R \tag{14}$$

$$S_g = \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} G(x, y) \tag{15}$$

$$S_{gx} = \sum_{y=0}^{L-1} \sum_{x=0}^{L-1} G(x, y) (x - c)^2 \tag{16}$$

$$S_{gy} = \sum_{x=0}^{L-1} \sum_{y=0}^{L-1} G(x, y) (y - c)^2 \tag{17}$$

where,  $G(x,y)$ : is the original values of the approximation (LL) coefficients for each (4x4) blocks and L is the width and height of the block.

After determination of the polynomial coefficients, then they used to construct the polynomial part of the blocks, and it is subtracted from the original LL block values to produce the residue part of each LL blocks.

To get good compression gain, the determined polynomial coefficients are quantized using uniform scalar quantization

$$b_i = \text{round} \left( \frac{a_i}{q_i} \right) \rightarrow \tilde{a}_i = b_i q_i \tag{18}$$

Where,

$$i=0,1,\dots,5,$$

$a_i$  is the  $i^{\text{th}}$  polynomial coefficient,

$q_i$  is the  $i^{\text{th}}$  corresponding quantization step,

$b_i$  is the corresponding  $i^{\text{th}}$  quantization step index,

and  $\tilde{a}_i$  is the corresponding dequantized value of the  $i^{\text{th}}$  polynomial coefficient.

For image quality preservation the error due to polynomial approximation was not ignored; it is considered as residue (R) and determined using the following equations:

$$R(x, y) = G(x, y) - \tilde{G}(x, y) \quad \forall x, y \quad (19)$$

Where:

$$\tilde{G}(x, y) = \tilde{a}_0 + \tilde{a}_1\hat{x} + \tilde{a}_2\hat{y} + \tilde{a}_3\hat{x}^2 + \tilde{a}_4\hat{y}^2 + \tilde{a}_5\hat{x}\hat{y} \quad (20)$$

$$\hat{x} = (x - c) \quad (21)$$

$$\hat{y} = (y - c) \quad (22)$$

$R(x,y)$  is the residue value at  $(x,y)$ .

#### D. Quantization

In order to raise the system compression gain the uniform quantization was applied to quantize the residue part (R) and the detail subbands.

Since the human vision system has low sensitivity to the local variation in detail bands (i.e., LH, HL, and HH) and in the residue part, so quantization doesn't cause significant subjective distortion in image appearance; but it will be very useful to get good compression gain. For each component (LH, HL, HH and residue) different quantization step ( $q_1$ ,  $q_2$ ,  $q_3$ , and  $q_4$ ) could adopted. The quantization step values depend on to the subjective significance of each wavelet band.

#### E. Quadtree

In this stage each quantized subband is divided into  $8 \times 8$  blocks; and a quadtree-based search is applied to check the emptiness of the tested blocks, in case the tested block is not empty then the search is repeated upon its four daughter quadrants ( $4 \times 4$  blocks).

The emptiness test is repeated for each daughter, in a hierarchal manner, to produce other quadrants when the tested daughter quadrant is found not empty. Once the produced quadrant size become  $(2 \times 2)$  and it is found non empty then its contents are saved. So, quadtree coding will be used to address the empty blocks whether they have size  $8 \times 8$ ,  $4 \times 4$  or  $2 \times 2$ , and the contents of non empty  $(2 \times 2)$  blocks are saved in a buffer beside to the quadtree coding sequence.

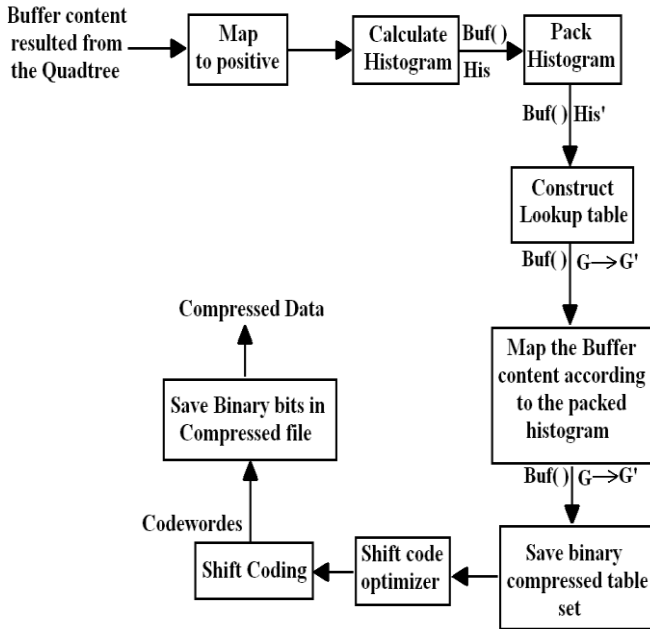


Fig. 2. The Entropy Encoder

**F. Entropy Encoding**

Entropy encoder (as explain in Fig.2 is used to encode the content of non empty (2x2) blocks, which are saved in the temporary buffer. To raise the compression gain the following steps are applied:

1. Mapping to positive: to avoid coding complexity due to existence of positive/negative values the buffer elements values are mapped to be always positive, by applying the following:

$$X_i = \begin{cases} 2X_i & \text{if } X_i \geq 0 \\ -2X_i - 1 & \text{if } X_i < 0 \end{cases} \tag{23}$$

Where  $X_i$  is the  $i^{\text{th}}$  element value registered in the buffer.

According to the above equation all negative values are mapped to set odd while the positive values are set even.

2. Calculate the histogram of the mapped elements values. Fig.3 shows the histogram for the original elements values (i.e. before mapping to positive), and also shows the histogram of the buffer elements after mapping. It is clear that the histogram has long tails due to the existence of few large element values. The existence of such long tail causes a decrease in compression gain. In order to handle this problem the histogram was packed to include

only the elements have non-zero frequency of occurrence. This histogram packing process is implemented using the following algorithm:

```

L=-1
For i= 0 to All Histogram Elements
  { if His(i) > 0 then
    { Seq(i)=1; L+ +
      nHis(L)= His(i); Tbl(i)=L
    }
    Else Seq(i)=0
  }
}

```

Where L is the length of the packed histogram, nHis is the histogram array (after packing), Tbl(i) is a lookup table used to map to positive elements values to packed positive values and Seq( ) is an array of bits required to make histogram expanding process (in the decoding stage). Fig.4 shows the histogram shape before and after packing process.

3. Map the buffer positive elements values to packed values using the lookup table Tbl( ).
4. As a final step, the shift-encoder was adopted as entropy encoder due to:
  - a) Its ease of implementation in both encoding and decoding stage.
  - b) The size of its corresponding over head information is small (i.e., only two numbers: the length of short and long codewords in terse of bits).

To determine the optimal values of short ( $n_S$ ) and long ( $n_L$ ) codewords the following steps are used

- (1) Set  $n_L = \lceil \log_2(L) \rceil$
- (2) For n that leads to lowest possible values for  $N_T$ , where
 
$$N_T = \sum_{i=0}^L n \text{ His}(i) + \sum_{i=2^{n-1}}^L n_L \text{ His}(i) \quad (24)$$
- (3) set  $n_S=n$

5. Save the output codewords in the binary stream of the compression file.

To reconstruct the decompressed image all the above steps will be reversed as shown in Fig.5.

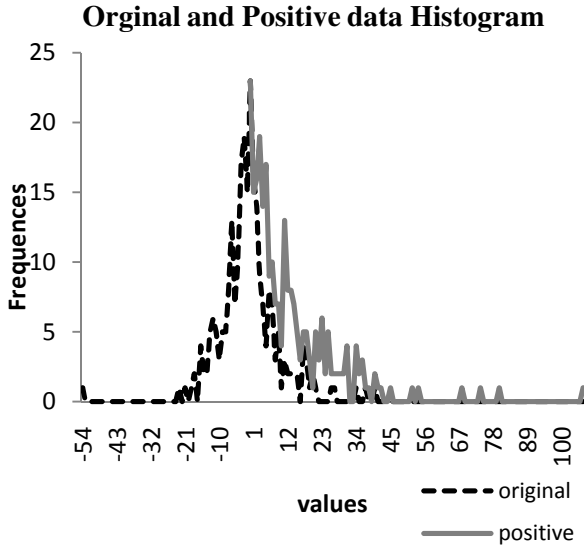


Fig. 3. The histogram of the original buffer values and the buffer values after converging them to positive

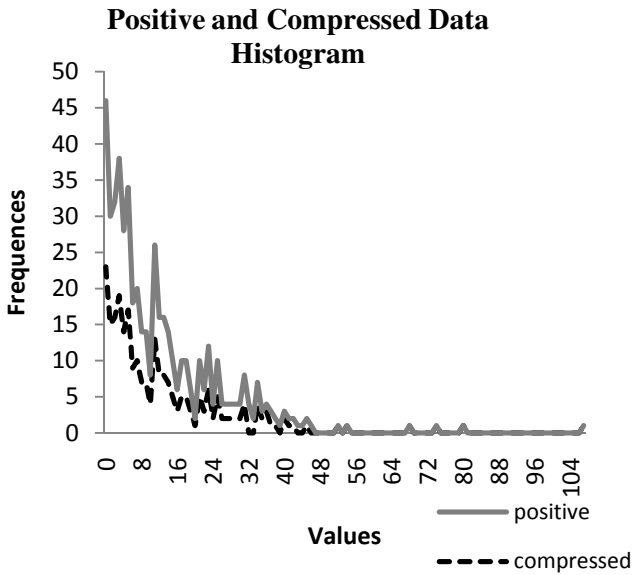


Fig. 4. The histogram of the positive and the compressed values

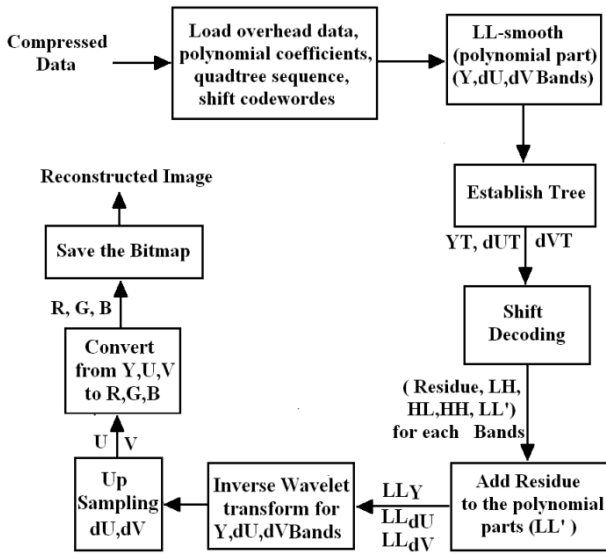


Fig. 5. The Decompressed System Layout

### 3 Tests Results

Many sets of tests were conducted to evaluate the compression performance of the proposed compression scheme. As image test samples Lena and Pepper image (with specifications: size=256x256 pixel, pixel color depth=24 bit) have been used. Table (1) lists the compression results for different values of quantization step, where  $q_1$  is the quantization step applied of the LL-residue component,  $q_2$  is the quantization step applied on LH,HL bands ; while the quantization step of the HH band is set equal to  $(2q_2)$ . The results listed in Table (1) are obtained where the quantization steps for the polynomial coefficients are set equal to 0.3. Table (2) shows the corresponding results from selected case (a) in Table (1) which are marked by small letters, this table shows the effect of the increase of the quantization steps for the polynomial coefficients on both PSNR and CR. The polynomial quantization steps are taken in the range [0.2-2.0] table 2.

The results in Table (1) shows that the set  $(q_1=7 \text{ and } q_2 \in \{20-30\})$  lead to acceptable compression results in terms of compression gain (CR) and fidelity measures (PSNR). While the results in Table (2) show that the most suitable polynomial quantization step is (0.3) lead to best compression gain. Table (3) lists the compression results that compare between applying different degree of polynomial coding of the proposed system in this paper degree2 and the system proposed in [previous paper] degree1, where the values of quantization step of the LL-residue, LH, HL, HH the same as the value of table 1 but the polynomial quantization steps of the degree one and two is set to one. Table (3) shows that the polynomial of degree two lead to best compression gain (CR) than polynomial of degree 1. Fig. 6 shows that Comparison between the compression ratios and PSNR of the polynomial coding of degree one and degree two use the shaded data of Table (3) where (the quantization step of the wavelet bands residue, LH, HL, HH is  $q_1=7, q_2=\{10,20,30,50\}, q_3=q_2, q_4=2q_2$  and the polynomial quantization step of degree one and two is set equal to one).

**Table 1.** Comparison between the PSNR and the CR in different value of Quantization steps

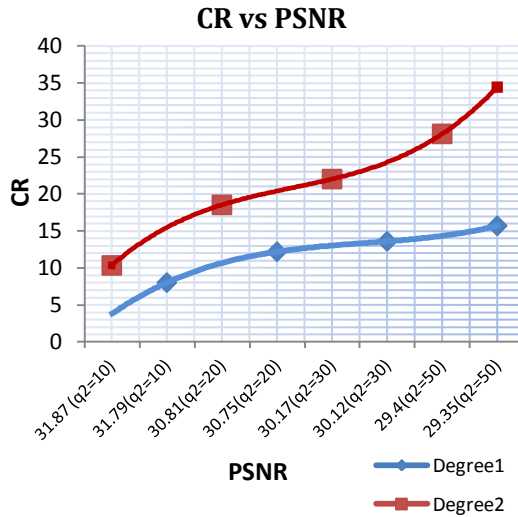
Quantization steps		Lena		Peppers		cases
q1	q2	PSNR	CR	PSNR	CR	
2	10	32.75	7.475	33.04	7.348	
	20	31.48	10.98	31.83	10.14	
	30	30.75	12.12	31.16	11.27	a
	40	30.26	13.2	30.64	11.87	b
	50	29.88	13.76	30.25	12.2	
3	10	32.62	8.137	32.89	8.358	
	20	31.38	12.47	31.718	12.17	
	30	30.67	13.96	31.061	13.83	c
	40	30.19	15.41	30.557	14.74	d
	50	29.81	16.19	30.175	15.27	
4	10	32.46	8.951	32.718	9.194	
	20	31.27	14.49	31.58	14.02	
	30	30.57	16.54	30.944	16.28	e
	40	30.1	18.62	30.453	17.56	
	50	29.73	19.77	30.078	18.31	
5	10	32.3	9.432	32.537	9.737	
	20	31.15	15.8	31.444	15.33	f
	30	30.47	18.27	30.822	18.06	g
	40	30.01	20.83	30.348	19.65	
	50	29.64	22.28	29.977	20.6	
6	10	32.15	9.843	32.327	10.2	
	20	31.03	16.98	31.276	16.51	h
	30	30.36	19.87	30.679	19.73	i
	40	29.91	22.94	30.217	21.64	
	50	29.55	24.71	29.862	22.8	
7	10	31.99	10.22	32.119	10.67	
	20	30.9	18.12	31.107	17.77	j
	30	30.25	21.45	30.537	21.55	k
	40	29.81	25.08	30.091	23.85	
	50	29.46	27.2	29.741	25.26	

**Table 2.** The effect of between the PSNR and the CR for different value of Polynomial Coefficients of Lena image

case	Pq <sub>0</sub>	Pq <sub>1</sub>	PSNR	CR
a	0.2	0.2	30.7508	12.1049
	0.3	0.3	30.7534	12.1198
	0.4	0.4	30.7498	12.0960
	0.5	0.5	30.7484	12.0389
	1.0	1.0	30.7479	11.3234
	1.2	1.2	30.7496	11.2463
	1.4	1.4	30.7493	11.5747
	1.5	1.5	30.7461	11.5570
	1.6	1.6	30.7480	11.5496
	1.8	1.8	30.7482	11.5076
	2.0	2.0	30.7472	11.4915

**Table 3.** Comparison between the PSNR and the CR of polynomial of degree 1 and 2 applied on Lena image in different value of Quantization steps

Quantization steps		Degree 1		Degree 2	
q1	q2	PSNR	CR	PSNR	CR
2	10	32.74	6.19	32.75	7.164
	20	31.48	8.41	31.48	10.32
	30	30.75	9.07	30.75	11.32
	40	29.87	9.96	29.87	12.75
	50	29.79	11.5	29.8	14.66
3	10	32.59	6.75	32.59	7.732
	20	31.36	9.49	31.36	11.55
	30	30.65	10.3	30.65	12.81
	40	29.79	11.5	29.8	14.66
	50	29.79	11.5	29.8	14.66
4	10	32.42	7.11	32.42	8.448
	20	31.22	10.2	31.23	13.22
	30	30.54	11.2	30.54	14.9
	40	29.7	12.6	29.7	17.47
	50	29.7	12.6	29.7	17.47
5	10	32.21	7.45	32.23	8.972
	20	31.07	10.9	31.09	14.55
	30	30.4	12	30.42	16.62
	40	29.59	13.7	29.6	19.87
	50	29.59	13.7	29.6	19.87
6	10	32.01	7.77	32.06	9.401
	20	30.92	11.6	30.95	15.71
	30	30.27	12.9	30.3	18.15
	40	29.48	14.8	29.5	22.1
	50	29.48	14.8	29.5	22.1
7	10	31.79	8.02	31.87	10.34
	20	30.75	12.2	30.81	18.53
	30	30.12	13.6	30.17	22.02
	40	29.35	15.7	29.4	28.13
	50	29.35	15.7	29.4	28.13



**Fig. 6.** Comparison between the compression performance for the cases of polynomial coding with the first and second order representation (quantization step of the wavelet bands residue, LH, HL, HH are set  $q_1=7, q_2=\{10, 20, 30, 50\}, q_3=q_2, q_4=2q_2$  and the polynomial quantization steps for both cases are set equal to one).

## 4 Conclusions

From the conducted test results on the proposed system, the following remarks are stimulated:

- The use of second higher order polynomial become had improved to compression performance (i.e., increase the compression gain while preserving the image quality).
- If we take into consideration the preservation PSNR level to be above the acceptable level. The best attained compression is around (21.5).
- The increase in quantization step causes an increment in the attend compression ratio and a decrease the PSNR value.
- As a future extension to the introduced system, we think that the used of hierarchal partitioning instead of fixed block size partitioning scheme could more improved the compression performance.

