

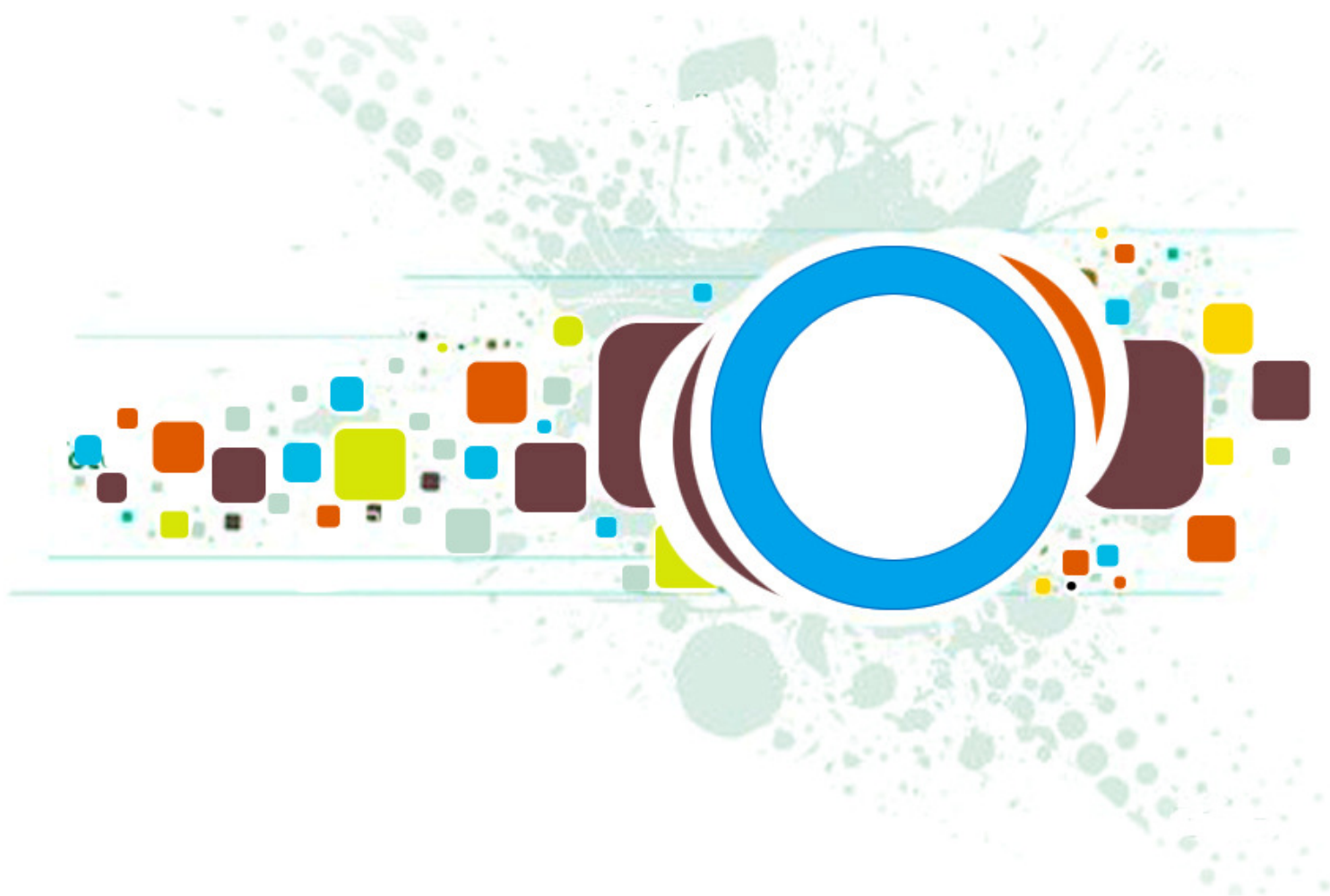
Volume 5 ■ Issue 4 ■ October 2011

Editor-in-Chief  
Professor Hu, Yu-Chen

INTERNATIONAL JOURNAL OF  
**IMAGE PROCESSING (IJIP)**

ISSN : 1985-2304

Publication Frequency: 6 Issues Per Year



CSC PUBLISHERS  
<http://www.cscjournals.org>

# **INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)**

**VOLUME 5, ISSUE 4, 2011**

**EDITED BY  
DR. NABEEL TAHIR**

ISSN (Online): 1985-2304

International Journal of Image Processing (IJIP) is published both in traditional paper form and in Internet. This journal is published at the website <http://www.cscjournals.org>, maintained by Computer Science Journals (CSC Journals), Malaysia.

IJIP Journal is a part of CSC Publishers

Computer Science Journals

<http://www.cscjournals.org>

# INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

Book: Volume 5, Issue 4, October 2011

Publishing Date: 05-10- 2011

ISSN (Online): 1985-2304

This work is subjected 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 of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers.

IJIP Journal is a part of CSC Publishers

<http://www.cscjournals.org>

© IJIP Journal

Published in Malaysia

Typesetting: Camera-ready by author, data conversion by CSC Publishing Services – CSC Journals, Malaysia

**CSC Publishers, 2011**

## EDITORIAL PREFACE

The International Journal of Image Processing (IJIP) is an effective medium for interchange of high quality theoretical and applied research in the Image Processing domain from theoretical research to application development. This is the forth issue of volume four of IJIP. The Journal is published bi-monthly, with papers being peer reviewed to high international standards. IJIP emphasizes on efficient and effective image technologies, and provides a central for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing. IJIP comprehensively cover the system, processing and application aspects of image processing. Some of the important topics are architecture of imaging and vision systems, chemical and spectral sensitization, coding and transmission, generation and display, image processing: coding analysis and recognition, photopolymers, visual inspection etc.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 5, 2011, IJIP appears in more focused issues. Besides normal publications, IJIP intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

IJIP give an opportunity to scientists, researchers, engineers and vendors from different disciplines of image processing to share the ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development. This journal is helpful for the researchers and R&D engineers, scientists all those persons who are involve in image processing in any shape.

Highly professional scholars give their efforts, valuable time, expertise and motivation to IJIP as Editorial board members. All submissions are evaluated by the International Editorial Board. The International Editorial Board ensures that significant developments in image processing from around the world are reflected in the IJIP publications.

IJIP editors understand that how much it is important for authors and researchers to have their work published with a minimum delay after submission of their papers. They also strongly believe that the direct communication between the editors and authors are important for the welfare, quality and wellbeing of the Journal and its readers. Therefore, all activities from paper submission to paper publication are controlled through electronic systems that include electronic submission, editorial panel and review system that ensures rapid decision with least delays in the publication processes.

To build its international reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP. We would like to remind you that the success of our journal depends directly on the number of quality articles submitted for review. Accordingly, we would like to request your participation by submitting quality manuscripts for review and encouraging your colleagues to submit quality manuscripts for review. One of the great benefits we can provide to our prospective authors is the mentoring nature of our review process. IJIP provides authors with high quality, helpful reviews that are shaped to assist authors in improving their manuscripts.

### **Editorial Board Members**

International Journal of Image Processing (IJIP)



## **EDITORIAL BOARD**

### **EDITOR-in-CHIEF (EiC)**

**Professor Hu, Yu-Chen**  
Providence University (Taiwan)

### **ASSOCIATE EDITORS (AEiCs)**

---

**Professor. Khan M. Iftekharuddin**  
University of Memphis  
United States of America

**Assistant Professor M. Emre Celebi**  
Louisiana State University in Shreveport  
United States of America

**Assistant Professor Yufang Tracy Bao**  
Fayetteville State University  
United States of America

**Professor. Ryszard S. Choras**  
University of Technology & Life Sciences  
Poland

**Dr. Huiyu Zhou**  
Queen's University Belfast  
United Kingdom

**Professor Yen-Wei Chen**  
Ritsumeikan University  
Japan

**Associate Professor Tao Gao**  
Tianjin University  
China

### **EDITORIAL BOARD MEMBERS (EBMs)**

---

**Dr. C. Saravanan**  
National Institute of Technology, Durgapur West Bengal  
India

**Dr. Ghassan Adnan Hamid Al-Kindi**  
Sohar University  
Oman

**Dr. Cho Siu Yeung David**  
Nanyang Technological University  
Singapore

**Dr. E. Sreenivasa Reddy**

Vasireddy Venkatadri Institute of Technology  
India

**Dr. Khalid Mohamed Hosny**

Zagazig University  
Egypt

**Dr. Chin-Feng Lee**

Chaoyang University of Technology  
Taiwan

**Professor Santhosh.P.Mathew**

Mahatma Gandhi University  
India

**Dr Hong (Vicky) Zhao**

Univ. of Alberta  
Canada

**Professor Yongping Zhang**

Ningbo University of Technology  
China

**Assistant Professor Humaira Nisar**

University Tunku Abdul Rahman  
Malaysia

**Dr M.Munir Ahamed Rabbani**

Qassim University  
India

**Dr Yanhui Guo**

University of Michigan  
United States of America

## TABLE OF CONTENTS

Volume 5, Issue 4, October 2011

### Pages

- 371 - 381      Compressed Medical Image Transfer in Frequency Domain  
*Aree Ali Mohammed, Haval Mohammed Sidqi*
- 382 - 389      Securing Biometric Images Using Reversible Watermarking  
*Sabu m Thampi, Ann Jisma Jacob*
- 390 - 402      A High Performance Modified SPIHT for Scalable Image Compression  
*Bibhuprasad Mohanty, Abhishek Singh, Sudipta Mahapatra*
- 403 - 416      An Efficient Thresholding Neural Network Technique for High Noise  
Densities Environments  
*M.A.Abdou, Mazhar B. Tayel, Azza M.Elbagoury*
- 417 - 424      Dynamic Threshold in Clip Analysis and Retrieval  
*Satishkumar L Varma, Sanjay N Talbar*
- 425 - 434      Statistical Feature-based Neural Network Approach for the Detection of Lung Cancer  
*K.A.G. Udeshani, R.G.N. Meegama, T.G.I. Fernando*
- 435 - 445      Detection of Cranial- Facial Malformations: Towards an Automatic Method  
*Hiba Chelbi, Manel Laajimi, Wala Touhami, KHLIFA Nawrès*
- 446 - 455      Unsupervised Multispectral Image Classification By fuzzy Hidden Markov Chains  
Model For SPOTHRV Images  
*Faiza DAKKA, Ahmed HAMMOUCH, Driss ABOUTAJDINE*

- 456 - 468 Fusion Based Gaussian noise Removal in the Images Using Curvelets and Wavelets With Gaussian Filter  
*Naga Sravanthi Kota, G. Umamaheswara Reddy*
- 469 - 478 An Efficient Multiplierless Transform algorithm for Video Coding  
*Geetha.K.S, Pushpa.M.K, M.Uttarakumari, S.Sethu Selvi*
- 479 – 503 Video Audio Interface for recognizing gestures of Indian sign Language  
*E.Kiran Kumar, S.R.C.Kishore , P.V.V.Kishore, P.Rajesh Kumar*
- 504 – 511 An Analysis and Comparison of Quality Index Using Clustering Techniques for Spot Detection in Noisy Microarray Images  
*Sri Nagesh, G.P.Saradhi Varma, A.Govardhan, B.Raveendra Babu*
- 512 - 520 Two Methods for Recognition of Hand Written Farsi Characters  
*Mohammad Reza Jenabzadeh, Reza Azmi, Boshra Pishgoo, Samanesadat Shirazi*

# Compressed Medical Image Transfer in Frequency Domain

**Aree Ali Mohammed**

*College of Science, Computer Dept.  
University of Sulaimani  
Sulaimani, Iraq*

*aree.ali@univsul.net*

**Miran Taha Abdullah**

*College of Science, Computer Dept.  
University of Sulaimani  
Sulaimani, Iraq*

*miran.abdullah@univsul.net*

---

## Abstract

A common approach to the medical image compression algorithm begins by separating the region of interests from the background of the medical images and then lossless and lossy compression schemes are applied on the ROI part and background respectively. The compressed files (ROI and background) are now transmitted through different media of communications (local host, Intranet and Internet) between the server and clients. In this work, a medical image transfer coding scheme based on lossless Haar wavelet transforms method is proposed. The test results have indicated that the performance of the proposed medical image transfer coding via Intranet is much better than via Internet in terms of transferring time, while the quality of the reconstructed medical image remains constant despite the medium of communication. For best adopted parameters, a compressed medical image file (760 KB → 19.38 KB) is transmitted through Internet (bandwidth= 1024 kbps) with transfer time = 0.156 s while the uncompressed file is sent with transfer time = 6.192 s.

**Keywords:** Haar Wavelet, Medical Image Compression, Region of Interest Coding, Adaptive Quantization.

---

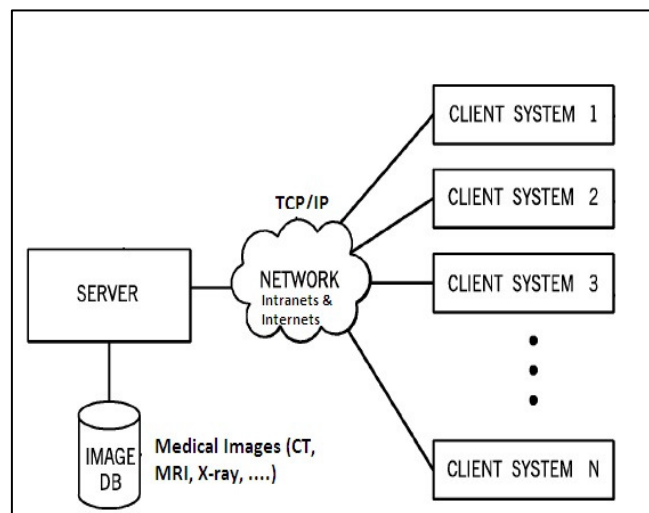
## 1. INTRODUCTION

In recent years, advances in information technology and telecommunications have played a crucial role as catalysts for significant developments in the sector of healthcare. These technological advances have played a very strong role in the field of medical imaging. The number of digital medical images has increased rapidly on the Internet. The necessity of fast and secure diagnosis is vital in the medical world.

Nowadays, the transmission of images has become routine, and it is imperative to find an efficient form by which to transmit them over the Internet or Intranet. To this end, image compression is an important research issue. For instance, in medical image compression applications, diagnosis is effective only when compression techniques preserve all the relevant and important image information needed. This is the case with lossless compression techniques. Lossy compression techniques, on the other hand, are more efficient in terms of storage and transmission needs but there is no warranty that they can preserve the characteristics needed in medical image processing and diagnosis. In this latter case, of lossy compression, image characteristics are usually preserved in the coefficients of the domain space in which the original image is transformed. That is, for instance, in the DWT based medical image compression, the wavelet coefficients keep all the information needed for reconstructing the medical image [1]. Several lossless and lossy techniques for the compression of the medical image data have been proposed [2-4]. Since there is a trade-off between reconstructed image quality and achieved compression ratio, an optimum compression ratio is important. [5]. Several studies have focused

on different quality levels for medical applications and a good survey on lossless compression of a region of interest can be found in [6]. The general theme is that diagnostically important regions must be preserved, whereas the rest of the image is only important in a contextual sense, helping the viewer to observe the ROI within the image [7-8]. A lossless image compression over lossy packet networks which is robust in an ATM (Asynchronous Transfer Mode) environment to packet loss and to non-sequential packet receipt is presented in [9]. A context-based and regions of interest (ROI) based approach to compress medical images in particular vascular images is implemented [10], where a high spatial resolution and contrast sensitivity is required in areas such as stenosis. A comparative analysis of image compression techniques is proposed [11-12] which is a framework for ROI based compression of medical images using wavelet based compression techniques (i.e. JPEG2000 and SPIHT). In [13] techniques for determining and extracting the ROI of an MRI brain image are presented.

Figure (1) presents general diagrams of the medical image transfer between the server and the clients in any healthcare centers. The process may include communicating the medical image data from the server to the client at the desired compression level in response to the client's request for the image data [14-15].



**FIGURE 1:** Medical image data transfer diagram.

The paper is organized as follows: in section 2, an outline framework of medical image transfer is given. Section 3 is a review of the wavelet transform and the description with perfect reconstruction of the selected ROI's. The section also gives information on the quantization and coding of the coefficients. Section 4 gives the results obtained with the proposed method and compares it in both Intranet and Internet. Finally, conclusions are drawn in section 5.

## 2. MEDICAL IMAGE TRANSFER FRAMEWORK

### Server side:

1. Input uncompressed medical image.
2. Coding process
  - a. Color space conversion RGB to YCbCr.
  - b. Apply forward wavelet transform.
  - c. Select ROI and indicate a set of coefficients which is sufficient for lossless ROI reconstruction.
  - d. Perform adaptive quantization on ROI and uniform quantization on background.

- e. Encode the resulting coefficients stream by entropy shift coding algorithm.

**Transmission side:**

3. Guarantee network connection (Intranet or Internet).
4. Transmit compressed image data for both (ROI and Background) through socket interface.

**Client side:**

5. Decoding process
  - a. Apply inverse entropy shift algorithm.
  - b. Perform inverse quantization.
  - c. Apply inverse wavelet transform.
  - d. Inverse color conversion YCbCr to RGB.
6. Output compressed medical image.

The performance test evaluation is conducted in both Intranet and Internet according to the quality measure PSNR and for different network bandwidth. The transferring time is also determined between server and clients.

### **3. LOSSY AND LOSSLESS (ROI) CODING**

Firstly, the medical image is divided into two parts. The ROI is selected using cropping algorithm which is using two points ((x1,y1) as starting point, (y1,y2) as ending point) to create a rectangle of a given region in the image with a variable block size ((20x20), (40,40), etc.). The remaining part of the image (background) is separately coded. The steps of the coding process at the server side for both parts consist of four stages.

#### **3.1 Forward Color Space Conversion**

This stage converts pixels color representation from RGB to YCbCr color space data. Then, (Cb, Cr) bands are downsampled by (2) to be at lower resolution, the brightness (luma) information is preserved.

#### **3.2 Forward Wavelet Transformation**

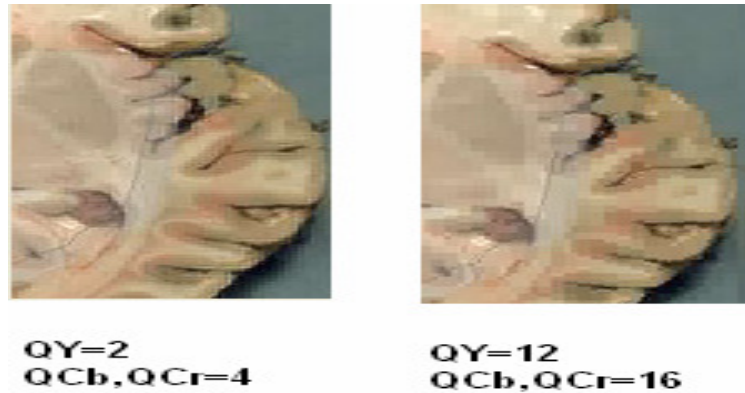
The filter used in this research to apply wavelet transform is the reversible transform (Haar Wavelet). This transform can be used for lossy and lossless coding. Unlike conventional transforms, wavelet decomposition produces a family of hierarchically organized decompositions. The selection of a suitable level for the hierarchy depends on signal nature and experience.

#### **3.3 Forward Quantization**

The retained (i.e., details) coefficients obtained by wavelet transformation must be quantized for the purpose of compression. The luminance component (Y) requires a small quantization step, while Cb and Cr can be quantized by a larger quantization step. The result of this step is a set of quantization indices that consists of a large sequence of zeros, especially the quantization indices of (HL, LH, HH) subbands.

##### **3.3.1 ROI Quasi Quantization**

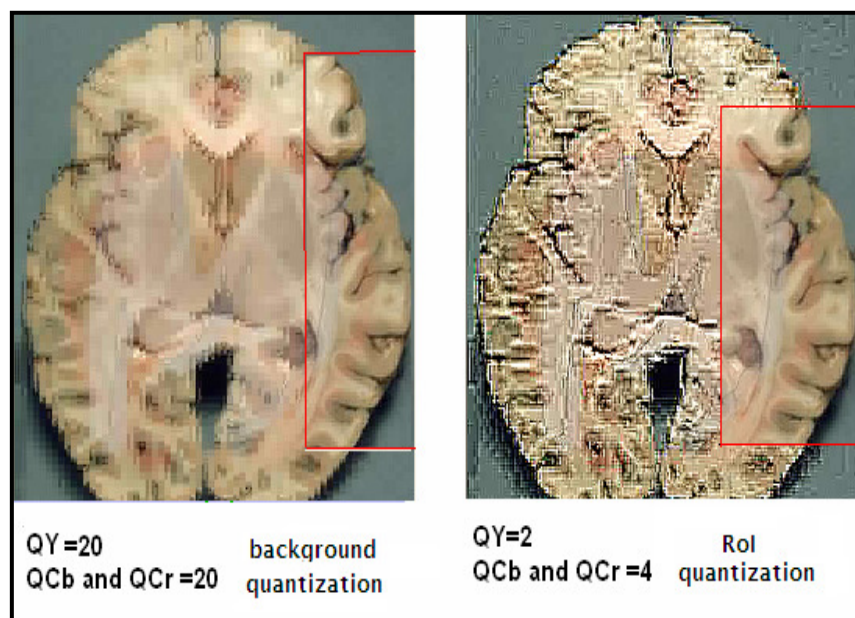
Since the ROI is the most important part of the image for medical diagnostic purposes, therefore this part must be coded with a non lossy compression algorithm. In this research, the quasi lossless compression algorithm is used by tuning quantization parameters (Q\_Y, Q\_Cb, Q\_Cr) for different wavelet bands (LL, HL, LH and HH). The quasi-lossless compression comes to a high speed. The quality of the reconstruction image under the quasi-lossless algorithm must remain with a higher compression ratio. In figure (2) the effect of different quantization steps according to the Y and Cb,Cr channels is shown.



**FIGURE 2:** Effect of quasi lossless quantization.

### 3.3.2 Background Uniform Quantization

The background of the medical image has not had a considerable effect of medical diagnostic. For this reason, this part is compressed by using lossy compression technique. This technique needs uniform quantization steps by dividing QY and QCb and QCr channels by the same number of steps. Figure (3) shows the effect of ROI and background quantization steps.



**FIGURE 3:** Effect of lossy and quasi lossless quantization.

### 3.4 Forward Color Space Conversion

After performing the quasi and uniform quantization on the subbands of the image (ROI and Background), the integer indices produced by the quantizer are entropy-encoded by the optimized variable encoder, which is developed in this work. This encoder is considered as a variable length encoder because it gives a small number of bits (short codewords) to represent the small numbers when they mostly occur and a larger number of bits (long codewords) to represent the less probable large numbers. The optimized (variable encoder) algorithm consists of two stages: optimizer and encoder. This algorithm is separately applied on the ROI and the background parts. The following pseudo code presents the shift encoder algorithms step.



Step1: Set Coef = positive wavelet coefficient  
Step2: Find MAX(Coef)  
Step3: Compute the number of bit to represent MAX{Coef} = n  
Step4: Compute the histogram of Coef = Hist{Coef}  
Step5: Optimizer  
Step6: Register Optimalm , n+1  
Step7: Encoder  
Step8: Output Coeff.

#### **4. EXPERIMENTAL RESULTS**

The results obtained and presented in this section are those of the tests conducted on different types of medical image samples. Normally, the medical image is taken by many medical devices (such as Dicom, CT, MRI, etc.). In this research, the test images are obtained from medical archiving systems through the Internet (Brain (192 KB), Cardiopulmonary Vasculature (768 KB)) MRI images.

The adopted test strategy was based on a determination of the effects of the involved coding parameters (no. of wavelet pass, quantization steps (QY,QCb,r), Internet or Intranet bandwidth) on compression factor C.F, transfer time and fidelity measure PSNR. The value of each coding parameter which led to the best coding results was used during the tests to investigate the effects of other coding parameters.

##### **4.1 Intranet Test**

In this part, the tests are performed via the Intranet with fixed speed (transfer rate) = 54 Mbps. Tables (1) and (2) show the test results of lossless ROI coding applied on the colorized 'brain' and 'Cardiopulmonary' image respectively, when the quantization parameters are varied for the same number of passes.

ROI block size (KB)	Transfer time (ms) before compression	Transfer time (ms) after compression	PSNR in dB
80x80 (18.8→1.13)	2.46	0.406	34.92
120x120 (42.2→2.4)	10.87	0.421	34.83
160x160 (75→4.10)	41.43	0.66	34.12
200x200 (117.18→6.42)	45.57	0.703	33.7

**TABLE 1:** PSNR and transfer time for lossless ROI (QY=2, QC=4, alpha=0.1 and No. pass= 2, 'brain.bmp').

ROI block size (KB)	Transfer time (ms) before compression	Transfer time (ms) after compression	PSNR in dB
100x100 (29.3→0.66)	2.55	0.407	34.92
200x200 (117→2.43)	8.45	0.42	33.36
300x300 (263→5.72)	26.34	0.625	32.1
400x400 (468→9.48)	62.79	0.905	32.9

**TABLE 2:** PSNR and transfer time for lossless ROI (QY=2, QC=4, alpha=0.1 and No. pass= 2, 'Cardiopulmonary.bmp').

To assess the performance tests of proposed scheme via the Intranet, the transfer time of the ROI part between the server and the clients is determined for different sizes of ROIs. In figure (4), the transfer time before and after compression is shown for different size of ROIs. The parameters are shown according to the table (1).

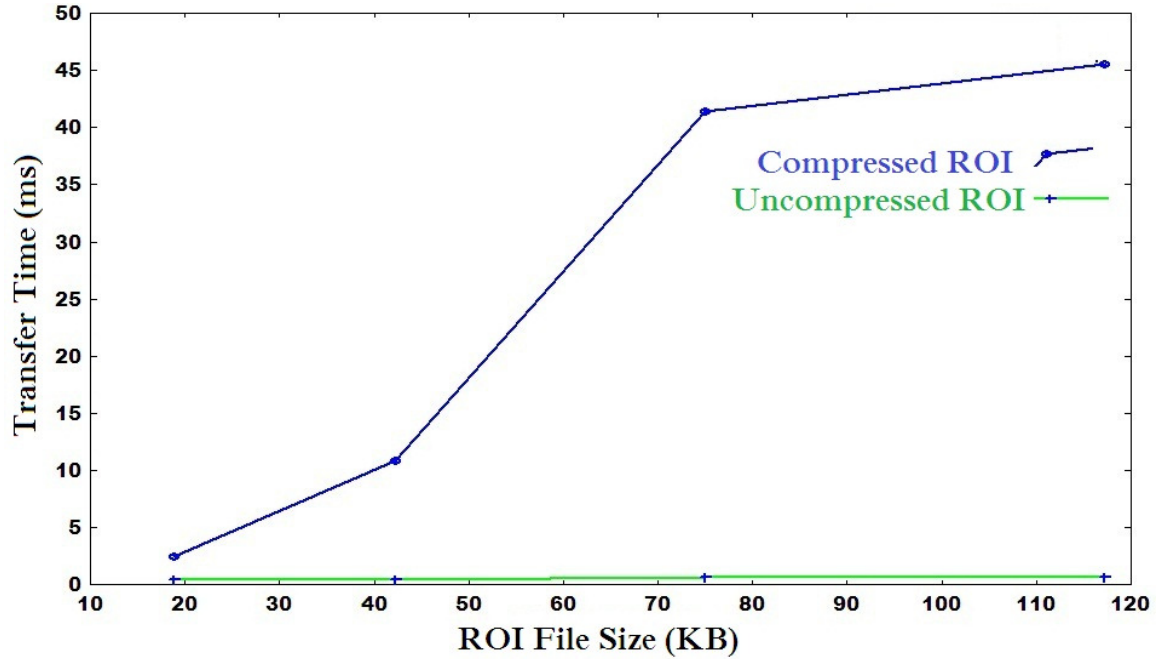


FIGURE 4: Transfer time versus ROI file size.