## References

- [1] Rawat, C.S.D., Meher, S.: A Hybrid Coding Scheme Combining SPIHT and SOFM Based Vector Quantization for Effectual Image Compression. *European Journal on Scientific Research* 38(3), 425–440 (2009)



- [2] Loganathan, R., Kumaraswamy, Y.S.: Medical Image Compression Using Biorthogonal Spline Wavelet with Deferent Decomposition. *IJCSE International Journal on Computer Science and Engineering* 2(9), 3003–3006 (2010)
- [3] Dhoub, D., Naft-Ali, A., Olivier, C., Naceur, M.S.: Performance Evaluation of Wavelet Based Color on Brain MRI Volumetric Medical Dataset for Storage and Wireless Transmission. *IJBLS* 3(3), 147 (2007)
- [4] Kharate, G.K., Patil, V.H.: Color Image Compression Based On Wavelet Paket Best Tree. *IJCSI International Journal on Computer Science Issues* 7(2), 31–35 (2010)
- [5] Chen, Y., Hao, P.: Integer Reversible Transformation to Make JPEG Lossless. In: *ICSP 2004 Proceedings, Center for Information Science, Peking University Beijing, 100871, China Department of Computer Science* (2004)
- [6] Sullivan, G.J., Baker, R.L.: Efficient Quadtree Coding of Images and Vidio. *IEEE Trans. Image Proce.* 3(3), 327–331 (1994)
- [7] Pavlidis, T.P.: *Algorithm for Graphic and Image Processing*. Springer, Heidelberg (1982)
- [8] Scharinger, J.: Image Compression by Multilevel Polynomial Interpolation and Wavelet Texture Coding. In: Moreno-Díaz, R., Pichler, F. (eds.) *EUROCAST 1997. LNCS*, vol. 1333, pp. 429–443. Springer, Heidelberg (1997)
- [9] EL-Kuhamy, S.E., Hadhoud, M.M., Dssouky, M.I., Salam, B.M., Abd EL-Samie, F.E.: A New Approach For Adaptive Polynomial Dased Image Interpolation. *International Journal of Information Aequisation* 3(2), 139–159 (2006)
- [10] Muthaian, R., NeelaKantan, K., Sharma, V., Arora, A.: Image Compression and Reconstruction Using Cubic Spline Interpolation Technique. *American Journal of Applied Sciences* 5(11), 1562–1565 (2008)
- [11] Krishnamoorthi, R., Kannan, N.: A New Integer Image Coding Technique Based on Orthogonal Polynomials. *Image and Vision Computer Archive* 27(8), 999–1006 (2006)
- [12] Goerge, L.E., Sultan, B.A.: Image Compression Bsed On Wavelet,Polynomial and Quadtree. Accepted for publication in issue No. 11 of the *Journal of Applied Computer Science & Mathematics* (October 2011) The journal is planned to appear in early, [jacs.usv.ro](http://jacs.usv.ro)
- [13] Ramakrishnan, B., Sriraam, N.: Optimal Wavelet Decomposition for Wavelet Based Image Compression Coders: An Empirical Study on Medical Images, Dept. of Biological Eng., Maniple Institute of Technology, India, [rama\\_bala@ieee.org](mailto:rama_bala@ieee.org) (2002)

# A Robust Method for Head Orientation Estimation Using Histogram of Oriented Gradients

Dinh Tuan Tran and Joo-Ho Lee

Ritsumeikan University, Japan

**Abstract.** We propose a new robust method to solve problem of estimating human head orientation using Histogram of Oriented Gradients (HOG). The proposed method is able to estimate orientation of the head with a camera even though when a user is not facing the camera. With this method, a head orientation and its rotation can be estimated precisely in all three axes: roll, yaw, pitch.

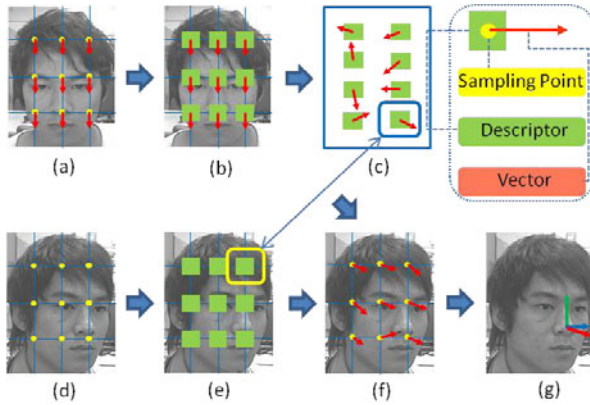
**Keywords:** Head Orientation Estimation, Histogram of Oriented Gradients, Approximate Nearest Neighbor Search, Least Median of Squares, Pattern Descriptor, Extended Template, Codebook.

## 1 Introduction

The estimation of human head orientation is an important area of research in computer vision and has a wide range of applications, including applications in human computer interaction which has been being developed in recent years and used intelligent environments to support human in various tasks. An example of these intelligent environments is iSpace in our laboratory, by estimating user head orientation the system is able to determine which position the user is focusing on and then provide services to the user [13].

As mentioned in [14], methods for human head orientation can be classified into two main categories as feature based and appearance based. Feature based methods [1,2,3,4] use a general approach that based on estimation of some facial features position in the input image to determine head orientation. Due to using facial features, near-frontal views are required and also some of these methods require manual initialization. On the other hand, appearance based methods [5,6,7] determine the relationship from a large number of training images with known face poses; therefore, these methods seem to be a classification problem using Neural Networks. Except for methods which use two or more than two cameras to estimate human head orientation, as mentioned above, the main limitation of feature based methods is that the estimating of head orientation is focused and applied for only near-frontal face. That means, if the image captured from camera only consists of the back side of the head without frontal face, the listed methods cannot estimate the approximate head orientation. On the other hand, output angle resolution of most of appearance based methods is restricted to very large steps of angle, often at least 25°.

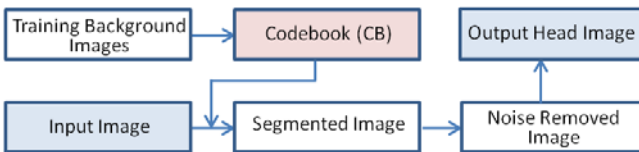
Due to these reasons, in this paper, with only one camera, we compose a new robust method to solve problem of estimating human head orientation using Histogram of Oriented Gradients (HOG) [8] that is able to estimate robustly a usable orientation of human head even when one’s face is not exposed to the camera. With this method, we are able to estimate output angle of head rotation with very small steps and the rotation of the head can be estimated in all three axes: roll, yaw, pitch. Overview of proposed method is shown in Fig. 1. In the next three sections, we will describe the method step by step. The experiments and results will be provided in Section 5 and Section 6 concludes this paper.



**Fig. 1.** Overview of the proposed method. (a) extracting sampling points for all training images. (b) calculating pattern descriptors at extracted sampling points and making extended templates. (c) a codebook which consists of all extended templates is made. (d) extracting sampling points for input image. (e) calculating pattern descriptors at extracted sampling points of the input image. (f) matching each calculated pattern descriptor of the input image with trained codebook to get the closest extended templates which provide the correspondent rotation angles. (g) from retrieved rotation angles, the final head orientation is estimated.

## 2 Head Region Extraction

The first process of every task about human head is head detection; therefore, detecting correct head region is an important task to get a correct final result. In this paper, the approach based on background-foreground segmentation is used to extract human head region for later process. Background subtraction, temporal differencing and optical flow are methods to segment background-foreground. However,



**Fig. 2.** Overview of head region extraction

Codebook (CB) model [9] is the method used in this research to segment background foreground because of its performance. After using CB to segment foreground of input image, noises are removed in the result image by applying Median Filter. Finally, head region is extracted based on its characteristics (Fig. 2).

### 3 Learning Codebook of Extended Templates from Training Images

The extended templates, sampling points and descriptors were used in [12] to locate facial feature points based on the concentration of directional vectors from sampling points on a face region. In this paper, some of these concepts will be used for the other purpose: the estimation of human head orientation.

In this section, the extended templates and the codebook which consists of all extended templates extracted from all training images are described. Unlike the extended template in [12], the extended template in this paper only consists of a pattern descriptor extracted from one sampling point and a vector which actually is rotation angles: roll, yaw, pitch. However, in real applications, estimation of only pitch axis and yaw axis is enough to determine head orientation of the user; therefore, in this paper, only these two rotation angles are used in the implementation of extended templates.

#### 3.1 Sampling Points

With every head image which was extracted and resized from all training images, sampling points are located by dividing the head image by a fixed number of vertical lines and horizontal lines. The sampling points are the intersection points of these lines. In reality, method of sampling points locating does not affect the accuracy of head orientation estimation; however, its accuracy depends on sampling point arrangement. Therefore, with this sampling points locating method, if the number of vertical and horizontal lines is large enough, we are able to get sufficient results.

#### 3.2 Pattern Descriptors

After locating sampling points for each training head image, pattern descriptors are computed by using Histogram of Oriented Gradients in the image areas which have

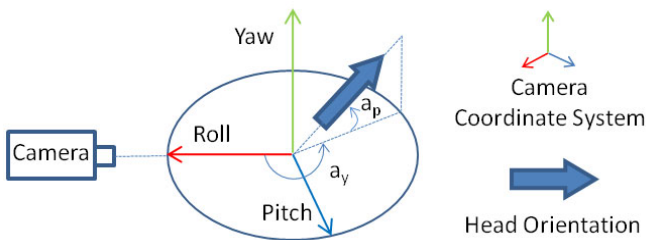


Fig. 3. Definition of two head rotation angles: yaw ( $a_y$ ), pitch ( $a_p$ )

center points at the sampling points. HOG are feature descriptors often used for object detection and computed on a dense grid of uniformly spaced cells.

Each descriptor is divided into a fixed number of blocks, and then each block is divided into a fixed number of cells. Each cell consists of a fixed number of pixels. First, at each pixel of descriptor, HOG is used to calculate gradient orientation. Then, from calculated gradient orientations at pixels of each cell, an orientation histogram with a fixed number of orientation bins is built. The built orientation histograms in cells of each block are combined into block orientation histograms; then, block orientation histograms are normalized by using one of block normalization methods of Dalal and Triggs [15]. Finally, normalized orientation histograms in blocks of each descriptor are combined into orientation histograms of descriptors. Since in this paper, only HOG of descriptor is used when mentioning about descriptor; therefore, from now, descriptor will be used as the meaning of its HOG. The calculation of descriptors is shown in Fig. 4 with values of the parameters: the number of orientation bins is three, cell size are 2 x 2 pixels, block size are 2 x 2 cells, and the number of blocks is four.

### 3.3 Extended Templates

Due to the assumption that the orientation of human head is manually known before, with each pattern descriptor which is calculated as described in Section 3.2, an extended template is made to consist of this descriptor and a vector which actually corresponds to two rotation angles: yaw and pitch rotation angles. Totally, an extended template in this paper consists of three elements, and the extended template  $f$  can be defined as follows:

$$f = (p, a_y, a_p). \tag{1}$$

Where  $p$  is the pattern descriptor,  $a_y$  and  $a_p$  are the yaw and pitch rotation angles of human head. In this paper, we assume that  $a_y = 0$  and  $a_p = 0$  are the orientation of human head looking at the camera, the range value of  $a_y$  is from 0 to 359 and the range value of  $a_p$  is from -90 to 90. In a training head image, all pattern descriptors have the same values of  $a_y$  and  $a_p$ .

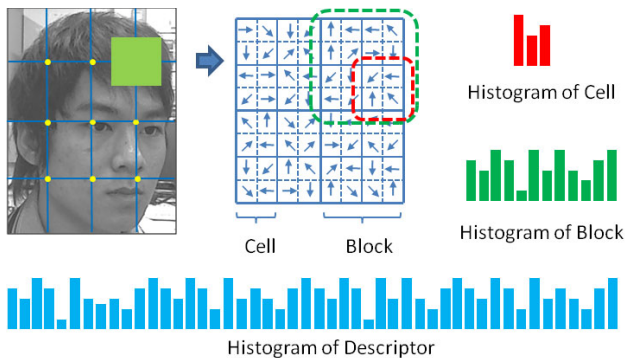


Fig. 4. Calculation of Descriptor HOG

After making all extended templates of all training head images, we obtain the codebook of extended templates. Now, human head orientation of an input image is estimated by using this learned codebook of extended templates.

## 4 Estimation of Human Head Orientation

In this section, human head orientation estimation of an input image is described. Firstly, head region extraction method proposed in Section 2 is used to extract only human head region, then extracted head image is resized to have the same size as the size used in Section 3. The sampling points are extracted as described in Section 3.1 and the pattern descriptors are calculated as described in Section 3.2. Our purpose is that from these calculated pattern descriptors, we perform matching them with the pattern descriptors in the codebook of extended templates and retrieve the closest extended templates as the matching results. Then, from the rotation angles of all these retrieved extended templates, the final human head orientation is estimated.

### 4.1 Matching

From the calculated pattern descriptors of the input image, matching is performed with the pattern descriptors in the codebook of extended templates. The approximate nearest neighbor search (ANNS) [10] is used in this research.

The results of the matching are the closest extended templates which provide the correspondent rotation angles and the distances between pattern descriptors of the input image and pattern descriptors of the closest extended templates. In the next section, two approaches of estimation step will be described: one approach uses weighted rotation angles and the other uses unweighted ones. With the approach using weighted rotation angles, we are able to use the distances between pattern descriptors of the input image and pattern descriptors of the closest extended templates or use result of the computing an inner product between pattern descriptor output of the input image and pattern descriptor output of the closest extended templates to calculate the weight.

### 4.2 Estimation from Unweighted and Weighted Rotation Angles

After matching process in Section 4.1, the closest extended templates which provide the correspondent rotation angles are retrieved. The distances between pattern descriptors of the input image and pattern descriptors of the closest extended templates can be achieved too. The new definition of the closest extended template  $f$  which includes weight  $w$  is as follows:

$$f = (p, a_y, a_p, w) . \quad (2)$$

With the approach that uses unweighted rotation angles, we assume that  $w$  is one. Before estimating the final human head orientation, some outliers caused by incorrect matches of the ANNS are removed. Those outliers are the closest extended templates

which have large matching distances when comparing with a threshold we choose. Matching distances are the distances between pattern descriptors of the input image and pattern descriptors of the closest extended templates

Then, the final human head orientation is estimated by using the least median of squares (LMedS) method [11] from the remaining rotation angles which correspond to the remaining closest extended templates. The estimation process has two steps: the estimation step of yaw rotation angle and the estimation step of pitch rotation angle. Both of them are done by the same way. The estimation step of yaw rotation angle is detailed as follows:

1. Choosing a fixed number of yaw rotation angles ( $m$  yaw rotation angles) randomly from the remaining yaw rotation angles which correspond to the remaining closest extended templates.
2. Estimating the yaw rotation angle using the chosen yaw rotation angles:

$$A_y = \frac{\sum_{i=1}^m w_i a_{yi}}{\sum_{i=1}^m w_i}, \quad (3)$$

3. Calculating  $E$ :

$$E = \text{med}_i(w_i | a_{yi} - A_y |), \quad (4)$$

where  $\text{med}$  is a median operator.

4. Iterating  $n$  times from step 1 and the yaw rotation angle of the input image is the result of step 2 which corresponds to the minimum  $E$ .

Outliers' rejection function is applied to estimate the final human head orientation more correctly.  $\sigma$  is the standard deviation which is calculated from minimum  $E$  and  $\alpha$  is a constant number. The outliers whose weighted angle errors are bigger than  $\alpha\sigma$  are rejected.

## 5 Experiments

In this section, some experiments are executed to evaluate the proposed method and to verify that the proposed method is robust when comparing with other ones. As mentioned in Section 1, the proposed method is able to estimate human head orientation with a high accuracy even though when the user is not facing the camera. Firstly, some necessary initializations are prepared for our experiments, and then, the experiments and the results are presented in detail.

### 5.1 Initializations

The proposed method was implemented in C++ programming language, and all experiments in this section were performed under a PC with Intel processor Core 2 Duo E8500 3.16GHz CPU and 2GB of RAM.

We used 147 head images with about  $5^\circ$  of angle step for training. With each image, the orientation of the head is manually trained before. Some training images



Fig. 5. Some training images

are shown in Fig. 5. The number of images we used in the experiments for testing was 152. As mentioned above, the proposed method is able to estimate human head orientation with a high accuracy even though when the user is not facing the camera; therefore in this dataset, the images do not consist of only frontal-view faces.

All head images were used for training and testing were resized to have a size of 33 x 33 pixels. The number of sampling points in each head image is 64. The HOG descriptor has three parameters: the number of orientation bins, cell size and block size. With each descriptor in this paper, four blocks were used with block size of 15 x 15 pixels. In each block, nine cells were used with cell size of 5 x 5 pixels. Fig. 6

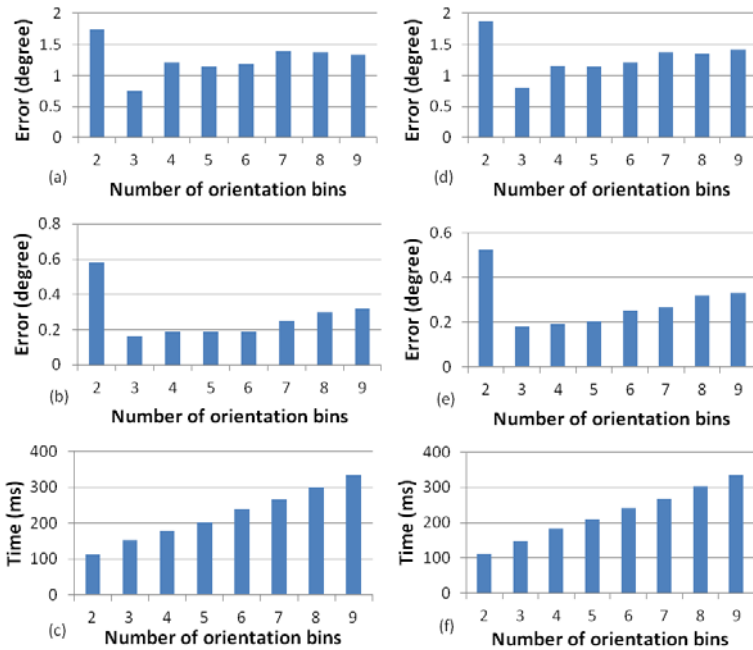


Fig. 6. Average head orientation estimation error and average estimation time with respect to changing the number of the orientation bins in each cell of the HOG descriptor. (a), (b), (c) for weighted rotation angle approach, (a) shows average error in yaw axis, (b) shows average error in pitch axis. (d), (e), (f) for unweighted rotation angle approach, (d) shows average error in yaw axis, (e) shows average error in pitch axis.

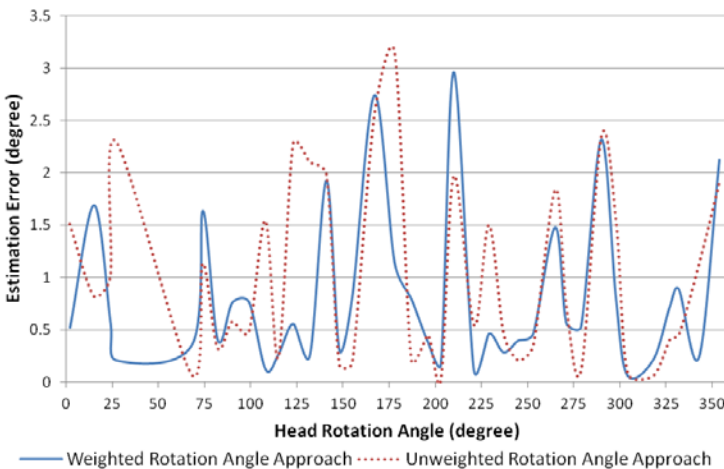


shows the average estimation error and average estimation time of head orientation of the selected dataset with respect to changing the number of orientation bins. The graph indicates that in both approaches, the average estimation time increases when the number of orientation bins increase. Since increasing the number of bins makes the dimension of the HOG descriptor larger, the ANN searching becomes therefore slower. In (a), (b), (d) and (e) of Fig. 6, since two orientation bins are not large enough for a good classification of pattern descriptors in the codebook of extended templates; therefore, results of the matching in Section 4.1 are not good and the average estimation error is the biggest. The average estimation error is the smallest when the number of bins is three. These graphs also indicate that the average estimation error tends to increase when the number of bins increases from three in both approaches. Due to these reasons, we chose three orientation bins to use in the experiments.

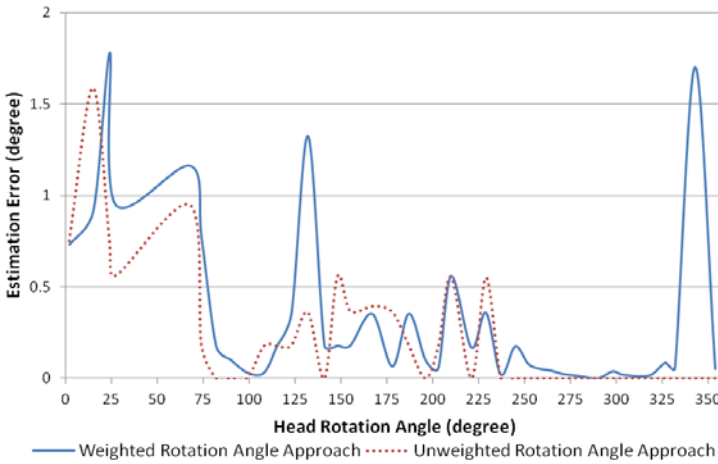
With each testing image, Codebook method was used to extract human head region. In our experiments, a video which has 460 frames and recorded at difference times was used for training codebook.

### 5.2 Results

When the number of orientation bins is three, Fig. 6 indicates that in both approaches, the average estimation error in yaw axis is smaller than  $1^\circ$  and the average estimation error in pitch axis is smaller than  $0.2^\circ$ . Since the range value of  $a_y$  is from 0 to 359 and the range value of  $a_p$  is from -90 to 90, this means that the range value of  $a_y$  is twice bigger than the range value of  $a_p$ ; therefore, the accuracy of ANNS in pitch axis is better than in yaw axis. Due to this reason, the average estimation error in pitch axis is smaller than in yaw axis. Fig. 6 also indicates that in two approaches, average estimation error of weighted rotation angle approach is smaller than unweighted one.



**Fig. 7.** Estimation error in yaw axis from  $0^\circ$  to  $359^\circ$  of yaw rotation angle (fixed pitch rotation angle to  $5^\circ$ ) for both approaches: weighted and unweighted rotation angle approaches



**Fig. 8.** Estimation error in pitch axis from 0° to 359° of yaw rotation angle (fixed pitch rotation angle to 5°) for both approaches: weighted and unweighted rotation angle approaches

Fig. 7 and Fig. 8 show estimation error in yaw and pitch axes when the human head rotation angle changes from 0° to 359° in yaw axis and the rotation angle in pitch axis is fixed to 5° for our both approaches. In Fig. 7 we can find that the estimation error in yaw axis of weighted rotation angle approach is better than unweighted one in most cases, but Fig. 8 indicates that the estimation error in pitch axis of unweighted rotation angle approach is better in most cases. Also, we can see that the estimation error of our both approaches is always smaller than 3.25° in yaw axis and 2° in pitch axis.

As mentioned above, the main limitation of feature based methods is that the estimating of head orientation is focused and applied for only near-frontal face. On the other hand, output angle resolution of most of appearance based methods is restricted to very large steps of angle, often at least 25°. Therefore, with the accuracy of our proposed method, it is unnecessary to do detailed experiment to compare with other methods.

## 6 Conclusion

We proposed a new robust method to solve problem of estimating head orientation using Histogram of Oriented Gradients that is able to estimate orientation of the head with a camera even though when a user is not facing the camera. With this method, a head orientation and its rotation can be estimated precisely in all three axes: roll, yaw, pitch.

The kind of codebook which was used in this paper is the offline codebook. This means that the codebook is made one time at the beginning of program running time, and it is not changed while the program is running. In the future, online codebook will be used to update the codebook; therefore, with a user, we can use the latest information of this person and the estimation accuracy will be better.

## References

1. Ballard, P., Stockman, G.C.: Controlling a Computer via Facial Aspect. *IEEE Trans. on Systems, Man and Cybernetics* 25(4), 669–677 (1995)
2. Horprasert, T., Yacoob, Y., Davis, L.S.: Computing 3-D Head Orientation from a Monocular Image Sequence. In: *Proc. Int. Conf. On Automatic Face and Gesture Recognition*, pp. 242–247 (1996)
3. Matsumoto, Y., Zelinsky, A.: An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement. In: *Proc. IEEE Int. Conf. on Automatic Face and Gesture Recognition*, pp. 499–504 (2000)
4. Dervinis, D.: Head Orientation Estimation using Characteristic Points of Face. *Electronics and Electrical Engineering* 72, 61–64 (2006)
5. Voit, M., Nickel, K., Stiefelhagen, R.: Neural Network-Based Head Pose Estimation and Multi-view Fusion. In: Stiefelhagen, R., Garofolo, J.S. (eds.) *CLEAR 2006*. LNCS, vol. 4122, pp. 291–298. Springer, Heidelberg (2007)
6. Zhang, Z., Hu, Y., Liu, M., Huang, T.: Head Pose Estimation in Seminar Room Using Multi View Face Detectors. In: Stiefelhagen, R., Garofolo, J.S. (eds.) *CLEAR 2006*. LNCS, vol. 4122, pp. 299–304. Springer, Heidelberg (2007)
7. Rae, R., Ritter, H.J.: Recognition of Human Head Orientation based on Artificial Neural Networks. *IEEE Tran. on Neural Networks* 9, 257–265 (1998)
8. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *Proceedings of the 2005 IEEE Computer Vision and Pattern Recognition*, pp. 886–893 (2005)
9. Kim, K., Chalidabhongse, T., Harwood, D., Davis, L.: Real-time Foreground-Background Segmentation using Codebook Model. *Real-Time Imaging in Special Issue on Video Object Processing* 11, 172–185 (2005)
10. Arya, S., Mount, D., Silverman, R., Wu, A.Y.: An Optimal Algorithm for Approximate Nearest Neighbor Searching. *Journal of the ACM* 45, 891–923 (1998)
11. Rousseeuw, P.J., Leroy, A.M.: *Robust Regression and Outlier Detection*. John Wiley and Sons, Inc., New York (1987)
12. Kozakaya, T., Shibata, T., Yuasa, M., Yamaguchi, O.: Facial Feature Localization using Weighted Vector Concentration Approach. *Journal Image and Vision Computing* 28, 772–780 (2010)
13. Lee, J.-H., Hashimoto, H.: Intelligent Space – Its Concept and Contents. *Advanced Robotics Journal* 16(4), 265–280 (2002)
14. Canton-Ferrer, C., Casas, J.R., Pardàs, M.: Head Orientation Estimation using Particle Filtering in Multiview Scenarios. In: Stiefelhagen, R., Bowers, R., Fiscus, J.G. (eds.) *RT 2007 and CLEAR 2007*. LNCS, vol. 4625, pp. 317–327. Springer, Heidelberg (2008)
15. Dalal, N., Triggs, B.: Histograms of Oriented Gradients for Human Detection. In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 886–893 (2005)

# Fusion of Gait and Facial Feature Using PCA

Sanjeev Sharma, Ritu Tiwari, Anupam Shukla, and Vikas Singh

ABV-Indian Institute of Information Technology and Management, Gwalior, India  
{sanjeev.sharma1868,tiwariritu2,dranupamshukla,  
vikas.singh97}@gmail.com

**Abstract.** Multimodal biometric has ability to improve the performance in biometric recognition system. In this paper we are using face biometrics to improve the performance of gait biometric. Here proposed multimodal biometrics use feature level fusion strategy. Here walking person face and gait data are captured in the form of image, and combination of this face and gait images PCA feature represent as combined feature vector. Principal component analysis algorithm is used to reduce the dimensionality of this PCA feature vector. At the testing phase we give the similar input vector which is the combination of both face and gait. Our experiments show that when we are using only gait feature of individual then its recognition rate is 67%. But when we combined gait with face then gait performance can be improved up to 90%. This system can be used in communities where automated method is require to determine the identity of individual.

**Keywords:** Human identification, Multimodal biometrics, Feature level fusion, Principal component analysis.

## 1 Introduction

Recognition of an individual is an important task to identify people. Identification through biometric is a better way because it associate with individual not with information passing from one place to another. Biometrics is a physiological or behavioral characteristic, which can be used to identify and verify the identity of an individual. Human identification based on biometrics refers to the automatic recognition of the individuals based on their physical and/or behavioral characteristics such as face, fingerprint, gait and spoken voice. Biometrics is getting important and widely acceptable nowadays because they are really personal / unique that one will not lose or forget it over time.

Biometrics is used as a way to tie a claim of identity to an individual in a way that is not easy to spoof. Historically passwords or PIN codes were used to authenticate a person's claim of identity; you entered your username as a claim of identity, and proved it by entering your password or PIN. Passwords have many problems associated with them, not least is the ease with which they can be shared and intercepted.

A biometric authentication does not require any physical token or memorized knowledge.

Gait is a special behavioral biometric used to authenticate people by the way they walk. The attractiveness of this technique relies in its unobtrusive properties, since individual are authenticate at certain distance without any need to big co-operation efforts. Although of its benefits, gait is not supposed to be very distinctive across individual and therefore it is not well suited for high security. Such types of biometric system that rely on the evidence of a single source of information for authentication, is known as unimodal biometric system.

In order for the biometrics to be ultra secure and to provide more than average accuracy more than one form of biometrics requires. Hence the need arises for the use of multimodal biometrics. This uses a combination of different biometrics recognition technologies. Multimodal biometrics technology uses more than one biometric identifier to compare the identity of the person. When designing a multimodal biometrics system, two factors should be considered: (a) the choice of biometric traits; (b) the level at which information should be fused. These systems are able to meet the stringent performance requirements imposed by various applications. They address the problem of non – universality, since multiple traits ensure sufficient population coverage [9].

## 2 Related Work

Multimodal biometric systems, are expected to be more reliable due to the presence of multiple, independent pieces of evidence.

Arun Ross and Anil Jain [1] describe the information fusion in biometrics. They describe the initial result on a multimodal biometrics system that uses face, fingerprint and hand geometry features for biometric verification purposes. They show that the multimodal biometrics may become even more evident in the case of a larger database of users.

In [2], face and palmprint feature level fusion for single sample biometrics recognition is describe, they extract discriminate feature using Gabor based image preprocessing and principal component analysis techniques. This experiment conclude three thing, first is normalized face and palmprint biometric features are suitable for fusion, because they have strong supplement, second is discriminant feature extraction based on Gabor and PCA transforms is effective to process face and palmprint images, third is distance-based separability weighting strategy can further improve the classification effect of biometrics fusion, and this strategy is well connected with the classification method.

In [3], fusion of face and speech is describe, they use modular neural network for recognition to reduce the dimensionality of attribute. Here they divide various attributes among various modules of the modular neural network. This limits their dimensionality without much loss in information. The integrator collects the probabilities of the occurrences of the various classes as outputs from these neural networks The integrator averages these probabilities from the various modules to get the final

probability of the occurrence of each class. The integrator selects the class that has the highest probability of occurrence. This class is returned as the output class.

In [4], the adaptive combination of multiple biometrics is employed to determine the optimal fusion strategy and the corresponding fusion parameters. They use score level fusion to ensure the desired system performance using a hybrid partial swarm optimization. His experimental results also confirm that the proposed score-level approach generates fairly stable performance and requires a smaller number of iterations to generate better performance as compared to the decision-level approach.

In [10], identity verification using face and speech is described. They give new category of classification. where the decision boundary is fixed but constructed to take into account how the distributions of opinions are likely to change due to noisy conditions; compared to a previously proposed adaptive approach, the proposed classifiers do not make a direct assumption about the type of noise that causes the mismatch between training and testing conditions.

Yang et. al[11], performed the fusion of biometrics based on fingerprint, hand geometry, and palm print. They use wavelet transformation to extract the features and use feature fusion and match score fusion to establish identity.

In [12], describe the multimodal biometrics fusion at feature level using face and palm print. They use Gabor based image processing is utilized to extract discriminant features, while principal component analysis (PCA) and linear discriminant analysis (LDA) are used to reduce the dimension of each modality. The output features of LDA are serially combined and classified by a Euclidean distance classifier.

### 3 Face and Gait Fusion Approach

For recognition process fusion of various biometrics traits is very important because it can improve performance and to provide more reliability. Biometrics fusion can be done at many levels. (a) Fusion at data or feature level: Here either the data itself or the feature extracts from various source are fused, (b) fusion at match score level: The score generated by multiple classifier pertaining to different modalities are combined, (c) fusion at decision level: Here different technique like majority voting are applied to take the decision. In proposed method we are using feature level fusion. In feature level fusion the data obtained from different is used to compute a feature vector, the feature extracted from one biometric trait are independent are of those extracted from other. So we can create one biometric feature vector for both biometrics traits.

In the case of our system we are using two different traits of biometrics, gait and face. Here we are improving gait recognition performance by fusion it with face. Feature level fusion technique is using here to combine the feature of gait and face. We are using principal component analysis method for recognition, which is used to reduce the large dimensionality of the data space to the smaller dimension of the feature space. Here both face image and gait image of walking person to extract. The Eigen value of these images is extracted using principal component algorithm. During recognition process gait and face image are given as input to identify the individual.

## 4 Principal Component Analysis

Recognition process is depends on the extracted features. Two basic feature extraction techniques are classified as feature based and holistic method [5]. Feature based method select individual features and characterization of geometrical relationship. Holistic method such as principal component, linear discriminant analysis and independent component analysis used appearance information extracted from entire image. Holistic feature extraction methods find feature with reduced dimensionality by projecting and original data onto the basic vectors. These extracted features can improve classification performance by reducing irrelevant feature from the data set. Selection of important features and eliminate irrelevant features play important role in recognition and computation process.

Principal component analysis is one method to reduce the dimensionality of variable. It can also be called Karhunen Loeve transform, named after kari Karhunen and michell Loeve. It is a feature extraction and data reduction methods which extract data, remove redundant information, highlighting hidden feature, and show the relationship that exist between observations [6]. A simple approach is here to extracting the information contained in image of face and gait, this extracted information show the variance of image, this information is needed to encode and compare individual face images. Mathematically we wish to find the principal component of the distribution of image, or the of the covariance matrix of the set of images [7]. Extracted information through principal component analysis is called eigen vector. These Eigen vector are ordered, which show the different amount of variance of images. Recognition through PCA can be described as follow [8].

1. First takes sets of training images. Each of these images in a form of 2-d matrix, where each matrix point represents the pixel value of image. In PCA these 2-d matrix can be represent in 1-d vector by concatenation of each row(of column) into a long vector. Lets we have image of size y rows and z column. This image can be represent as p size vector, where P=y×z.

Let us taking K subject image, so set of sampled image, can be represent as

$$S_i = [p_1, p_2, \dots, p_p]^T \quad i=1,2,\dots,K \tag{1}$$

2. Computing mean image by using the following formula

$$m = \frac{1}{K} \sum_{i=1}^K S_i \tag{2}$$

3. Computing the deviation of each image from mean image by subtracting mean image from each image

$$w_i = S_i - m \tag{3}$$

4. Next we are trying to find a set of Eigen value  $e_{ij}$  which has the largest projection on to each of  $w_i$ 's. For finding set of  $K$  orthonormal vector  $e_i$ , the following quality to be maximized.

$$\lambda_i = \frac{1}{K} \sum_{n=1}^K (e_i^T w_n)^2 \tag{4}$$

Orthonormality constraints is

$$e_i^T e_k = \delta_{lk} \tag{5}$$

So the convergence matrix can be defined as-

$$C = WW^T \tag{6}$$

Where  $W$  is the matrix composed placing  $w_i$  side by side. Size of  $W$  is  $P \times P$ . Let  $d_i$  and  $\mu_i$  be the eigen vector and eigen value of  $W^T W$

$$W^T W d_i = \mu_i d_i \tag{7}$$

Multiply  $W$  at both sides

$$WW^T (W d_i) = \mu_i (W d_i) \tag{8}$$

Which shows that first  $K-1$  eigen value and its corresponding eigen vector is given by

$$e_i = W d_i \tag{9}$$

$$\lambda_i = \mu_i \tag{10}$$

This Eigen vectors to be normalize. So the Eigen vector corresponding to non zero Eigen value of covariance matrix represent the most relevant data of image. Put these Eigen vector in decrease order, which show highest to lowest variance in image.

5. A facial image can be projected onto  $K' (<< K)$  dimensions by computing.

$$\Omega = [v_1 v_2 \dots v_{K'}] \tag{11}$$

Where  $v_i = e_i^T w_i v_i$  is the  $i^{th}$  coordinate of the facial image in the new space, which came to be the principal component. describes the contribution of each eigenface in representing the facial image by treating the eigenfaces as a basis set for facial images. The simplest method for determining which face class provides the best description of an input facial image is to find the face class  $l$  that minimizes the Euclidean distance.



$$\varepsilon_l = \|(\Omega - \Omega_l)\| \quad (12)$$

Where  $\Omega_l$  is a vector describing the  $l^{\text{th}}$  face class. If  $\varepsilon_l$  is less than some predefined threshold  $\theta_\varepsilon$  a face is classified as belonging to the class  $l$ .

## 5 Fusion of Feature

Feature is defined as a function of one or more quality of objects that are capable to distinguish objects from each other. Gait image feature vector is comprised of parameters of moment features in image regions containing the walking person. Face features represents the certain point on face that can classify individual.

At preprocessing stage all face and gait images were converted to gray scale images. Here gait image is in silhouette form, the silhouette appears to be a good feature to utilize since it captures the motion of most of the body parts and also encodes structural as well as transitional information. Particularly, it is independent of the clothing, illumination and textures etc. Since now we have database in silhouette form, silhouette show most of the body part so we can extract feature from this silhouette.

Here we are combining feature of face and gait. Let a face image  $f(x,y)$  and gait image  $g(x,y)$  be a two dimensional  $N_1 \times N_2$  array of 8 bit intensity value. So each of these images can be considered as  $N_1 \times N_2$  vector. Here image size is  $320 \times 240$ . So created vector size for 1 image is 76800. For calculating combined feature we are taking group of image one for face, one for gait. One face image and one gait image by principal component analysis represent the one column vector to train. In our recognition process both of this one face and gait image represent in 1-d vector format and represent one person. Image dimension is same for both face and gait so fusion based feature vector size is  $320 \times 240 \times 2$ . For each subject we will create 1-d vector, create matrix for  $n$  subject, where each column correspond to the combined feature of face and gait. These vectors are mapped into subspace using principal component analysis method. During the testing faces we are using the combined face and gait image and project it with the existing vector.

## 6 Recognition System

Figure shows the overall working of the proposed recognition system. Firstly face and gait database is collected. In our multimodal method four group of face and gait is taking as training database set and one group of face and gait using as testing dataset. Preprocessing operation are applied on this dataset to improve the quality of image. For training dataset extract the PCA feature, which is the most relevant feature of face and gait, and extracts the feature vector for training data set. Testing data set can be projected with existing image PCA feature and find the feature vector for testing dataset itself. For recognition process PCA employ classifier, this classifies data based on Euclidean distance.

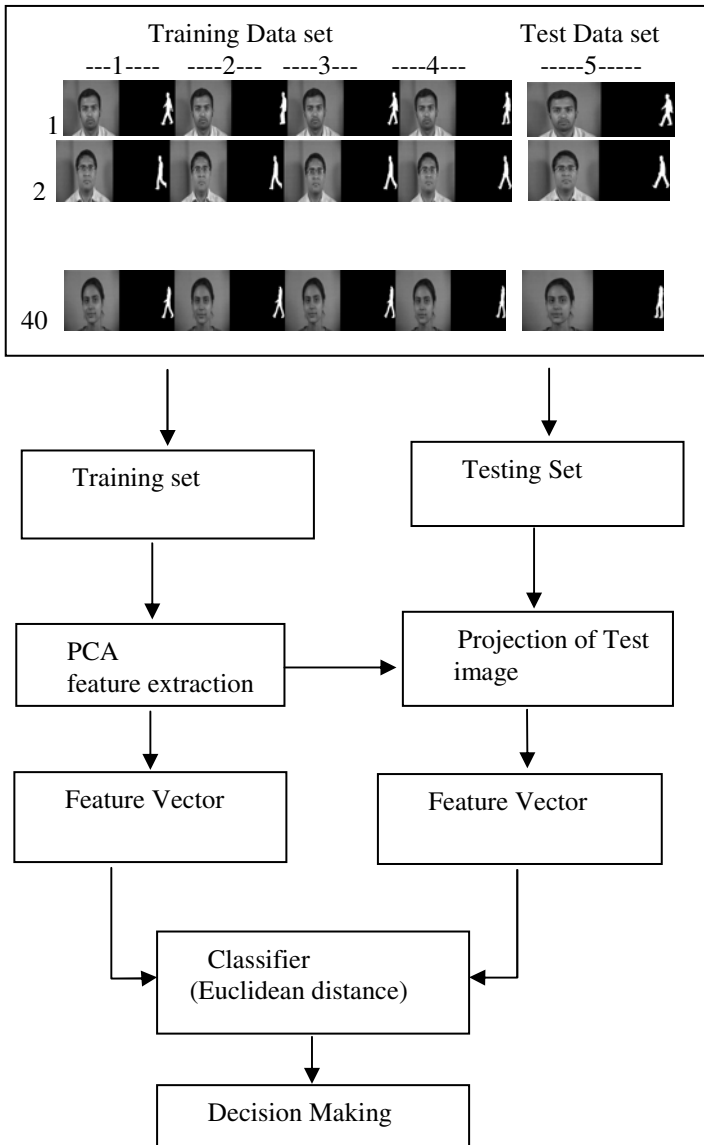


Fig. 1. recognition system

## 7 Experiment and Result

we are using face database prepared in Indian institute of Information Technology and Management Gwalior India, this data base contain 403 subject, which consist 325 male

subject database and 78 female subject database. This database is available in different angle. For our work we are taking 40 subjects database of them are male subject and remaining are female subject. We are taking five front view of face image.