**4.2 Internet Test**

In Internet, the conducted tests are performed using the Internet with a dynamic range of bandwidth speed varied from 128 Kbps to 1024 Kbps for upload and download.

To evaluate the performance of the proposed method tested on the Internet, the transfer time of the ROI part for different values of bandwidth between the server and the clients is determined for different sizes of ROIs. In figure (5), the transfer time versus bandwidth rates before compression is shown for different sizes of ROIs.

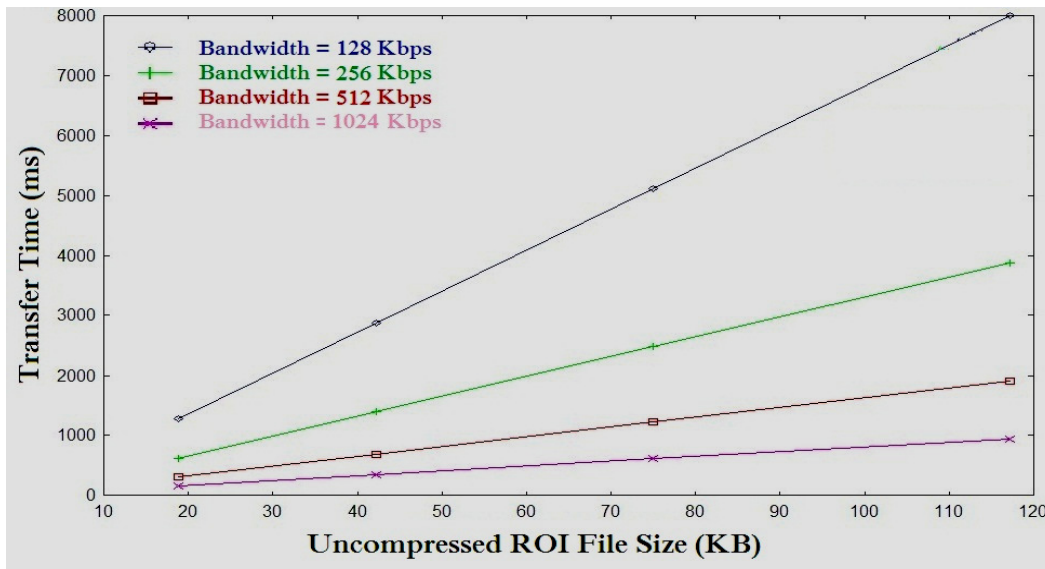
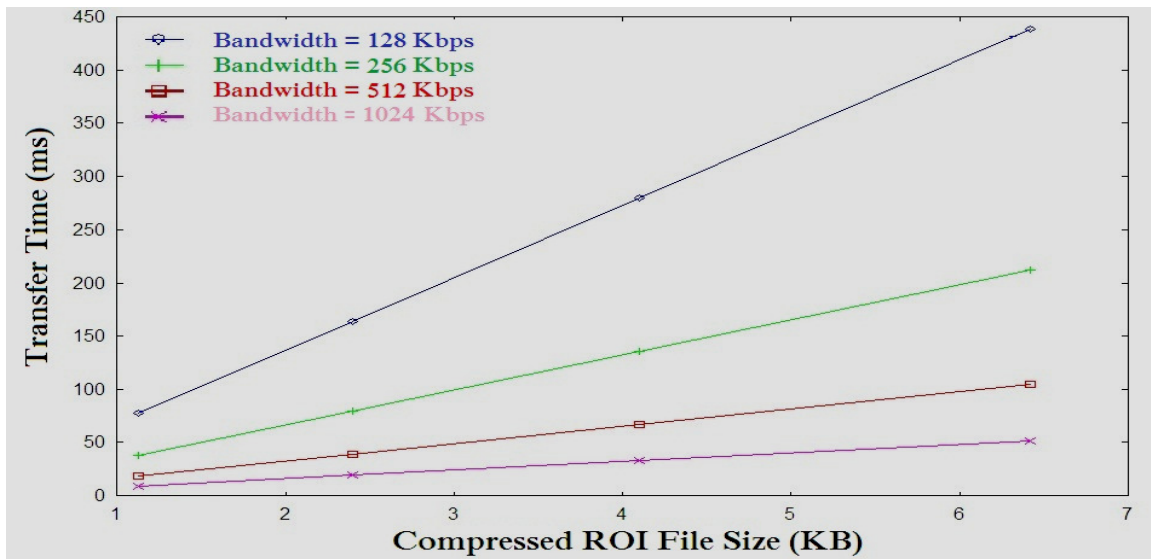
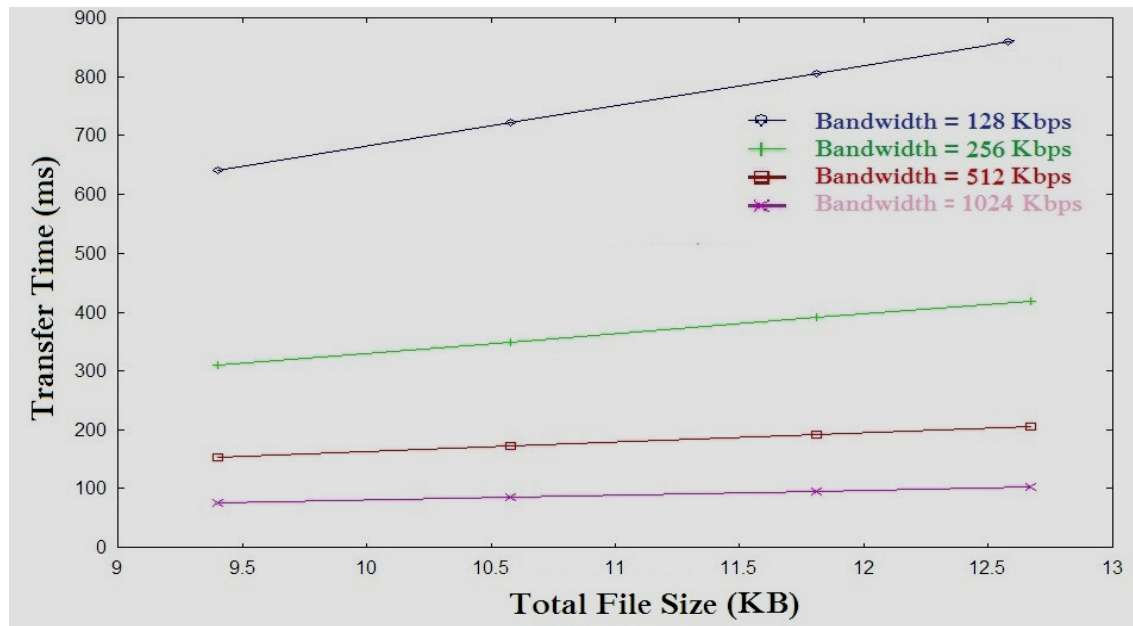


FIGURE 5: Uncompressed ROI file transfer time for different bandwidth.

In figure (6), the transfer time versus bandwidth rates after compression is shown for different sizes of ROIs. To make a difference between compressed and decompressed total file size (ROI and background) in terms of transfer time for different bandwidth rates, figure (7) shows the test results of Brain color image.



**FIGURE 6:** Compressed ROI file transfer time for different bandwidth.



**FIGURE 7:** Compressed file transfer time for different bandwidths.

The same tests as presented in figures (6) and (7) have been performed for larger medical image file size ('Cardiopulmonary.bmp'=768 KB). Another test is performed on a very large color image encoding with RGB system (2.59 MB). The transfer time is calculated for different bandwidth rates as follows:

128 kbps → 3.50 minutes : 256 kbps → 1.74 minutes  
 512 kbps → 0.83 minutes : 1024 kbps → 0.41 minutes

## 5. CONCLUSIONS

The well-known wavelet transform and variable entropy coding have been used to allow transmission of a medical image with perfect reconstruction of selected ROIs in the image. The proposed scheme calculates a mask which specifies which coefficients are needed for a certain region rather than another to obtain perfect diagnoses of a disease. In this research, the quasi lossless image compression algorithm has been designed and implemented to encode the ROI, while the lossy compression algorithm is used to encode the background of the image. The algorithms are tested in both Intranet and Internet to assess the performance of the proposed scheme.

1. The intranet test results with a fixed bandwidth for both ROI and background have reflected the following:

a. The transferring time before and after compression is decreasing depending on the ROI block size.

b. The intranet communication does not affect the PSNR.

2. The internet test results with a dynamic range of bandwidth rate for both ROI and background have reflected the following:

a. The transferring time before and after compression is decreasing depending on the ROI block size and the bandwidth rates.

b. The larger medical image file size led to larger transferring time when the ROI block size = 200x200.

- Uncompressed file size = 192 KB → transferring time = 1548.1 ms for bandwidth=1024 kbps.
- Compressed file size = 12.67 KB → transferring time = 102.56 ms for bandwidth=1024 kbps.
- Uncompressed file size = 768 KB → transferring time = 6192.3 ms for bandwidth=1024 kbps.
- Compressed file size = 19.38 KB → transferring time = 156.26 ms for bandwidth=1024 kbps.

## 6. REFERENCES

- [1] A. Amine, "Compression of Biomedical Images and Signals", British Library Cataloguing-in-Publication Data, ISBN: 978-1-84821-028-8, 1st Edition, 2008.
- [2] S. Miaou, S. Chen and S. Chao, "Wavelet-Based Lossy to Lossless Medical Image Compression using Dynamic VQ and SPIHT Coding", Journal of Biomedical Engineering Applications, Basis & Communications, Vol. 15, No. 6, pp. 235-242, 2003.
- [3] S. Takamura and M. Takagi, "Lossless Image Compression with Lossy Image Using Adaptive Prediction and Arithmetic Coding", IEEE Proceeding on Data Compression, UT, 1994.
- [4] G. Wang and B. Jackson, "Compression and decompression of medical images", Patent No. US 7,697,741 B2, USA, 2010.
- [5] O. Pianykh, "Digital Imaging and Communications in Medicine (DICOM): A practical introduction and survival guide", Springer-Verlag Berlin Heidelberg, ISBN 978-3-540-74570-9, 2008.
- [6] D. Nister and C. Christopoulos, "Lossless region of interest with a naturally progressive still image coding algorithm", IEEE International Conference on Image Processing, pp. 856-860, 1998.
- [7] J. Ström, P. C. Cosman, "Medical image compression with lossless regions of interest", Signal Processing 59, No. 2, pp. 155-171, 1997.
- [8] J. Hou and X. Song, "Region of Interest Coding for JPEG2000", International Conference on Measuring Technology and Mechatronics Automation, Vol. 2, pp.1004-1007, China, 2010.
- [9] L. Merriam et al, "Lossless image compression over lossy packet networks", International Conference on Signal Processing Applications and Technology, 1997.
- [10] M. Firoozbakht et al, "Compression of digital medical images based on multiple regions of interest", IEEE Computer Society, International Conference on Digital Society pp. 260-263, 2010.
- [11] P. Bharti, S. Gupta and R. Bhatia, "Comparative Analysis of Image Compression Techniques: A Case Study on Medical Images", IEEE Computer Society, 2009 International Conference on Advances in Recent Technologies in Communication and Computing, pp. 820-822, 2010.
- [12] S. Shenbaga Devi and K.Vidhya. "Development of Medical Image Compression Techniques." ICCIMA. 2007.
- [13] Onsy Alim, Nadder Hamdy and Wesam Gamal. "Determination of the Region of Interest in the Compression of Biomedical Images". 24th National Radio Science Conference. March 2007.

- [14] R. Cheryl, T. Toan and M. Rakesh, "Networked image visualization, image quality enhancement method and system", Patent Application Publication, Pub. No. US-0300167A1, 2009.
- [15] S. Ghrare, M. Ali, M. Ismail and K. Jumari, "The Effect of Image Data Compression on the Clinical Information Quality of Compressed Computed Tomography Images for Teleradiology Applications", European Journal of Scientific Research, Vol. 23, No. 1, pp. 6-12, 2008.

## SECURING BIOMETRIC IMAGES USING REVERSIBLE WATERMARKING

**Sabu M. Thampi**

Rajagiri School of Engineering and  
Technology, Kochi, India

sabum@rajagiritech.ac.in

**Ann Jisma Jacob**

Indira Gandhi National Open University,  
India

annjisma@yahoo.com

---

### ABSTRACT

Biometric security is a fast growing area. Protecting biometric data is very important since it can be misused by attackers. In order to increase security of biometric data there are different methods in which watermarking is widely accepted. A more acceptable, new important development in this area is reversible watermarking in which the original image can be completely restored and the watermark can be retrieved. But reversible watermarking in biometrics is an understudied area. Reversible watermarking maintains high quality of biometric data. This paper proposes Rotational Replacement of LSB as a reversible watermarking scheme for biometric images. PSNR is the regular method used for quality measurement of biometric data. In this paper we also show that SSIM Index is a better alternate for effective quality assessment for reversible watermarked biometric data by comparing with the well known reversible watermarking scheme using Difference Expansion.

**Keywords :** Biometric, Reversible Watermarking, Difference Expansion, SSIM Index, Image Processing

---

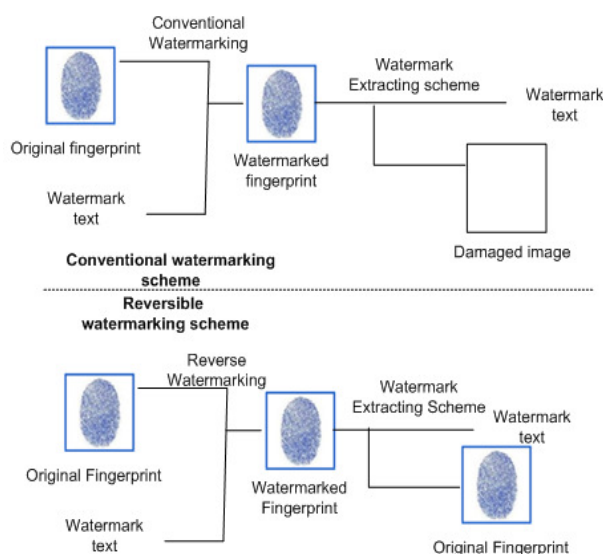
### 1. INTRODUCTION

Applying security along with privacy is challenging in this digital age. Conventional methods are defeated by biometric authentication as it is more convenient for privacy protection. But along with convenience it has introduced new threats. It is prone to attacks since the data will not be renewed so often. Thus securing biometric data has become crucial. Though there are different methods being introduced for this, watermarking plays an important role here in protecting data.

Watermarking hides a message or image in some data to obtain a new data so that the hidden data is indistinguishable from the original data so that an attacker cannot remove or replace the message from the new data. Thus it has become a major part of information hiding. Watermarking is used to pass hidden messages in communications. Watermarking can be done in different domains like spatial or frequency domain. In Spatial domain actual pixel values are changed where as in the frequency domain, coefficients used in transformed image representations are changed. There are some techniques, which uses LSBs for embedding watermark data. In most of these, the changes made to the original data become irreversible. There is yet another way in which the original image and the embedded payload can be perfectly recovered later by watermarking in some particular way. This technique is called *reversible watermarking*.

In Reversible watermarking, a watermark is embedded into an image and from the watermarked image; original image and the watermark are retrieved. The retrieval process does not need the original image and so this technique is said to be blind [7]. There are different types of reversible watermarking based on the technique they use such as Data Compression, Difference expansion, Histogram operation and LSB Replacement [8]. Figure 1 shows the difference between the conventional watermarking techniques and the reversible watermarking techniques. This paper focuses on reversible watermarking using difference expansion and LSB replacement.





**FIGURE 1:** Difference between conventional and Reversible watermarking system

Quality of watermarked images plays an important role in biometrics. Since watermarking adds noise to the data, it should be done in such a way that it will not defeat the purpose of watermarking. There are quite a few number of techniques introduced for watermarking but the quality of watermarking is usually measured using PSNR. PSNR is a simple measure of image quality based on average error. This technique is known to be not very accurate [5], [6]. Despite its drawbacks, however, PSNR is still heavily reported because it is easy to compute. This paper introduces SSIM Index as a better way of quality measurement for biometric watermarked images. For this, the proposed rotational replacement of LSB scheme is being compared with the well known reversible watermarking scheme using Difference Expansion.

**2. PREVIOUS WORK**

Earliest Reversible watermarking scheme was invented by Barton [3]. DE is the most established method for reverse watermarking [2]. Reversible Watermarking using DE was first proposed by Tian [2]. Tian used regular images for his experiments. This paper uses DE method for watermarking biometric image.

PSNR is the common quality measurement technique used for watermarked images as it is easy to calculate and optimize. But they are not very well matched to perceived visual quality [1]. SSIM (Structural SIMilarity) Index is based on Full-reference image quality assessment [1]. SSIM Index compares local patterns of pixel intensities that have been normalized for luminance and contrast [1]. Since SSIM Index is designed to improve on traditional methods like PSNR, measuring the quality using SSIM Index will add value to the reversible watermarking technique.

In this paper, reversible watermarking is done using proposed rotational replacement of LSB and also using Tian’s reversible watermarking technique. The quality of watermarked images is assessed using SSIMIndex and compared.

**3. WATERMARKING USING DIFFERENCE EXPANSION**

Tian [2] introduced the Difference Expansion technique. In this technique, pairs of pixels of the host image *HI* are identified and they are converted to a low-pass image *LI* using integer transformation with integer average  $\alpha$  and a high-pass image *HI* containing the pixel differences  $\delta$ . If *x* and *y* be the intensity values of a pixel-pair, then  $\alpha$  and  $\delta$  are defined as

$$\alpha = \left\lfloor \frac{(x+y)}{2} \right\rfloor \quad (1)$$

$$\delta = x - y \quad (2)$$

x and y can be computed from  $\alpha$  and  $\delta$  as

$$x = \alpha + \left\lfloor \frac{(\delta+1)}{2} \right\rfloor \quad (3)$$

$$y = \alpha - \frac{\delta}{2} \quad (4)$$

By appending an information bit  $i$  to the LSB of the difference  $\delta$ , a new LSB can be created. The watermarked difference is

$$\delta_w = 2\delta + i \quad (5)$$

The resulting pixel gray-levels are calculated from the difference ( $\delta_w$ ) and integer average  $\alpha$  using (3) and (4).

For an image with  $n$ -bit pixel representation, the gray levels satisfy  $x, y \in [0, 2^n - 1]$ , if and only if  $\alpha$  and  $\delta$  satisfies the following condition:

$$|\delta| \in R_d(\alpha) = [0, \min(2(2^n - 1 - l), 2\alpha + 1)] \quad (6)$$

Where  $R_d(\alpha)$  is called the invertible region. Combining (5) and (6) we obtain the condition for a difference  $\delta$  to undergo DE.

$$|2\delta + i| \in R_d(\alpha) \text{ for } i = 0, 1 \quad (7)$$

This condition is called the *expandability condition* for DE. If an integer average is given and the difference satisfies the expandability condition, it is called an *expandable difference*.

Another method used by Tian in his Difference Expansion technique, other than embedding which is already discussed is replacing LSB. Here, an information bit will be used to replace the LSB of the difference. Since the LSB is replaced in the embedding process, this cannot be considered as lossless as the first embedding technique. However, here the information about the true LSBs of the differences that are embedded by replacing LSB are saved and embedded with the payload, to ensure that it is not lossy.

The LSB of a difference can be flipped without affecting its ability to invert back to the pixel domain if and only if

$$\left\lfloor 2 \left\lfloor \frac{\delta}{2} \right\rfloor + i \right\rfloor \in R_d(\alpha) \text{ for } i=0, 1 \quad (8)$$

This is called the *changeability condition*. A difference satisfying the changeability condition, given a corresponding integer average, is called a *changeable difference*. An expandable difference is also a changeable difference. A changeable location remains changeable even after its LSB is replaced, whereas an expandable location may not be expandable after DE, but it remains changeable.

#### 4. WATERMARKING USING ROTATIONAL REPLACEMENT OF LSB

In common LSB replacement techniques, the LSB of the original image is replaced with the MSB of the watermark. For this, first the number of bits needs to be required is calculated and those many LSB are replaced with MSB of watermark. To retrieve the watermark, first get the LSB according to the number used and then use them to create new image by changing them to MSB. Here, the restored image will be low in quality. In Rotational replacement of LSB, same bit will be used to store secondary LSB of the original image and the watermark.

For embedding watermark using rotational replacement of LSB, first create matrices of original image and watermark. Assuming original image is  $X \times Y$  matrix and watermark is  $x \times y$  matrix. The ratio  $\gamma = (M \times N) / m \times n$  should be greater than or equal to 8, so that watermark can be embedded using this technique. The original image matrix is divided into blocks. Thus if  $\gamma \geq 64$ , the block size will be  $8 \times 8$ . Only one byte of watermark data can be embedded to such a block. Consider such a block. The last row of the original host image will be replaced by the last but one row data. This will be continued rotationally so that the first row of the original matrix will be placed at the second row and first row will be free. One byte of the watermark data will be placed in this row.

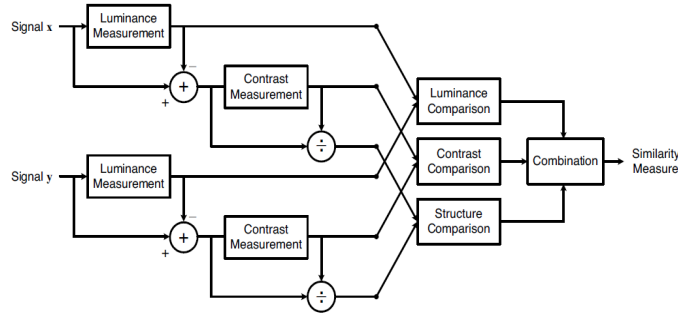
$$\begin{bmatrix} 11111111 \\ 10101010 \\ 01010101 \\ 11011011 \\ 00110100 \\ 11100010 \\ 10001000 \\ 11000001 \end{bmatrix} + [10101010] \rightarrow \begin{bmatrix} 10101010 \\ 11111111 \\ 10101010 \\ 01010101 \\ 11011011 \\ 00110100 \\ 11100010 \\ 10001000 \end{bmatrix}$$

**FIGURE 2:** Rotational replacement of LSB while embedding

Extraction will be similar to embedding process where, using the  $\gamma$  value, blocks are determined. Based on the size of the watermarked data, first row LSB will be used to reconstruct the watermark. After removing the first row, the remaining rows are rotationally replaced in the reverse order of embedding. Last row will replace the data of the same row.

#### 5. STRUCTURAL SIMILARITY INDEX (SSIM INDEX)

SSIM Index is a framework for quality assessment based on the degradation of structural information introduced by Zhou Wang [1] *et al.* For human visual system a calculation of structural information difference can provide a good approximation to the image distortion perceived. The product of the illumination and the reflectance gives the luminance of the surface of an object, but the structures of the objects in the scene are independent of the illumination [1]. SSIM Index defines the structural information in an image as those attributes that represent the structure of objects in the scene, independent of the average luminance and contrast [1].



**FIGURE 3:** Diagram showing calculation of SSIM Index<sup>[1]</sup>

In SSIM Index, the similarity measurement is done by making comparison of luminance, contrast and structure. Let x and y be two image signals.

Luminance of each signal is calculated as:

$$\mu_x = \frac{1}{N} \sum_{i=1}^n x_i \quad (1)$$

Signal contrast is calculated as:

$$\sigma_x = \left( \frac{1}{N-1} \sum_{i=1}^n (x_i - \mu_x)^2 \right)^{\frac{1}{2}} \quad (2)$$

The structure comparison is calculated as:

$$(x - \mu_x) / \sigma_x \text{ and } (y - \mu_y) / \sigma_y \quad (3)$$

Combining these three, the SSIM index is calculated as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}$$

Where  $C_1$  &  $C_2$  are included to avoid instability when  $\mu_x^2 + \mu_y^2$  is very close to zero.

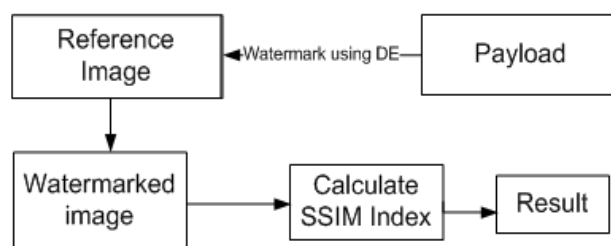
$C_1 = (K_1L)^2$  where  $K_1 = 0.01$

$C_2 = (K_2L)^2$  where  $K_2 = 0.03$

L = dynamic range of the pixel values.

## 6. DESIGN OVERVIEW

The novelty of this paper is in introducing Rotational Replacement of LSB as a reversible watermarking scheme and proving SSIM Index as a better alternate for measuring quality of biometric data. Since PSNR is easy to calculate, it is the common method used everywhere, where as the experiments conducted show that SSIM Index can be used as a better and effective alternate. The system uses Biometric image (fingerprint) for watermarking using Rotational Replacement of LSB and also using Difference Expansion. These reversible watermarking techniques reduce the noise while watermarking which is highly recommended for biometric images. The watermarked fingerprint image is used for calculating SSIM index. SSIM Index is calculated for watermarked images with different payload.

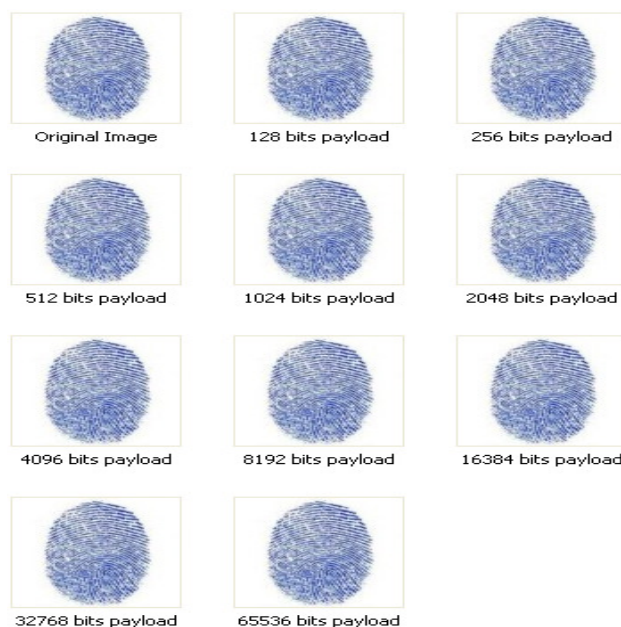


**FIGURE 4 :** Calculation of SSIM Index of watermarked image

## 7. EXPERIMENTAL SETUP AND RESULTS

The experiment was conducted using a single fingerprint image with different payloads since the quality of image is affected more as the payload increases. The finger print was collected from internet [<http://www.nist.gov>] and it is of the resolution 512\*512. The experiment was implemented in Java using Image APIs. Watermarking was done using Rotational Replacement of LSB and DE for different payload. SSIM index for the watermarked images are found. The goal is to prove that the SSIM Index based quality measurement is a better alternate than PSNR for biometric image watermarking and also that Rotational replacement of LSB is a better reverse watermarking scheme than DE.

Fig 5 shows the original image and the watermarked images used for the experiment using rotational replacement of LSB. As it can be seen, the images are visually not tampered by watermarking.



**FIGURE 5:** Watermarked samples with different payload

Experimental results show that Rotational replacement of LSB is a better reverse watermarking scheme than DE by measuring the quality using SSIMIndex. It is also proved that the execution time for Rotational replacement of LSB is much less than DE and payload capacity of Rotational replacement of LSB is more than DE

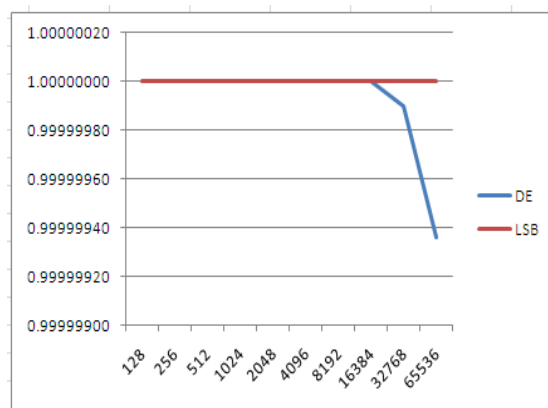
The SSIM Index for all the watermarked images mentioned here were found out. Figure 6 shows that SSIMIndex for DE watermarked images has a lower value indicating more noise and Figure 7 shows that the Rotation LSB replacement is faster than DE for a given payload.

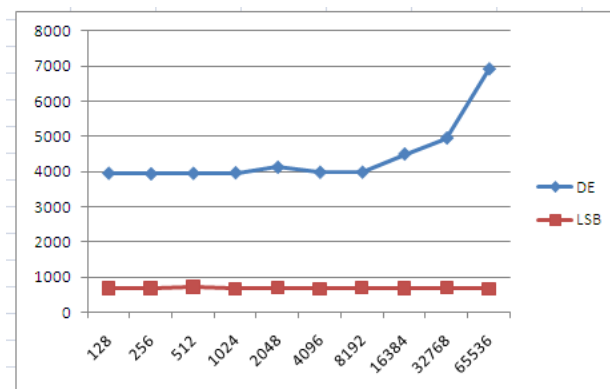
Payload	DE	LSB
128	1.00000000	1.00000000
256	1.00000000	1.00000000
512	1.00000000	1.00000000
1024	1.00000000	1.00000000
2048	1.00000000	1.00000000
4096	1.00000000	1.00000000
8192	1.00000000	1.00000000
16384	1.00000000	1.00000000
32768	0.99999990	1.00000000
65536	0.99999936	1.00000000

**TABLE 1:** Value of SSIMIndex for different payloads using DE and RRL

Payload	DE	LSB
128	3954	688
256	3937	687
512	3953	719
1024	3968	672
2048	4125	703
4096	3984	672
8192	3985	703
16384	4500	687
32768	4953	703
65536	6906	672

**TABLE 2:** Value of execution time for different payloads using DE and RRL



**FIGURE 6:** Graph corresponding to Table 1**FIGURE 7:** Graph corresponding to Table 2

### 8. CONCLUSIONS AND FUTURE WORK

Biometric images can be secured using watermarking. Reversible watermarking helps us to have images with minimum distortion. At the same time, the original image can be retrieved for any further use. Rotational Replacement of LSB is a high performing scheme for moderate payload. SSIM Index is proved to be more accurate for measuring the quality of watermarked image. Consequently, in the biometric image security field SSIM Index can be used instead of PSNR to get better result. For reversible watermarking, there is a limitation of maximum payload, depending on the capacity of original image being used. Our future work will be focussing on calculating the performance of feature level extraction for a watermarked image using rotation replacement of LSB.

### REFERENCES

- [1] Z. Wang, A. Conrad, H. Rahim, and E. Simoncelli, "Image Quality Assessment: From Error Visibility to Structural Similarity," *IEEE Trans. Image Processing*, Vol. 13, No. 4, April 2004.
- [2] J.M.Barton, "Method and apparatus for embedding authentication information within digital data," U.S.Patent 5 646 997.1997.
- [3] J. Tian, "Reversible data embedding using a difference expansion," *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 13, pp. 890–896, Aug.2003.
- [4] A. M. Alattar, "Reversible watermark using the difference expansion of a generalized integer transform," *IEEE Trans. Image Process.*, Vol.13, no. 8, pp. 1147–1156, Aug. 2004.
- [5] Multibiometric Authentication- An Overview of Recent Developments -Term Project CS574, Spring 2003, Available: <http://www.ub-net.de> on April 2010.
- [6] Ya-Hui Shiao, Tzong-Jer Chen, Keh-Shih Chuang, Cheng-Hsun Lin and Chun-Chao Chuang, "Quality of Compressed Medical Images," *Journal of Digital Imaging*, Vol. 20, No. 2, 149-159, DOI: 10.1007/s10278-007-9013-z.
- [7] Jen-Bang Feng, luon-Chang Lin, Chwei-Shyong Tsai, and Yen-Ping Chu, "Reversible Watermarking: Current Status and Key Issues," *International Journal of Network Security*, Vol.2, No.3, PP.161–171, May 2006.
- [8] Lixin Luo, Zhenyong Chen, Ming Chen, Xiao Zeng, and Zhang Xiong, "Reversible Image Watermarking Using Interpolation Technique," *IEEE Transactions on information forensics and security*, Vol.5, No.1, March 2010.

# A High Performance Modified SPIHT for Scalable Image Compression

**Bibhuprasad Mohanty**

*Assistant professor,  
Institute of Technical Education and Research  
SoA University, Bhubaneswar*

*bmohanty.iit07@gmail.com.*

**Abhishek Singh**

*Assistant professor  
Faculty of Science and Technology,  
ICFAI University, Tripura*

*abhishek.uor@gmail.com.*

**Dr. Sudipta Mahapatra**

*Associate Professor,  
Department of E &ECE,  
Indian Institute of Technology, Kharagpur, India*

*sudipta.mahapatra@gmail.com*

---

## Abstract

In this paper, we present a novel extension technique to the Set Partitioning in Hierarchical Trees (SPIHT) based image compression with spatial scalability. The present modification and the preprocessing techniques provide significantly better quality (both subjectively and objectively) reconstruction at the decoder with little additional computational complexity. There are two proposals for this paper. Firstly, we propose a pre-processing scheme, called Zero-Shifting, that brings the spatial values in signed integer range without changing the dynamic ranges, so that the transformed coefficient calculation becomes more consistent. For that reason, we have to modify the initialization step of the SPIHT algorithms. The experiments demonstrate a significant improvement in visual quality and faster encoding and decoding than the original one. Secondly, we incorporate the idea to facilitate resolution scalable decoding (not incorporated in original SPIHT) by rearranging the order of the encoded output bit stream. During the sorting pass of the SPIHT algorithm, we model the transformed coefficient based on the probability of significance, at a fixed threshold of the offspring. Calling it a fixed context model and generating a Huffman code for each context, we achieve comparable compression efficiency to that of arithmetic coder, but with much less computational complexity and processing time. As far as objective quality assessment of the reconstructed image is concerned, we have compared our results with popular Peak Signal to Noise Ratio (PSNR) and with Structural Similarity Index (SSIM). Both these metrics show that our proposed work is an improvement over the original one.

**Key Words :** Zero Shifting, 2D SPIHT, Lifting Wavelet Transform, Context Modeling, Resolution Scalability, SSIM

---

## 1. INTRODUCTION

Image compression algorithms manipulate images to dramatically reduce the storage and bandwidth required while maximizing perceived image quality. Image compression based on discrete wavelet transform (DWT) has emerged as one of the popular tool. The wavelet domain provides a natural setting for many applications which involve real world signals. The presence of blocking artifacts at low bit rate in the entropy coding schemes (e.g., Discrete Cosine Transform) is completely avoided in this wavelet based coding. This transform achieves better energy compaction than the DCT and hence can help in providing better compression for the same quality measurement, like peak-signal-to-noise-ratio (PSNR) [1] and structural similarity (SSIM) [2].



The widespread deployment of diversified and heterogeneous digital networks requires different bandwidth range for various user nodes. This heterogeneity property demands for scalability to optimally serve each user according to the available bandwidth and computing capabilities. The goal of the scalable compression is to enable clients to reconstruct the image from a single compressed bit stream, according to their specific capabilities. A scalable coder generates a bit stream which consists of a set of embedded parts that offer better SNR and/or greater spatial resolution. Different types of decoders with different complexity and various access bandwidths can be accommodated in the case of an entirely scalable bit stream,

Based on DWT, several competitive image compression coders with high coding efficiency have been developed. Among those coders, Shapiro's Embedded Zero tree Wavelet (EZW) compression [3], Said and Pearlman's set partitioning in Hierarchical Trees (SPIHT) [4], and Tubman's Embedded Block Coding with Optimized Truncation (EBCOT) [5] are the benchmark contribution. The 2-D SPIHT coder performs competitively with most of the other coders published in the literature [6]. The encoded and compressed SPIHT bit stream is further compressed by applying adaptive arithmetic encoder which provides an additional PSNR gain of 0.2 to 0.6 dB [4] over the original one. The low margin PSNR improvement shows the fact that SPIHT itself is exploiting the redundancies across the sub band so efficiently that it leaves little room for further compression by the arithmetic coder. However, use of arithmetic coder increases the computational complexity and hence the scheme sacrifices its speed for a meager gain.

Further improvements of SPIHT have been published in [7-12]. In [13], the authors proposed a novel modification in exploiting the correlation of coefficients by lifting based DWT with 1D addressing method instead of traditional 2D method. Also, for the purpose of hardware implementation, they modified the SPIHT algorithm by interchanging the refinement pass and the sorting pass. Although almost all of the state-of-the-art zero tree-based image compression methods are SNR scalable and provide bit streams for progressive (by quality) image compression, they do not explicitly support spatial scalability and do not provide a bitstream which can be adapted easily to a heterogeneous environment. PROGRES [12] brings the concept of resolution scalability as well as faster encoding and decoding into the SPIHT, but it loses desirable feature like embeddedness and SNR scalable decoding.

In this paper, we present a new scheme which provides significant improvement in the quality of the reconstructed image in terms of PSNR and SSIM without using the adaptive arithmetic encoder. The method, called zero shifting, is applied on the spatial domain of the image before it is wavelet transformed [18]. To accommodate the zero shifting schemes without sacrificing the inherent property of the SPIHT codec, we have modified the original algorithm in [4]. Further, we modeled the transformed coefficient according to its significance in a 2x2 adjacent offspring group as a fixed context. The systematic trend in probability distribution of such a context has been utilized for further compression of the image with the help of Huffman coding which is of lower complexity than the arithmetic coding [19]. We also propose an approach to resolution scalability by grouping together the bit streams generated for each resolution at one location. Both the sorting and refinement pass for SPIHT encoder is performed for each resolution and certain header information are generated which indicate the boundary condition for that resolution. The paper is organized as follows: Section 2 reviews the Lifting based wavelet transform and SPIHT coder briefly. The motivations for the proposed schemes are discussed in Section 3. The codec structure is described in Section 4 and we have discussed the results elaborately in Section 5. The conclusions of this paper are reported in Section 6.

## **2. BRIEF REVIEW**

### **2.1 Lifting Based Wavelet Transform (WT)**

Wavelet transforms provide both spatial and frequency-domain information about an image or a sequence of frames [14]. Such multiresolution representations of images/frames are often a more convenient format for image coding. This is particularly important since Human Vision System

(HVS) is less sensitive to high-frequency distortions than low-frequency ones. WT analyzes a signal at different frequency bands with different resolutions by decomposing it into a coarse approximation and detail information. The classical 2D transform is performed by two separate 1D transforms along the rows and the columns of the image data, resulting at each decomposition step in a low pass image or the coarse scale approximation, and three detail images. Because of their inherent multiresolution signal representation, wavelet-based coding schemes have the potential to support both SNR and spatial scalability [15].

The lifting scheme is a fast and efficient method to realize biorthogonal wavelet transforms [16]. It is simple, fast and reversible in computation due to the use of integer transformation. The algorithm has the advantages of using lifting, instead of convolution, which reduces the memory requirement. This influences the lossless compression algorithm profoundly because of its much wider dynamic range and hence JPEG 2000 uses the scheme. Even 3D wavelet decomposition is computed by applying three separate 1D transforms along the video data: rows and columns of the image as two directions and time as the third direction.

### **2.2 Set Partitioning In Hierarchical Trees (SPIHT)**

SPIHT is an embedded coding technique. In embedded coding algorithms, encoding of the same signal at lower bit rate is embedded at the beginning of the bit stream for the target bit rate. Effectively, bits are ordered in importance. This type of coding is especially useful for progressive transmission using an embedded code; where an encoder can terminate the encoding process at any point. SPIHT algorithm is based on following concepts [4]:

1. Ordered bit plane progressive transmission.
2. Set partitioning sorting algorithm.
3. Spatial orientation trees.

SPIHT keeps three lists: LIP, LSP, and LIS. LIP stores insignificant pixels, LSP stores significant pixels and LIS stores insignificant sets. At the beginning, LSP is empty, LIP keeps all coefficients in the lowest sub band, and LIS keeps all tree roots which are at the lowest sub band. SPIHT starts coding by running two passes. The first pass is the sorting pass. It first browses the LIP and moves all significant coefficients to LSP and outputs its sign. Then it browses LIS executing the significance information and following the partitioning sorting algorithms.

The second pass is the refining pass. It browses the coefficients in LSP and outputs a single bit alone based on the current threshold. After the two passes are finished, the threshold is divided by 2 and the encoder executes the two passes again. This procedure is recursively applied until the number of output bits reaches the desired number.

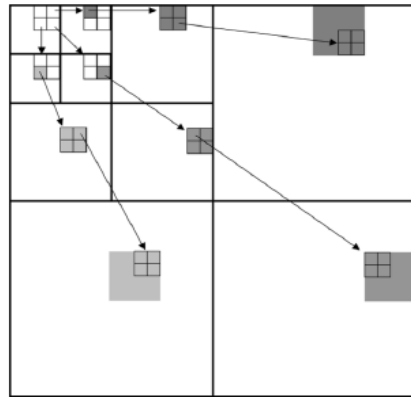
## **3. THE PROPOSED SCHEME**

### **3.1 The Zero Shifting**

The method of zero shifting is a simple and easy-to-implement technique which preserves the embedded property of the SPIHT coding. By downward scaling of the pixel values by  $2^{(N-1)}$ ,  $N$  being the number of bits representing the original pixel, the spatial domain magnitude becomes bipolar ranging from  $(-2^{(N-1)})$  to  $(+2^{(N-1)}-1)$ , with almost half of the maximum absolute spatial value. In the transform domain, the wavelet lifting scheme produces the high pass sub band by taking the weighted differences of the pixel values after prediction. Similarly, the low pass sub band is produced by taking the weighted average of the pixel values after the updating step. This low pass sub band is further decomposed iteratively for each level of decomposition to finally provide, one lowest frequency band, the rest being high frequency bands.

SPIHT coding is a method of bit plane coding technique. The highest number bit plane is decided by the maximum absolute value of the transformed coefficients. As the maximum coefficient decreases by the method of zero shifting, SPIHT algorithm takes less time for encoding as well as for decoding. Further, the lower absolute image values in the spatial domain (which originally ranges from  $0 \leq p_{i,j} \leq 2^{(N-1)}$ ) are shifted to higher absolute range in the bipolar sense. The lower

image values are detailed information pertaining to the boundaries, edges, etc. These coefficients become significant at a higher threshold value and accordingly are ordered earlier in the bit plane progressive transmission of the SPIHT bit stream. Since SPIHT algorithm always encodes a sign bit for significant coefficients, no extra bit budget is necessary for being bipolar. By terminating the decoding process even at a lower bit rate, the quality improves, both subjectively and objectively, because of the inclusion of some detail information at that rate.



**FIGURE 1.** Spatial Orientation tree in 2D SPIHT (source [8])

### 3.2 Modified SPIHT

The parent offspring dependency and corresponding spatial orientation trees for the SPIHT algorithm is as shown in Fig. 1 for three levels of decomposition. Each node of the tree is represented by its coordinates in the algorithm. The tree is defined in such a way that each node has no offspring (leaves) or four off springs at the same spatial location in the next finer level [4].

In original SPIHT, the pixels in the coarsest level e.g., LL sub band of the 3rd level for Fig. 1, are the tree roots and they are grouped into blocks of 2x2 adjacent pixels. In each block one (the left top corner) pixel has no descendant. In the initialization stage, the coordinates of all coefficients in the coarsest level are put into LIP list and the coordinates of coefficients having descendants only into the LIS list as D-type sets. We have modified the initialization stage of the SPIHT algorithm to accommodate the principle of zero shifting as follows: if the decomposition level is such that the coarsest level LL subband consists of a single pixel (e.g., for an image of 512x512 size decomposed into 9 level), then we set the LIS list with the HL, LH, and HH coordinates of the coarsest level (each of which is also a single pixel). Otherwise, we initialize the LIS list with all the coordinates of the LL sub band as of the LIP list. Accordingly, in the sorting pass for LIS, the SPIHT first tests the significance in the LH, HL, and HH band of the lowest subband. For each root in LL, it compares the maximum value between the other three bands at the same spatial orientation and sorts it in the output bit stream. Once this test is over, the set partition rule of SPIHT [6] divides the remaining descendants into either D-type set or L-type set. The decoder duplicates the path of execution of the encoder. At the encoder we are at a liberty to encode the entire bit plane without any bit budget constraint. Experiments carried out over different type of images with spatially different features show that the proposed modification provides a faster codec than the original one while resulting almost the same PSNR values.

### 3.3: Fixed Context Modeling

SPIHT algorithm generates significant information during encoding of LIS and LIP. LIP is not supposed to generate any redundancies as they are distinctly decorrelated. When LIS of type A becomes significant at certain threshold, the significance information of offspring of 2x2 blocks are generated. We exploited the possibility of such four offspring being significant. Assigning 0 to insignificant and 1 to significant coefficient, we created 16 possible combination of such configuration, starting from 0000 to 1111. We called each configuration a context and found the probability of occurrence of each context model (Fig. 6A and 6B). The systematic trend found in

probability distribution of such context has been utilized for further compression of the image with the help of Huffman coding which has a lower complexity than the arithmetic coding.

### 3.4 Fixed Context Resolution Scalability (FCRS)

SPIHT algorithm provides SNR (quality) scalability due to its bit plane coding strategy but fails to provide resolution scalability. Each bit plane is coded in two passes – Significant pass and then Refinement pass. Each pass generates the bitstream as a whole in an interleaved fashion without distinguishing between any resolutions so needed (Fig 2(a)). If the decoder wants to decode an image at a given resolution then all the information carrying bits of that resolution should be available at one place. In addition, the knowledge of boundary of separation of different resolution should be available. We have modified the bitstream structure in a way so as to gather all the information pertaining to a particular resolution at a place as shown in Fig.2 (b), where we have shown only three resolutions as per our simulation. It is to be noted that we have incorporated the context modeling into the SPIHT algorithm. Here, the bitstream for both the passes are adjacent to each other corresponding to different individual resolution. Thus, coding of each bit plane should be split in parts depending upon the specific resolution required and each part should contain the corresponding information generated by both the significance and refinement passes for that resolution.

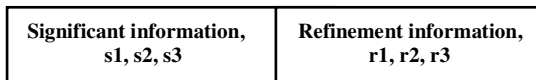


FIGURE 2(a). Structure of output bit stream

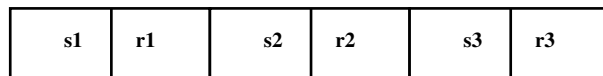
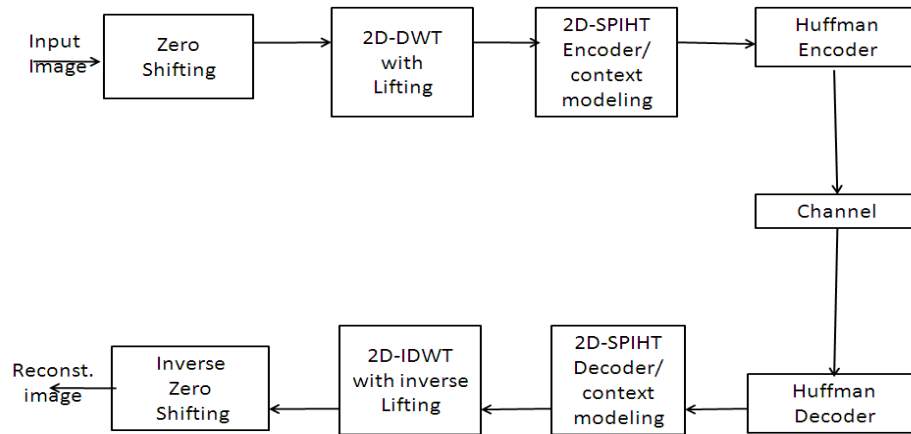


FIGURE 2(b). Modified Structure

For each resolution, the sorting pass is carried out following the refinement pass for all the threshold values and kept in the output bitstream as per the progressive bit plane coding. Starting from the resolution 1, the algorithm encodes the entire image and generates a header which carries the information regarding the boundary for a particular resolution. Say, for resolution 1, we process those LIS of type A whose offspring are also lying in the resolution level 1 only. Similarly, while processing LIP, we are processing only those coefficients which belong to the same resolution. The header, which stores all the boundary information, is sent to the decoder. The reconstruction at the decoder at a particular resolution level is done using the header information.

## 4. THE CODEC STRUCTURE

The well known SPIHT algorithm encodes the wavelet transformed coefficient in a bit plane manner and outputs the significant information with respect to a specified threshold and its sign. We pre processed the pixel values by shifting downwards by  $2^{N-1}$ , N being the no of bits in the original image, and then applied the 2D wavelet on it. The transformed coefficient is then encoded with 2D SPIHT coder (Fig.3).



**FIGURE 3.** Proposed image codec with zero shifting

In the decoder, the execution path of the encoder is duplicated and the values of inverse transformed coefficients are shifted upwards by  $2^{(N-1)}$  to get back the original image.

## 5. RESULTS AND DISCUSSION

### 5.1 Quality Measurement Metrics

During the process of said compression, the reconstructed image is subject to a wide variety of distortion. Subjective evaluations emphasizes on the visual image quality, which is too inconvenient, time consuming, and complex. The objective image quality metrics like Peak Signal to Noise Ratio (PSNR), or Mean Squared Error (MSE) and structural Similarity (SSIM) [2] are thought to be the best for the image processing application.

The MSE metric is most widely used for it is simple to calculate, having clear physical interpretation and mathematically convenient. MSE is computed by averaging the squared intensity difference of reconstructed image,  $\hat{x}$  and the original image,  $x$ . Then from it the PSNR is calculated. Mathematically,

$$MSE = \frac{1}{M \times N} [x(i, j) - \hat{x}(i, j)]^2, \tag{1}$$

Where,  $M \times N$  is the size of the image and assuming the grey scale image of 8 bits per pixel (bpp), the PSNR is defined as,

$$PSNR = 10 \log_{10} \frac{255^2}{MSE} \tag{2}$$

But, they are not very well matched to human visual perception. The authors in [2] have shown that even for the same PSNR values, the visual quality for two reconstructed images is different at different bit rate. It is because the pixels of the natural images are highly structured and exhibit strong dependencies. The SSIM system separates the task of similarity measurement into three comparisons: luminance, contrast, and structure. The overall similarity measure is defined as [2,17]

$$S(x, \hat{x}) = f(l(x, \hat{x})) \cdot c(x, \hat{x}) \cdot s(x, \hat{x}) \tag{3}$$

Where,  $x$  is the original assuming to be completely perfect and  $\hat{x}$  is the reconstructed image such that both are non negative images.  $S=1$  implies a perfect similarity between the images.

- $l(x, \hat{x})$  = luminance comparison function, which in turn the function of the mean values of  $x$  and  $\hat{x}$ ;
- $c(x, \hat{x})$  = contrast comparison function and they are function of the standard deviation of  $x$  and  $\hat{x}$ ; and

$s(x, y)$  = structure comparison function

### 5.2 Algorithm Complexity in Terms of Encoding/decoding Time

We will present some evidence of our implementation with Matlab in this section. Experiments are performed with grey scale ‘Lena’ and ‘Ship’ image for their established popularity. Ship image contains more details than Lena image. All experiments are done with Pentium IV, 1.76 GHz, 512 MB RAM. We have considered two type of filters for their establish popularity in the image compression application. They are ‘bior4.4’ and ‘cdf9/7’ with half point symmetry at the boundary.

Level of decomposition	With BM	With CM	With BMZ	With CMZ	With BO	With CO	With BOZ	With COZ
3	3.4735	3.2595	3.1445	2.9640	35.6672	34.6788	40.9529	41.2126
4	2.1118	2.0357	1.8985	1.9531	68.7433	739693	75.9793	86.8388
5	1.6166	1.5989	1.5829	1.6727	87.8300	88.0677	102.1387	88.9418
6	1.4674	1.5062	1.4425	1.4883	90.3934	92.3333	89.8692	96.6292
7	1.4110	1.5085	1.4085	1.4189	87.7200	93.4404	94.7028	93.2454
8	1.3989	1.5181	1.3745	1.4009	95.7797	93.2782	97.9520	94.9666
9	1.8483	1.4252	1.4139	1.4035	98.4183	104.0956	95.5076	110.1410

B (C) M (O) : With Bior4.4 (CDF 9/7) filter for Modified ( Original) SPIHT algorithm with 1bpp,  
 Z : with method of Zero-shifting

**TABLE 1:** Complexity Comparison in terms of Encoding Time (sec)

It is evident from the Table 1 that, with modification of the algorithm, the time of encoding decreases significantly, i.e., the codec becomes faster irrespective of the type of the filter used. The modified algorithm takes 10 to 90 times less time than that for the original one. Interestingly, the encoding time goes on decreasing as the number of decomposition increases from 3 to 9 (for square image of resolution size 512x512) in steps of 1, whereas the time increases almost linearly with the original one. By the method of zero shifting, the original algorithm takes even more time for encoding than that for without the said pre processing steps. It is found that for the modified algorithm, the encoding time remains almost constant for 6th level of decomposition or more. So for other experiments with the modified algorithm, we have spatially decomposed the image for 8 levels, unless otherwise stated. After verifying the complexity in terms encoding time, we are conducting the following experiment to verify the quality of the reconstructed image with the proposed scheme of zero shifting. The following Table 2 and Table 3 provide some detail of decoding time, the PSNR value and SSIM for the reconstructed LENA image at different bit rate for both the algorithms with ‘bior4.4 filter.

Bit rate in (bpp)	Decoding time (sec) without Zero shifting	Decoding time (sec) with Zero shifting	SSIM without Zero shifting	SSIM with Zero shifting	PSNR without Zero shifting	PSNR with Zero shifting
0.01	0.06	0.03	0.4574	0.5298	17.77	20.89
0.11	0.14	0.05	0.8425	0.8464	26.98	28.67
0.21	0.22	0.07	0.9077	0.9125	29.95	31.59
0.31	0.26	0.10	0.9340	0.9435	32.46	32.87
0.41	0.33	0.13	0.9550	0.9532	33.69	34.78
0.51	0.35	0.13	0.9638	0.9671	35.31	35.51
0.61	0.44	0.16	0.9674	0.9716	35.85	36.07
0.71	0.53	0.20	0.9705	0.9744	36.35	36.53
0.81	0.60	0.25	0.9790	0.9784	37.31	37.98
0.91	0.62	0.24	0.9818	0.9826	38.28	38.44

**TABLE 2 :** Decoding time, PSNR and SSIM for reconstructed LENA with Modified SPIHT

It is to be noted that the modified algorithm encodes the image for all the bitplane; whereas the original is restricted to the specified bit budget. However, it is possible to encode with the modified algorithm with the same bit budget constraints. It is evident from the observations from both these tables that the modification in the algorithm makes the decoding faster. The objective metric like PSNR also improves due to modification in the initialization of the image pixel coordinate and hence in the scanning process. However, the SSIM metric is better for the original one at some bit rate. In the following we are showing the comparison curves for LENA and SHIP images. The experiments are performed with 6 level of spatial decomposition with bior4.4 filter.