Other database we are taking here is casia gait database[13]. This database is divided into three sets, Dataset A, Dataset B (multiview dataset) and Dataset C (infrared dataset). Here we are using data set B. Data set B contain 124 subjects and the gait data was captured in 11 views. In this experiment we are taking 4 subject gait Data. For each subject we are taking 5 frame moving from right to left.

We are using matlab 8 to code principal component analysis algorithm. Our result shows that recognition rate only for gait is 67%, but when we combined gait with face then its recognition rate is 90%, which is reasonably good, as compare to use only Gait. Table shows the result of both unimodel and multimodal biometrics.

**Table 1.** Result Comparrison

Traits	Algorithm	Result
Only Gait	PCA	67%
Only face	PCA	92%
Gait Fusion with Face	PCA	90%

## 8 Conclusion and Future Work

Recognition from gait is very effective, because it has the ability to recognition individual from distance, and without their cooperation, but recognition only through gait not always give good result. So we can use multimodal biometrics trade where we can combine gait with other biometric like face. This paper presents a improved gait recognition approach using PCA, Gait performance can be improved by using multimodal biometrics and improved recognition result is 90%.

This work can be extended to develop new multimodal biometric system, where we can combine gait other biometric traits like finger, hand, speech and iris. We can also use other recognition algorithm.

## References

1. Ross, A., Jain, A.: Information fusion in biometrics. *Pattern Recognition Letters* (24), 2115–2125 (2003)
2. Yaoa, Y.-F., Jingb, X.-Y., Wong, H.-S.: Face and palmprint feature level fusion for single sample biometrics recognition. *Neuro Computing* 70(7-9) (March 2007)
3. Kala, R., Vazirani, H., Shukla, A., Tiwari, R.: Fusion of Speech and Face by Enhanced Modular Neural Network. In: Prasad, S.K., Vin, H.M., Sahni, S., Jaiswal, M.P., Thipakorn, B. (eds.) *ICISTM 2010. CCIS*, vol. 54, pp. 363–372. Springer, Heidelberg (2010)
4. Kumar, A., Kanhangad, V., Zhang, D.: A New frame work for adaptive multimodel biometrics management. *IEEE Transactions of Information Forensics and Security* 5(1) (March 2010)

5. Na, J.-H., Park, M.-S., Choi, J.-Y.: Pre-clustered principal component analysis for fast training of new face databases. In: Proc. Int. Conf. Control, Autom. Syst., pp. 1144–1149 (2007)
6. Zhao, G., Liu, J.: Application of principal component analysis and neural network on the information system evaluation. In: Pacific –Asia Conference on Circuits, Communications and System, pp. 785–788 (2009)
7. Turk, M., Pentland, A.: Eigenfaces for recognition. *Journal of Cognitive Neuroscience* 3, 71–86 (1991)
8. Kim, K.: Face Recognition using principal component analysis, USA (June 2000)
9. Ross, A., Jain, A.K.: Multimodal Biometrics: An Overview. In: Proceedings of the 12th European Signal Processing Conference, pp. 1221–1224 (2004)
10. Sanderson, C., Paliwal, K.K.: Identity verification using speech and face information. *Digital Signal Processing* 14(5), 449–480 (2004)
11. Yang, F., Ma, B., Wang, Q., Ya, D.: Information Fusion of Biometrics Based-on Fingerprint, Hand-geometry and Palm-print. In: 2007 IEEE Workshop on Automatic Identification Advanced Technologies, pp. 247–252 (2007)
12. Ahmad, M.I., Woo, W.L., Dlay, S.S.: Multimodal biometric fusion at feature level: Face and palmprint. In: 2010 7th International Symposium on Communication Systems Networks and Digital Signal Processing (CSNDSP), July 21-23, pp. 801–805 (2010)
13. CASIA Gait Database,  
<http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp>

# Shockwave Velocity Estimation from Laser Induced Breakdown Images

Jaemyoung Lee

Korea Polytechnic University  
2121 Jungwang Shihung Kyunggi, Korea (ROK)  
lee@kpu.ac.kr

**Abstract.** Laser-induced breakdown at early time is analyzed in terms of the shockwave velocity using an Nd:YAG laser. The shockwave velocities are estimated by capturing the shockwave propagation images of a 20  $\mu\text{m}$  glass microsphere through the Schlieren method. The velocity estimation from images shows that the forward shockwave is faster than the backward. The forward and backward velocities within 30 ns from the breakdown are approximately 5.5 Km/s and 3.5 Km/s, respectively.

**Keywords:** laser-induced breakdown, ablation, shockwave, Schlieren image.

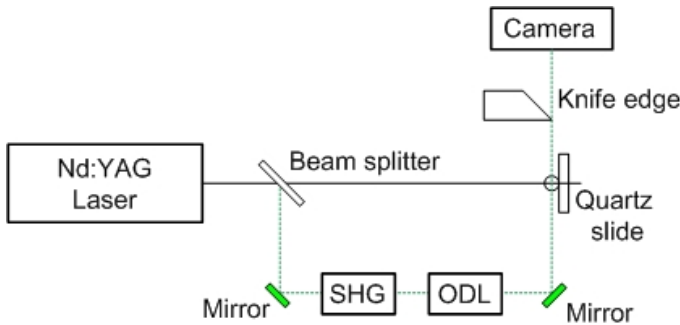
## 1 Introduction

Pulsed laser ablation is a key technology in many areas such as micro-machining, nanoparticles fabrication, biological tissue removal, etc. [1–3] However, the mechanism of the laser-induced breakdown is not well known. Lots of researches have been conducted in both numerical analysis and experiment to understand the laser ablation dynamics by investigating shockwave propagation generated by the laser-induced breakdown. Analyzing the shockwave propagation has given us some clues regarding the initial energy distribution of laser-induced breakdown for certain material, energy expansion after the breakdown, shockwave structure, temperature transition after the breakdown and behind the shockwave propagation. In laser-induced ablation, a large amount of energy is dumped on a material for the applied laser's temporal bandwidth, establishing a high temperature and high pressure region in or outside of the material. The high temperature and high pressure region generates high temperature and pressure plasma which initiates the breakdown. The temperature and the pressure of the plasma are known to be larger than 10,000 K and 1 GPa. [4]. However, detailed information about the shockwave from the breakdown has not been fully analyzed.

To examine the initial breakdown process, we conducted laser-induced breakdown of glass microsphere of 20  $\mu\text{m}$  diameter and obtained shockwave propagation images through the Schlieren method. The measured forward and backward velocities of the shockwave from the glass microsphere are about 5.5 Km/s and 3.5 Km/s, respectively.

## 2 Shockwave Experiment

We investigated laser-induced breakdown of glass microsphere of 20  $\mu\text{m}$  diameter at 6  $\text{J}/\text{cm}^2$  using an Nd:YAG laser of 1.06  $\mu\text{m}$  with a 10 ns pulse width. The experiment setup for catching the shockwave propagation is shown in Fig. 1. A beam splitter divides the beam into the fundamental Nd:YAG laser beam to ablate the glass microsphere and the second harmonic generated probe beam. The probe beam of 533 nm was sent through a variable length optical delay line (0 - 30 ns) and illuminated the microsphere from the side. The probe beam projects the shockwave propagation onto a CCD camera using the Schlieren optical setup.



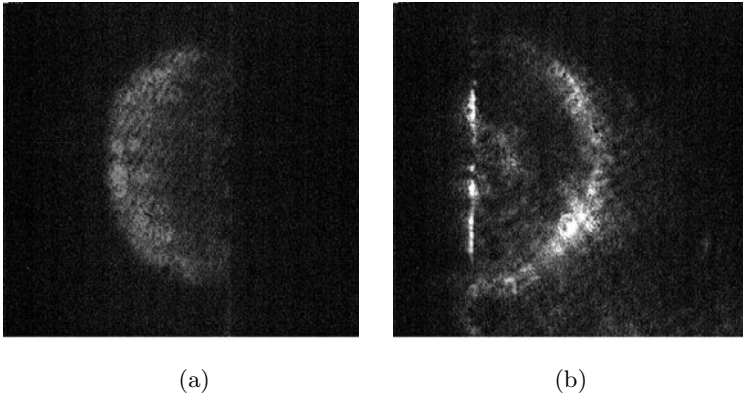
**Fig. 1.** Optical setup to catch the shockwave propagation through Schlieren method. SHG: Second harmonic generation, ODL: Optical delay line.

## 3 Shockwave Velocity Estimation

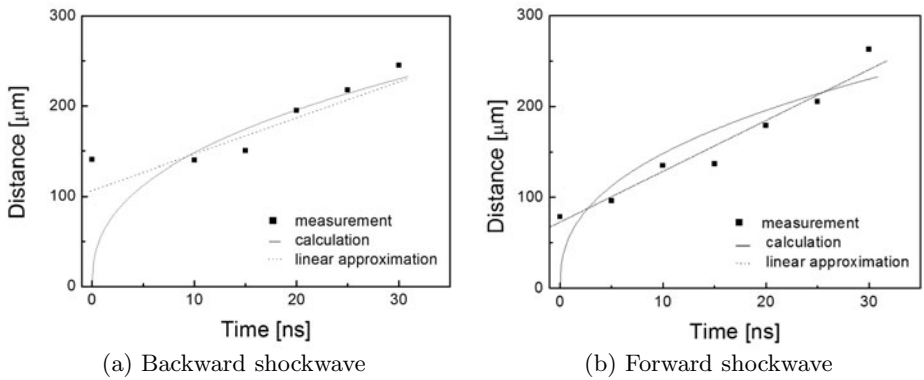
In our experiment, the captured forward and backward shockwave propagation images through the Schlieren method were used to estimate the shockwave velocities for a laser fluence of 6  $\text{J}/\text{cm}^2$ , Fig. 2. To estimate the shockwave velocities, we measured the distance traveled by the shockwave from the initial breakdown point to the shockwave front along the laser axis from 0 ns to 30 ns with a 5 ns time interval, Fig. 3. The initiation point for breakdown was assumed to be the backside of particle as determined from the energy enhancement calculation of Lorenz-Mie theory [5]. Figure 3 shows the measured distance of the shockwave front from the initiation point with calculation of Zel'dovich's equation [5], and linear approximations. The measured experiment data does not match with the following Zel'dovich equation,

$$R \propto t^{2/5}. \quad (1)$$

where  $t$  is time. While other groups' results of [6, 7] which are measured on a larger than micro second time scale show the shockwave radius dependency on



**Fig. 2.** Shockwave propagations of (a) backward direction and (b) forward direction at 30 ns after the laser induced breakdown for  $6 \text{ J/cm}^2$  (The laser beam propagates from left to right)



**Fig. 3.** Shockwave propagation distance vs. time for  $6 \text{ J/cm}^2$

time with the power of  $2/5$ , it is hard to observe this tendency within our 30 ns time scale. The measured data in our experiment roughly agree with linear approximations. on an early time scale, about constant velocity expansion are observed in Ref. [4, 8, 9]. Distances traveled by the forward shockwave were larger than for the backward propagating shockwave and the measured forward and backward velocities are  $5.5 \text{ Km/s}$  and  $3.5 \text{ Km/s}$ , respectively, in our experiment.

## 4 Conclusion

We examined the shockwave propagation in early time, conducting laser-induced breakdown of a glass microsphere of  $20 \mu\text{m}$  diameter and obtained shockwave propagation images through the Schlieren method from 0 ns to 30 ns. The

measured distance from initiation position of the breakdown to the front end of the shockwave along the laser axis were used to get the shockwave velocities, which can be linearly approximated. The measured forward and backward shockwave velocities are 5.5 Km/s and 3.5 Km/s, respectively, in our experiment. Further research is required for better understanding of the laser-induced breakdown dynamics on a larger time scale.

## References

1. Lee, J., Becker, M.F., Brock, J.R., Keto, J.W., Walser, R.M.: Permalloy nanoparticles generated by laser ablation. *IEEE Trans. Magnetics* 32, 4484–4486 (1996)
2. Dubowski, J.J., Mazumder, J., Migliore, L.R., Roychoudhuri, C.S., Schaeffer, R.D.: Lasers as Tools for Manufacturing of durable Goods and Microelectronics. In: *Proc. SPIE*, vol. 2703 (1996)
3. Thalhammer, S., Lahr, G., Clement-Sengewald, A., Heckl, W.M., Burgemeister, R., Schutze, K.: Laser Microtools in Cell Biology and Molecular Medicine. *Laser Physics* 13, 681–691 (2003)
4. Wen, S.-B., Mao, X., Liu, C., Greif, R., Russo, R.: Expansion and radiative cooling of the laser induced plasma. *J. Physics* 59, 343–347 (2007)
5. Zel'dovich, Y.B., Raizer, Y.P.: *Physics of Shock Waves and High-Temperature Hydrodynamic Phenomena*. Academic Press, New York (1966)
6. Hauera, M., Funk, D.J., Lippert, T., Wokaun, A.: Time-resolved techniques as probes for the laser ablation process. *Optics and Lasers in Engineering* 43, 545–556 (2005)
7. Koch, J., Heiroth, S., Lippert, T., Günther, D.: Femtosecond laser ablation: Visualization of the aerosol formation process by light scattering and shadowgraphic imaging. *Spectrochimica Acta Part B* 65, 943–949 (2010)
8. Celliers, P., Ng, A., Xu, G., Forsman, A.: Thermal equilibration in a shock wave. *Phys. Rev. Lett.* 68, 2305–2308 (1992)
9. Gonzalo, J., Afonso, C.N., Madariaga, I.: Expansion dynamics of the plasma produced by laser ablation of BaTiO<sub>3</sub> in a gas environment. *J. Appl. Phys.* 81, 951–955 (1997)



# A Temporal Item-Based Collaborative Filtering Approach

Lei Ren<sup>1,2</sup>, Junzhong Gu<sup>2</sup>, and Weiwei Xia<sup>2</sup>

<sup>1</sup> Department of Computer Science & Technology,  
Shanghai Normal University

<sup>2</sup> Department of Computer Science & Technology,  
East China Normal University, Shanghai, China  
renlei@shnu.edu.cn, {jzgu, wwxia}@ica.stc.sh.cn

**Abstract.** Item-based collaborative filtering is becoming the most promising approach in recommender systems. It can predict an active user's interest for a target item based on his observed ratings. With the user's interests changing during interacting with collaborative filtering, the issue of concept drift is becoming a main factor impacting the accuracy of recommendation. Aiming at the issue of concept drift, we propose a temporal item-based collaborative filtering approach, in which the temporal weight is employed in both similarity computing and rating prediction. As the experimental result shows, the proposed approach improves the quality of recommendation in contrast to the classic item-based collaborative filtering.

**Keywords:** Recommender system, Item-based collaborative filtering, concept drift, temporal similarity, temporal prediction.

## 1 Introduction

With the rapid development of web applications, numerous information systems have emerged during the last two decades. Such systems as E-Commerce stores, digital libraries and social networking sites provide an overwhelming quantity of information. Facing those numerous information, users have to spend a lot of time and efforts on selecting information according to their individual interests, and this is termed as "information overload" [1]. To address the issue, recommender systems have been proposed in academia and industry, and have been becoming a core component in adaptive information systems [1, 2]. Recommender systems can produce individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful objects in a large space of possible options [3].

Item-based collaborative filtering (IBCF) is the most promising approaches employed in recommender systems [4]. IBCF follows the assumption that users will prefer items similar to those which they appreciated previously. To find similar items, similarities between items are computed offline based on the historical ratings by users. Various IBCF approaches have been adopted in many fields, yet in most of those existing approaches, with the active user's interests and preferences changing

radically or gradually over time, all his observed ratings are treated equally in the process of similarity computing and rating prediction and the outdated ratings deteriorate the quality of recommendation. This issue is coined as concept drift firstly proposed in machine learning [5].

Some works have been done to deal with concept drift in the context of recommender systems [6-8]. Most of the existing works improve the accuracy of collaborative focusing on time-weighted rating prediction, yet the improvement of similarity computing is not taken into full consideration. In this work, we argue that the similarity considering the temporal feature can also play an important part in improving the quality of recommendation.

Aiming at the issue of concept drift in IBCF, a temporal item-based collaborative filtering approach (TIBCF) is proposed in this work. For the proposed approach, a linear time weight scheme of users' observed ratings is employed in both the similarity computing and rating prediction. As the experimental result shows, the recommendation quality of TIBCF is more accurate than the classic IBCF.

The rest of this paper is organized as follows: related works are reviewed in section 2. Next, section 3 elaborates the proposed approach. Then in section 4, TIBCF is evaluated experimentally. Finally, conclusions are made in section 5.

## 2 Related Works

As an intelligent system, recommender system can work as an independent system, but more often than not it is integrated with the host information system as an intelligent component. IBCF takes each column of the user-item matrix as item profile based on which the similarity between items can be computed. The similarity between two given items indicates their distance in rating space with respect to those users who have rated both involved items. Based on the derived similarity, the  $k$  nearest items to the target item are selected from the active user's visited items as its neighborhood. For an active user, the rating for each of his unvisited items is estimated by a weighted sum of the observed ratings for its neighbors [4].

The item-based similarity is a key reference in the steps of neighbor selection and rating prediction. In existing IBCF, two common similarity measures have been used frequently, including Pearson correlation coefficient (PCC) and Cosine similarity (COSIM) [4, 9]. The item-based COSIM takes the ratings for an item as a vector in the user space, and it is the cosine of the angle between two involved items' rating vectors. Whereas as the most frequently used similarity measure, the item-based PCC indicates the linear correlation between the ratings for two involved items, and has been proved to perform more effective than COSIM [10]. The item-based PCC similarity  $sim(t_i, t_j)$  can be defined as

$$sim(t_i, t_j) = \frac{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_i)(r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U_{ij}} (r_{ui} - \bar{r}_i)^2} \sqrt{\sum_{u \in U_{ij}} (r_{uj} - \bar{r}_j)^2}}, \quad (1)$$

where  $U_{ij}$  is the user set in which each user has rated both the item  $i$  and  $j$ , and  $r_{ui}$  denotes the rating for the item  $i$  by the user  $u$ ,  $\bar{r}_i$  is the average rating of the item  $i$ .

The rating prediction for a user’s unvisited items is done based on his historical ratings. The unobserved rating for an active user’s unvisited item is usually estimated as a weighted aggregate of his observed ratings. For an active user  $u$ , the rating for his unvisited item  $i$  is denoted as  $r_{ui}$ , and the classic item-based rating prediction can be defined as

$$r_{ui} = \frac{\sum_{j \in T_{ui}} r_{uj} \cdot \text{sim}(t_i, t_j)}{\sum_{j \in T_{ui}} |\text{sim}(t_i, t_j)|}, \tag{2}$$

where  $T_{ui}$  is the neighbor item set for the item  $i$ , including the  $k$  most similar items to the target item  $i$  according to their similarities. To measure the significance of each observed rating in prediction, the derived similarity  $\text{sim}(t_i, t_j)$  is taken as its weight.

The user’s concept drift is a crucial issue in real-world recommender systems. The issue means that a user’s interest or preference may change depending on some hidden context, such as job, residence and marriage etc., but for the classic collaborative filtering the change is not traced and perceived. The corresponding interest model derived from existing ratings still follows the outdated examples, and is not updated in time. Then the delay will deteriorate the quality of recommendation. Fig. 1 shows an example of the impact by concept drift in collaborative filtering.

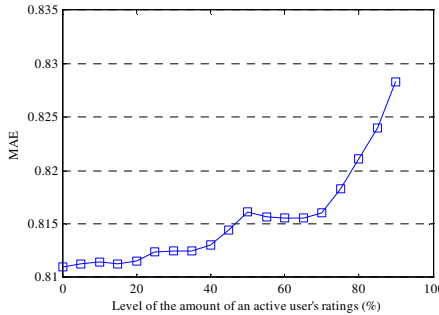


Fig. 1. An example of the impact by concept drift

To address the issue of concept drift in collaborative filtering, some works have been done [6-8, 11]. Ding and Li [7] propose an algorithm to compute the time-based weight for different items’ ratings, and the weight of each rating follows an exponential function of the produced time of the rating. In addition the decay factor of their proposed exponential function is cluster-dependent which is optimally selected from a range of possible value. Finally the weight is employed in rating prediction. Sugiyama et al. [6] use the contents of the active user’s historical visited pages to perform personalized web search, and the weights of terms gradually decay as days pass. In the proposed approach, they also adopt an exponential function of time as the forgetting factor, in which the importance of each user’s preference is assumed to be reduced by

1/2 every 7 days. Lee et al. [11] propose a refined collaborative filtering approach using a pseudo rating matrix. During the construction of the pseudo rating matrix, such temporal information as the item launch time and user rating time is incorporated with rating data. To estimate the ratings for unvisited items, the procedures of similarity computing and rating prediction are achieved based on the pseudo rating matrix.

### 3 The Proposed Approach

As discussed in section 2, the issue of concept drift in IBCF is a key factor influencing the quality of recommendation. Focusing on the improvement of recommendation accuracy, our proposed TIBCF takes the produced time of ratings as temporal feature which is employed in both similarity computing and rating prediction. TIBCF can be divided into four phases:

- Phase 1. Based on the historical ratings and their produced time, similarities between items are measured by a temporal similarity;
- Phase 2. For a target item, those  $k$ -percentage-nearest items in the active user’s visited items are selected as its neighbors according to their similarities;
- Phase 3. The rating for each of the active user’s unvisited items is predicted by a weighted sum of the observed rating for its item neighbor, in which both the derived temporal similarity and temporal weight of each neighbor act as the measure of importance for each neighbor.
- Phase 4. For an active user, a ranked unvisited item list with corresponding top- $n$  estimated ratings will be recommended to him.

In TIBCF,  $U=\{u_1, u_2, \dots, u_i, \dots, u_m\}$  denotes the user set with  $m$  users, and  $T=\{t_1, t_2, \dots, t_j, \dots, t_n\}$  is the item set with  $n$  items. All users’ observed ratings with their produced time are stored in a user-item rating matrix. The user-item rating matrix with  $m$  users and  $n$  items is represented as  $M(m \times n)$ , in which the element  $\langle r_{ij}, d_{ij} \rangle \in M$  indicates the rating for item  $t_j$  by user  $u_i$  and the corresponding produced time respectively. The value of rating  $r_{ij}$  is usually defined as an integer varying in a given range such as 1 to 5, or a continuous value is optional. The higher the rating is, the more the user prefers the item. Specifically, the minimum of 1 means “disliking the item”, on the contrary, the maximum of 5 means “preferring the item”. The rating value of 0 means the user has not rated the item which will be predicted by recommender system. Table 1 demonstrates an example of user-item rating matrix.

**Table 1.** An example of user-item rating matrix

	$t_1$	...	$t_j$	...	$t_n$
$u_1$	5 (1999/10/01)	...	4 (2003/01/15)	...	4 (1995/10/24)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u_i$	3 (1999/11/12)	...	5 (2005/01/21)	...	3 (2002/06/16)
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$u_m$	1 (1998/10/10)	...	2 (1999/02/03)	...	0 ( )

As the discussion in section 2, most of the existing collaborative filtering approaches deal with concept drift only in the phase of rating prediction [6, 7], yet we argue that the similarity play an important role in IBCF, so the temporal feature of ratings should be also employed in similarity computing. On the other hand, the time weight usually decay exponentially [7], but other decay functions of the time weight have not been tested by existing works and should be tested.

During the phase of similarity computing in TIBCF, a time-weighted PCC similarity measure is proposed, in which a time weight factor is integrated with the classic PCC. The time-weighted PCC similarity can be formulized as

$$sim(t_i, t_j) = \frac{\sum_{u \in U_{ij}} w_{ui} \cdot (r_{ui} - \bar{r}_i) \cdot w_{uj} \cdot (r_{uj} - \bar{r}_j)}{\sqrt{\sum_{u \in U_{ij}} w_{ui} \cdot (r_{ui} - \bar{r}_i)^2} \sqrt{\sum_{u \in U_{ij}} w_{uj} \cdot (r_{uj} - \bar{r}_j)^2}}, \quad (3)$$

where the weight  $w_{ui}$  denotes the temporal factor of the rating  $r_{ui}$ . The temporal factor is a real value in the range between 0 and 1, and it indicates the temporal importance of the corresponding rating  $r_{ui}$ . The temporal factor is associated with a linear decay function of the rating's produced time, and it is defined as

$$w_{ui} = 1 - \frac{RK(R_u, d_{ui}) \cdot (n-1)}{n \cdot |R_u|}, \quad (4)$$

Where the rating set  $R_u$  composes all the ratings of user  $u$ , and  $RK(R_u, d_{ui})$  is the sequence number of the rating time  $d_{ui}$  according to the ranked  $R_u$  in descending order.

In the phase of neighbor selection, most of existing CF approaches set the size of neighborhood as an absolute threshold. Whereas the amount of different users' visited items does not distribute identically, i.e. the neighborhood size is user-dependent. So an identical absolute threshold can't fit all users, and we set the neighborhood size  $k$  as a relative percentage of the amount of each user's visited items.

After the neighborhood of each item is identified, a temporal prediction method is employed to estimate the rating for each unvisited item. For an active user  $u$ , the rating for his unvisited item  $i$  can be predicted as

$$r_{ui} = \frac{\sum_{j \in T_{ui}} w_{ij} \cdot sim(t_i, t_j) \cdot r_{uj}}{\sum_{j \in T_{ui}} |w_{ij} \cdot sim(t_i, t_j)|}. \quad (5)$$

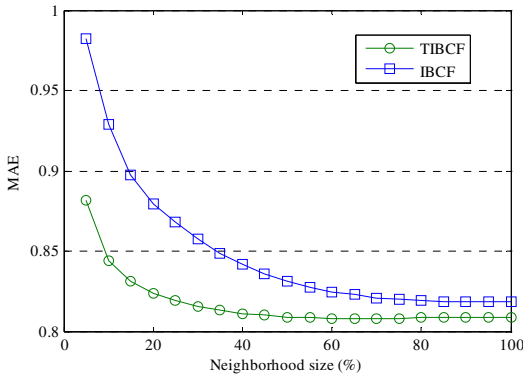
where  $T_{ui}$  is the neighbor item set of the item  $i$ . To measure the significance of the observed rating  $r_{uj}$  for each neighbor in  $T_{ui}$ , both the derived temporal similarity  $sim(t_i, t_j)$  and temporal factor  $w_{ij}$  are taken as its weight.

Finally, for the active user, TIBCF can produce a ranked item list with  $n$  top-ranked items according to their estimated ratings and recommend it to him.

## 4 Experimental Result

To empirically optimize the parameter and prove the effectiveness of TIBCF, three experiments are conducted. Focusing on the improvement of recommendation quality, TIBCF is mainly compared with the classic IBCF. The well-known dataset MovieLens [12] is adopted as the test data. The MovieLens dataset are composed of 100,000 anonymous discrete ratings scaling from 1 to 5 for approximately 1682 movies by 943 users, and the ratings for unvisited items is represented as 0. Each user in the data set has at least 20 ratings, and each item has at least one rating. In following experiments, mean absolute error (MAE) is employed as predictive accuracy metric, and the classic item-based collaborative filtering approaches based on PCC is set as the benchmark.

To optimize the size of neighborhood, a neighborhood size experiment compares the benchmark and the proposed approach at different levels of neighborhood size. The top-100 users are selected from the user list sorted by the amount of ratings in descending order, and the latest 20 ratings of those selected users are set as the test set. Both of the examined approaches are repeated with the neighborhood size varying from 5% to 100% by step of 5%. The result is demonstrated in Fig. 2.



**Fig. 2.** The comparison of neighborhood size

As Fig. 2 indicates, with the size of neighborhood increasing, MAE of both involved approaches tends to decrease. The proposed approach reaches the lowest MAE around 65% of neighbors, and then more neighbors help nothing in improving accuracy. In contrast, the classic IBCF need more neighbors to achieve prediction. So the proposed approach is more effective and the neighborhood size is specified as 65% in the following two experiments.

To test the impact of the quantity of rating, two experiments with different set are conducted, including the user experiment and the rating experiment. In the user experiment, by varying the quantity of test user, the proposed approach is compared with the classic IBCF, in which the latest 20 ratings of the top- $n$  users are set as the test set with the amount of test user  $n$  varying from 100 to 900 by step of 50. The result is illustrated in Fig. 3.

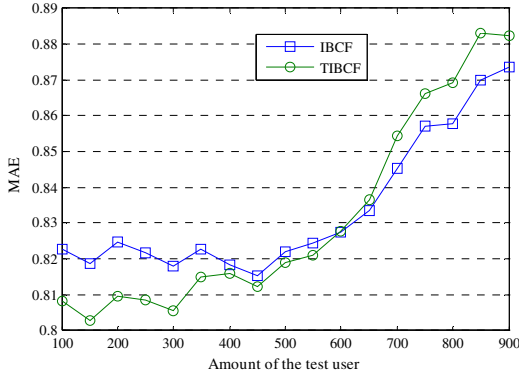


Fig. 3. The impact of the quantity of test user

The result shows that the proposed TIBCF is more effective than IBCF when the amount of test user is less than 600, and reaches the lowest MAE around 150 users, but after the amount exceeds 600, IBCF performs better than TIBCF. This means that TIBCF can produce recommendation for those users with sufficient observed ratings.

The rating experiment is set similarly to the user experiment, but the amount of test user is fixed at 150. For those selected users, their  $n$  latest items are selected as test data from the item list sorted by the ratings' produced time in descending order, and the parameter  $n$  varies from 20 to 150 by step of 10. The result is presented in Fig. 4.

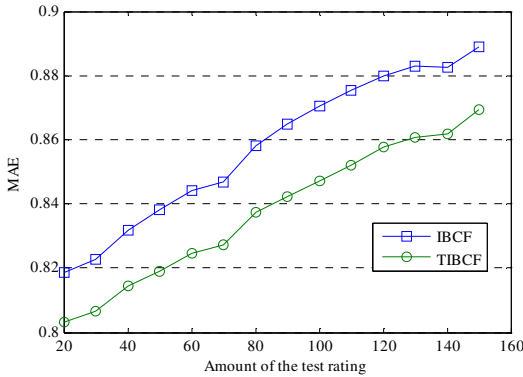


Fig. 4. The impact of the quantity of test rating

As the result shows, with the amount of test rating increasing, MAE of both examined approaches tend to increase, and TIBCF is always more accurate than IBCF at any level of the quantity of test rating, so more observed ratings will help improving the accuracy of recommendation. Moreover it is noticed that the variance between two involved approaches tend to increase, this indicates that TIBCF is less sensitive to the quantity of test rating than the benchmark.

## 5 Conclusions

Aiming at the issue of concept drift, a temporal item-based collaborative filtering approach is proposed in this work. In TIBCF, a linear temporal factor with respect to the produced time of ratings is employed in both similarity computing and rating prediction. The experimental results indicate that TIBCF needs a neighborhood size of 65% to reach its optimal recommendation, which is less than the one of the classic PCC. Compared with the classic IBCF, the proposed approach is suitable for making recommendation for those users producing sufficient ratings, and is more stable to the amount of available ratings.

**Acknowledgment.** This work is partly supported by the grant from STCSM project (No. 09510703000 and No. 08511500303) and SHNU project (No. A-3101-10-033). The authors wish to thank GroupLens for providing the MovieLens data set.

## References

1. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., Riedl, J.: GroupLens: an open architecture for collaborative filtering of netnews. In: Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, pp. 175–186. ACM, Chapel Hill (1994)
2. Adomavicius, G., Tuzhilin, A.: Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions. *IEEE Trans. on Knowl. and Data Eng.* 17, 734–749 (2005)
3. Burke, R.: Hybrid Recommender Systems: Survey and Experiments. *User Modeling and User-Adapted Interaction* 12, 331–370 (2002)
4. Sarwar, B., Karypis, G., Konstan, J., Reidl, J.: Item-based collaborative filtering recommendation algorithms. In: Proceedings of the 10th International Conference on World Wide Web, pp. 285–295. ACM, Hong Kong (2001)
5. Widmer, G., Kubat, M.: Learning in the Presence of Concept Drift and Hidden Contexts. *Machine Learning* 23, 69–101 (1996)
6. Sugiyama, K., Hatano, K., Yoshikawa, M.: Adaptive web search based on user profile constructed without any effort from users. In: Proceedings of the 13th International Conference on World Wide Web, pp. 675–684. ACM, New York (2004)
7. Ding, Y., Li, X.: Time weight collaborative filtering. In: Proceedings of the 14th ACM International Conference on Information and Knowledge Management, pp. 485–492. ACM, Bremen (2005)
8. Koren, Y.: Collaborative filtering with temporal dynamics. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 447–456. ACM, Paris (2009)
9. Linden, G., Smith, B., York, J.: Amazon.com recommendations: item-to-item collaborative filtering. *IEEE Internet Computing* 7, 76–80 (2003)
10. Breese, J.S., Heckerman, D., Kadie, C.M.: Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In: Cooper, G.F., Moral, S. (eds.) Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43–52 (1998)
11. Lee, T.Q., Park, Y., Park, Y.-T.: A time-based approach to effective recommender systems using implicit feedback. *Expert Syst. Appl.* 34, 3055–3062 (2008)
12. GroupLens: MovieLens data set. GroupLens Research (2011)



# A New Fusion Algorithm for Dim Target Detection Based on Dual-Wave Infrared Images

Jin Liu\*, Shao-Hua Wang, and Hong-Bing Ji

School of Electronic Engineering, Xidian Univ., Xi'an, China  
jinliu@xidian.edu.cn

**Abstract.** In this paper, a new approach to dim targets detection based on dual-wave infrared images is proposed. Based on the salient features and regional relativity of dual-wave infrared images, the algorithm gives relative fusion detection for target. Firstly, the source dual-wave infrared images are decomposed by wavelet transform, and the salient features of the images of each band are used to determine the potential target area. Secondly, based on the relativity of dual-wave images, the resulted wavelet coefficients are weighted. Finally, a threshold segmentation method is used for dim target detection in the restructured fusion image. Compared with traditional algorithms, the experimental results show that the proposed algorithm makes full use of the dual-wave infrared images, inhibits the background carrier and improves the SNR and detects dim target in higher probability.

**Keywords:** image fusion, wavelet transforms, salient feature, regional relativity, dim target detection.

## 1 Introduction

In infrared images, since the attenuation caused by long-distance transmission, background clutter and sensor noise, the available target information in the image is very limited, so traditional detection methods based on single-band image can hardly achieve an ideal result[1][2][3]. In recent years, in order to improve the viability of infrared surveillance system itself, and expect to detect enemy targets early, many countries have launched target detection methods based on a dual-band imaging system. Currently, most of dual-band infrared surveillance systems use two atmospheric transmission windows with medium-wave (3~5 $\mu$ m) and long-wave (8~12 $\mu$ m). As a result of using two channels which are from different electromagnetic areas, the information contained has a greater degree of complementary and redundancy. Using the appropriate fusion method and the information from different sensors, we can greatly improve the ability of target detection and target identification of the imaging system[4][5].

Amount the current available image fusion methods, the method based on wavelet multi-resolution analysis shows good results and is obtained widespread

---

\* Corresponding author.

approval[6][7][8]. The current research focuses on the selection of fusion rules and fusion operator in the fusion process. The proposed fusion rules contain weighted average, local gradient, local energy methods and so on. For the purpose of improving the visual effects of images, most of these methods strive to retain as much image details as possible in the fusion process. In the paper, the proposed image fusion method based on salient feature and relativity of images is applied to the clutter suppression of the dual-band infrared images. In the selection of fusion rules, we focus on as greater degree as possible to enhancement targets and background suppression.

## 2 Wavelet Multi-resolution Analysis

Multi-resolution analysis which is also called Multi-scale analysis or Multi-resolution approximation is a beautiful theory of the Mallat development[9][10]. Multi-resolution wavelet transform makes the image decompose into approximate signals of different frequency bands and detail signals in multi-resolution layers, and the resolution in the decomposition of the index decreases by 2. The different resolution describes the different features of physical structure, and reflects different demands for details of images.

### 2.1 The Decomposition Principle of Two-Dimensional Wavelet

The corresponding two-dimensional image space  $L^2(\mathbb{R} \times \mathbb{R})$ , two-dimensional scaling function  $\phi(x, y)$ , considering the two-dimensional scaling function is separable, that is

$$\phi(x, y) = \phi(x)\phi(y) \tag{1}$$

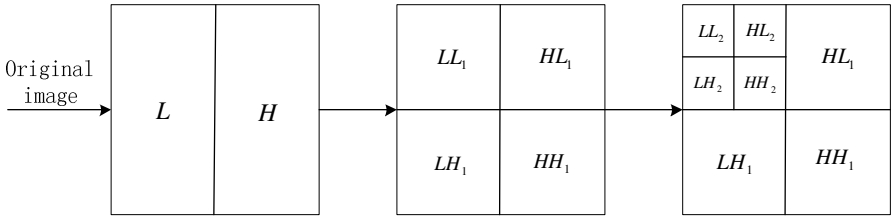
where  $\phi(x)$  is one-dimensional scaling function. If  $\Psi(x)$  is the corresponding wavelet  $\Psi(x) = \sum_k h_1(x)\phi(2x - k)$ , so the three two-dimensional wavelets as

follows

$$\begin{cases} \Psi^1(x, y) = \phi(x)\Psi(y) \\ \Psi^2(x, y) = \Psi(x)\phi(y) \\ \Psi^3(x, y) = \Psi(x)\Psi(y) \end{cases} \tag{2}$$

The foundation of two-dimensional wavelet transform is established, and the three wavelets extract image details in different scales and directions. At low horizontal frequency and high vertical frequency  $|\hat{\psi}^1|$  is large, at high horizontal frequency and low vertical frequency  $|\hat{\psi}^2|$  is large, and at high horizontal and vertical  $|\hat{\psi}^3|$  is large.

Based on two-dimensional wavelet, images can be expanded as follows. In the every level of the transform, the image is decomposed into four parts one-fourth the size of the image. As in the Fig. 1.