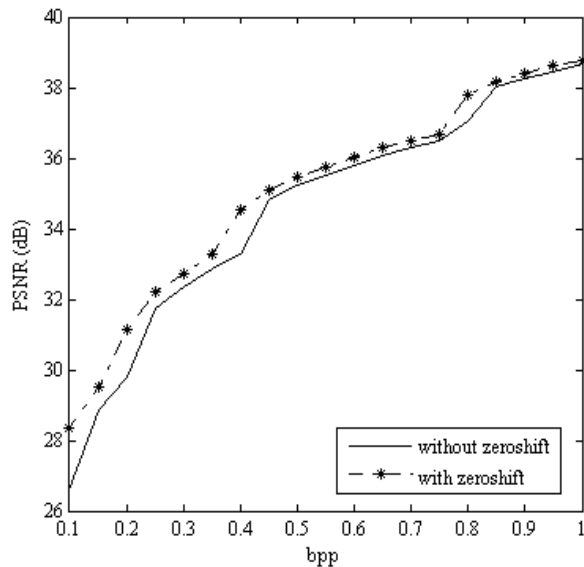
### 5.3 Results with Proposed Image Codec

We have found that 6 or more level of spatial decomposition provides almost constant PSNR for any bit rate. Fig. 4 and 5 shows the performances of PSNR of the Lena and Ship image at different bpp. It is quite evident from the curves that the modified algorithm significantly improves the performances for all bit rates. The improvement is more pronounced in Lena than in Ship, because Ship contains more detail information which are not properly exploited by the SPIHT algorithm. At the lowest possible decoding rate (i.e., 0.1 to 0.2), where the perceptual quality is significantly good, our zero shifting scheme provides almost 0.5 to 1dB higher PSNR than the original one. Experiments show that at 0.15 bpp the visual quality is quite good. Further, on wavelet transforming the shifted pixel values, coefficients are generated which are mostly decorrelated in sign. Experimental observations reveal that these correlations are more pronounced in the lowest frequency sub band.

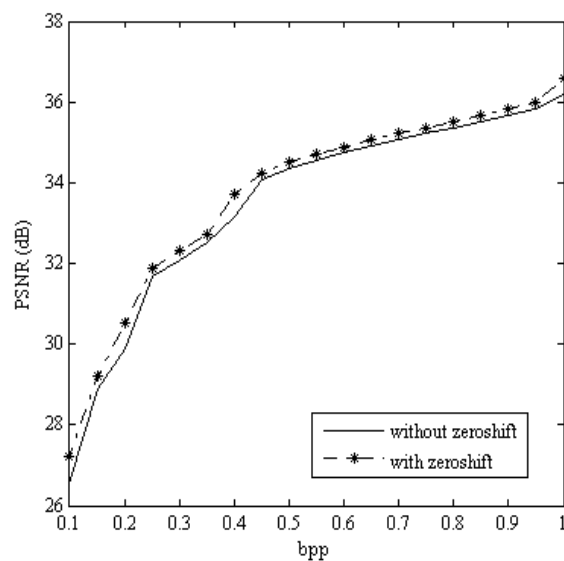
Bit rate in (bpp)	Decoding time (sec) without Zero shifting	Decoding time (sec) with Zero shifting	SSIM without Zero shifting	SSIM with Zero shifting	PSNR without Zero shifting	PSNR with Zero shifting
0.01	0.039	0.043	0.4697	0.4093	16.72	20.59
0.11	1.005	0.953	0.8590	0.8622	28.30	29.4
0.21	2.501	2.64	0.9272	0.9221	31.61	32.38
0.31	4.483	5.739	0.9536	0.9517	33.69	34.11
0.41	9.178	9.002	0.9639	0.9629	35.03	35.60
0.51	12.995	17.53	0.9729	0.9728	36.37	36.62
0.61	21.854	24.955	0.9776	0.9765	37.12	37.39
0.71	33.991	38.809	0.9804	0.9795	37.83	38.05
0.81	47.440	49.256	0.9830	0.9828	38.46	38.74
0.91	58.591	65.62	0.9855	0.9857	39.18	39.37

**TABLE 3:** Decoding time, PSNR and SSIM for reconstructed LENA with Original SPIHT

Since in the process of SPIHT encoding we are at a liberty to invest an extra bit always for testing any significance, the reconstruction at the lower bit rate is better. Also, due to efficient grouping of the insignificant information in LIS with our proposed scheme, the no of bits to be transmitted to decoder also reduced to a factor of 4. Hence truncating the bit stream at lower bit rate provides much of the significant information to the decoder and the reconstruction quality also improves.



**FIGURE 4.** Comparison of Lena image



**FIGURE 5.** Comparison of Ship image



### 5.4 Result With Context Modeling and Scalability

Extensive experiments are carried out with lot many images to find the pattern of the probability distribution of the significant coefficient in the 2x2 blocks. We found that the distribution of almost all the images of varying characteristics show a fixed pattern. The distribution for two characteristically different grey scale images are shown; Lena image with much less detail within the sub band (Fig. 6A), and Ship image with much high details (Fig. 6B). By taking the average of this probability distribution we model that context.

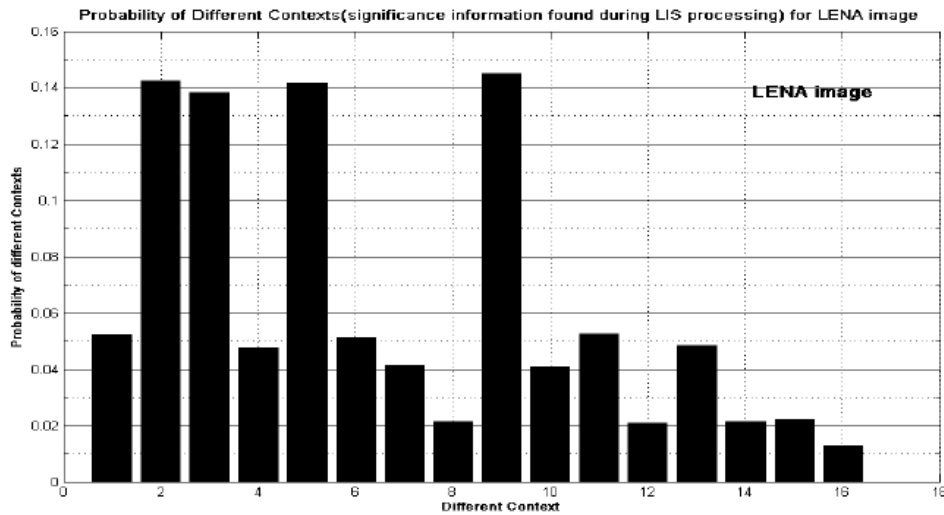


FIGURE 6A. Context Count for Lena

We prepared a table to store the generated Huffman codes corresponding to each model and at the time of encoding, we utilized this table directly. We compared the result of this context modeling with that of the arithmetic encoding. We define 128x128 sizes as resolution level 1, 256x256 as resolution level 2 and 512x512 as resolution level 3. All results are obtained by applying 6 level of spatial wavelet decomposition with 'bior 4.4' filter. The modified SPIHT encoder with the method of zero shifting is set to produce a bit stream that supports three number of spatial scalability levels as mentioned above. The PSNR calculation is carried out with the following formula:

$$PSNR = \frac{10 \log_{10} (255 \times 2^{(s-resolution)})^2}{MSE}, db \tag{4}$$

Where, MSE is the Mean Squared Error calculated between reconstructed image at a particular resolution and the original image of the same size as in eqn (2).

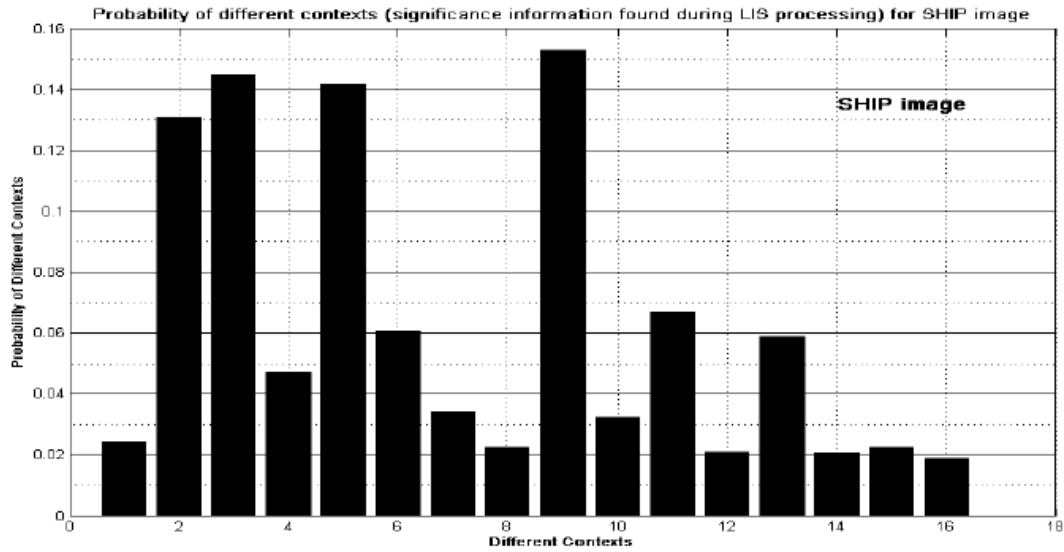


FIGURE 6B. Context Count for Ship

The rate distortion curves for each resolution in Fig.7 and the corresponding reconstructed LENA images in Fig.8 clearly shows that our proposed encoder completely keeps the progressiveness (by SNR) property of the original SPIHT algorithm. For resolution level 1 and 2, our resolution scalable decoder obtained the proper bit stream tailored by the encoder for that resolution level.

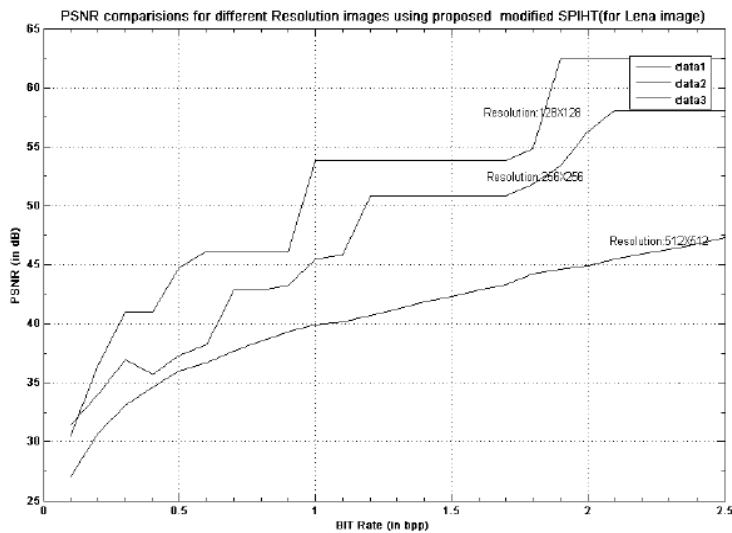


FIGURE 7. Performance comparison of FCRS SPIHT for different resolution



(a) Resolution 1 (128x128)



(b) Resolution 2 (256x256)



(c) Resolution 3(512x512)

**FIGURE 8.** FCRS reconstructed Lena at 0.5 bpp for 3 different level

As expected, the performance of scalable decoder is much better than for SPIHT and as the resolution level becomes smaller (hence the scale becomes larger), the difference becomes more significant. All the results are obtained with the context modeling of the encoded bit stream.

## 6. CONCLUSION

In this paper, we presented a novel extension of wavelet based SPIHT coding scheme. The scheme pre processed the pixel values around zero before any wavelet transform takes place. Higher PSNR and more compression are obtained with the scheme without using the popular arithmetic coding. Experimental results show that it performs better measurably and also at lower bit rate its performance is well suited for the purpose of surveillance and monitoring system. We have presented an extension to the original SPIHT by incorporating spatial scalability feature to the encoded bit stream. The bit stream can be stopped at any point to meet a bit budget during the process of coding and decoding for a particular resolution level. To lessen the burden of computation on the codec, we proposed a Huffman coding based context modeling to achieve further compression at each resolution level. This technique can be extended for combined SNR, spatial and frame rate scalable video coding.

## 7. REFERENCES

- [1] G. Wallace, "The JPEG still picture compression standard," *Communications of ACM*, vol. 34, no. 4, 1991.
- [2] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600-612, Apr., 2004.
- [3] J.M. Shapiro, "Embedded image coding using Zero trees of wavelet coefficients," *IEEE Trans. on Signal Processing*, vol.41, pp. 3445-3462, Dec.,1993.
- [4] A.Said and W.A.Permalink, "A New, Fast, and Efficient Image Codec Based on Set Partitioning in Hierarchical Trees," *IEEE Trans. on Circuits and Systems for Video Technology*, vol.6, pp. 243-250, June, 1996.

- [5] D. Taubman, "High Performance Scalable Image Compression with EBCOT," *IEEE Trans. on Image Processing*, vol. 9, pp. 1158-1170, July, 2000.
- [6] B.-J. Kim and W. A. Pearlman, "An embedded video coder using three-dimensional set partitioning in hierarchical trees (SPIHT)," in *proc. IEEE Data Compression Conf.*, pp. 251-260, Mar., 1997.
- [7] J. Karlenkar and U. B. Desai, "SPIHT video coder," in *Proc. IEEE Region 10 International Conference on Global Connectivity in Energy, Computer, Communication and Control, TENCON'98*, vol. 1, pp. 45-48, 1998.
- [8] B.-J. Kim, Z. Xiong, and W. A. Pearlman, "Low bit-rate scalable video coding with 3-d set partitioning in hierarchical trees (3-D SPIHT)," *IEEE Trans. Circ. and Syst. for Video Technology*, vol. 10, no. 8, pp. 1374-1387, Dec. 2000.
- [9] J. Zhu and S. Lawson, "Improvements of the SPIHT for image coding by wavelet transform," *Proc. IEEE Seminar on Time-scale and Time-Frequency Analysis and Applications (Ref. No. 2000/019)*, pp. 24/1 -24/5, 2000.
- [10] E. Khan and M. Ghanbari, "Very low bit rate video coding using virtual spiht," *IEE Electronics Letters*, vol. 37, no. 1, pp. 40-42, Jan. 2001.
- [11] H. Cai and B. Zeng, "A new SPIHT algorithm based on variable sorting thresholds," in *Proc. IEEE Int. Symp. Circuits and Systems*, vol. 5, pp. 231-234, May 2001.
- [12] Yushin Cho, W.A. Pearlman and A. Said, "Low complexity resolution progressive image coding algorithm: progres (progressive resolution decompression)," *IEEE International Conference on Image processing*, Volume 3, pp. 49-52, 2005.
- [13] J. Jyotheshwar and S. Mahapatra, "Efficient FPGA implementation of DWT and modified SPIHT for lossless image compression," *Journal of System Architecture*, vol. 53, pp. 369-378, 2007.
- [14] S. Mallat, "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Patt. Rec. and Mach. Int.*, Vol. PAMI-11, No. 7, pp. 2091-2110, July 1989.
- [15] J.R. Ohm, "Three dimensional subband coding with motion compensation," *IEEE Trans. Image Processing*, vol. 3, no. 5, pp. 559-571, Sept. 1994.
- [16] W. Sweldens, "The Lifting Scheme: A Custom-design Construction of Biorthogonal Wavelets," *Journal of Applied Computation Harmonic Anal.*, vol. 3, issue 2, pp. 186-200, 1996.
- [17] <http://www.ece.uwaterloo.ca/~z70wang/research/ssim>
- [18] B. Mohanty, P. Verma and S. Mahapatra, "A High Performance SPIHT based Image and Video codec for surveillance," *Int. Journal of Signal and Imaging System Engg.* (in press).
- [19] A. Singh, B. Mohanty, P. Verma and S. Mahapatra, "Fixed context resolution scalable SPIHT: a novel extension," *Proc. of International conference on Communication, Computation, Control and Nanotechnology (ICN 2010)*, pp. 28-32, Oct. 2010.

# An Efficient Thresholding Neural Network Technique for High Noise Densities Environments

## **Mazhar B. Tayel**

*Faculty of engineering / Electrical engineering department /  
communication division Alexandria University  
Alexandria, Egypt*

## **Mohamed A.Abdou**

*IRI, City of Scientific Research & Technology  
Applications  
Alexandria, Egypt*

*doc\_abdelrahman@yahoo.com*

## **Azza M.Elbagoury**

*Faculty of engineering / Electrical engineering department /  
communication division Alexandria University  
Alexandria, Egypt*

*Azza\_Elbagoury@hotmail.com*

---

## **Abstract**

Medical images when infected with high noise densities lose usefulness for diagnosis and early detection purposes. Thresholding neural networks (TNN) with a new class of smooth nonlinear function have been widely used to improve the efficiency of the denoising procedure. This paper introduces better solution for medical images in noisy environments which serves in early detection of breast cancer tumor. The proposed algorithm is based on two consecutive phases. Image denoising, where an adaptive learning TNN with remarkable time improvement and good image quality is introduced. A semi-automatic segmentation to extract suspicious regions or regions of interest (ROIs) is presented as an evaluation for the proposed technique. A set of data is then applied to show algorithm superior image quality and complexity reduction especially in high noisy environments.

**Keywords:** Thresholding Neural Networks, Image Denoising, High Noise Environments, Wavelet Shrinkage.

---

## **1. INTRODUCTION**

According to American Cancer Society, breast cancer is the second most common form of cancer in women. The chance of dying with breast-cancer is one in 33 however that number is decreasing as new forms of treatment and early detection are being implemented. Magnetic Resonance Imaging (MRI) is the state-of-the-art medical imaging technology which allows cross sectional view of the body with unprecedented tissue contrast. MRI provides a digital representation of tissue characteristic that can be obtained in any tissue plane. The images produced by an MRI scanner are best described as slices through the breast [1]. There has been considerable effort aimed at developing computer-aided diagnosis (CAD) systems that might provide a consistent and reproducible second opinion to a radiologist. Currently, most CAD systems are designed to prompt suspicious regions [2]. Early detection of breast cancer starts with qualitative analysis of medical imaging data by radiologists. But Diagnosing using MRI is a time consuming task even for highly skilled radiologists because MRI are noisy images. Thus, it is crucial to apply further work in denoising areas. This paper uses a hybrid denoising-semi automatic segmentation algorithm to solve this problem.

## **2. IMAGE DENOISING**

Denoising means suppress or remove noise from image data while preserve the image quality. Imperfect instruments, problems with acquisition process, and transmission media can all

degrade the data of interest. High noise density images might lose information if not applied to a robust denoising preprocessing toolkit before analysis either manually or via CAD systems.

### 2.1. Statistical Filters

Early methods introduced for image denoising were based on noise statistical in spatial domain [3]. Among are the histogram modification, mean filters, Gaussian filters, unsharp masking, median filters, and morphological filters [4].

### 2.2. Denoising Problem and Objective

The general denoising problem can be formulated as follows.

$$\mathbf{y} = \mathbf{x} + \mathbf{n} \quad (1)$$

Where:

$\mathbf{y}$ : is a noisy image in wavelet domain, which can be represented as finite data sample:

$$\mathbf{y} = [y_0, y_1, \dots, y_{N-1}]^T \quad (2)$$

$N$ : is the number of wavelet coefficients;

$\mathbf{x}$ : is the noise free image in wavelet domain can be represented as finite data sample:

$$\mathbf{x} = [x_0, x_1, \dots, x_{N-1}]^T \quad (3)$$

$\mathbf{n}$ : is a noise data vector. Gaussian white noise with distribution  $N(0, \sigma^2)$ .

And the objective of noise reduction is to reduce the noise in  $\mathbf{y}$  and make the denoised image  $\hat{\mathbf{x}}$  as close to the original image  $\mathbf{x}$  as possible. The measurement of the closeness is the mean square error (MSE) risk which is calculated:

$$J_{\text{MSE}} = \frac{1}{2} E \left\| \hat{\mathbf{x}} - \mathbf{x} \right\|^2 = \frac{1}{2N} \sum_{i=0}^{N-1} (x_i - \hat{x}_i)^2 \quad (4)$$

Where  $\hat{\mathbf{x}} = [\hat{x}_0, \hat{x}_1, \dots, \hat{x}_{N-1}]^T$  is denoising image or the output of thresholding function in wavelet domain.

### 2.3. Wavelet Transform and Wavelet Shrinkage

Wavelet transform (WT) has become an important tool to suppress noise [5-9]. In WT decomposition, coefficients could be categorized as those that contain the main energy of the signal and others that can be ignored [10]. Moreover, noise is spread among all the coefficients in the wavelet domain, and the WT of a noisy signal could be considered as a linear combination of the WT of the signal and noise. Research in this area aims to suppress noise significantly while preserving the original data. Noise reduction methods using WT, or wavelet shrinkage, can be summarized as follows:

First, a noisy image is decomposed in the wavelet domain. Selecting a proper threshold value to filter the coefficients, according to thresholding function, is to be applied. Finally, using the inverse wavelet transform (IWT), the reconstructed image is obtained.

### 2.4. Thresholding Function

Donoho and Johnstone [11-14] had proposed a denoising method based on thresholding in the wavelet domain. The wavelet coefficients of a noisy image usually divided into important coefficients "keeping (shrinking)" and non-important coefficients "killing". These groups are modified according to certain rules. There are two thresholding rules, namely soft and hard thresholding.

#### 2.4.1 Soft Thresholding Function

If an amplitude is smaller than a predefined threshold, it will be set to zero (kill); otherwise it will shrunk in the absolute value by an amount of the threshold, i.e.

$$\eta_s(y,t) \equiv \Delta \operatorname{sgn}(y)(|y|-t)_+ = \begin{cases} y+t, & y < -t \\ 0, & |y| \leq t \\ y-t, & y > t \end{cases} \quad (5)$$

Where  $t \geq 0$  is the threshold value.

### 2.4.2 Hard Thresholding Function

Same as soft thresholding, if its amplitude is smaller than a predefined threshold, it will be set to zero (kill); otherwise it will be kept unchanged

$$\eta_h(y,t) \equiv \begin{cases} y, & |y| > t \\ 0, & |y| \leq t \end{cases} \quad (6)$$

Soft and hard functions are shown in figure (1).

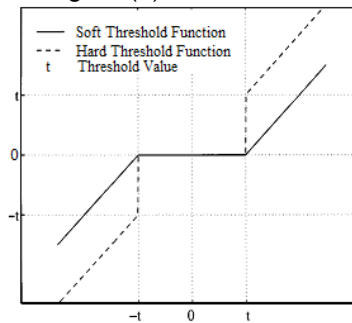


FIGURE (1): Stander soft and hard thresholding functions

## 2.5. Threshold Value

The key decision in the denoising using thresholding technique is to select an appropriate threshold. If this value is too small, the recovered image will remain noisy. On the other hand, if the value is too large, important image details will be smoothed out. Threshold values are divided into three main groups. The first group is universal-threshold in which the threshold value is the same for all wavelet detail subbands of the noisy image. The main method of this group (Visu Shrink) is presented with the hard and soft thresholding function in [11-14]. The second group is SureShrink, where the threshold value is selected differently for each detail subband [16 and 17]. In the third group each detail wavelet coefficient has its own threshold value [18].

### 2.5.1 Universal Thresholding

Donoho proposed the universal threshold [15], given by

$$t_{universal} = \sigma \sqrt{2 \log(M)} \quad (7)$$

Where:  $M$ : is the sample size.

$\sigma$ : is the noise standard deviation of noisy image. And if it is not known, a robust median estimator is used from the finest scale wavelet coefficients [12]:

$$\sigma = \frac{\Delta \operatorname{median}(|y_{ij}|)}{0.6745}, y_{ij} \in \text{subband } HH_1 \quad (8)$$

## 2.6. Thresholding Neural Network

Artificial neural networks (ANN) are composed of simple elements operating in parallel. One can train a neural network to perform a particular function by activation function and adjusting the values of the connections (weights) between internal elements. The network is adjusted or trained, based on a comparison of the output and the target essentially to minimizing the mean square error between the output and the target, until the network output matches the target. Iteration through the process of providing the network with an input and updating the network's weights is an epoch. Typically many epochs (iteration) are required to train the neural network. Zhang [19, 20] constructed a **new type of neural network** known as **thresholding neural network** (TNN) to seek the optimal threshold value and to perform the thresholding in wavelet

domain for better image denoising. A TNN is different from the conventional ANN. In TNNs, a fixed linear transform is used and the nonlinear activation function is adaptive, while in conventional multilayer neural networks the activation function is fixed and the weights of the linear connection of input are adaptive [21]. However, a TNN has some basic elements similar to ANN, i.e., interconnection of input, nonlinear activation functions, epochs, learning (threshold value), training and finding optimal MSE etc. Most ANN learning algorithms employ gradients and higher derivatives of the activation function. High-order differentiable activation functions make a neural network have better numerical properties. However, the standard soft-thresholding function is only differentiated and does not have high order derivatives. The standard hard-thresholding function is a discontinuous function and cannot be differentiated at all.

We will present new types of smooth soft thresholding and hard-thresholding functions which are infinitely differentiable. They make many gradient-based learning algorithms feasible. The input of the TNN is noisy image samples in wavelet domain  $\mathbf{y}=\mathbf{x}+\mathbf{n}$ , where  $\mathbf{x}$  is the noise free image and  $\mathbf{n}$  is additive noise and the output of the TNN is the denoised image  $\hat{\mathbf{x}}$ .

## 2.7. High-Order Differentiable Thresholding Functions

### 2.7.1 Zhang Thresholding Function

Two functions were proposed to overcome the discontinuously derivative of soft thresholding function. The first function [22] is a type of soft-thresholding function which has second order weak derivatives and proved to be useful, where  $\mathbf{k}$  is a positive integer. Note that the limit of  $\eta_{\mathbf{k}}(\mathbf{y}, \mathbf{t})$  when  $\mathbf{k} \rightarrow \infty$  is just the commonly used soft-thresholding function

$$\eta_{\text{Zhang (1998)}}(\mathbf{y}, \mathbf{t}, \mathbf{k}) = \begin{cases} \mathbf{y} + \mathbf{t} - \frac{\mathbf{t}}{2\mathbf{k} + 1} & \mathbf{y} < -\mathbf{t} \\ \frac{1}{(2\mathbf{k} + 1)\mathbf{t}^{2\mathbf{k}}} \mathbf{y}^{2\mathbf{k}+1} & |\mathbf{y}| \leq \mathbf{t} \\ \mathbf{y} - \mathbf{t} + \frac{\mathbf{t}}{2\mathbf{k} + 1} & \mathbf{y} > \mathbf{t} \end{cases} \quad (9)$$

The second [19] is smooth soft-thresholding function which is infinitely differentiable. Where  $\lambda$  is positive value when  $\lambda=0$ ,  $\eta_{\mathbf{k}}(\mathbf{x}, \mathbf{t})$  is just the standard soft thresholding function

$$\eta_{\text{Zhang (2001)}}(\mathbf{y}, \mathbf{t}, \lambda) = \mathbf{y} + 0.5(\sqrt{(\mathbf{y} - \mathbf{t})^2 + \lambda} - \sqrt{(\mathbf{y} + \mathbf{t})^2 + \lambda}) \quad (10)$$

It can be seen that the Zhang shrinkage functions perform the similar operations to the standard soft-thresholding function, as shown in figure (2). Therefore, the similar smoothness property of the estimate using this thresholding functions can be expected.

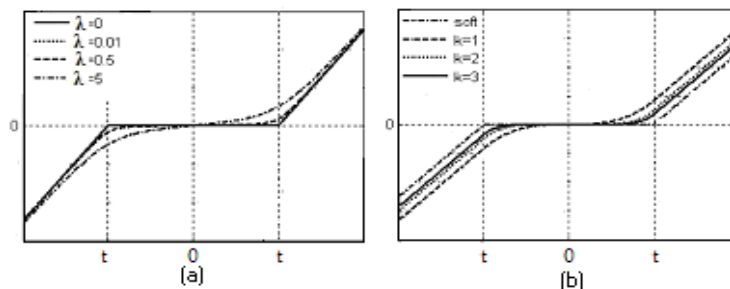


FIGURE (2): Zhang thresholding function with different values of shape tuning factor (a) Zhang2001 (b) Zhang1998.



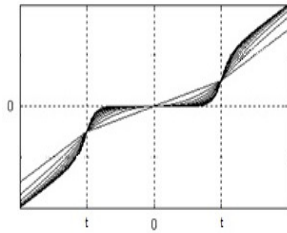
**2.7.2 Nasri and Nezamabadi Thresholding Function**

In [24], a new nonlinear thresholding function with a high capability has been presented by Nasri and Nezamabadi based on adding shape tuning factors. The differentiability property is valid through these functions and has high-order derivatives. The main difference between this function and other thresholding functions is in the non-important coefficients. Classic functions set the coefficients below the threshold value to zero, but in this proposed method these coefficients are tuned by a polynomial function.

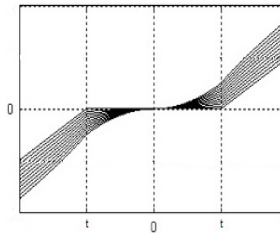
$$\eta(y, t, m, k) = \begin{cases} y + (k - 1)t - 0.5 \frac{k * t^m}{y^{m-1}} & y > t \\ 0.5 \frac{k * |y|^{m + [(2-k)/k]}}{t^{m + [(2-2k)/k]}} \text{sign}(y) & |y| \leq t \\ y - (k - 1)t - 0.5 \frac{k * (-t)^m}{y^{m-1}} & y < -t \end{cases} \quad (11)*$$

Where: **m**, **k** are shape tuning factors.

Parameter **m** determines the shape of the function for coefficients that are less and bigger than absolute threshold value. By tuning the parameter **k**, the thresholding function can be somewhere between hard and soft functions. In other words, for **k=1**, the function tends to hard thresholding function and when **k→0** it tends to soft thresholding. Figures (3 and 4) show the function for several values of the tuning parameters.



**FIGURE (3):** Nasri and Nezamabadi thresholding function for **K=1** and several values of **m** in the range [2, 10].



**FIGURE (4):** The Nasri and Nezamabadi class of thresholding function for **m=2** and **k=[0,1]**.

Note that for **k→0** the function tends to soft thresholding function.

**2.8. TNN Learning Algorithms**

One of the most practical and efficient methods in neural networks learning is least mean squares (LMS) algorithm [23]. Using LMS algorithm and thresholding function presented in [24], in each step the threshold value is adjusted along with gradient descent of the MSE risk. The equations of the algorithm as follows:

$$t_p(j+1) = t_p(j) + \Delta t_p(j) \quad (12)$$

$$\Delta t_p(j) = -\alpha \left. \frac{\partial J_{MSE}}{\partial t} \right|_{t=t_{p(j)}} \quad (13)$$

$$= -\alpha \sum_{i=0}^{M-1} \epsilon_i * \left. \frac{\partial x}{\partial t} \right|_{t=t_{p(j)}} \quad (14)$$

$$\hat{x}_i = \eta(x_i) \Big|_{t=t_{p(j)}} \quad (15)$$

$$\left. \frac{\partial x}{\partial t} \right|_{t=t_{p(j)}} = \left. \frac{\partial \eta(x_i)}{\partial t} \right|_{t=t_{p(j)}} \quad (16)$$

$$\epsilon_i = \hat{x}_i - x_i \quad (17)$$

Where **α** is Learning rate, **j** is Learning step count **M** is Length of subband **p**, and  $\epsilon_i$  is Thresholding error or the difference between denoising and original wavelet coefficients.

\* We have made a limited correction in this equation

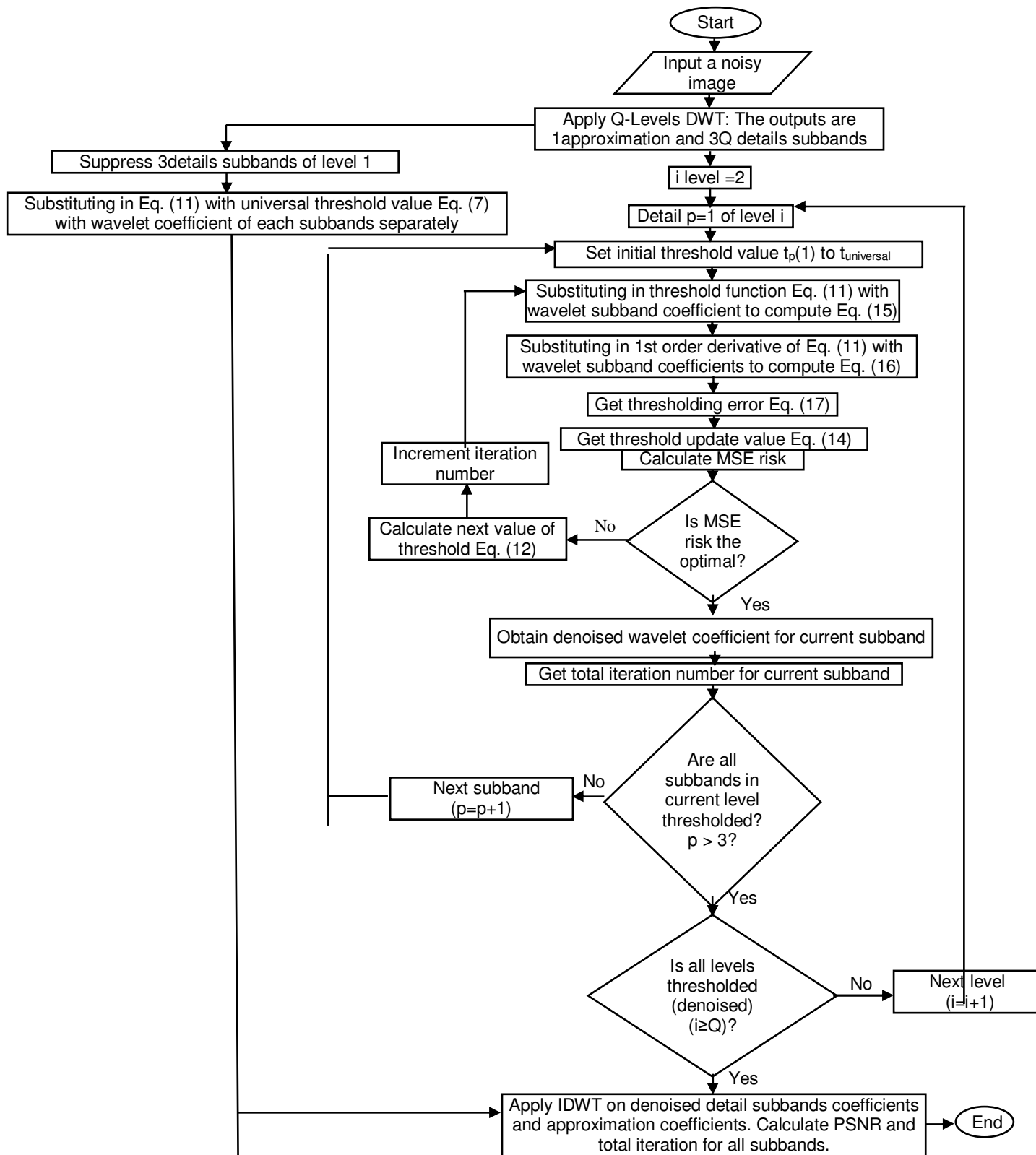


FIGURE (5): flow chart of proposed technique

### 2.9. The Proposed TNN With Modified Thresholding Concept

After studying and implementing remarkable existing TNN algorithms, it could be observed that the threshold in the highest frequency subbands when obtained by TNN does not differ from universal threshold equation (7). The work presented in this paper assumes universal threshold value in the wavelet decomposition level (1). Time saving and iterations reduction could both be obtained as consequences of this proposed idea. Figure (5) shows the flow chart of image denoising by proposed technique. The use of peak signal to noise ratio (PSNR) is very common in image processing, probably because it gives better-sounding numbers than other measures. The PSNR is given in decibel units (dB)

$$PSNR = 20 \text{Log}_{10} \left[ \frac{255}{\sqrt{\frac{1}{N} \sum_{i=0}^{N-1} (x - \hat{x})^2}} \right] \quad (18)$$

### 3. SEGMENTATION ALGORITHM

Segmentation is to distinguish important regions of interest (ROI) from the background after the denoising phase. The main idea for applying the segmentation technique is to measure the performance of the proposed denoising algorithm at different noise densities. Segmentation techniques can be classified into two main categories: edge-based segmentation (locating object's boundary using image gradient) and region-based segmentation (identifying all pixels that belong to the object based on the intensity of pixels) techniques. Edge-based techniques such as Roberts, Prewitt, Robinson, Kirsch, Laplacian and Frei-Chen [25] have been well studied. Region-based techniques such as Region growing [26], Watershed algorithm [27], and Thresholding [28] are more suitable for breast tumors extraction since suspicious regions are belonging to the same texture class, while surrounding tissues are belonging to others.

#### 3.1. Proposed Algorithm

In this paper semi-automatic region based thresholding method is used. The word 'Semi-automatic' means squaring the expected suspicious regions to remove the top body edges (for example, the shoulders, pectoral muscle and breastbone). Breast tissues (for example, ducts, fat, lobules, lymph nodes and lymph vessels) are also excluded. The infected tissues are then selected automatically by selecting a threshold based on the image histogram or local statistics such as mean value, standard deviation and the local gradient [29]. For a given image, the binarization can be done using the pixel intensity values and is given

$$\text{bin}(x, y) = \begin{cases} 1 & I(x, y) \geq T \\ 0 & I(x, y) < T \end{cases} \quad (19)$$

Where  $\text{bin}(x, y)$  is the resulting binary image and  $T$  is the threshold value.

The threshold value can be estimated based on the mean of the pixel intensity ( $M$ ) and the standard deviation ( $\sigma$ ) for each region of interest. It is given by

$$T = M + \alpha \cdot \sigma + M \cdot K$$

Where  $\alpha$  and  $K$  are the constants.

### 4. RESULTS AND DISCUSSION

In image denoising, six selected algorithms: soft, hard, Zhang (1998), Zhang (2001), Nasri-Nezamabadi with universal threshold values, and Nasri-Nezamabadi with TNN are all implemented in MATLAB. Then, the proposed TNN method is coded. Finally, a comparison is held between these seven different algorithms. To compare accurately, we have to deal with Lena image, the reference chosen in their papers. Furthermore, these algorithms are applied to the target medical images, where the WT is applied using db8 four levels of decompositions.

#### 4.1. Image Denoising

##### 4.1.1. Universal Case

While implementing the Nasri-Nezamabadi thresholding function [24], it is used in universal Visu Shrink method, and compared with conventional functions such as hard, soft, Zhang(1998) and Zhang(2001) functions classes with their optimized values ( $k=3$  and  $\lambda=0.01$ ).

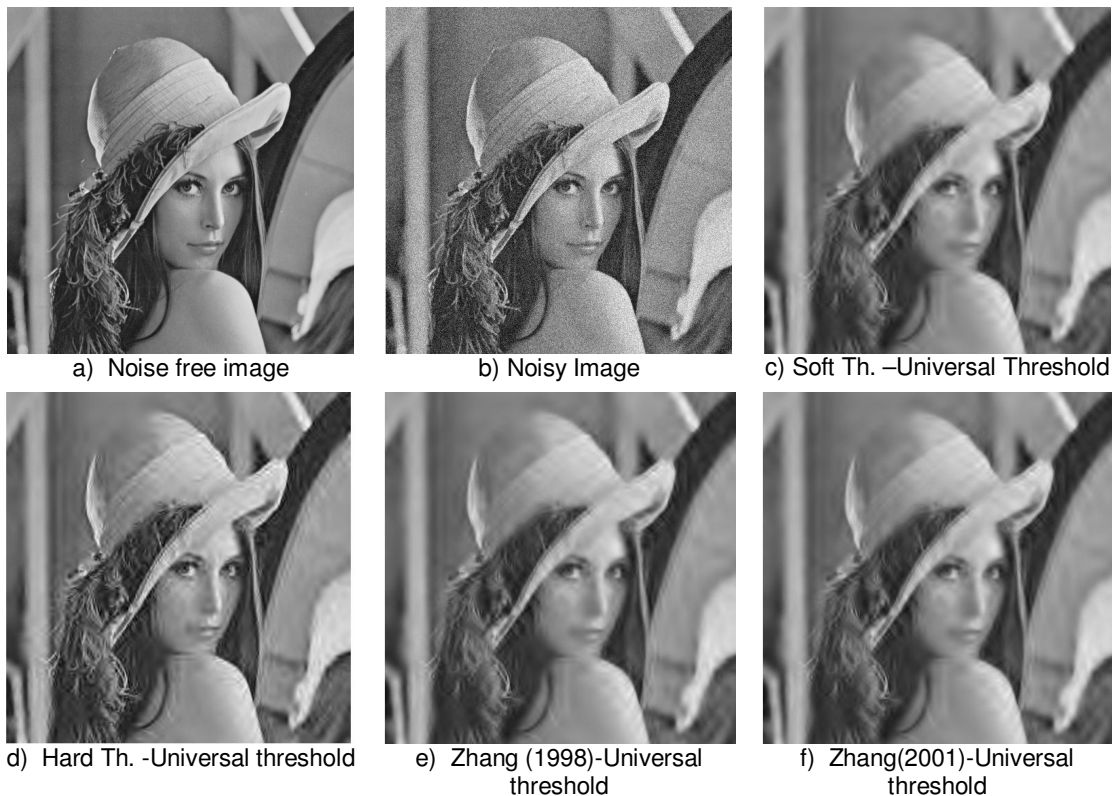
#### 4.1.2 TNN with Threshold Learning Method

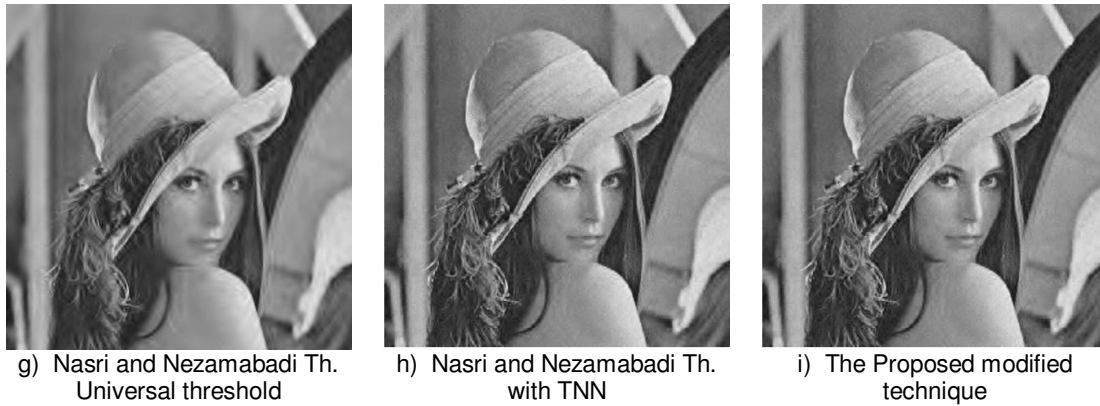
In the second step, the Nasri and Nezamabadi thresholding function [24] is used in the subband-adaptive threshold with TNN method. In this method, the threshold value is learnt and shape tuning factors of function are  $k=1$  and  $m=2$ . The learning step is set to  $\alpha = 1e - 6$  and the convergence criteria is considered as  $\Delta thr(i)/thr(i) < 1e-6$ .

#### 4.1.3 Proposed Modified TNN Technique

As mentioned before, one of our main concerns when introducing our innovative TNN is to save time. This will be achieved by reducing the total number of epochs with preserving the image quality.

Figure (6) and table (1) show the results of denoised Lena image with PSNR = +23.98 dB using the seven different algorithms. From figures (6-h) and (6-i), we can see how the last two algorithms offer good visualization success. Applying the proposed TNN, we have obtained approximately similar PSNR $\approx$  +31 but we have reduced the number of iterations significantly as will shown next. Selecting one medical image from the used data set in five different noise densities, results are shown in table (2) and figure(7). Figure (8) is histogram summary for one of the cases shown in table (2) case (5).

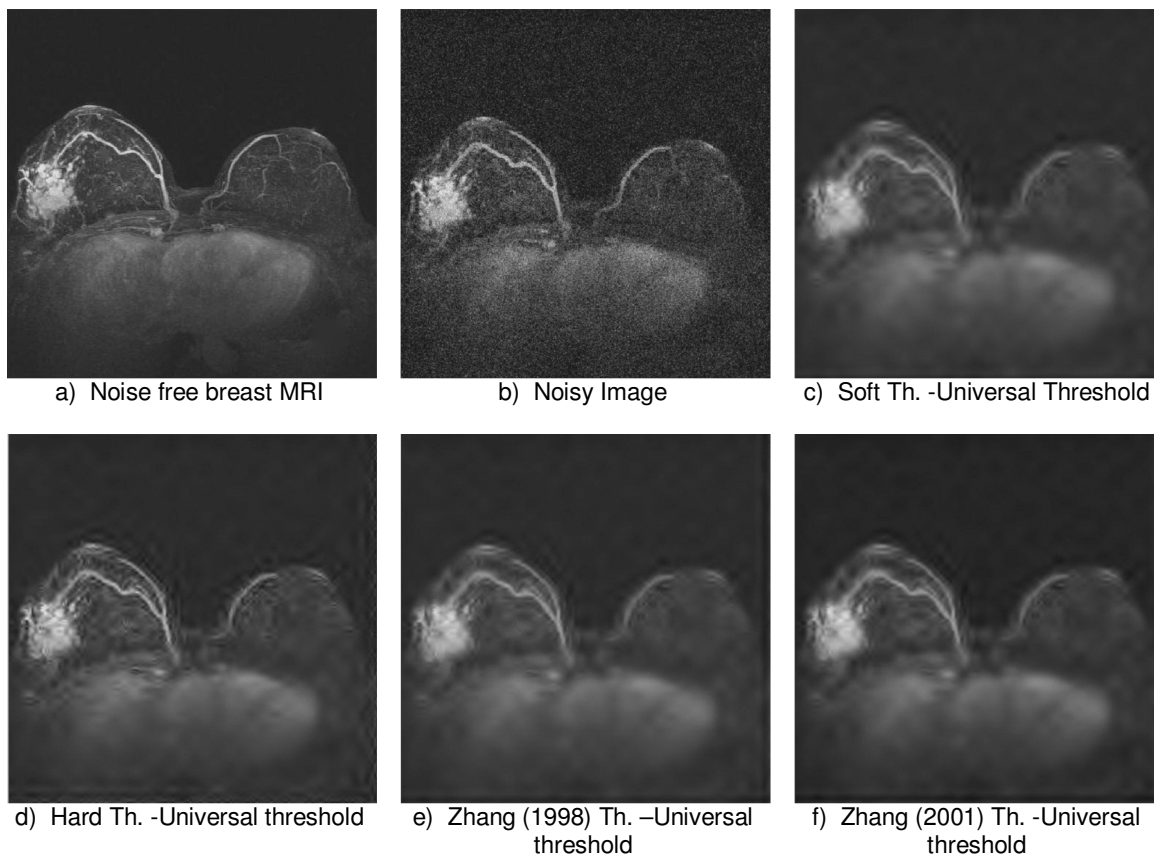


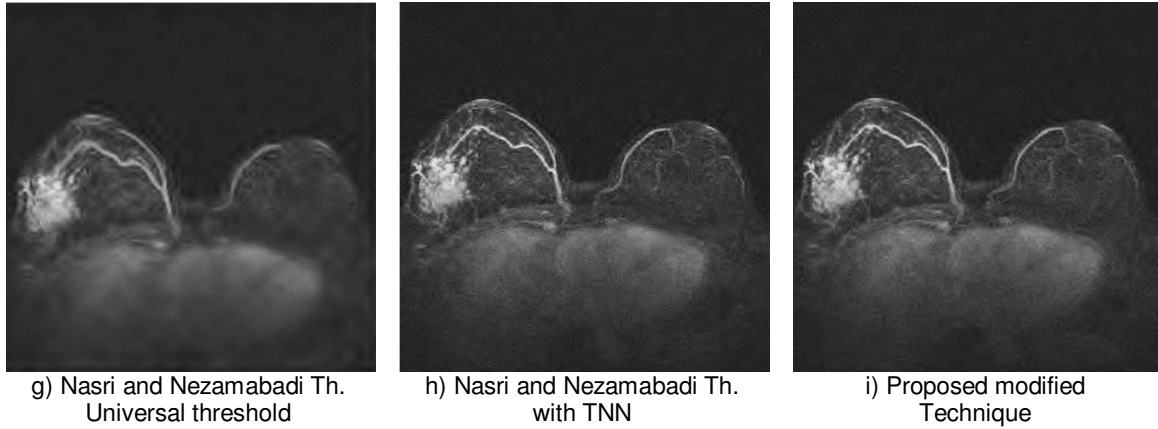


**FIGURE (6):** The results of denoised Lena image with PSNR=+23.98 dB

Noisy image	Universal Threshold					TNN	
	Soft Th.	Hard Th.	Zhang (1998) Th.	Zhang (2001) Th.	Nasri and Nezamabadi Th.	Nasri and Nezamabadi Th.	Proposed Technique
+23.98	+26.24	+28.15	+26.83	+26.73	<b>+29.34</b>	<b>+31.56</b>	<b>+31.50</b>

**TABLE (1):** The PSNR (in dB) results for Lena test image with Gaussian noise PSNR=+23.98, using different thresholding functions

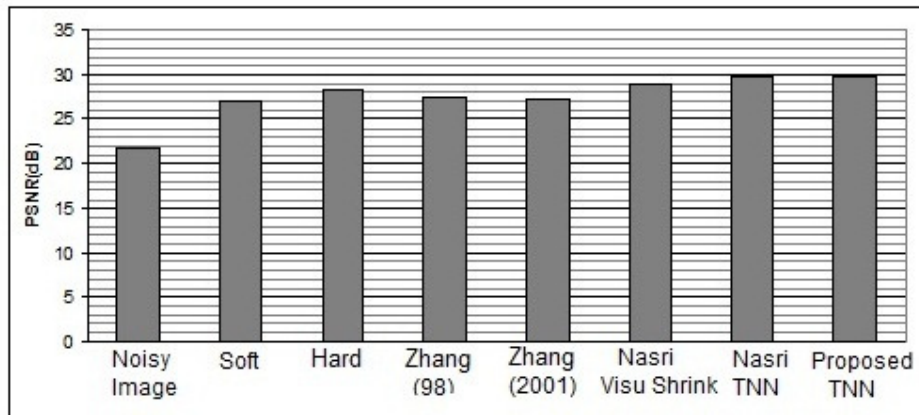




**FIGURE (7):** The results of denoised breast MRI test image for case (5) (PSNR=+21.82 dB)

	Noisy image	Universal Threshold				TNN		
		Soft Th.	Hard Th.	Zhang (1998) Th.	Zhang (2001) Th.	Nasri and Nezamabadi Th.	Nasri and Nezamabadi Th.	Proposed Technique
Case (1)	+30.59	+30.99	+33.12	+32.21	+31.61	<b>+34.25</b>	<b>+36.18</b>	<b>+35.96</b>
Case (2)	+27.35	+29.67	+31.56	+30.45	+30.24	<b>+32.56</b>	<b>+33.98</b>	<b>+33.87</b>
Case (3)	+25.04	+28.65	+30.35	+29.17	+29.17	<b>+31.18</b>	<b>+32.32</b>	<b>+32.22</b>
Case (4)	+23.27	+27.79	+29.33	+28.27	+28.18	<b>+30.02</b>	<b>+30.95</b>	<b>+30.90</b>
Case (5)	+21.82	+27.03	+28.39	+27.47	+27.33	<b>+29.00</b>	<b>+29.79</b>	<b>+29.75</b>

**TABLE (2):** The PSNR (in dB) results for breast MRI test image with different Gaussian noise densities, using different thresholding functions



**FIGURE (8):** The PSNR histogram of different thresholding function to MRI breast image Case (5) (PSNR=+21.82 dB)

From table (3), the efficiency of both the proposed modification technique, and Nasri-Nezamabadi function with TNN method [24] is demonstrated. When comparing total number of iteration, a decreasing percentage is remarkable while preserving the PSNR. In figure (9) the numbers of

iterations (execution time) are compared in two algorithms: Nasri-Nezamabadi [24] and the proposed TNN method for all MRI breast noisy cases.

From figure (10) it is clearly seen that the reduction in number of iterations is 22.29244 at PSNR= +30.59 (Case (1)); however it is increased to 38.75152 at PSNR= +21.82 (case (5)). This means that as the noise density increases, the proposed TNN gives better results compared to [24]. The proposed algorithm surpassed [24] in execution time by about 40%. Resultant image has a negligible difference output PSNR in all cases.

	PSNR of Noisy image (dB)	Nasri and Nezamabadi function with TNN			Proposed Modification Technique			
		Total iteration	Level 1 iteration	PSNR (dB)	Total iteration	Decreasing iteration%	PSNR (dB)	Decreasing PSNR%
Case (1)	+30.59	182,452	40,673	+36.18	141,779	22.29244	+35.96	0.608071
Case (2)	+27.35	229,912	68,626	+33.98	161,286	29.84881	+33.87	0.32372
Case (3)	+25.04	259,538	83,815	+32.32	175,723	32.29392	+32.22	0.309406
Case (4)	+23.27	295,000	95,095	+30.95	199,905	32.23559	+30.9	0.161551
Case (5)	+21.82	360,293	139,619	+29.79	220,674	38.75152	+29.75	0.134273

TABLE (3): Comparison between Nasri and Nezamabadi function with TNN and Proposed Modification Technique performance.