**Fig. 1.** Schematic diagram of the wavelet decomposition

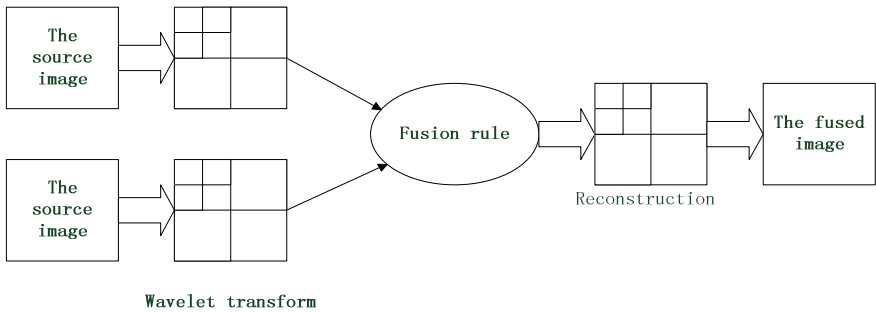
where L is low frequency , H is high frequency, subscript 1, 2 is the first or the second decomposition. In each layer, the image is decomposed into LL, LH, HL and HH bands. We only deal with LL, the low frequency components, in the next layer of decomposition. Each of four sub-images is an inner product between a source image and a wavelet basis function, and then generated after the 2 times sampling interval in the x and y directions. Inverse transform, also called the image reconstruction, is achieved by increasing frequency sampling and convolution of images.

**2.2 Image Fusion Based on Wavelet Decomposition**

In general, the basic steps of the decomposition fusion algorithm based on wavelet transform are as follows:

- Step 1:** conduct wavelet decomposition on the two source images using the selected wavelet basis, and then establish the wavelet decomposition pyramid.
- Step 2:** Using appropriate fusion rules, fuse coefficients of every level and various types of the image which is wavelet decomposed, and we can get the wavelet pyramid of the fusion image.
- Step 3:** Conduct inverse transformation of wavelet pyramid which has been fused, then we can get the restructured fusion image.

Fig. 2. describes the process of fusion method.



**Fig. 2.** Image fusion process based on wavelet transforms

### 3 Image Fusion Based on Energy and Region Relativity

#### 3.1 Algorithm Theory

Through the analysis of dual-band infrared image characteristics, we can find that the target features are reflected more on the high frequency components of the image which has been wavelet decomposed. Specific performance: relative to its surrounding areas, the target of gray value is higher and with stronger significant features. Therefore, we can determine the possible areas of the target by calculating the significant features of high frequency components in the infrared images. Through further investigation of the infrared dual-band image, two images with high frequency region that are nearby the target map has a strong correlation, and because of the random noise, the correlation of high frequency points generated by the noise is lower in the two images. Therefore, we can distinguish between noise and objectives by the related characteristics of the image, so as to achieve the purpose of noise suppression and enhancing target.

#### 3.2 Fusion Rules

Conducting  $N$  layer wavelet decomposition with the image, we can get  $3 \times N$  high frequency sub-graphs and one low sub-graph. Images in different frequency bands reflect different types of image information, so we use different rules to process the two types of sub-graphs respectively. The low frequency sub-graph reflects mainly about the slowly varying features of the image and the background, but we only concern with the target area and the high frequency characteristics, therefore, a simple principle of small is taken for the low frequency sub-graph.

For the high frequency sub-graph, the method to determine fusion coefficients is as follows:

**Step 1:** Calculate the significant features point by point in the high frequency sub-graphs of the two source images.

Significant features of the image are defined as the energy generated by the coefficient of neighborhood

$$S(i, j, k) = \sum_m \sum_n c(i+m, j+n, k)^2 \quad (3)$$

where  $(i, j)$  is the coordinates of wavelet coefficients  $c(i, j)$ ,  $k$  is decomposition series,  $(m, n)$  defines a window, contains the wavelet coefficients around the coefficients which is being processed. In general, the window can be  $3 \times 3$  or  $5 \times 5$ . By calculating the significance features of each pixel in the image, we can extract potential target points to prepare for the next fusion.

**Step 2:** Take high frequency sub-graphs at all levels of the image into block diagram, and calculate the correlation within the corresponding window in the high frequency coefficients of the two images.

The correlation coefficient between two images A and B in the  $m \times n$  window is written as

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2\right)\left(\sum_m \sum_n (B_{mn} - \bar{B})^2\right)}} \quad (4)$$

where  $\bar{A}$  and  $\bar{B}$  are the mean of two images windows.

**Step 3:** According to the correlation of the corresponding block of the image to determine the weighted way. For the area with large correlation that is related to potential target area, is weighted according to regional significant features. For the less relevant area, we think that the significant features are mainly generated by the noise, and inhibit them by selecting a smaller factor of the source image as the fusion coefficient.

$$\begin{cases} f(i, j) = \frac{1}{2} \left[ \left( c_a(i, j) \times \frac{S_a(i, j)}{S_a} \right) + \left( c_b(i, j) \times \frac{S_b(i, j)}{S_b} \right) \right] & \text{if } r(i, j) > \bar{r} \\ f(i, j) = \min(c_a(i, j), c_b(i, j)) & \text{else} \end{cases} \quad (5)$$

where  $c_a, c_b$  are the high frequency coefficients of two source images that have been decomposed.  $\bar{S}_a, \bar{S}_b$  are the significant features mean of the two images,  $\bar{r}$  is the mean of correlation coefficient.

**Step 4:** Reconstruct the low frequency coefficients and the fusion image coefficients to obtain the fusion image.

## 4 Experimental and Analysis

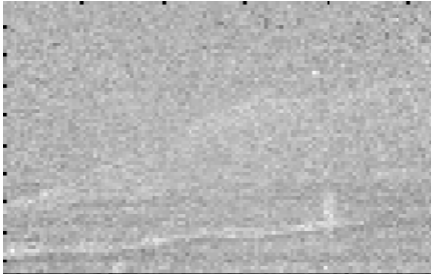
We use the measured dual-band infrared image sequence to simulate the proposed method of clutter suppression. Experiments use MATLAB 7.0 programming, and choose the sym4 wavelet to decompose the source image by degree 3-layer, the fusion results are as follows



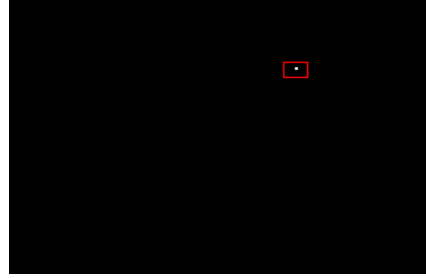
(a) Original long-wave sequences



(b) Original median-wave sequences



(c) Fusion image



(d) Dim target detection result

Based on the simulation experiment for the measured data, we use the image SNR and SNR gain as the indicators to inspect the performance of the algorithm. Indicators are defined as follows

$$SNR = (G_i - G_b) / \sigma \tag{6}$$

where  $G_i$  is the gray value of target,  $G_b$  represents the gray background,  $\sigma$  is the standard deviation of the background.

$$G_{SNR} = SNR_0 / SNR_I \tag{7}$$

where  $SNR_0$  is the output SNR,  $SNR_I$  is the input SNR of image.

We use the maximum median filter [11] and Top-hat morphological filter [12] to suppress clutter with the measured images, and compare the results with the proposed method in the paper, the results are as follows

**Table 1.** Performance of fusion algorithm

	original images		maximum median filter		morphological filter		proposed method	
	LW	MW	LW	MW	LW	MW	LW	MW
SNR	1.3999	2.3132	1.0507	0.2109	2.3094	1.7764	5.2146	5.2146
$G_{SNR}$	—	—	0.7506	0.0912	1.6497	0.7679	3.7250	2.2543

where LW represents long-wave, MW represents median-wave.

From Table 1, As the image contains lots of outline information, strong noise and clutter, the largest median filtering and morphological filtering are useless, and sometimes they even reduce SNR. The method proposed in the paper utilizes the dual-band image information, and can suppress image noise and improve SNR much better. Based on the proposed method, we detect the top 32 frame infrared dual-band image sequences frame by frame, and can find that the method greatly improves the detection probability. The results are shown as Table 2.

Firstly, we define as

$$\left\{ \begin{aligned}
 \text{Detection rate} &= \frac{\text{the right detected frame number of target}}{\text{the total number of frames of image sequences}} \\
 \text{Missed rate} &= \frac{\text{the undetected frame number of target}}{\text{the total number of frames of image sequences}} \\
 \text{False detection rate} &= \frac{\text{the wrong detected frame number of target}}{\text{the total number of frames of image sequences}}
 \end{aligned} \right. \quad (8)$$

**Table 2.** Probability of target detection

	<i>Detection rate</i>	<i>Missed rate</i>	<i>False detection rate</i>
original long-wave image sequence	0.4375	0.1875	0.5000
original median-wave image sequence	0.4375	0.4375	0.1250
fusion image sequence	<b>0.8750</b>	<b>0.0312</b>	<b>0.0937</b>

From Table 2, we can see that the small target detection method based on image fusion can greatly improve the detection rate and reduce the missed rate and false detection rate.

## 5 Conclusions

As a medium-wave infrared and a long-wave infrared are spectral characteristics distinct, and each image contains different information, so more information about the target can be obtained by choosing appropriate rules for image fusion. Based on the wavelet multi-resolution analysis, the paper proposes a detection method for image fusion using the significant features and regional correlation of images. The method takes full use of dual-band infrared image information, and overcomes the difficulties such as low SNR and small amount information of an image. All these set very good bedding for the subsequent processing. The measured data in the experiment shows that compared with the traditional single-sensor target detection, the method can extract the target information of a dual-band image effectively, and suppress noise and the interference of strong clutter much better. In the follow-up study, we may consider using the motion characteristics of the target in the multi-frame images to further enhance the stability and reliability of detection.

**Acknowledgement.** This paper is supported by the National Natural Science Foundation of China (NSFC) under Grant Number 61101246 and the Fundamental Research Funds for the Central Universities (2011).

## References

- Desai, U.B., Merchant, S.N., Zaveri, M., Ajishna, G., Purohit, M., Phanish, H.S.: Small Object Detection and Tracking: Algorithm, Analysis and Application. In: Pal, S.K., Bandyopadhyay, S., Biswas, S. (eds.) PReMI 2005. LNCS, vol. 3776, pp. 108–117. Springer, Heidelberg (2005)

2. Liu, W., Liu, Z.-h., Xiong, R.-s.: A Method to Detect IR Dim and Small Objects Based on Window. *ACTA Photonica Sinica* 37(3), 618–620 (2008) (in chinese)
3. Deng, H., Liu, J., Li, H.: EMD Based Infrared Image Target Detection Method. *Journal of Infrared, Millimeter and Terahertz Waves* 30(11), 1205–1215 (2009)
4. Li, Q.-h., Du, Y.: A Method For Two Color IR Weak Target Detection Based On The Asymmetric Fusion Strategies. *Signal Processing* 25(5), 713–719 (2009) (in Chinese)
5. Nunez, J., Otazu, X., Fors, O., et al.: Multiresolution-based Image Fusion with Additive Wavelet Decomposition. *IEEE Transactions on Geoscience and Remote Sensing* 37(3), 1204–1211 (1999)
6. Tran, M.D.-J., Abeynayake, C.: Evaluation of the Continuous Wavelet Transform for Feature Extraction of Metal Detector Signals in Automated Target Detection. In: Nakamatsu, K., Phillips-Wren, G., Jain, L.C., Howlett, R.J. (eds.) *New Advances in Intelligent Decision Technologies*. SCI, vol. 199, pp. 245–253. Springer, Heidelberg (2009)
7. Onur Karali, A., Erman Okman, O., Aytac, T.: Adaptive Image Enhancement Based on Clustering of Wavelet Coefficients for Infrared Sea Surveillance Systems. *Infrared Physics & Technology* (2011) (in press, Corrected Proof)
8. DelMarco, S., Agaian, S.: The Design of Wavelets for Image Enhancement and Target Detection. In: *Proceedings of SPIE, the International Society for Optical Engineering*, Proceedings, vol. 7351 (2009)
9. Mallat, S.G.: A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 11(7), 674–693 (1989)
10. Mallat, S.G.: A wavelet tour of signal processing, pp. 302–310. Academic Press, San Diego (1998)
11. Liu, S., Wei, X., Gao, X.: A Target Detection Method Based on Median Filter and Multi-Grad Search Algorithm. *Electronics Optics & Control* 18(2), 81–84 (2011) (in chinese)
12. Bai, X., Zhou, F.: Analysis of New Top-hat Transformation and The Application for Infrared Dim Small Target Detection. *Pattern Recognition* 43(6), 2145–2156 (2010)



# Fractal Analysis and the Effect of Aging on the Heart Rate and Breathing Frequency Relationship

Wilson Bucaoto, Han Jong Kim, and Artem Lenskiy

School of Electrical, Electronics & Communication Engineering,  
Korea University of Technology and Education,  
1800 Chungjeol, Byeongcheon, Cheonan, 330-708, Korea  
{wilson.bucaoto, a.a.lensky}@gmail.com,  
hjkim@kut.ac.kr

**Abstract.** Simultaneous observation of the Electrocardiogram (ECG) and respiratory cycle over long period has been proven clinically useful. Classical methods of analysing time-series, such as time and frequency domain analysis have been widely used in the area of physiological signal analysis. However, such characteristics as mean, standard deviation or location of peaks in the frequency domain are not capable of finding nonlinear relationship between signals. More sophisticated properties are based on estimation of multifractal spectrum. This spectrum is capable of confirming the existence of nonlinear relationships in analysed signals. In this study we estimated the spectrum of scale exponents for heartbeat dynamics and respiration dynamics to investigate the cardiorespiratory relationship under supine resting condition between young and elderly people. Using two border and one peak scale exponents selected from each spectrum we constructed 3-dimensional feature space. The distance and the angle of a line connecting features estimated from heartbeat and respiratory dynamics for each patient were calculated and compared in both age group. Analysing the distances and angles we found that in the case of elderly people there is a degradation of long-range dependency in heartbeat dynamics that leads to a large difference between dynamics of respiratory and heartbeat dynamics. On the other hand in the case of young people there is no significant difference between heartbeat and respiratory dynamics, which implies a possibility of existence of a nonlinear relationship between cardiac and respiratory systems in young people.

**Keywords:** heartbeat, respiration, dynamics, fractal.

## 1 Introduction

Heartbeat dynamics and respiration dynamics are two of the most studied physiological processes primarily because of their clinical importance and complexity. Health specialists use Heart Rate Variability (HRV) as an indirect index of health or disease in the autonomic nervous system[1], while respiration signal is directly obtained or derived from the Electrocardiogram (ECG) signal for intensive care monitoring.

Heart rate and respiration rate has been associated with each other. It is known that respiration can modulate heart rate during a breathing cycle such that the heart rate increases during inhalation and decreases during exhalation, a phenomenon called Respiratory Sinus Arrhythmia (RSA). The complementary effect is called cardioventilatory coupling wherein temporal heart beat timing influences the onset of inhalation. Respiratory influences on heart rate have been studied by system analysis techniques in the time and frequency domains and results showed that RSA was a function of both breathing frequency and tidal volume. [2]

While time-domain and frequency-domain analysis is useful in deriving the mean, variance and peaks in power spectrum of physiological signals, they are not suitable to be used for non-stationary signals such as heart beat and respiration dynamics. They also seldom uncover complex and fractal behaviour of the underlying processes.

A more sophisticated method of analyzing those signals involves using scale exponents. By estimating scale-exponents for physiological signals, the effects of gender, age and diseases can be studied.

There are numerous studies involving the observation of the scaling properties to analyze heartbeat dynamics and respiration dynamics. In those studies it was shown that both processes demonstrate fractal characteristics.

Respiration dynamics was shown to have long-range correlation, although this correlation degrades in the case of elderly man. Peng et al. [3] computed the scaling exponents, or self-similarity parameter, of the inter-breath interval (IBI) signals from supine resting humans using Detrended Fluctuation Analysis (DFA) and Fourier analysis. DFA provided evidence for the fractal organization in physiologic human breathing cycle dynamics, and their degradation in elderly men. However, spectrum analysis did not show significant age or gender difference in the power spectrum exponents, owing to the fact that the signals of IBI are nonstationary.

These findings were confirmed using a different method of fractal analysis [4], and in addition, demonstrated fractal fluctuations in two other respiratory parameters. The power law relationships in the Allan Factor, Fano factor and Dispersional Analysis plots reflected the long-range correlations among the fluctuations in the number of breaths and breathe amplitude, as well as the peak-to-peak breath interval. Whereas there was a fractal component of the fluctuations in all three of these parameters in 9 of 20 spontaneously breathing subjects, there were subjects in whom only the fluctuations in two, one or none of the parameters exhibited time-scale invariance, suggesting that they are controlled by different fractal processes.

Iyengar et al.[5] conducted a separate study on the effect of aging in the alteration of scale-exponents in cardiac interbeat interval. They found a decrease in the complexity of heartbeat dynamics for elderly subjects. Lenskiy et al. [6] found a multifractal behaviour of HRV in both young and elderly subjects, and confirmed that HRV is less correlated for elderly people. This decrease of correlation, also considered as loss of complexity, was associated with the loss of long-term memory in heartbeat dynamics of elderly people.

A study on the relationship between the heartbeat dynamics and the dynamics of respiration under rest, exercises and increased atmospheric pressure showed no cross-correlation [7]. Using Holder exponent, a measure of the fluctuations related to the fractal dimension, they computed local fractal dimensions for the HRV and IBI variability and determined the cross-modal correlation across the population of

subjects. They found a significant decrease in fractal dimension for both heartbeat and respiration variabilities with increasing levels of exercise. Although qualitatively similar, the heartbeat and respiration dynamics were not significantly correlated and are concluded to be statistically independent with regards to the response to exercise. Also, they did not see statistically significant effects of age on fractal dimension under any of the conditions. They recommended the search for non-linear interdependency.

One RSA study attempted to determine if the origin of fractal heart beat dynamics was due to the fractality of respiration.[2] Yamamoto et al. mentioned that HRV is better modeled as fractal process rather than sum of harmonics. The spectral exponent was determined using Coarse graining spectral analysis. They examined the cross-correlation of the spectral indices for the inverse power law spectra of HRV and IBI variability. They concluded that the fractal behavior of the two processes were not correlated in any statistically significant way for subjects at rest, supporting the claim that the fractal component of heartbeat dynamics is not totally due to the respiratory modulation of heart rate. However, they suggested the possibility that respiratory influence on heart rate is nonlinear.

In this paper we are trying to fill the gap and investigate the possible non-linear relationship of heartbeat and respiration dynamics in supine resting conditions and compare the variability among elderly and young people. We hypothesize that there is a non-linear relationship between the heartbeat and respiration dynamics in one age group or both age groups.

## 2 Methods

### 2.1 Subject and Signal Acquisition

The researchers selected and obtained two groups of records, 10 elderly (range 68-81) and 10 young (range 21-34) from an online database of physiological signals [8]. The subjects are rigorously screened healthy adults who laid supine for 120 min while watching a movie to maintain wakefulness. A set of records for a single subject contains the ECG with beat annotations and respiration signals.

Signals for the heartbeat interval for each subject were derived from the ECG beat annotations by subtracting the previous index to the current index and dividing by the sampling frequency. Conversely, respiratory annotations were not available, and so a Local Extrema Detection algorithm [9] was applied to extract the breathing points from the respiratory signals. The algorithm takes the signal and a delta threshold, which is the difference between a peak and its surrounding in order to identify it as an extrema. The function returns local maxima and minima, which are the exhalation and inhalation points. Effectively, the interval between two exhale points was extracted as the IBI signal. Misidentifications were addressed by visually inspecting the signal and ensuring that the peak threshold is optimally tuned such that it avoids identifying small bumps but at the same time correctly detecting true respiration peaks. The correction rules that we applied are based from a study [10] that also extracted interbreath interval from the same dataset. IBI were then derived as in heartbeat intervals.