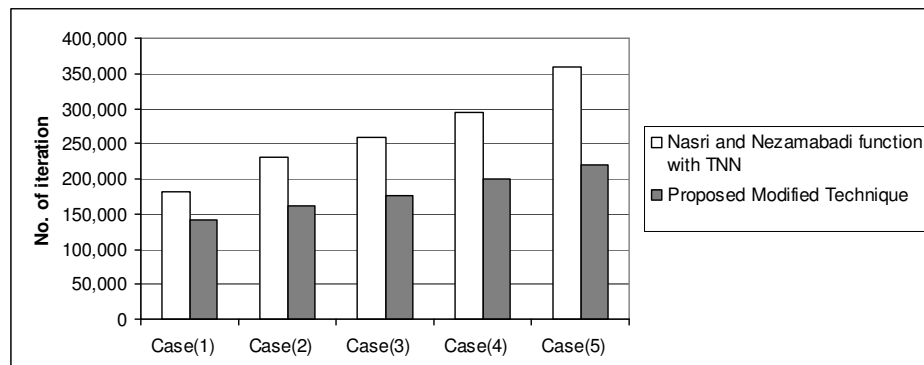
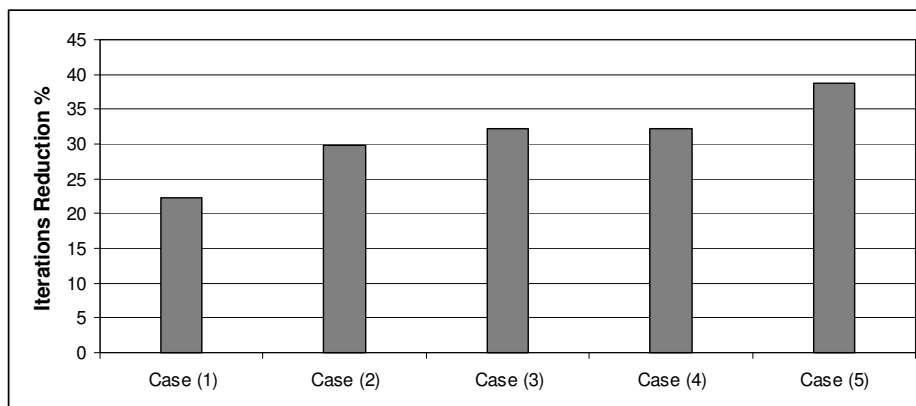


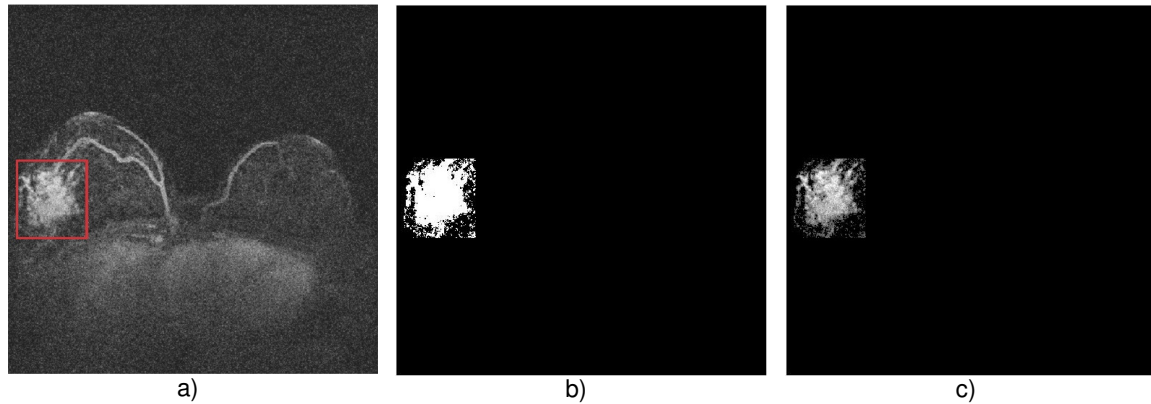
FIGURE (9): Number of iteration comparison between Nasri and Nezamabadi function with TNN method and proposed modified technique for MRI breast image.



**FIGURE (10):** Iteration reduction percentages in proposed technique for MRI breast image

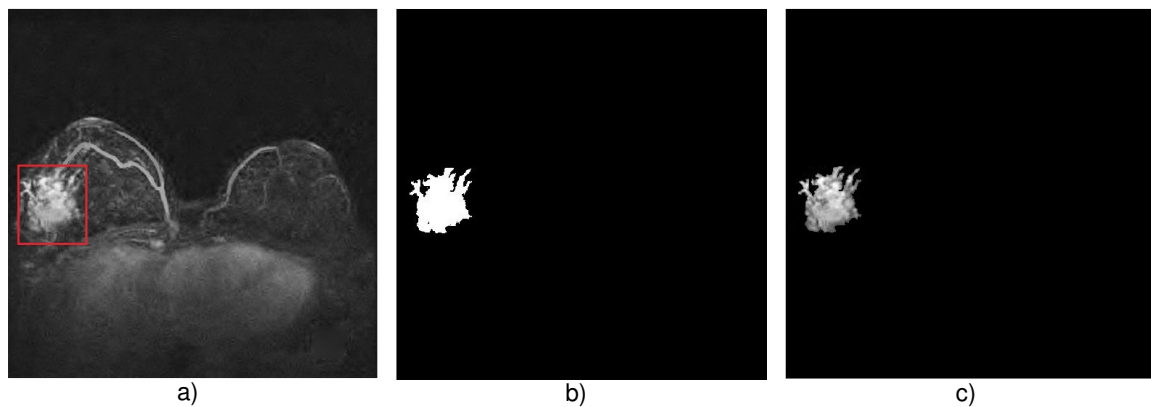
#### 4.2 Segmentation Phase

The processed images enter a segmentation procedure to extract ROI. To evaluate the performance of the proposed TNN algorithm, a comparison between extracting ROIs from noisy and denoising image is presented. Figures (11) and (12) show the semi automatic segmentation procedures for case (5).



**FIGURE (11):** Segmentation noisy breast MRI case (5).

- a) Square the bright region to remove the top body edges of noisy image.
- b) Binary segmentation mask of noisy image.
- c) ROI noisy image. Original noisy image after applying the segmentation mask.



**FIGURE (12):** Segmentation denoised breast MRI case (5).

- a) Square the bright region to remove the top body edges of denoised image by proposed technique.
- b) Binary image segmentation mask image of denoised image by proposed technique.
- c) ROI image by proposed both denoising and segmented technique. i.e. Original denoised image after applying the segmentation mask.

## 5. DISCUSSION

The Nasri-Nezamabadi function [24] has produced better results in noise reduction compared to classic functions such as soft and hard and improved functions such as Zhang's class of functions in the universal threshold case. The hybrid method presented in this paper saves time and complexity without significant reduction in PSNR. Normally in wavelet subbands, as the level increases the coefficients of the subband becomes smoother. For example, horizontal- Level2 is smoother (less noise) than the corresponding subband in the first level, i.e., horizontal- Level1. So the threshold value for of horizontal- Level2 should be smaller than for horizontal- Level1. The



relation between levels and PSNR is summarized as follows: As the level increases, PSNR increases, noise is reduced, number of iterations increases, threshold value decreases. The wavelets act as filter of noise. We have found that the value of threshold by learning is near of universal threshold value only in first level; this is because first level is highest noise and fine tuning is needed in higher wavelet levels due to lower noise. The proposed TNN has shown remarkable execution time reduction especially in higher noisy environments, where the PSNR is low.

By means of the semi-automatic segmentation phase a clearer ROI is obtained. Very small tumor structure becomes more visible and features of the calcifications become more evident. Thus, CAD systems, radiologists, or physicians are towards better classifications and better decisions.

## 6. CONCLUSION

In this paper, an efficient detection of breast cancer tumor algorithm from medical images in noisy environment has been introduced. Based on a novel thresholding neural network in wavelet domain, and a semi automatic segmentation technique breast images are classified. The proposed method combines the Visu Shrink and the subband-adaptive cases with thresholding neural network. Comparison to other well-known existing algorithms shows that the proposed algorithm yields significantly superior image quality besides a reduced neural network learning time. This means a total reduction in complexity and execution time; while maintaining good image quality.

## 7. REFERENCES

- [1] Dr. H. B. Kekre et. al.. "Detection Of Tumor In MRI Using Vector Quantization Segmentation". International Journal of Engineering Science and Technology, Vol. 2, no.8, 3753-3757, 2010.
- [2] M.Suganthi and M.Madheswaran."A Novel Approach towards Segmentation of Breast Tumors from Screening Mammograms for Efficient Decision Support System". World Academy of Science Engineering and Technology, Vol.64, 398-403, 2010.
- [3] Isaac N. Bankman ."Handbook of Medical Imaging". Academic Press, 2000.
- [4] R.C. Gonzalez, R.E. Woods. "Digital Image Processing", Prentice-Hall, Inc. second ed., 2002.
- [5] Olivier Rioul and Pierre Duhamel."Fast Algorithms for Discrete and Continuous Wavelet Transforms". IEEE transactions on information theory, Vol.38, no. 2, march 1992.
- [6] I. Daubechies."Ten Lectures on Wavelets". CBMS-NSF Regional Conference Series in Applied Mathematics, no. 61, SIAM, Philadelphia, PA, 1992.
- [7] Mallat, S.." A wavelet tour of signal processing". San Diego (CA) Academic Press, 1999.
- [8] Chui, C.K.. "An introduction to wavelets". Boston (MA), Academic Press, 1992.
- [9] Hernandez E., Weiss G.. "A Course on Wavelets", CRC Press, Boca Raton, 189p, 1996.
- [10] M.S. Crouse, R.D. Nowak, R.G. Baraniuk. "Wavelet-based signal processing using hidden Markov models". IEEE Trans. Signal Process., Vol.46,886–902, 1998.
- [11] D.L. Donoho. "De-noising by soft thresholding". IEEE Trans. on Info. Theory, 933-936, 1993.
- [12] D.L. Donoho, I.M. Johnstone. "Ideal spatial adaptation via wavelet shrinkage". Biometrika, Vol. 81, 425-455, 1994.

- [13] D.L. Donoho, I.M. Johnstone. "Adapting to unknown smoothness via wavelet shrinkage". Journal of American statistical assoc., Vol. 90, no. 432, 1200-1224, 1995.
- [14] D.L. Donoho, I.M. Johnstone." Wavelet shrinkage: Asymptopia". J.R. Stat. Soc., series B, Vol. 57, no. 2, 301-369, 1995.
- [15] G.Sambasiva Rao,C. NagaRaju,Dr.L.S.S. Reddy and Dr.E.V. Prasad."A Novel Thresholding Technique for Adaptive Noise Reduction using Neural Networks". IJCSNS International Journal of Computer Science and Network Security, Vol.8, No.12, 315-320, December 2008.
- [16] S. Chang, B. Yu, M. Vetterli. "Adaptive wavelet thresholding for image denoising and compression". IEEE Trans. Image Process, Vol.9, 1532–1546, 2000.
- [17] M.S. Crouse, R.D. Nowak, R.G. Baraniuk. "Wavelet-based signal processing using hidden Markov models".IEEE Trans. Signal Process, Vol.46, 886–902, 1998.
- [18] M.K. Mihcak, I. Kozintsev, K. Ramchandran, P. Moulin." Low-complexity image denoising based on statistical modeling of wavelet coefficients". IEEE Signal Process, Vol.6, 300–303, 1999.
- [19] X.-P. Zhang. "Thresholding neural network for adaptive noise reduction". IEEE Trans. Neural Networks,Vol. 12,no. 3, 567–584, 2001.
- [20] Xiao-Ping Zhang and M.Desai."Nonlinear Adaptive Noise Suppression Based On Wavelet Transform". Proc. of ICASSP'98, Seattle, Washington, May 12-15, 1998.
- [21] X.-P. Zhang." Space-scale adaptive noise reduction in images based on thresholding neural networks". Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, 1889–1892, 2001.
- [22] X.-P. Zhang, M.D. Desai." Adaptive denoising based on SURE risk". IEEE Signal Process, Vol. 5, no.10265–267, 1998.
- [23] S. Haykin, Prentice-Hall. NJ."Neural Network: A Comprehensive Foundation", second ed., 1999.
- [24] Mehdi Nasri, Hossein Nezamabadi-pour." Image denoising in the wavelet domain using a new adaptive thresholding function". Elsevier Journal of Neurocomputing, Vol.72, 1012-1025, 2009.
- [25] [25] R.M. Haralick, L.G. Shapiro. "Survey: Image segmentation techniques," Comp. Vision Graph Image Proc., Vol. 29, 100-132, 1985.
- [26] I.N. Bankman, T. Nizialek, I. Simon, O.B Gatewood, I.N. Weinberg, W.R. Brody. "Segmentation algorithms for detecting microcalcifications in mammograms". IEEE Trans. Inform. Techn. Biomed., Vol. 1, no.2, 161-173, 1993.
- [27] S.Bernsen. "Segmentation tools in mathematical morphology". SPIE, Image algebra and morphology- ical image processing, vol. 1350, 70-84, 1990.
- [28] J.S.Weszka."A Survey of threshold selection techniques". Computer Graphics and Image Proc., Vol.7, 259-265, 1978.
- [29] Nobuyuki Otsu." A Threshold Selection Method From Gray-Level Histograms ".IEEE transacnons on systems, man, and cybernencs, Vol. smc-9, no. 1, January 1979.

# Dynamic Threshold in Clip Analysis and Retrieval

**Satishkumar L Varma**

*Department of Computer Engineering  
PIIT, Mumbai University  
Mumbai, 410206, India*

*varmasl@yahoo.co.in*

**Sanjay N Talbar**

*Department of Electronics and Telecommunication  
SGGS IE&T,  
Nanded, 431606, India*

*sntalbar@yahoo.com*

---

## Abstract

Key frame extraction can be helpful in video summarization, analysis, indexing, browsing, and retrieval. Clip analysis of key frame sequences is an open research issues. The paper deals with identification and extraction of key frames using dynamic threshold followed by video retrieval. The number of key frames to be extracted for each shot depends on the activity details of the shot. This system uses the statistics of comparison between the successive frames within a level extracted on the basis of color histograms and dynamic threshold. Two program interfaces are linked for clip analysis and video indexing and retrieval using entropy. The results using proposed system on few video sequences are tested and the extracted key frames and retrieved results are shown.

**Keywords:** Video Clip, Key Frames Extraction, Entropy, Indexing, Retrieval.

---

## 1. INTRODUCTION

The growing prevalence of digital images and videos increases the need for effective and efficient searching techniques. In traditional databases the stored data is searched through alphanumeric matching. Each entry in the database has several key fields by which a query is matched. However images and video cannot be characterized by alphanumeric strings.

To save these video records we need large memory space. The videos are consists of key frames. So, instead of storing the whole video we can save only the key frames in which significant change is noted. This will save the memory space. Also accessing them will be easier. Key-frames are still images which best represent the content of the video sequence in an abstracted manner, and may be either extracted or reconstructed from original video data. Each key frame is usually described by low level features such as color, texture, sketch, shape, spatial constraints, annotation etc. A combination of features is always needed because there is no single best feature that gives accurate description in any general setting. In video databases each key frame is stored with its corresponding features. The features are chosen in the hope of capturing salient semantically information about the video key frame.

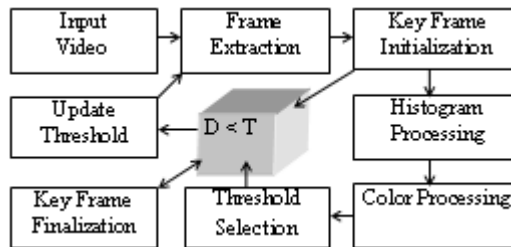
The research work done on video summarization methods has only focused on the uncompressed domain [1]-[6]. The shot detection techniques are also mentioned in [7][8]-[10]. The TRECVID Rushes task and data are described in Over et al. [9]. The scene analysis [11] and video content based frame and scene extraction [12] have been already worked out. Systems for summarizing unedited BBC footage were developed in TRECVID which was organized by NIST [13] [14]. A number of algorithms for extraction of spatio-temporal features from shots are reported in the literature [15]-[19]. Alternative approach was carried out by Vermaak et al. [20] to find the optimal set of key frames such that the frames are maximally distinct. Han and Yoon [21], describes a technique for key frame extraction using temporal sampling. In case of content based

video indexing and retrieval, the various systems have been developed and reported, These systems include: the iMATCH [22], the IRMoment [23], the QBIC system [24], the Retrieval Ware system [25], the Virage system [26], the VisualSEEK and WebSEEK system [27], the Blobworld system [28], the Photobook system [29], the Mars system [30], the Video-Q [31], and the USC system [32].

The clip analysis is overviewed in section 2. In section 3, we describe the system of content-based video indexing and retrieval using entropy. The video retrieval interface and experimental results are discussed in section 4. The summary and conclusion of the paper is presented in Section 5.

**2. OVERVIEW OF CLIP ANALYSIS**

In this paper we describe the user interface that tries to compare frames in the Video and extract the key frames as well as do video segmentation by segmenting the video into shots using dynamic threshold depending on the number of key frames extracted. Further these key frames are used for indexing and retrieval of videos in the given database. The prepared system architecture is given in Fig. 1.

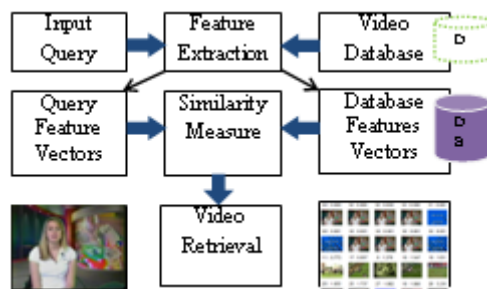


**FIGURE 1:** Key Frame Extraction Using Dynamic Threshold.

Shot segmentation: A shot is defined as the consecutive frames from the start to the end of recording in a camera. It shows a continuous action in an image sequences. There are two different types of transitions that can occur between shots, abrupt (discontinuous) also referred as cut, or gradual (continuous) such as fades, dissolves and wipes.

The cut boundaries show an abrupt change in image intensity or color, while those of fades or dissolves show gradual changes between frames. These transitions are defined as follows [33]:

- A cut is an instantaneous transition from one scene to the next and it occur over two frames
  - A fade is a gradual transition between a scene and a constant image (fade out) or between a constant image and a scene (fade in)
  - A dissolve is a gradual transition from one scene to another, in which the first scene fades out and the second fades in
  - A wipe occurs as a line moves across the screen, with the new scene appearing behind the line.
- There are different approaches used to detect the shot in a video and some are outlined here.



**FIGURE 2:** Key Frame Indexing and Retrieval.

### 3. VIDEO INDEXING AND RETRIEVAL

The indexing is carried out using color feature and texture entropy feature. Key frame indexing is carried out using color feature. The HSV color space is used to indexing the key frames to generate feature vectors. The block diagram is shown in Fig. 2. The indexing is also carried out using entropy. Entropy is a scalar value representing a statistical measure of randomness that can be used to characterize the texture of the input images. The value of entropy is also an invariant that is neither affected by rotation nor scaling. Entropy is defined as

$$H = -\sum_{i=1}^{\Omega} P_i \log(P_i) \tag{1}$$

#### 3.1 Video Retrieval Measure

The video retrieval system is evaluated using two common measures, recall and precision. The recall measures also known as the true positive function or sensitivity, which corresponds to the ratio of correct experimental detections over the number of all true detections. It measures the ability of a system to present all relevant items:

$$recall = \frac{\text{number of } r}{\text{number of } rel} \tag{2}$$

The Precision measure defined as the ratio of correct experimental detections over the number of all experimental detections. It measures the ability of a system to present only relevant items.

$$precision = \frac{\text{number of } r}{\text{total number}} \tag{3}$$

#### 3.2 Key Frame Retrieval

The query key frame is provided as input to the system. The query feature vector is compared with every key frame feature vector in the database under similarity measure process as shown in Fig. 2. The similarity measure is sorted in ascending or descending order. The close matches are found. The key frames also called as candidate key frames are displayed using user interface. The retrieved key frame belongs to a video in the database. So the respective video is retrieved.

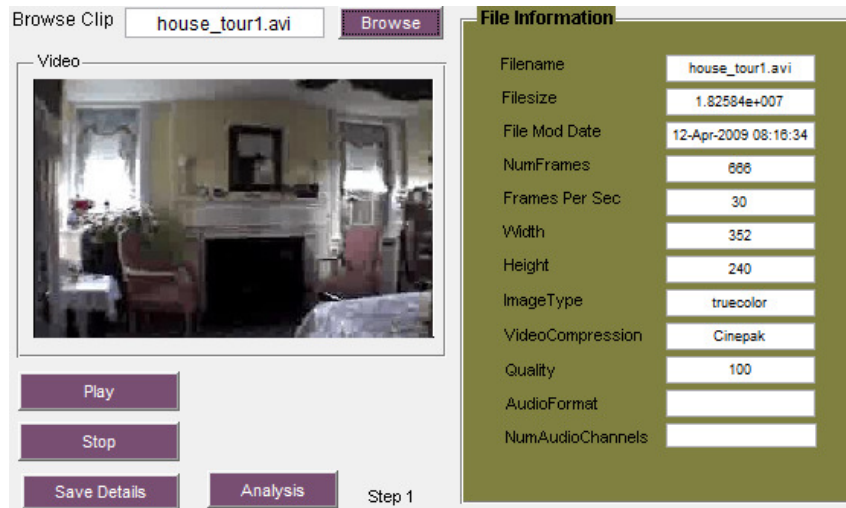


FIGURE 3: User Interface for Key Frame Extraction.

#### 4. EXPERIMENTAL RESULTS

There are two different user interfaces. The user interface helps for clip analysis. The program interface is shown in Fig. 3. The interface also displays the information about the video being browsed in the display area. The second user interface shown in Fig. 4 is used for video indexing and retrieval. The first number above the frame is video number. Next to the video number, the distance measure is displayed. The program interface shows the provision of first 20 candidate video frames to be displayed.

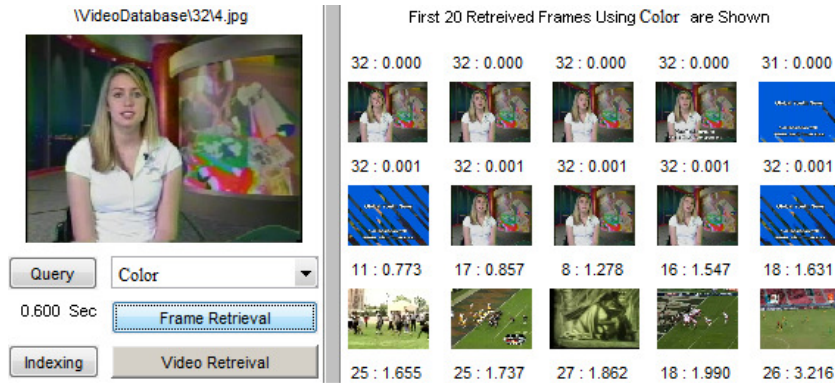


FIGURE 4: User Interface for Video Indexing and Retrieval.

##### 4.1 Clip Analysis

The key frames are then packed into one summary file called shots.avi in this file; the size of the video frames is determined by its relative length of the shots, in another word, determined by how important that shot is. The output file format for segmented video is .avi. The file format of extracted key frames is .jpg. The system stores extracted key frames as key0.avi, key1.avi, and so on. The extracted key frames using user interface is shown in Fig. 5.

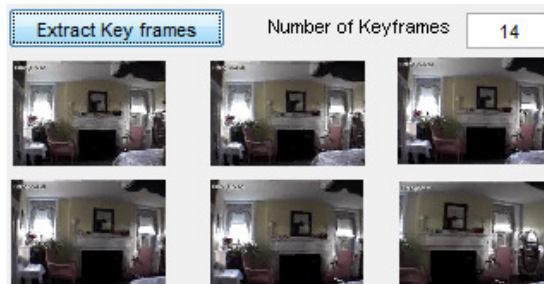


FIGURE 5: Interface Extracted Key Frames (Scene.avi).

Clip Details	Duration (sec)	Frames	Key Frames	Position of Key Frames	Tolerance
BladeRunner	3	103	8	1,4,20,25,33,36,37,41	2.064
Rollcall2	8	194	9	1,10,20,27,33,40,176,183,189	1.950
Scene	6	92	4	1,17,41,73	2.476
Vip	5	46	6	1,2,5,11,26,45	2.476
house_tour	91	666	14	1,2,14,31,39,41,73,77,96,121,153,211,229,250,419,447	2.125

TABLE 1: Clip Analysis.



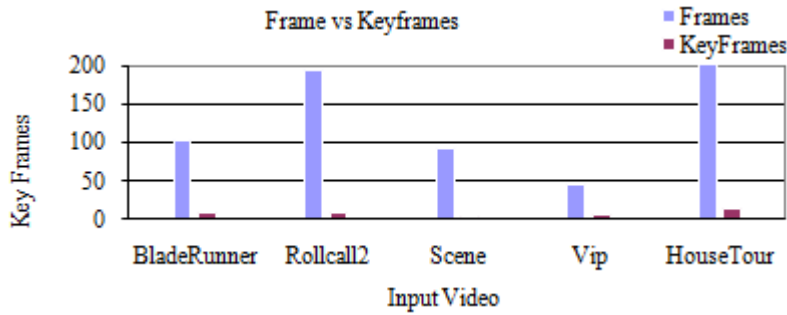


FIGURE 6: Number of Key Frames in Videos.

The number of key frames with respect to given frames in a video is shown in Fig. 6. The first and last frame of any given video is always a key frame in the given video. The detail of the key frames is stored in text file. Different length of input videos is used here to compare the results. The positions of the key frames extracted are shown in Table 1. The dynamic threshold is automatically selected based on the difference between current and next key frame of a given videos. The tolerance and key frames are also shown in Table 1.

### 4.2 Video Retrieval Results

The key frames retrieved using HSV and Entropy are shown in Fig. 7 and Fig. 8 respectively.



FIGURE 7: Video Retrieval Using HSV Color Planes (a) Space (b) TV Show.

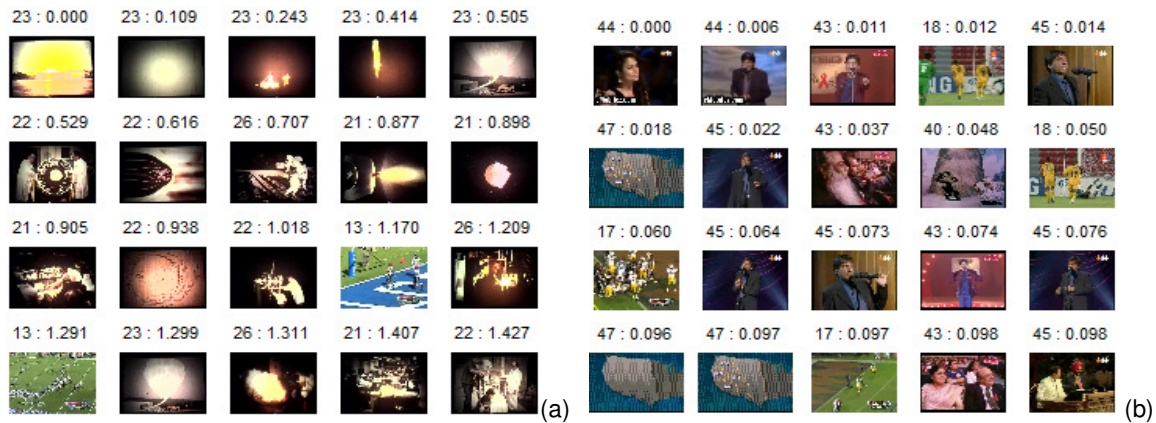


FIGURE 8: Video Retrieval Using Entropy Features (a) Space (b) TV Show.

In this result class of input videos is space and TV Shows. The result is measured and compared using precision and recall curve. The estimated precision recall curve using color features and entropy features for video retrieval are shown in Fig. 9.

It is observed that the video retrieval results using entropy features are better than the color features results.

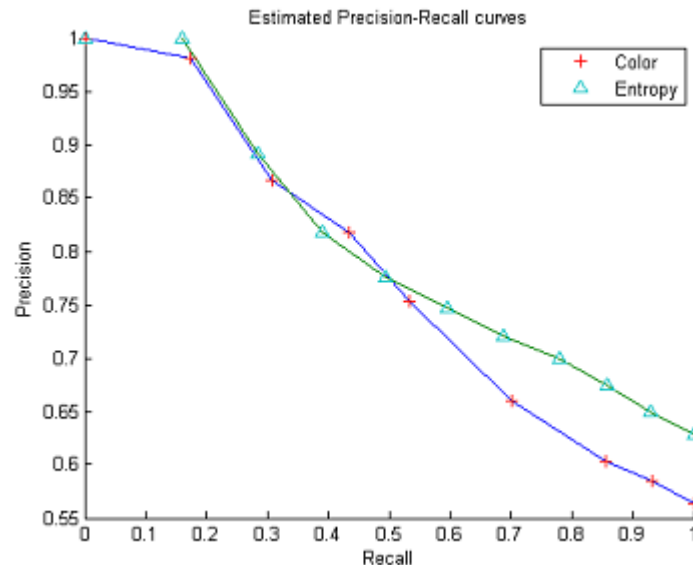


FIGURE 9: Estimated Precision-Recall curve.