## 2.2 Fractal scaling Analysis

The next step after extracting HRV and IBI variability for each person, is to estimate multifractal spectrum. The applied algorithm for estimation of multifractal spectrum[6] is similar to the Detrended Fluctuation Analysis (DFA) [11] and multifractal DFA [12], however instead of polynomial subtraction, we apply a linear operator to suppress unwanted harmonics outside of the analysed frequency range defined by the scale  $l$ . The estimation algorithm is summarized in the following four steps:

1. Apply sub-band filtering using a kernel obtained by taking a derivative of the Gaussian. The spread parameter in the Gaussian depends on the scale  $l$ ;
2. Split the signal into  $N_l$  overlapping sub-windows of length  $2l$  and calculate variance  $\sigma_k^2(l)$  in each sub-window.
3. Calculating the the partition function as a sum of variances in each window powered by moment  $q$ :

$$\chi(q, l) = \sum_{k=1}^{N_l} [\sigma_k^2(l)]^q \quad (1)$$

4. Estimating multifractal spectrum as follows:

$$f(q) = \lim_{l \rightarrow 0} \frac{\sum_{k=1}^{N_l} \mu_k(q, l) \ln(\mu_k(q, l))}{\ln l} \quad (2)$$

$$\alpha(q) = \lim_{l \rightarrow 0} \frac{\sum_{k=1}^{N_l} \mu_k(q, l) \ln(\sigma_k^2(l))}{\ln l} \quad (3)$$

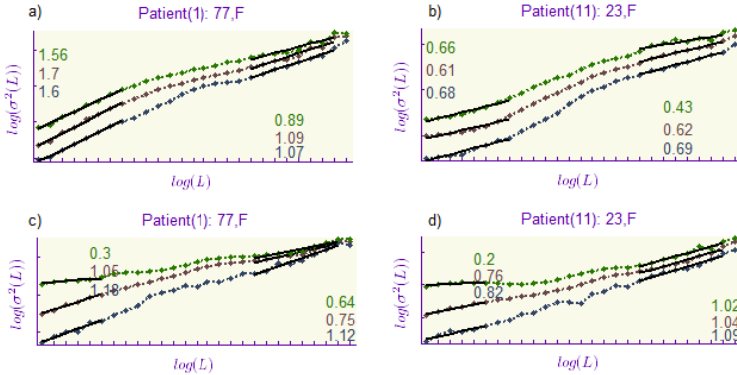
$$\text{where } \mu_k(q, l) = \frac{[\sigma_k^2(l)]^q}{\chi(q, l)}$$

The analyzed process is divided into number of overlapping fragments each of them is treated as a standalone process. The estimated spectra are then averaged and the resulting spectrum is the one that characterize the whole process. Compared to DFA and MF DFA the above described algorithm is free of underestimation and overestimation problems due to linear operation of sub-filtering.

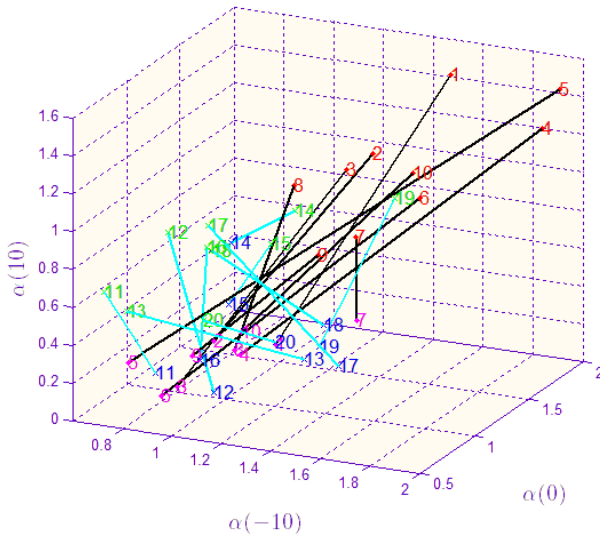
## 3 Results

The first step on this project is to confirm that the signals, HRV and IBI variability alike, are indeed fractal. Although, it has been proven that fluctuations in heart rate and respiration rate are both fractal, we still need to test our gathered data for validation. Figure 1 shows the power law curve in a double logarithmic plot for window length  $L$  and partition function  $\chi(q, l)$  for the respiratory and hearbeat variability for two patients.

The most interesting result was given by the feature space of the fractal spectrum estimated for short range scales, as shown in Figure 2. The ranges we used in the algorithm, i.e.  $L$ , were different for the HRV signals and IBI signals, since the latter is shorter compared to the former. The ranges that we chose covered the same duration of signals for cardiac and respiration signals. We also chose the best ranges by optimizing the classification function of the algorithm.



**Fig. 1.** Examples of power plot scaling of heartbeat interval signal for a) an elderly patient b) young patient and the power law scaling of their corresponding interbreath interval signals c) and d), respectively. The different colors correspond to the different values of  $q$  in the DFA algorithm,  $q = -10$  (green),  $q = 0$  (red) and  $q = 10$  (blue).



**Fig. 2.** Lines connecting the MFS of heartbeat interval (red for elderly[68-81], green for young[21-32]) and interbreath interval (violet for elderly and dark green for young)

**Table 1.** Distance and spherical angles between multifractal spectrum exponents of heartbeat and respiration dynamics for a) elderly and b) young people. Mean and standard deviation were also computed.

$N$	age(gender)	$R$	$\theta, deg$	$\phi, deg$
1	77(F)	1.47	31	56
2	73(F)	1.09	39	55
3	73(M)	1.23	37	57
4	81(M)	1.58	52	51
5	76(M)	2.09	54	44
6	74(F)	1.37	45	37
7	68(M)	0.40	28	115
8	73(F)	0.83	25	74
9	71(M)	0.68	51	53
10	71(F)	1.00	44	48
mean $\pm$ std	73.7 $\pm$ 3.6	1.177 $\pm$ 0.48	41 $\pm$ 10	59 $\pm$ 21

(a)

$N$	age(gender)	$R$	$\theta, deg$	$\phi, deg$
11	23(F)	0.50	114	-46
12	28(F)	0.79	108	154
13	34(M)	0.69	164	-12
14	31(M)	0.29	62	46
15	23(M)	0.44	28	-159
16	30(M)	0.62	95	-140
17	21(M)	0.83	137	164
18	30(F)	0.54	148	176
19	32(F)	0.77	33	71
20	21(F)	0.34	149	-51
mean $\pm$ std	27.3 $\pm$ 4.8	0.585 $\pm$ 0.19	104 $\pm$ 48	20 $\pm$ 122

(b)

The distance of the MFS for HRV and IBI is calculated and compared in both age groups. The spherical angles of the lines,  $\theta$  and  $\phi$ , were obtained. The values are tabulated in Table 1. The average distance between the multifractal spectrum of HRV and IBI for elderly is significantly higher (mean: 1.177075 std: 0.485178) than for the young group (mean: 0.58528 std: 0.191234). The angle of the lines for the elderly shows more uniformity (mean theta: 41.02 std: 10.43, mean pi: 59.55 std: 21.91) than in the young group (mean theta: 104.39 std: 48.83, mean pi: 20.52 std: 122.64).

## 4 Analysis

The plot hints for a non-linear relationship of the fractal properties of heartbeat and respiration dynamics for young subjects. The set of exponents for heartbeat and

respiration dynamics for young are clustered in one region, with short distances, as compared to those for elderly patients. This indicates that the dimensions of the two sets of quantities, for young people can be regarded as ‘the same’ and are more likely to have a non-linear relationship. By virtue of the proximity in the dimensions and the Hausdorff dimension corollary[13], we infer that heartbeat and respiration dynamics can have one-to-one bi-Lipschitz mapping, while the dissociation and the consistent variability in the exponents for heartbeat and respiration dynamics in elderly patients provides no evidence for any nonlinear cardiorespiratory variability relationship.

The direction of the lines connecting HRV and IBI exponents reveals the presence of a uniform shift directed towards higher dimensions in all degree of fluctuations. This is consistent with the fact that for elderly subjects, long-range correlation is degraded and there is less memory in heartbeat dynamics[14].

## 5 Conclusion

In this study we investigated the relationship of the heartbeat dynamics and respiration dynamics for 10 elderly and 10 young subjects. We aimed to know if there is a non-linear relationship between the fluctuations in heartbeat and respiration, using cardiac interbeat interval and inter-breath interval as materials. We also wished to determine the effect of aging in this relationship.

The results showed the evidence of non-linear relationship between the heartbeat and respiration dynamics for young subjects, while having non- relationship for the elderly subjects. We also saw a uniform shifting of dimensions for elderly subjects, signifying less correlation especially for heartbeat dynamics.

## References

1. Camm, A., Malik, M., Bigger Jr., J., et al.: Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology (93), 1043–1065 (1996)
2. Yoshiharu Yamamoto, J.-O.F., Hughson, R.: On the Fractal Nature of Heart Rate Variability in Humans: Effects of Respiratory Sinus Arrhythmia (1995)
3. Peng, C.-K., et al.: Quantifying fractal dynamics of human respiration: age and gender effects. *Annals of Biomedical Engineering* 30(5), 683–692 (2002)
4. Fadel, P.J., Barman, S.M., Phillips, S.W., Gebber, G.L.: Fractal fluctuations in Human Respiration. *J. Appl. Physiol* (2004)
5. Iyengar, N., et al.: Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *American Journal of Physiology, Regulatory, Integrative and Comparative Physiology* 217(4) (1996)
6. Lenskiy, A.A., Lee, J.S., Cichocki, A.: Multifractal scaling in cardiac interbeat interval dynamics of elderly and young healthy people (in preparation)
7. West, B.J., et al.: The independently fractal nature of respiration and heart rate during exercise under normobaric and hyperbaric conditions. *Respiratory Physiology & Neurobiology* 145(2-3) (2005)

8. Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., Mietus, J.E., Moody, G.B., Peng, C.-K., Stanley, H.E.: PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals, pp. e215–e220 (2000)
9. Billauer, E.: Peakdet: Peak detection using MATLAB (April 8, 2011), <http://billauer.co.il/peakdet.html>
10. Peng, C.-K., Mietus, J.E., Liu, Y., Lee, C., Hausdorff, J.M., Stanley, H.E., Goldberger, A.L., Lipsitz, L.A.: Quantifying Fractal Dynamics of Human Respiration: Age and Gender Effects. In: *Annals of Biomedical Engineering*, pp. 683–692 (2002)
11. Peng, C.-K., et al.: Finite-size effects on long-range correlations: Implications for analyzing DNA sequences. *Phys. Rev. E* 47, 3730–3733 (1993)
12. Kantelhardt, J.W., et al.: Multifractal Detrended Fluctuation Analysis of Nonstationary Time Series. *Physica A: Statistical Mechanics and its Applications* 316(1-4), 87–114 (2002)
13. Falconer, K.: Hausdorff measure and dimension. In: *Fractal Geometry: Mathematical Foundations and Application*. John Wiley & Sons (2003)
14. Ivanov, P.C.: Scaling behaviour of heartbeat intervals obtained by wavelet-based time-series analysis. *Nature* 383, 323–327 (1996)



# Relational Features for Texture Classification

Wan Nural Jawahir Hj Wan Yussof and Hans Burkhardt

Department of Computer Science, Universiti Malaysia Terengganu,  
21030 Kuala Terengganu, Terengganu, Malaysia  
wannurwy@umt.edu.my, Hans.Burkhardt@informatik.uni-freiburg.de

**Abstract.** Texture features play an important role in facilitating various applications, for instance, image retrieval and object recognition. In this work, we investigate the relational features as a texture descriptor in classifying materials and visual textures from their appearance. The relational features used in this paper are constructed by histogramming the values extracted for each point within an image with fuzzy histogram. To test the performance of relational features, two benchmarks were used which have a variety of poses and conditions. Despite the challenging occurrence in both benchmarks, impressive results were achieved by using the relational features.

**Keywords:** relational features, texture descriptor, fuzzy histogram.

## 1 Introduction

Image data contains meaningful information that has to be automatically extracted using computers or electronic devices. Such image information are called image features. Depending upon the particular task, the extracted features capture morphological properties, color properties, or certain textural properties of the image.

There are many texture features proposed in the past and some of them gained popularity to solve texture analysis problems in various domains. For example, Gabor filter and co-occurrence matrices have been successfully applied in image retrieval [5], texture classification [4] and image segmentation. Among the various texture computing method, the relational features are probably not widely used. However, in medical application [9], it has been proven to be efficient.

The aim of this work is to investigate the performance of relational features in recognizing materials and visual textures from a variety of poses with different conditions and scales. This is particularly a challenging task as the appearance of the image can vary considerably because of the differences in scale, orientation and illumination. Certainly, fine details either become visible or disappear with the changes and the object under consideration is subjected to a transformation from its current situation. The relational features is developed to be invariant under certain group of transformation [7]. In this work, we construct the relational features using a histogramming approach where fuzzy histogram is used instead of the traditional histogram. The performance of relational features are compared with other texture methods.

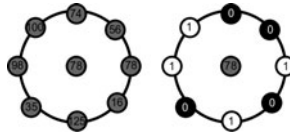
The organization of this paper is as follows: In Section 2, we will describe the relational features and fuzzy histogram. The relational features will be compared to four texture descriptors that will be described in Section 3. Section 4 gives an overview of the SVM classifier used for the classification task. Section 5 presents and discusses the results and finally we give the conclusion in Section 6.

## 2 Relational Features

The relational features are calculated similar to Local Binary Pattern (LBP) texture features [6]. LBP thresholds the neighborhood with the gray value of its center pixel and represents the result as a binary pattern (0 and 1). Applying this to all pixels in a circular neighborhood of the center pixel, the binary pattern is then transformed into a unique number as follows:

$$\text{LBP} = \sum_{i=0}^{n-1} s(v_i - v_c)2^i, \quad s(x) = \begin{cases} 1 & x \geq 0, \\ 0 & x < 0 \end{cases} \quad (1)$$

where  $v_i$  and  $v_c$  are the gray values at a neighboring pixel and at the center pixel, respectively. The number of the pixels in the circular neighborhood is denoted by  $n$ . The drawback of LBP is that the discontinuity of the LBP operator (the  $s$  function), since it maps to 0 or 1 which makes them sensitive to noise. A small disturbance in the image may cause a big deviation of the feature.



**Fig. 1.** Illustration of neighborhood pixels and center pixel of LBP. On the right is the representation in a binary pattern.

Using the following form

$$f(X) = \text{rel}(X(x_1, y_1) - X(x_2, y_2)) \quad (2)$$

we construct an invariant feature (detail on invariant feature can be found in [18]) by applying Eq. 2 into the following formula [8]:

$$F(\mathbf{X}) = \mathbf{HIST}(\{f(g(t_0, t_1, \theta = j \frac{2\pi}{q}) | 0 \leq t_0 < M, 0 \leq t_1 < N, 0 \leq j < q)\} \quad (3)$$

[7] has introduced a ramp function that extends the step function in Eq. 1 giving values in the range of  $[0, 1]$  as follows:

$$\text{rel}(\eta) = \begin{cases} 1 & \text{if } \eta < -\varepsilon \\ \frac{\varepsilon - \eta}{2\varepsilon} & \text{if } -\varepsilon \leq \eta \leq \varepsilon \\ 0 & \text{if } \varepsilon < \eta \end{cases} \quad (4)$$

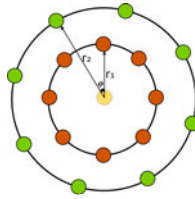
where  $\varepsilon$  is a threshold parameter. Using a ramp function, it is now more robust to image noise. Note that, if  $\varepsilon$  is set to zero, then the rel function will reduce to the simple LBP operator  $s$ .

As opposed to LBP, relational features use two circular sets. Let  $(x, y)$  be the coordinates of central pixel in two-dimension, taking into consideration a phase shift,  $\phi$ , the coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$  in Eq. 2 are given by:

$$(x_1, y_1) = (x + r_1 \cos(\theta), y + r_1 \sin(\theta)) \tag{5}$$

$$(x_2, y_2) = (x + r_2 \cos(\theta + \phi), y + r_2 \sin(\theta + \phi)) \tag{6}$$

where  $r_1$  and  $r_2$  are the radii of the first and second circle, respectively. Local information at different scales and orientations can be captured with different combinations of  $r_1$ ,  $r_2$  and  $\phi$ . The calculation of two dimensional relational features is illustrated in Fig. 2



**Fig. 2.** Calculation of a set of relational features. A feature is formed by applying the relational function to the gray-value difference of the pixels lying on the specific distance and phase to the reference point (i.e. center of the circles).

### 2.1 Histogramming of Features

To calculate the histogram, the range of values of the feature space  $\mathcal{F}$  is partitioned into  $K$  regions. Let  $\mathcal{F}^{(k)} \subset \mathcal{F}$  with with  $\bigcup_{k=0}^{K-1} \mathcal{F}^{(k)} = \mathcal{F}$  and  $\mathcal{F}^{(k)} \cap \mathcal{F}^{(k')} = \emptyset, \forall k \neq k'$ . The probability for data points falling into one of these regions is determined by counting the number of data points within it. Let  $L^k$  be the number of data points in region  $\mathcal{F}^k$  from a total of  $N$  points, the estimated probability is given by:

$$\mathcal{P}_k := \mathcal{P}(\mathbf{x} \in \mathcal{F}^{(k)}) = \frac{L^k}{N} \tag{7}$$

The value  $\mathcal{P}_k$  forms the histogram and the histogram regions are called bins. However, this standard histogram strategy depends on the number of bins and the origin of the bins. If the number of bins and the origin of the bins are chosen wrongly, then there is a discontinuity at the bin boundaries. Due to this discontinuity very small variations can cause a jump in the assignment (see illustration in Fig. 3a). Because of that, [8] proposed a fuzzy histogram to

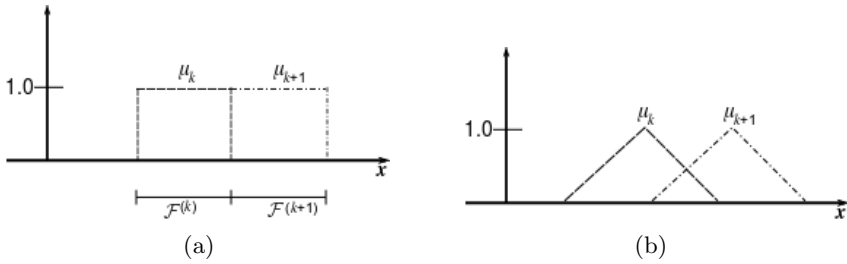
remove the discontinuous bin assignment of the standard histogram. Formally, the probability of an individual bin is defined as:

$$\mathcal{P}_k = \frac{1}{N} \sum_{n=0}^{N-1} \mu_k(\mathbf{x}^{(n)}), k = 0, 1, \dots, K - 1. \quad (8)$$

$$\text{with } \mu_k(\mathbf{x}) = \begin{cases} 1 & \text{for } x \in \mathcal{F}^{(k)} \\ 0 & \text{else} \end{cases} \quad (9)$$

$$\text{and } \mu_k(x)\mu_{k'}(x) = 0 \forall x \in \mathcal{F}, k \neq k' \quad (10)$$

where  $\mu_k$  denotes a membership function. The illustration of this continuous bin assignment can be seen in Fig. 3b.



**Fig. 3.** Examples of membership functions. (a) membership function for standard histogram and (b) membership function for fuzzy histogram.

### 3 Other Texture Descriptors

For the comparison study, the relational features are compared with four other texture descriptors : Gray-level Co-occurrence Matrices (GLCM), Haar wavelet, Gabor filter and Tamura texture features. In this section, we provide the description of these methods.