## 5. REFERENCES

- [1] F. Idris and S. Panchanathan. "Review of Image and Video Indexing Techniques," *Journal of Visual Communication and Image Representation*. 8(2). 146-166. 1997.
- [2] N. Dimitrova, T. McGee and H. Elenbaas. "Philips Research," 345 Scarborough Rd. Briarcliff Manor NY, 10510.
- [3] J. Oh, A. Kien and Hua. "An Efficient Technique for Summarizing Videos Using Visual Contents," *Multimedia and Expo, ICME 2000. IEEE International Conference*, Vol. 2, pg 1167 – 1170, Jul 2000.
- [4] Y. Gong and X. Liu. "Generating Optimal Video Summaries," *Multimedia and Expo, ICME 2000, IEEE International Conference*, Vol. 3, pg 1559 – 1562, Jul 2000.
- [5] D. DeMenthon, V. Kobla and D. Doermann. "Video Summarization by Curve Simplification," *Technical Report LAMP-TR-018, CS-TR-3916, University of Maryland, College Park*, 1998.
- [6] R. Leinhardt, S. Pfeiffer and W. Effelsberg. "Video Abstracting," *Communications of the ACM*, Vol. 40, No. 12, Dec 1997.
- [7] A. Girgensohn and J. Boreczky. "Time-Constrained Keyframe Selection Technique," in *IEEE Multimedia Systems '99, IEEE Computer Society*, vol. 1, pp. 756-761, 1999.
- [8] N. Gamaz, X. Huang, and S. Panchanathan. "Scene Change Detection in MPEG Domain," *Image Analysis and Interpretation, IEEE Southwest Symposium*, pg. 12-17, 1998.



- [9] D. M. Ali and M. Ghanbari. "Clear Scene Cut Detection Directly from MPEG Bit Streams," IEEE IP and its Applications, No.465, Vol. 1, pg. 285-289, Jul 1999.
- [10] J. Nang, S. Hong and Y. Ihm. "An Efficient Video Segmentation Scheme for MPEG Video Stream using Macroblock Information," 7th ACM international conference on Multimedia, pg 23 – 26, Oct 1999.
- [11] B. L. Yeo and B. Liu. "Rapid scene analysis on compressed videos," IEEE Trans. Circuits Systems Video Technol. 5, 1995, 533–544.
- [12] N. Doulamis, A. Doulamis, Y. Avrithis and S. Kollias. "Video content representation using optimal extraction of frames and scenes," in Proc. of IEEE Int. Conference on Image Processing (ICIP), Chicago USA, Oct. 1998.
- [13] F. Chen, M. Cooper and J. Adcock. "Video summarization preserving dynamic content," proceeding of the TRECVID video summarization, pages 40-44. Germany 2007.
- [14] P. Over, A. F. Smeaton and P. Kelly. "The TRECVID 2007 BBC rushes summarization evaluation pilot," In Proceedings of the TRECVID Workshop on Video Summarization (TVS'07), pages 1-15, New York, NY, September 2007, ACM Press.
- [15] X. F. Yang, Q. Tian and P. Xue. "Short Video Repeat Identification With Application to News Video Structure Analysis," IEEE Transaction on Multimedia, 9(3)(2007), pp. 600-609.
- [16] N. M. Loccoz, E. Bruno, and S. M. Maillet. "Interactive Retrieval of Video Sequences from Local Feature Dynamics, Lecture Notes in Computer Science, 3877(2006), pp. 128-140.
- [17] H. Lu, B. C. Ooi, H. T. Shen, and X. Xue. "Hierarchical Indexing Structure for Efficient Similarity Search in Video Retrieval," IEEE Transaction on Knowledge and Data Engineering, 18(11), 2006, pp. 1544-1559.
- [18] J. Shao, Z. Huang, H. T. Shen, X. Zhou, E. P. Lim and Y. Li. "Batch Nearest Neighbor Search for Video Retrieval," IEEE Transaction on Multimedia, 10(3)(2008), pp. 409-420.
- [19] C. G. M. Snoek, B. Huurnink, L. Hollink, M. D. Rijke, G. Schreiber and M. Worring. "Adding Semantics to Detectors for Video Retrieval," IEEE Transaction on Multimedia, 9(5)(2007), pp. 975-986.
- [20] T. T. Sato, T. Kanade, E. K Hughes and M. A. Smith. "Video OCR for digital news archive," Proceedings of IEEE International Workshop on Content-Based Access of Image and Video Databases, Jan. 3rd, pp. 52-60, 1998.
- [21] S. H. Han, K. Yoon and I. S. Kweon. "A new technique for shot detection and key frames selection in histogram space," 12th Workshop on Image Processing and Image Understanding, pp. 475-479, 2000.
- [22] S. L. Varma and S. N. Talbar. "iMATCH Image Matching and Retrieval for Digital Image Libraries," ICETET10, pp. 196-201, December 2009.
- [23] S. L. Varma and S. N. Talbar. "IRMoment Image Indexing and Retrieval by Combining Moments," IET Digest, Volume 2009, Issue 1, pp. 38, 2009.
- [24] W. Niblack, R. Berber, W. Equitz, M. Flickner, E. Glasman, D. Petkovic and P. Yanker. "The QBIC project. querying images by content using color, texture and shape," SPIE Storage and Retrieval for Image and Video Database II, pp. 173-187, 1993.

- [25] J. Dowe. "Content-based retrieval in multimedia imaging," in SPIE Storage and Retrieval for Image and Video Databases II, pp.164-167, 1993.
- [26] J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain and C. F. Shu. "The Virage image search engine. an open framework for image management," in SPIE Storage and Retrieval for Image and Video Databases V, pp 76-87, 1996.
- [27] J. R. Smith and S. F. Chang. "An image and video search engine for the World-Wide Web," in Proc. of SPIE, vol. 3022, pp 85-95, 1997.
- [28] C. Carson, S. Belongie, H. Greenspan and J. Malik. "Region-based image querying," IEEE CVPR'97 Workshop on Content-Based Access of Image and Video Libraries, pp. 42-49, 1997.
- [29] T. P. Minka and R. W. Picard. "Interactive learning with a society of models," in Pattern Recognition, 30(4), pp. 565-581, Apr. 1997.
- [30] Y. Rui, T. Huang and S. Mehrotra. "Content-based image retrieval with relevance feedback in MARS," IEEE International Conference on Image Processing, pp. 815-818, Oct. 1997.
- [31] S. F. Chang, W. Chen, H. J. Meng, H. Sundaram and D. Zhong. "A fully automated content-based video search engine supporting spatiotemporal queries," IEEE Trans. Circuits Syst. Video Technol., vol. 8, no. 5, pp. 602-615, Sep. 1998.
- [32] Z. Yang, X. Wan and C. C. J. Kuo. "Interactive image retrieval. concept, procedure and tools," in IEEE 32nd Asilomar Conference, Monterey, CA, pp. 261-265, Nov. 1998.
- [33] R. Zabih, J. Miller and K. Mai. "A feature based algorithm for detecting and classifying scene breaks," Proceedings of the 3rd ACM International Conference on Multimedia, pp. 189-200, 1995.

# Statistical Feature-based Neural Network Approach for the Detection of Lung Cancer in Chest X-Ray Images

**K.A.G. Udeshani**

*Excel Technology Lanka Ltd  
Thimbirigasyaya Road  
Colombo 05, Sri Lanka*

*gayaudeshani@gmail.com*

**R.G.N. Meegama**

*Department of Statistics and Computer Science  
Faculty of Applied Science  
University of Sri Jayewardenepura  
Gangodawila, Nugegoda, Sri Lanka*

*rqn@dscs.sjp.ac.lk*

**T.G.I. Fernando**

*Department of Statistics and Computer Science  
Faculty of Applied Science  
University of Sri Jayewardenepura  
Gangodawila, Nugegoda, Sri Lanka*

*gishantha@dscs.sjp.ac.lk*

---

## Abstract

Lung cancer, if detected successfully at early stages, enables many treatment options, reduced risk of invasive surgery and increased survival rate. This paper presents a novel approach to detect lung cancer from raw chest X-ray images. At the first stage, we use a pipeline of image processing routines to remove noise and segment the lung from other anatomical structures in the chest X-ray and extract regions that exhibit shape characteristics of lung nodules. Subsequently, first and second order statistical texture features are considered as the inputs to train a neural network to verify whether a region extracted in the first stage is a nodule or not. The proposed approach detected nodules in the diseased area of the lung with an accuracy of 96% using the pixel-based technique while the feature-based technique produced an accuracy of 88%.

**Keywords:** Lung Nodule, Computer Assisted Diagnostic, Artificial Neural Network, Chest Radiography, Medical Imaging

---

## 1. INTRODUCTION

Lung cancer is the growth of a tumor, referred to as a nodule, that arise from cells lining the airways of the respiratory system. These cells are often in bright contrast in chest X-rays and take the shape of a round object. However, these nodules that can be seen in a chest X-ray may not necessarily be a lung cancer; it can be due to some other disease such as pneumonia, tuberculosis or calcified granuloma. As such, the detection of lung cancer has been a tedious task in medical image analysis over the past few decades. If lung nodules can be identified accurately at an early stage, the survival rate of the patients can be increased by a significant percentage. In the health industry, chest X-rays are considered to be the most widely used technique for the detection of lung cancer. However, because it is difficult to identify lung nodules using raw chest X-ray images, analysis of such medical images has become a tedious and complicated task. This paper presents a novel technique that can be used to detect lung cancer in early stages.

With the evolution of various technologies, several projects have been carried out to identify lung cancer at early stages. In these projects, digital photographs of chest X-rays and CT scans have been used to identify lung cancer. In recent research literature, it is observed that principles of neural networks have been widely used for the detection of lung cancer in medical images with simulated lung nodules [1, 5], massive training artificial neural networks [11], two level neural classifiers [23], hybrid lung nodule detection [24], and ladder structured decision trees [6, 25]. A co-occurrence matrix

and texture measures have been used in [3] to detect lung nodules. A rectangular area of a chest X-ray was considered in [4] as the input to a neural network.

Apart from using chest X-rays, researchers have used an ensemble of artificial neural networks to detect cancer cells in specimen images of needle biopsies [2]. A template matching algorithm [7] and genetic algorithms [8] have been used to detect lung nodules. These are quite fundamental approaches to detect nodules where the similarity between an unknown object (observed image) and a template (reference image) are compared.

One of the most common problems during identification of lung nodules is overlapping of the ribs and clavicles with nodules. Such overlapping of images results in the detection of suspicious areas difficult. Several rib suppression techniques have been proposed in recent research literature such as blind source separation [9], texture analysis [12], watershed segmentation [10], Gaussian filters [14], active shape modeling [15] and quasi-gabo filters [12].

In [13], a set of features that contain translation invariant wavelet based features and co-occurrence features of mammograms were used to classify images. Features extracted from a multi scale Gaussian filter bank and some specific features that are readily calculated from the blob detector scheme have been used to detect nodules [15]. The local curvature of the image data were considered when viewed as a relief map [22].

An image subtraction approach was proposed in [16] where a test image and a comparison image were used. After applying several widely used pre-processing techniques, the two images were subtracted to obtain a resultant image which can be used to identify cancers with candidate nodules. This research, however, did not verify the utility of the study using many chest X-ray films.

A heuristic approach using artificial intelligence tools have been used in [26] to detect cancer nodules.

## **2. METHODOLOGY**

### **2.1 Overview**

The proposed algorithm consists of two stages as depicted in Fig. 1. The first stage involves a pipeline of image processing routines to separate the lung from other structures of original chest X-rays and identifying the region that is suspected of being a nodule. At this stage, the system extracts 65x65 square areas considering the suspicious point as the center. Because it involves a pixel-based technique, all the pixels in the square region are considered as the inputs to the system. The intensity values of the pixels that fall within this region are extracted and are stored in a database which is used to train the system at the next stage. The database is then divided into several sub categories, and the data available in these sub categories will be used for the training as well as for testing the results.

In the second stage, a neural network is trained based on two types of inputs: pixel-based inputs in which we consider the intensity levels of pixels within the suspected region and statistical feature-based inputs where we take into account first and second order statistical features.

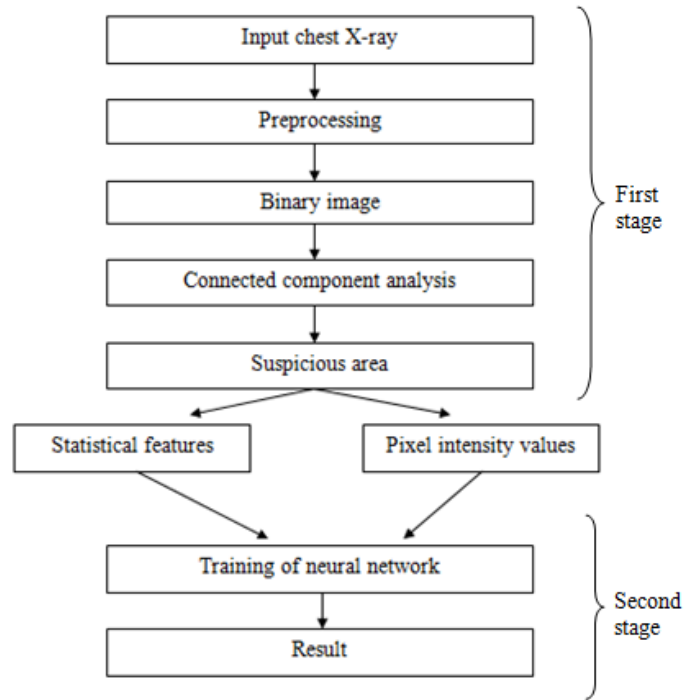


FIGURE 1: Stages of the proposed algorithm.

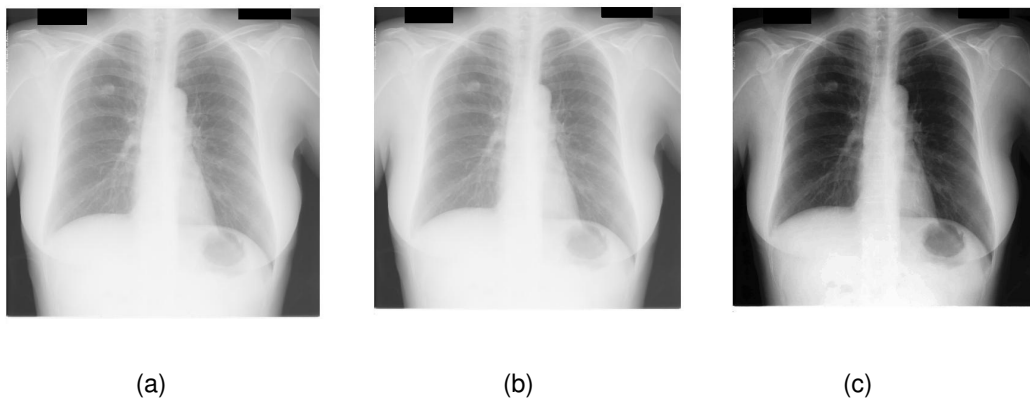
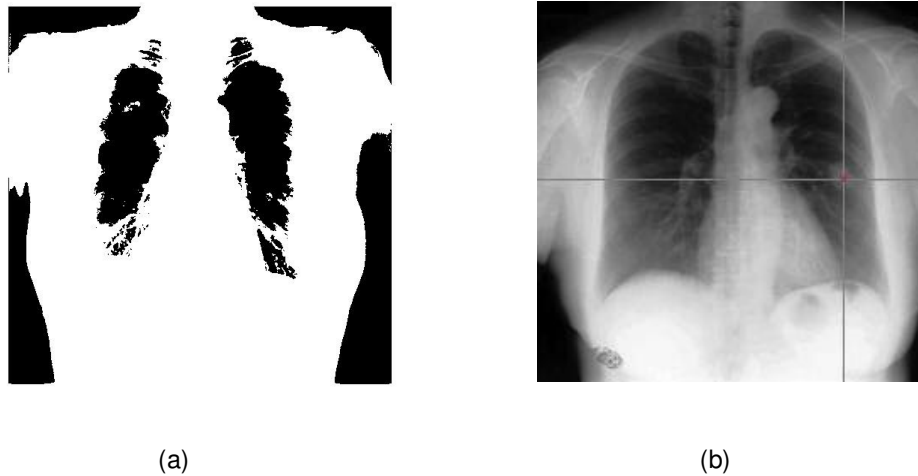


FIGURE 2: Pre-processing of chest X-rays. (a) original image, (b) median filtering and (c) histogram equalization

## 2.2 Pre-processing

As shown in Fig. 2, Median filtering is required to remove the effect of poor contrast due to glare, noise and effects caused by poor lighting conditions during image capture. A low-frequency image was generated by replacing the pixel value with a median pixel value computed over a square area of 5x5 pixels centered at the pixel location. Sharpening and histogram equalization methods were used to enhance the contrast of the images.



**FIGURE 3:** Lung separation. (a) binary thresholded image and (b) suspicious region.

The image was then converted to binary image by using a thresholding technique. It computes a global threshold that can be used to convert an intensity image to a binary image. It utilizes Otsu's method, which chooses the threshold to minimize the intra-class variance of black and white pixels [17].

### 2.3 Lung Region Segmentation

Lung masks were prepared using active shape models that are available with the JSRT database [27]. These images can be used to segment the lung region while the user can identify the scope when selecting suspicious points [15]. These images consist of 154 lung nodules (100 malignant cases, 54 benign cases), and 93 non-nodules. The chest radiographs were digitized with a 0.175mm pixel size, a matrix size of 2048 x 2048 and a 12-bit grayscale level. The criteria for inclusion of radiographs in the database were: (1) absence of nodules larger than 35 mm, (2) absence of suspicious nodules that were not confirmed by CT examination, (3) no more than one nodule per patient, and (4) absence of nodules with margins that could not be confirmed by radiologists. The subtlety of the images contained in this database are grouped into five categories, namely, obvious, relatively obvious, subtle, very subtle and extremely subtle. These categories have been defined by expert radiologists which takes into account the size, contrast, and anatomical position of the lesions.

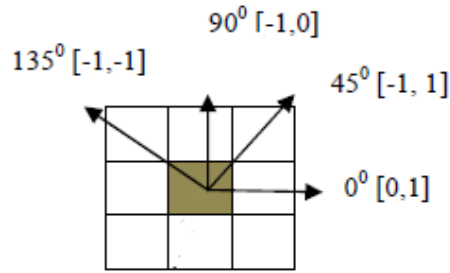
We used connected components labeling to scan the binary thresholded image and group its pixels into components based on pixel connectivity, i.e.. all the pixels in a connected component share similar pixel intensity values and are in some way connected with each other as in Fig. 3 [17].

We define the roundness index of each connected component  $C_i$  as

$$C_i = \frac{4\pi A_i}{R_i^2}$$

where  $A_i$  is the area,  $R_i$  is the perimeter of the  $i^{\text{th}}$  component. This index, that gives a measure of irregularity of a connected component, is equal to 1 only for circles and less than 1 for any other shape. As such, if a connected component exhibits a roundness index closer to 1, the probability of it being a nodule can be considered to be high.

After identifying a region that shows a high probability of being a nodule, we proceed to the second stage of our algorithm to train the neural network.



**FIGURE 4:** Gray level co-occurrence matrix with different offsets. Shaded region at the middle indicates the pixel of interest.

Six features based on first order statistical textures measured with the gray level count of the image and four second order statistical texture features based on how often pairs of pixel with specific values and in a specified spatial relationship occur in an image were considered as inputs to a neural network [18, 19]. Average gray level, standard deviation, smoothness of intensity in a region, skewness (third moment) of the histogram, uniformity and entropy were the first order statistical texture features that we considered. First order statistical features consider statistical moments of the region of interest where the  $n^{\text{th}}$  statistical moment is given by

$$\mu_n = \sum_{i=0}^{L-1} (z_i - m)^n p(z_i)$$

where  $L$  is the possible number of intensity levels,  $m$  is the mean of intensity values,  $z_i$  is the intensity value and  $p(z_i)$  is the percentage of pixels with intensity value  $z_i$ . Thus, skewness is considered to be the 3<sup>rd</sup> statistical moment. The entropy  $e$ , which provides a measure of randomness in the pixels, is given as

$$e = - \sum_{i=0}^{L-1} p(z_i) \log_2 p(z_i)$$

The second order statistical texture features measure: (a) the local variations in the gray level co-occurrence matrix (GLCM), (b) correlation that measures the joint probability occurrence of the specified pixel pairs, (c) energy that provides the sum of squared elements in the GLCM and (d) homogeneity that measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

The GLCM, also known as the gray-level spatial dependence matrix, is a statistical method of examining texture and considers the spatial relationship between pixels. First, we create a GLCM with different offsets, expressed as an angle, as seen in see Fig. 4. To create this GLCM, we consider the above directions which specify the offset values with the common angles and the pixel distance. A 1-by-2 array of integers specifies the distance between the pixel of interest and its neighbor. Subsequently, we extract the statistical measures from these matrices as shown in Fig. 5 to obtain an average.

A neural network with one hidden layer of 1000 neurons and an input layer of 10 neurons to accommodate the ten first and second order textures was used at the training stage [19, 21].

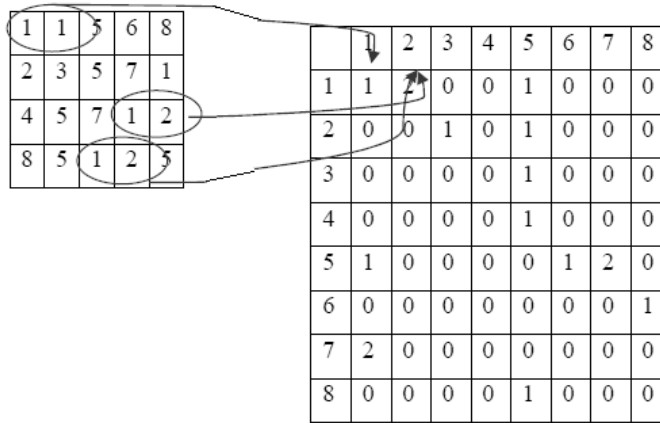


FIGURE 5: Creating the GLCM matrix.

Suspicious region	Recognition rate (%)	
	Pixel-based	Feature-based
Nodule	96	88
Non-nodule	88	7

TABLE 1: Recognition rates for the two types of inputs to the neural network.

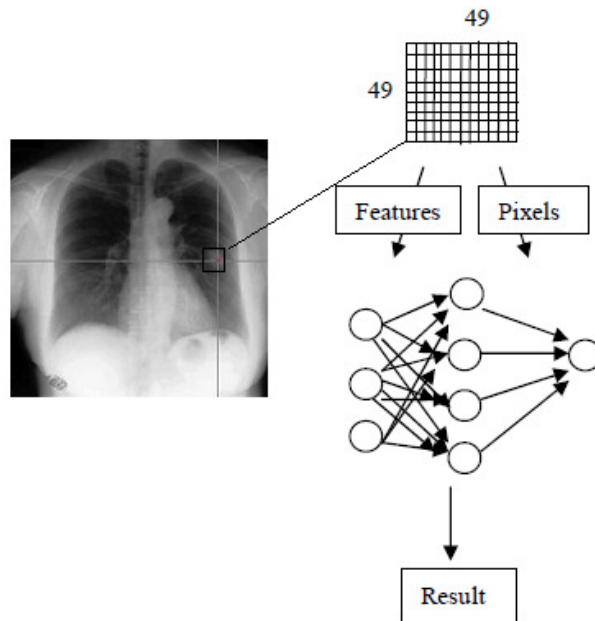
### 3. RESULTS AND DISCUSSION

A 49x49 mask was selected within a suspicious region via an interactive user interface that we have developed for this purpose. After selecting the suspicious region during the first stage of our algorithm, we train the neural network until convergence is achieved. Two types of inputs were considered as inputs to the neural network. One type of inputs consists of intensity levels of pixels within the 49x49 mask. The other contains the ten statistical texture features within this mask as described in Section 2.3.

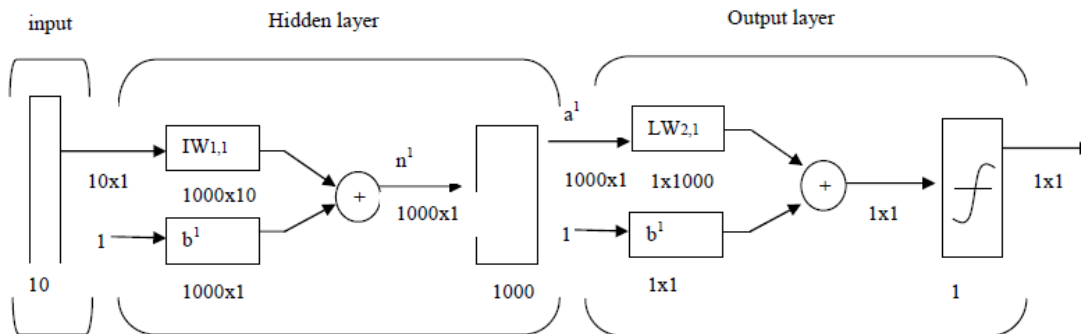
This system uses a neural network with one hidden layer containing 1000 neurons, and an output layer with 1 neuron as in Fig. 7. For the pixel-based intensity input vectors, we used purelin and tansig transfer functions whereas two tansig transfer functions were used for the statistical feature-based inputs vectors. Backpropagation was created by generalizing the Widrow-Hoff learning rule to multiple-layer networks and non-linear differentiable transfer functions. Fig. 8 demonstrates the performance of the neural network for these two types of inputs considering the training approaches and the mean squared error (MSE). It was observed that during pixel-based technique, the R value was 1 and the MSE was 1.2682e-009. The feature-based training resulted in an R value of 0.737 and an MSE of 0.2684.

As seen in Table 1, we managed to achieve a high recognition rate for a nodule when the neural network was trained using pixel-based intensity values. Recognizing a non-nodule was 16% lower with statistical feature-based training of the neural network. We have used a pipeline of image processing routines at the first stage to identify regions suspected to be of lung nodules from the visible regions of chest X-rays. The second stage involves a neural network trained using backpropagation to verify whether the suspected region is a nodule or not. The reason for applying a neural network for the verification is due to detection of false-positives during the image processing stage.





**FIGURE 6:** Selecting a suspicious region on the image to determine statistical features or consider pixel intensity values.

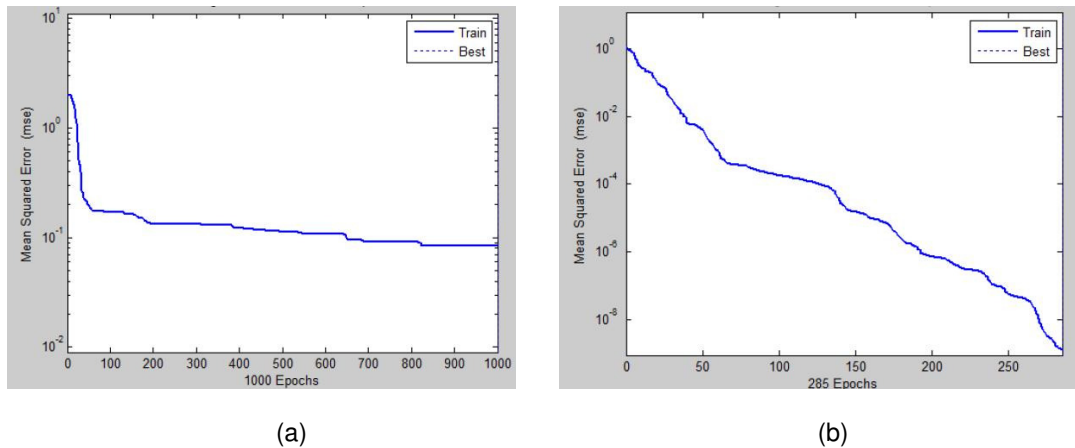


**FIGURE 7:** Architecture of the neural network.

The selection of suspicious region involves a 49x49 square mask which include nodules with 13 mm diameter with a resolution of 96 pixels per inch. In the JSRT database, the size of a nodule varies between 8.9 mm and 29.1 mm with an average size of 17.4 mm. Because in medical imaging literature, the size of a lung nodule if detected at an initial stage, is considered to be between 5mm and 20 mm.