#### 3.1 Gray-Level Co-occurrence Matrices

Texture features derived from GLCM are so-called second order texture calculations because they are based on the joint co-occurrence of gray values for pairs of pixels  $(i, j)$  at a given distance,  $d$  and direction,  $\theta$ . The original GLCM reference [3] suggested 14 statistical measures to evaluate the properties of GLCM. However, some of them are highly correlated.

#### 3.2 Gabor Filters

The Gabor function and its Fourier transform can be written as follows:

$$g(x, y) = \left( \frac{1}{2\pi\sigma_x\sigma_y} \right) \exp \left[ -\frac{1}{2} \left( \frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) + 2\pi j W x \right] \quad (11)$$

$$G(u, v) = \exp \left[ -\frac{1}{2} \left( \frac{(u - W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right) \right] \tag{12}$$

where  $\sigma_u = -\frac{1}{2\pi\sigma_x}$  and  $\sigma_v = -\frac{1}{2\pi\sigma_y}$ ,  $\sigma_x$  and  $\sigma_y$  are the standard deviations of the Gaussian function in the  $x$  and  $y$  directions, respectively.  $W = U_h$  is the upper center frequency, and  $u$  and  $v$  define the position of the function in Fourier domain.

A class of self-similar functions can be obtained by dilations and rotations of the Gabor functions. These self-similar function called Gabor wavelets are given in the following form:

$$g_{mn}(x, y) = a^{-m}G(x', y') \tag{13}$$

with  $x' = a^{-m}(x \cos \theta + y \sin \theta)$      $y' = a^{-m}(x \sin \theta + y \cos \theta)$

where  $a > 1$ ,  $m = 0, \dots, S - 1$ ,  $n = 0, \dots, K - 1$ ,  $S$  and  $K$  are the number of scales and orientations and  $(x', y')$  are the rotated coordinates by the angle  $\theta$ .

A filter dictionary is therefore designed to reduce this redundancy [5]. The aim of this design is to ensure that the half-peak magnitude of the filter responses touch each other. This can be achieved by selecting the following parameters:

$$a = (U_h)^{-\frac{1}{S-1}} \tag{14}$$

$$\sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\ln 2}} \quad \sigma_v = \tan\left(\frac{\pi}{2K}\right)(U_h - 2\left(\frac{\sigma_u^2}{U_h}\right)) \tag{15}$$

where  $U_l$  is the lower center frequency. The Gabor filter features can be obtained by calculating the mean and standard deviation of the sub-images.

### 3.3 Haar Wavelets

The Haar wavelet is the simplest orthogonal wavelet transform given by

$$\psi_i^j(x) = \psi(2^j x - i), \quad i = 0, \dots, 2^j \tag{16}$$

where

$$\psi(x) = \begin{cases} 1 & \text{for } 0 \leq x < 1/2 \\ -1 & \text{for } 1/2 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

Its scaling function is given by

$$\phi(x) = \begin{cases} 1 & 0 \leq x < 1 \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

For two dimensional Haar wavelets, the construction proceeds by first defining a two-dimensional scaling function,  $\phi(x, y) = \phi(x)\phi(y)$  and three wavelet functions  $\psi^H(x, y) = \psi(x)\phi(y)$ ,  $\psi^V(x, y) = \phi(x)\psi(y)$ , and  $\psi^D(x, y) = \psi(x)\psi(y)$ . The resulting feature vector is  $f = (\mu_n, \sigma_n)$  extracted from the detail images at different scale  $n$ .

### 3.4 Tamura Features

In [10], the authors propose six texture features corresponding to human visual perception: coarseness, contrast, directionality, line-likeness, regularity, and roughness. However, the first three features are very important.

The coarseness measure is calculated as follows:

1. For every point  $(n_0, n_1)$  calculate the average over neighborhoods. The size of the neighborhoods are powers of two, e.g.: 1x1, 2x2, 4x4, ..., 32x32 :

$$A_k(n_0, n_1) = \frac{1}{2^{2k}} \sum_{i=1}^{2^{2k}} \sum_{j=1}^{2^{2k}} X(n_0 - 2^{k-1} + i, n_1 - 2^{k-1} + j) \quad (19)$$

2. For every point  $(n_0, n_1)$  calculate differences between the not overlapping neighborhoods on opposite sides of the points in horizontal and vertical direction:

$$E_k^h(n_0, n_1) = |A_k(n_0 + 2^{k-1}, n_1) - A_k(n_0 - 2^{k-1}, n_1)| \quad (20)$$

and

$$E_k^v(n_0, n_1) = |A_k(n_0, n_1 + 2^{k-1}) - A_k(n_0, n_1 - 2^{k-1})| \quad (21)$$

3. At each point  $(n_0, n_1)$  select the size leading to the highest difference value:

$$S(n_0, n_1) = \arg \max_{k=1, \dots, 5} \{ \max_{d=h, v} \{ E_k^d(n_0, n_1) \} \} \quad (22)$$

4. Finally take the average over  $2^S$  as a coarseness measure for the image:

$$F_{crs} = \frac{1}{N_0 N_1} \sum_{n_0=1}^{N_0} \sum_{n_1=1}^{N_1} 2^{S(n_0, n_1)} \quad (23)$$

The contrast of an image is calculated by

$$F_{con} = \frac{\sigma}{\alpha_z^4} \quad \text{with } \alpha_4 = \frac{\mu_4}{\sigma_4} \quad (24)$$

where  $\mu_4 = \frac{1}{N_0 N_1} \sum_{n_0=1}^{N_0} \sum_{n_1=1}^{N_1} (X(n_0, n_1) - \mu)^4$  is the fourth moment about the mean  $\mu$ ,  $\sigma^2$  is the variance of the gray values of the image, and  $z$  has experimentally been determined to be  $\frac{1}{4}$ .

To calculate the directionality the horizontal and vertical derivatives  $\Delta_H$  and  $\Delta_V$  are calculated by the convolution of the image  $X(n_0, n_1)$  with the following 3x3 operators respectively

$$\begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline \end{array} \quad \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline -1 & 0 & 1 \\ \hline \end{array}$$

and then for every position  $(n_0, n_1)$

$$\theta = \frac{\pi}{2} + \tan^{-1} \frac{\Delta_V(n_0, n_1)}{\Delta_H(n_0, n_1)} \quad (25)$$

is calculated. These values are then histogrammed in a 16 bin histogram  $H_D$ . The directionality can be calculated as the sum of second moments around each peak from valley to valley.

## 4 Experimental Results

The different texture descriptors were evaluated on two benchmarks: VisTex [11] and KTH-TIPS [12]. The description for both benchmarks is given below.

### 4.1 VisTex Database

The VisTex database is a collection of texture images. The database was created with the intention of providing a large set of high quality textures for computer vision applications. In particular, the set was made as an alternative to the Brodatz texture library, which is not freely available for research use. The VisTex database in our experiment consists of 19 categories with 168 128x128 gray-scale images in total. 74 of them were used for training set and the balance 94 were used as a test set.

### 4.2 KTH-TIPS Database

The KTH-TIPS image database (KTH is the abbreviation of Royal Institute of Technology in Stockholm, Sweden, and TIPS stands for **T**exture under varying **I**llumination, **P**ose and **S**cale) contains a subset of CURET database (10 out of 61 materials). The database is designed to provide variations in scale as well as variations in pose and illumination [4]. Images were taken at different scale spanning two octaves. At the central scale, the distance between the camera and the target was 28cm. At each scale 9 images were taken in a combination of three poses (frontal, rotated 22.5° left and right) and three illumination conditions (from the front, from the side at roughly 45° and from the top at roughly 45°).

### 4.3 Parameters Setting

For each of the texture descriptors, we set the following parameters:

- Relational Features (RF)
 

We used three set of parameters  $\{r_1, r_2, \phi\} - \{2, 5, 0\}, \{3, 6, \pi/4\}$  and  $\{4, 8, \pi/2\}$ . The sampling number was set to 20. We integrated values obtained from these three sets for each sampling point and 20 length of features were obtained for each point within the image with  $M \times N$  size. The size was reduced by histogramming the features extracted from all points and put into 20 bins.
- Gray Level Co-occurrence Matrices (GLCM)
 

For the GLCM, the co-occurrence matrix was created using a window size  $5 \times 5$  by scanning each image in the  $0^\circ, 45^\circ, 90^\circ$  and  $135^\circ$  directions with unit pixel distance  $d = 2$ . Six Haralick's texture features (as described in Section 3) were calculated from the matrix.
- Gabor Filters
 

The following parameters were selected:  $K = 4, S = 3, U_h = 0.4$  and  $U_l = 0.1$ . It means the filter bank was created with 4 orientations and 3 combinations of frequency. This operation produced a total of 24 filters covering the mean and standard deviation of the magnitude images.
- Haar Wavelet
 

Haar wavelet features were extracted from three image details at six scales. So, for each image we have  $3 \times 6 \times 2$  size of Haar features vector.
- Tamura
 

For Tamura texture features,  $20 \times 3$  length of features were extracted from coarseness, contrast and directionality. This features is a modification of the original features where we replaced the averaging process with the histogramming.

### 4.4 Classification Results and Discussion

The classifier used in the classification process was the SVMs classifier [2]. For both experiments, we used Gaussian RBF kernel with  $\gamma = 2$ . The classification results are shown in Table 2 and Table 1 for VisTex and KTH-TIPS, respectively.

**Results for KTH-TIPS.** It can be observed from the table, the average accuracy for RF is 95.56% which ranked it at the top list and higher than Tamura by 1.16%. There are four materials that the relational features better from the others: Sponge, Brown bread, Orange peel and Linen. The lowest accuracy it gets is for Cotton where only 90.44% compared to Gabor which can recognize this material better than the other materials. The relatively low result achieved by RF could be because the confusion between Cotton and Linen.

**Results for VisTex.** On the whole, all methods demonstrate good performance on the VisTex database. The accuracy across different methods is not much



**Table 1.** Classification results of texture descriptors for KTH\_TIPS database

Material	Accuracy (%)				
	RF	Gabor	Haar	GLCM	Tamura
Sandpaper	93.78	90.44	95.11	90.44	90.22
Crumpled aluminum foil	92.44	92.44	100	100	96.22
Styrofoam	96.89	94	98.44	97.56	92.67
Sponge	96.89	94.67	97.11	92.22	94.67
Corduroy	96.22	90.44	98.89	85.11	94
Linen	92.66	39.33	90.67	90	91.33
Cotton	90.44	96.22	91.56	92.67	93.78
Brown bread	95.33	93.78	84.22	91.33	95.33
Orange peel	98.22	93.11	97.33	92.22	97.11
Cracker B	95.56	93.11	89.78	92.44	98.67
<b>Average</b>	<b>95.56</b>	<b>87.31</b>	<b>94.31</b>	<b>92.4</b>	<b>94.4</b>

**Table 2.** Classification results of texture descriptors for VisTex database

Categories	Accuracy (%)				
	RF	Gabor	Haar	GLCM	Tamura
Bark	88.3	91.49	86.17	89.36	91.49
Brick	93.62	94.68	88.3	91.49	93.62
Building	92.55	92.55	88.3	94.68	94.68
Cloud	97.87	98.94	98.94	97.87	97.87
Fabric	84.04	18.09	86.17	90.42	72.34
Flower	95.74	95.74	91.49	95.74	90.43
Food	91.49	92.55	89.36	90.43	90.43
Grass	97.87	97.87	94.68	95.74	96.81
Leaves	90.43	93.62	88.3	90.43	82.98
Metal	96.81	96.81	97.87	97.87	95.74
Misc	97.87	97.87	96.81	97.87	97.87
Painting	90.43	92.55	89.36	85.11	91.49
Sand	97.87	95.74	95.74	94.68	94.68
Stone	95.74	96.81	96.81	94.68	95.74
Terrain	93.62	94.68	90.43	97.87	98.94
Tile	94.68	93.62	94.68	92.55	91.49
Water	94.68	95.74	98.94	94.68	96.81
WheresWaldo	98.94	97.87	96.81	93.62	97.87
Wood	98.94	98.94	98.94	96.81	94.68
<b>Average</b>	<b>94.29</b>	<b>91.38</b>	<b>93.06</b>	<b>93.78</b>	<b>92.95</b>

different. However, in this experiment, the RF once again outperforms other methods with the average accuracy 94.29%. It is 0.51% higher than GLCM, 1.23% higher than Haar, 1.34% higher than Tamura and 2.91% higher than Gabor. This result is quite impressive since the numbers of images of different categories are imbalance.

## 5 Conclusion

In this work, relational features has been proposed for better representation of textures in real-world applications where the appearance is a challenging problem. From the experiment we had conducted on two benchmarks; the one for materials classification under different scales and orientations and the other one on a real-world texture benchmark, we find that the relational features outperform several existing methods in classifying texture images. This means that these features has potential to be used for many other crucial tasks where texture plays an important role.

## References

1. Burkhardt, H., Siggelkow, S.: Invariant Features in Pattern Recognition. In: Kotropoulos, C., Pitas, I. (eds.) *Nonlinear Model-Based Image/Video Processing and Analysis*, pp. 269–307. John Wisley and Sons (2001)
2. Burges, C.J.C.: A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery* 2, 121–167 (1998)
3. Haralick, R.M., Shanmugam, K., Dinstein, I.: Textural Features for Image Classification. *IEEE Transactions on Systems, Man, and Cybernetics SMC* 6, 610–621 (1973)
4. Hayman, E., Caputo, B., Fritz, M., Eklundh, J.-O.: On the Significance of Real-World Conditions for Material Classification. In: Pajdla, T., Matas, J(G.) (eds.) *ECCV 2004*. LNCS, vol. 3024, pp. 253–266. Springer, Heidelberg (2004)
5. Manjunath, B.S., Ma, W.Y.: Texture Features for Browsing and Retrieval of Image Data. *IEEE Transaction on Pattern Analysis and Machine Intelligent* 18(8), 837–842 (1996)
6. Ojala, T., Pietikäinen, M., Mäenpää, T.: Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns. In: Vernon, D. (ed.) *ECCV 2000*. LNCS, vol. 1842, pp. 404–420. Springer, Heidelberg (2000)
7. Schael, M.: Invariant grey scale features for texture analysis based on group averaging with relational kernel function, Internal Report 01/01, University of Freiburg (2001)
8. Siggelkow, S., Schael, M., Burkhardt, H.: SIMBA - Search IMages By Appearance. In: Radig, B., Florczyk, S. (eds.) *DAGM 2001*. LNCS, vol. 2191, pp. 9–16. Springer, Heidelberg (2001)
9. Setia, L., Teynor, A., Halawani, A., Burkhardt, H.: Radiograph annotation using local relational features. In: *Workshop on Cross Language Evaluation Forum, CLEF* (2006)
10. Tamura, H., Yamawaki, T.: Textural features corresponding to visual perception. *IEEE Transaction on Systems, Man, and Cybernetics* 8(6), 460–472 (1978)
11. MIT Media Lab Vision Textures, <http://vismod.www.media.mit.edu>
12. Fritz, M., Hayman, E., Caputo, B., Eklundh, J.-O.: The KTH-TIPS database, <http://www.nada.kth.se/cvap/databases/kth-tips>

# Author Index

- Abdullah-Al-Wadud, M. 300  
Abid, Mohamed 82  
Ahn, Kiok 10  
Ali, Yusnita Mohd 103  
Angamuthu, Tamilarasi 328  
Anitha, J. 356  
Aomori, Hisashi 225  
Awang, Saidatul Ardeenawatie 103
- Babu, B. Raveendra 320  
Baksi, Debanjana 195  
Barath, S. Raj 336  
Bhattacharyya, Debnath 113, 126, 140, 153, 204  
Bispo, Aline 56  
Bucaoto, Wilson 430  
Burkhardt, Hans 438
- Chae, Oksam 10  
Chakraborty, Rudrasis 179  
Chakravarty, Srishty 48  
Chen, Wu-Lin 310  
Choi, Tae-Sun 186  
Chung, Yoojin 300
- Dhar, Pranab K. 234  
Dhara, Bibhas Chandra 179  
Dharmaratne, Anuja T. 215, 254, 261
- El-Sonbaty, Yasser 26
- Fahmy, Aly A. 280  
Fathalla, Karma M. 26  
Fujimura, Makoto 1
- Galdia, Philipp 92  
Georgiadis, Anthimos 92  
Ghosal, Arijit 179  
Ghosh, Rajib 153  
Goerge, Loay E. 379  
Gorai, Pinky 195  
Govardhan, A. 320  
Gu, Junzhong 366, 414
- Haque, Mohammad Nazmul 300  
Hasan, D.M.H. 234
- Hasan, Md. Mehedi 10  
Hassanien, Aboul Ella 280  
Hemanth, D. Jude 356  
Hj Wan Yussof, Wan Nural Jawahir 438  
Hu, Yu-Chen 310  
Huang, Chih-Cheng 160  
Huang, Zhimin 270
- Imamura, Kousuke 1  
Irtaza, Aun 186
- Jaffar, M. Arfan 186  
Ji, Hong-Bing 20, 422
- Kandasamy, Megalai 328  
Kesh, Subhasis 48  
Khan, Mohammad I. 234  
Kim, Han Jong 430  
Kim, Jong-Myon 234  
Kim, Tai-hoon 113, 126, 140, 153, 204, 280  
Kisku, Dakshina Ranjan 195  
Koch, Carsten 92  
Kodikara, N.D. 254  
Kumari, L.L. Gayani 261  
Kuroda, Hideo 1  
Kurukulasooriya, Ayesha 215
- Lakshmi, S. 346  
Lassoued, Arij 82  
Lee, Gang-soo 113, 126, 140, 153  
Lee, Jaemyoung 410  
Lee, Joo-Ho 391  
Lee, Sungyoung 74  
Lee, Young-Koo 74  
Lenskiy, Artem 430  
Leung, C.H.C. 291  
Li, Yuanxi 291  
Liu, Jin 422  
Lu, Yue 270
- Maiti, Chandreyee 195  
Mat, Fauziah 103  
Minamoto, Teruya 170  
Momin, Hajra 244

- Mukherjee, Imon 204  
 Mukherjee, Nabanita 204  
  
 Ohura, Ryuji 170  
 Ouni, Tarek 82  
  
 Pandiyan, Paulraj Murugesa 103  
 Pelaes, Evaldo 56, 66  
 Prakash, A. Jaya 336  
 Prasomphan, Sathit 225  
  
 Qin, Yuzhu 270  
  
 Ramidi, Fadzly 103  
 Ren, Lei 414  
  
 Saha, Sanjoy Kumar 179  
 Salama, Mostafa A. 280  
 Sankaranarayanan, V. 346  
 Selvakumar, A. Immanuel 356  
 Sharma, Sanjeev 401  
 Shukla, Anupam 401  
 Singh, Vikas 401  
 Siravenha, Ana Carolina 56, 66  
 Sousa, Danilo 56, 66  
 Sri Nagesh, A. 320  
 Sultan, Bushra A. 379  
  
 Tanaka, Mamoru 225  
 Tapamo, Jules-Raymond 244  
 The Vinh, La 74  
 Tiwari, Ritu 401  
 Tran, Dinh Tuan 391  
 Tsai, Chun-Yi 160  
  
 Uddin, Mohammad Shorif 300  
  
 Varma, G.P. Saradhi 320  
 Vijila, C. Kezi Selva 336, 356  
  
 Wang, Shao-Hua 422  
 Wang, Yunqiong 36  
 Wickramasinghe, W.A.P. 254  
  
 Xia, Weiwei 414  
 Xu, Tianwei 36  
  
 Yaacob, Sazali 103  
 Yang, Danfeng 270  
 Yang, Horng-Chang 160  
 Yang, Jin-Long 20  
 Youssef, Sherin M. 26  
  
 Zamider, Ipsita 195  
 Zeng, Jun-Rong 310  
 Zhou, Juxiang 36