We have used the masking method for lung region segmentation as the JSRT database provides all the required masks for the chest X-rays. As the final results depends on the subtlety of the images, only images with a low subtlety have been used.

In contrast to other computer-aided lung cancer detection methodologies identified in research literature that are based on artificial neural networks, the proposed network uses 10 features as inputs to the system.



**FIGURE 8:** Training graph of the neural network: (a) pixel-based and (b) statistical feature-based inputs.

#### 4. CONCLUSION

In this research, we have successfully developed a solution for the detection of lung cancer nodules using image processing algorithms and neural networks. The system integrates a graphical user interface where the user is able to select a digital chest X-ray image as the input and the system will show suspected regions of the image that contain lung nodules.

The detection of nodules is confounded by overlapping shadows of blood vessels and ribs. Although researchers have proposed several techniques to reduce the effect of ribs in chest X-rays to eliminate false-positives, it is prudent to consider hidden regions too, such as behind the heart, spine and the diaphragm, and apply relevant filtering techniques to suppress these anatomical structures.

#### 5. REFERENCES

- [1] P.R. Snoeren, G.J.S. Litjens, B.V. Ginneken and N. Karssemeijer, Training a computer aided detection system with simulated lung nodules in chest radiographs, Proc. 3rd International Workshop on Pulmonary Image Analysis, Beijing, 2010.
- [2] Z.H. Zhou, Y. Jiang, Y.B. Yang and S.F. Chen, Lung cancer cell identification based on artificial neural network ensembles, *Artificial Intelligence in Medicine*, vol. 24, N 1, pp. 25-36, 2002.
- [3] G.S. Cox, F.J. Hoare and G. de Jager, Experiments in lung cancer nodule detection using texture analysis and neural network classifiers, *Third South African Workshop on Pattern Recognition*, 1992.
- [4] J.W. Lee, H.W. Lee, J.H. Lee, I.T. Kang and G.K. Lee, A study on lung nodule detection using neural networks, *Proc. IEEE Region 10 Conference*, pp. 1150 - 1153, 1999.
- [5] G. Coppini, S. Diciotti, M. Falchini, N. Villari and G. Valli, Neural networks for computer aided diagnosis: detection of lung nodules in chest radiographs, *IEEE Trans. on Information Technology in Biomedicine*, vol. 4, pp. 344-357, 2003.
- [6] M.G. Penedo, M.J. Carreira, A. Mosquera and D. Cabello, Computer aided diagnosis: A neural network based approach to lung nodule detection, *IEEE Trans. on Medical Imaging*, vol. 17, N 6, pp. 872-880, 1998.
- [7] M.N. Gurcan, B. Sahiner, N. Petrick, H.P. Chan, E.A. Kazerooni, P.N. Cascade and L. Hadjiiski, Lung nodule detection on thoracic computed tomography images: Preliminary evaluation of a computer-aided diagnosis system, *Medical Physics*, vol. 29, N 11, pp. 2552, 2002.

- [8] Y. Lee, T. Hara, H. Fujita, S. Itoh and T. Ishigaki, Nodule detection on chest helical CT scans by using a genetic algorithm, Proc. Intelligent Information Systems, pp. 67 - 70, 1997.
- [9] T. Rasheed, B. Ahmad, M. A. U. Khan, M. Bettayeb, S. Lee and T. S. Kim, Rib suppression in frontal chest radiographs: A blind source separation approach, Proc. Signal Processing and its Applications, pp. 1 - 4, 2007.
- [10] K. Le, Automated detection of early lung cancer and tuberculosis based on X-ray image analysis, Proc. WSEAS International Conference on Signal, Speech and Image Processing, pp. 1-6, 2006.
- [11] K. Suzuki, J. Shiraishi, H. Abe, H. MacMahon and K. Doi, False-positive reduction in computer-aided diagnostic scheme for detecting nodules in chest radiographs by means of massive training artificial neural network, Academic Radiology, vol. 12, N 2, pp. 191-201, 2003.
- [12] M. Park, J. S. Jin and L. S. Wilson, Detection of abnormal texture in chest X-rays with reduction of ribs, Proc. Pan-Sydney area workshop on visual information processing, 2004.
- [13] L. Zhang and X. Gao, Research on translation invariant wavelet transform for classification in mammograms, Proc. 3rd international conference on natural computation, 2007.
- [14] P. Campadelli, E. Casiraghi and D. Artioli, A fully automated method for lung nodule detection from postero-anterior chest radiographs, IEEE Trans. on Medical Imaging, vol. 25, N 12, pp. 1588-1602, 2006.
- [15] A.M.R. Schilham, B.V. Ginneken and M. Loog, A computer-aided diagnosis system for detection of lung nodules in chest radiographs with an evaluation on a public database, Medical Image Analysis, vol. 10, N 2, pp. 247-258, 2006.
- [16] R. Hayashibe, N. Asano, H. Hirohata, K. Okumura, S. Kondo, S. Handa, M. Takizawa, S. Sone and S. Oshita, An automatic lung cancer detection from X-ray images obtained through yearly serial mass survey, Proc. International conference on image processing, pp. 343 - 346, 1996.
- [17] R. C. Gonzalez , R.E. Woods and S. L. Eddins. Digital Image Processing, Addison-Wesley, 2002.
- [18] S. A. Patil and V. R. Udupi, Chest X-rays features extraction for lung cancer classification, Journal of Scientific and Industrial Research, vol. 69, pp. 271-277, 2010.
- [19] J. Wei, Y. Hagihara, A. Shimizu and H. Kobatake, Optimal image feature set for detecting lung nodules on chest X-ray images, Proc. Int. Workshop on Computer-Aided Diagnosis, 2002.
- [20] O. Ludwig and U. Nunes, Novel maximum- margin training algorithms for supervised neural networks, IEEE Trans. on Neural Networks, vol. 21, N 6, pp. 972-983, 2010.
- [21] R. N. Strickland, Tumor detection in nonstationary backgrounds, IEEE Trans. on Medical Imaging, vol. 13, pp. 491-499, 1994.
- [22] S.B. Lo, S.L. Lou, J.S. Lin, M.T. Freedman and S.K. Mun, Artificial convolution neural network techniques and applications for lung nodule detection, IEEE Trans. on Medical Imaging, vol. 14, pp. 711-718, 1995.
- [23] J.S. Lin, S.B. Lo, A. Hasegawa, M.T. Freedman and S.K. Mun, Reduction of false positives in lung nodule detection using a two-level neural classification, IEEE Trans. on Medical Imaging, vol. 15, pp. 206-216, 1996.
- [24] Y.S.P. Chiou, Y.M.F. Lure and P.A. Ligomenides, Neural networks image analysis and classification in hybrid lung nodule detection (HLND) system, IEEE Workshop on Neural Networks for Signal Processing, pp. 517-526, 1993.

- [25] D.H. Ballard and J. Sklansky, A ladder-structured decision tree for recognizing tumors in chest radiographs, IEEE Trans. on Computers, vol. C-25, pp. 503-513, 1976.
- [26] B.V. Ginneken, B.M.T.H. Romeny and M.A. Viergever, Computer Aided Diagnosis in chest radiography: A survey, IEEE Trans. on Medical Imaging, vol. 20, pp. 1228-1241, 2001.
- [27] Japanese Society of Radiological Technology, <http://www.jsrt.or.jp>

## Detection of Cranial- Facial Malformations: Towards an Automatic Method

**Hiba Chelbi**

*National Engineering School of Tunis  
University of Tunis ElManar*

*chelbi.hibatollah@yahoo.fr*

**Manel Laajimi**

*National Engineering School of Tunis  
University of Tunis ElManar*

*laajimimanel@yahoo.fr*

**Wala Touhami**

*National Engineering School of Tunis  
University of Tunis ElManar*

*wala.touhami@gmail.com*

**Nawrès Khelifa**

*National Engineering School of Tunis  
University of Tunis ElManar*

*khalifa\_nawres@yahoo.com*

---

### Abstract

This paper presents an original method to detect malformations in cranial facial 3D images. This is a hard task because of the complex structure of the cranium. The method is composed by several stages containing essentially a segmentation phase to localize the bone structure, a 3D image characterization to delineate automatically geometrical characteristics from the segmented images. These need to possess a mathematical definition in order to be calculated and an anatomical pertinence to be representative of the structures' shape. So, the characteristics retained are the crest lines. And finally a registration phase to identify the deformations place. This method is tested on real and synthetic data and has proved his performance.

**Keywords:** Cranial Malformations, HMM Segmentation Method, Crest lines, Mesh Simplification, ICIP Registration

---

### 1. INTRODUCTION

Whether they have a congenital or a traumatic origin, cranial traumas are a real health problem. Indeed, they represent one of the main causes of mortality in many countries. Of course, the continued technological development of scanner stations' keeps helping maxillary-facial surgery in the accurate analysis of these malformations. However, the complex structure of the human skull raises a problem in the detection of these malformations as well as in the diagnosis and planning of the surgical operations. To make easier this task, doctors generally used anatomical atlases which offer promising solutions.

An anatomical atlas represents a reference model taken from several individuals. It takes into account as much the resemblances as the diversities between the subjects. The atlas is generally similar to a map describing the shape, the size and the location of the structures, the identification of the deformation and the planning of interventions. Besides medicine, the atlas was used in other fields such as archeology to follow up the evolution of the skull from the prehistoric to the modern man. Unfortunately, the automatic creation of an atlas is not as obvious as it was thought. Indeed, its creation requires the manipulation of a large image base representing all possible variabilities. These images are, later on, represented by a reduced set of characteristics. We can, then, compute the average of the characteristics in order to represent the atlas. In practice, many researchers choose, with the help of an expert, an image

representing a normal anatomy and refer to it as atlas. This method could, however, be automated.

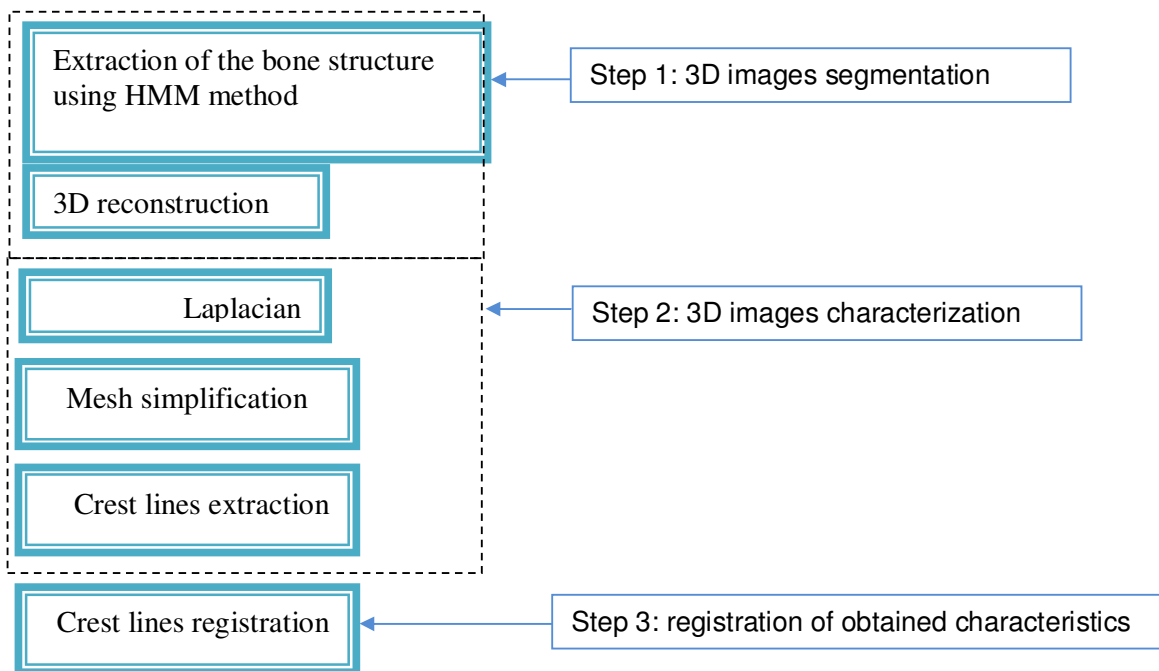
The common method to create automatically an atlas consists in few steps:

- 3D images characterization
- Identification of common characteristics.
- Modeling of common characteristics.
- Analysis of variability of common characteristics

In this work, we present the nucleus of a method aiming at the automatic creation of an atlas of the Tunisian population's skull. We validate the first stages of this method on real data in order to detect globally the maxillary-facial malformations. The following part of this paper will be organized as follows: We start with a presentation of the different stages of the chain we propose, then, we detail each stage separately before presenting and discussing the obtained results.

## 2. THE PROPOSED METHOD

For the creation of the atlas, we propose some stages which can be summarized in figure 1.



**FIGURE 1:** Global architecture of the proposed method

The method we propose, simple as it appears, involves some complexity at the level of choosing each technique. In what follows, we detail each step.

### 2.1 Extraction of the Bone Structure Using HMM Method

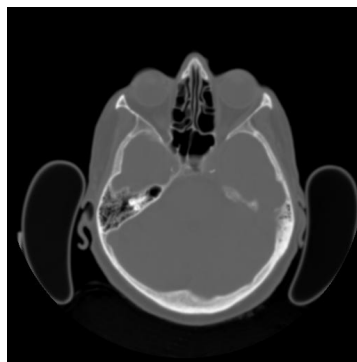
This first stage consists in isolating the bone structures from the remaining structures present in the skull. The manual location of these bones, though theoretically feasible, is in practice, inconceivable in reasonable times. It is, therefore, crucial to develop automatic techniques capable of providing results similar to those of an expert.

Generally speaking, the segmentation of the anatomic structures in the medical images is a particularly difficult task mainly because of both; the frequently indiscernible contours and the presence of noise. This hard aspect is increased, for the cranium CT data, due to the complex

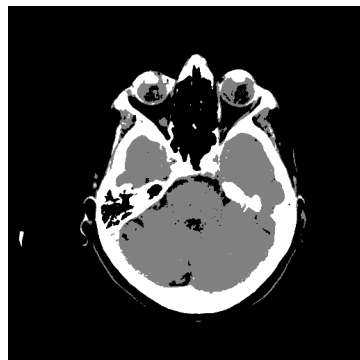
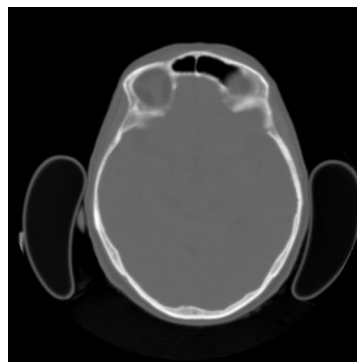
structure of this organ. This is why the conventional methods failed to isolate the bone-structure in these images. In order to get around this problem, we propose to explore the segmentation techniques integrating a priori knowledge to guide the segmentation processes.

Markov models seem to adapt well to the segmentation problem of skull MRI scans. In fact, they introduce spatial dependence between pixels allowing segmentation robust to noise. The segmentation is performed, in a Bayesian framework, using a maximum a posteriori criteria (MAP) where the data term is modeled using a finite gaussian mixture model and the prior term is modelled by means of a hidden markov random field (HMRF). Specifically, we use the HMRF-Expectation Maximization algorithm, proposed by Zhang et al. [17]. The segmentation is fully automatic

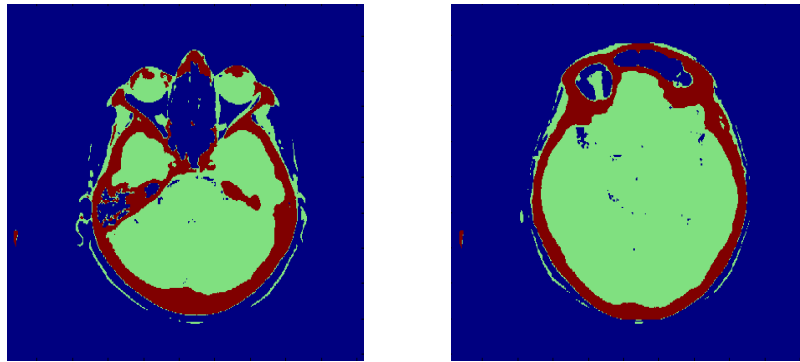
Figure 2, presents some important results obtained by the cited method. The method performance is compared with the Otsu thresholding method generally used in isolating the bone in skull images.



a. Original images



a. Results of Otsu method



b. results of the HMM method

**FIGURE 2:** Results of bone segmentation

### 2.2 Tridimensional Characterization of the Bone-Structures Using Crest Lines

For the creation of the atlas to be statistically significant, it is important to have a large skull image base available. Besides the problem of structural complexity related to these images, we would face the problem of the huge amount of tridimensional data to be processed. This would generate a striking complexity in the processing and analysis chain. It would so be important to represent these data through sparser but significant elements. The problem thus becomes one of tridimensional characterization with an anatomic significance.

The second stage of the proposed chain consists then in working out automatically geometrical characteristics from the segmented images. These need to possess a mathematical definition in order to be calculated together with an anatomical pertinence to be representative of the structures' shape. So, the choice of these characteristics is dependent on certain constraints. In fact, these characteristics need to be:

- Generic: They need to be present in each one of the images of the image base.
- Measurable: They allow a mathematical definition permitting them to be calculated.
- Invariable: they need to have the same position in all images of the base.
- Anatomically pertinent: they need to approve an anatomic significance as they are going to serve as a reference mark.

We finally point out that the choice of the characteristics is tightly linked to nature of the application.

Generally speaking, to characterize the volumic images, we can either characterize the volumes or the surfaces. The first case is about considering the intensity of each voxel. Given the huge number of these data (voxel), this is going to bring about very heavy and complicated calculations. In the second case, only the surface of the structures is considered. It is in this frame that our work is situated seeing that the surgical intervention is going to be operated on the external surface of the skull.

The punctual characterization consists in working out particular points which are significant and able to serve as reference. Divers points can be worked out, as it happened, the extreme ones [1] and the Harris corners [2].

The line characterization consists in calculating characteristic lines using differential geometry. As it happens, we could cite: the geodesic lines [3], the crest lines [4], the medial axes [5], the junction lines [5].

The surface characterization consists in decomposing the surface of the object into significant pieces. This method hasn't been quite used in practice, but its principle has been adopted by Clouchoux [7] to automatically define a robust fragmentation of cortical surface.

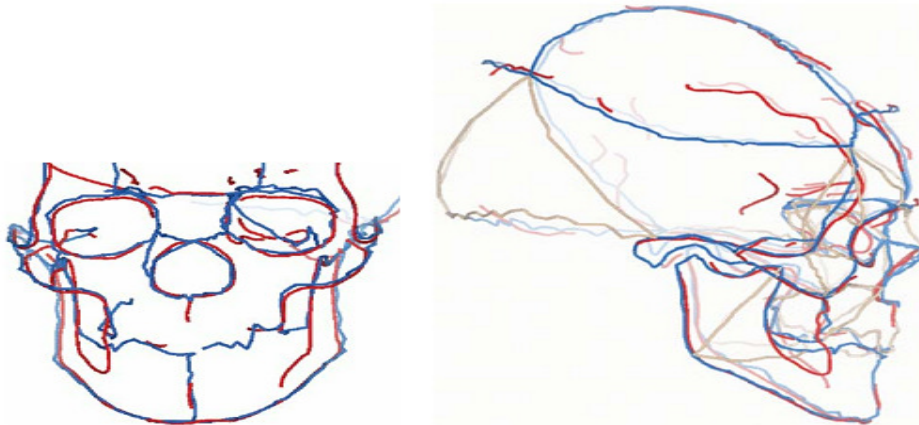
The study of the different methods of characterization leads us to deduce that:



- The surface characteristics are difficult to manage because they include much information as they contain all the structure's points.
- In contrast, the punctual characteristics provide no more than one disparate and little robust information as there is no relationship between the points.
- The line characteristics seem to be a good compromise. Indeed, they possess at the same time a strong reduction of the information quantity together with strong connexity constraints as a line is an orderly list of points. On the other hand, the line characteristics are not all efficient: The extraction algorithms of the geodesic lines raise some problems and the median axes are unstable.

Our choice has then turned to the crest lines in order to characterize the human skull surface. Indeed, in addition to their quite simple geometrical characteristics, the crest lines involve a strong anatomic pertinence as they are very close to the "Ridge lines"; the anatomic reference marks used by doctors and which correspond to the salient structures of the skull shape.

Indeed, the anatomists locate, on the cranial structures, very specific lines called the "Ridge lines" to serve as anatomic reference marks for the maxillary facial operations. These lines correspond to the salient structures of the skull form [8]. Figure 3 presents the crest lines (in red) and the "Ridge lines" (in blue) worked out on the same skull. Both line sets are very close. They indicate that the crest lines have a strong anatomic pertinence and so are efficient for the skull characterization.



**FIGURE 3:** Correspondence between the crest lines (in red) and the "Ridge lines" (in blue) [8]

To work out the crest lines we are going to follow these stages. This is to guarantee that the worked out lines be usable afterwards.

- Represent the reconstructed data by a triangular grid/meshing (we use the ITK-VTK environment).
- Smooth the data with a Laplacian filtering
- Simplify the obtained mesh
- Search for the crest lines on the obtained grids.
- Filtering the set of obtained lines to keep the significant ones only.

### 2.2.1 Data Smoothing

The reconstructed data are unusable. Indeed, the detection of the crest lines becomes erroneous because of the detection of many which are not significant. We propose to pre-treat the raw data with a Laplacian in order to regulate them. The laplacian triangular grids operator covers practically the whole spectrum of the 3D images treatment applications including the filtering, grid, compression, configuration, segmentation and interpolation applications [9].

The Laplacian operator reduces the high frequency information in the mesh geometry. With an excessive smoothing, the important details run the risk of disappearing. A good compromise between a strong smoothing and the detection of a sufficient number of crest lines is then necessary. For the moment, we control this compromise manually, thanks to a relaxation factor

$\alpha$  (equation (1)). This will allow the doctor to control the number of details he wants to characterize/represent in the skull structure.

The filter used consists in adjusting the vertex's position so as to 'relax' the mesh. The new position is function of the balanced average of its direct neighbors in an iterative procedure. Indeed, the algorithm functions this way: to a given iteration  $i$ , for each point  $\mathbf{p}_i^n$ , a topological analysis is carried out on the vicinity  $N(i)$  to determine the points  $\mathbf{p}_j^n$  which are linked to  $\mathbf{p}_i^n$ , and the faces which are connected to them. Then, the iterative process starts on all points. The new coordinates of the point  $\mathbf{p}_i^n$  are altered according to a balanced average of the connected points with reference to the equation (1):

$$p_i^{n+1} = p_i^n + \alpha \cdot \frac{\sum_{j \in N(i)} \alpha_{ij} (p_j^n - p_i^n)}{\sum_{j \in N(i)} \alpha_{ij}} \quad (1)$$

In this equation, alpha denotes the factor comprised between 0 and 1 which controls the displacement rate. The alpha weights can be constant or adaptive. Their calculation can be done according to several methods. For details, we cite the works [10,11]

Figure 4 gives the smoothing result with different relaxation parameters.



**FIGURE 4:** the smoothing result with different relaxation parameters

### 2.2.2 Mesh Simplification Phase

The obtained meshes are very dense. This could raise visualization problems and even influence the characterization quality. More generally, the grid's complexity conditions the quality of the treatments following from it. A mesh simplification stage, then, proves necessary. This amounts to a reduction of the number of the structure's polygons while preserving, in prime condition, the initial structure's topology. The mesh simplification problem is one of active research.

We remind here that we have opted for a triangular mesh this for their general aspect and their widespread use. The problem of triangular mesh simplification has been broadly studied in several works. These methods use different approaches: Edge Collapse, Vertex decimation, Vertex Clustering, Face Constriction Process.

In the first case, an edge is identified and broken down to form a new vertex. In vertex decimation, a vertex is deleted together with its borders and the whole resulting from that is re-triangulated. In vertex clustering a area is placed around the model and is divided into an equally spaced grid. The vertexes present in each cell are contracted together and

consequently the corresponding faces are updated. The process is quite fast and the exit grid's quality depends on the matrix's size. In the approaches, based on the face construction, a face is strangulated and consequently the adjacent faces degenerate and will be deleted [12-14]. For a detailed state of the art, we recommend the works of...and the references it includes. In the ITK-VTK environment, the pre-established method is that of [15] we have then opted for this solution.

### 2.2.3 Crest Lines Extraction Phase

The meshes are now ready to be analyzed. The following stage, then, consists in calculating the crest lines on this mesh. In the literature, several methods have been proposed to extract these lines on triangular grids.

Given that the crest lines are characterized by passing through zero of one of the curvature function's derivative. These methods are all based, in the first phase on the calculation of the curvatures. As an input, the data are meshes. All calculations are then done in the discrete domain. Several curvature estimation methods in a discrete field have been proposed. We cite by way of example: the methods using surface local parametric approximations, the curvatures are thus analytically deduced [16]. The methods considering the average at the grids' faces and which approximate the curvatures in the vicinity of each vertex. The methods using the average of the curvatures' tensors curvature. Indeed, for each vortex, the curvature is taken as the tensors average of all the arrests closely related to this vortex.[16]

In what follows, we detail the curvature estimation done in the ITK-VTK environment. To do this, we used the grid's geometrical parameters represented by figure 2 which presents a face F having P as vortex: It is about the vortex P, face F's angle  $\alpha_i$ , the length  $l_i$  of the arrest opposite the / opposed to the angle  $\alpha_i$  and  $N_i$  the averages to the face F and  $\beta_i$  which represents the angle between two averages with two faces neighboring the vertex P.

From these parameters, we can calculate for each grid's vertex, its Gaussian curvature  $K_{Dis}$  given by

$$K_{Dis} = \frac{2\pi - \sum_i \alpha_i}{1/2 \sum_i Aire(F_i) - 1/8 \sum_i \cot \alpha_i l_i^2} \quad (2)$$

and its average curvature  $H_{Dis}$  given by :

$$H_{Dis} = \frac{1/4 \sum_i \beta_i l_i}{1/2 \sum_i Aire(F_i) - 1/8 \sum_i \cot \alpha_i l_i^2} \quad (3)$$

From  $K_{Dis}$  and  $H_{dis}$  we can deduce the maximal curvature  $K_{max}$  of the vertex such as:

$$k_{max} = H_{Dis} + \sqrt{H_{Dis}^2 - K_{Dis}} \quad (4)$$

A vertex S is considered a "crest point" only if it admits a maximal curvature  $k_{max} > 0$ . In fact, knowing that certain grid vertexes can have maximal curvatures with a positive value close to zero, one has to choose a value  $\epsilon > 0$  which will serve as a threshold. So, a vertex S will be interpreted as "crest point" only if it admits  $k_{max} \geq \epsilon$ . The set of points found constitutes the grid's "crest points". The crest lines will be the arrises whose vertexes are the crest points.

### 2.3 Data Registration

In a subsequent stage, we will try to detect the potential malformations going solely by the characteristics calculated beforehand. This is done by registration the lines of each subject against those of the base subject.

The registration method used is the algorithm ICP (for "iterative closest point")[17.] This algorithm has been kept for its simplicity and its performance. It allows the registration of two

surfaces iteratively: to each iteration, the corresponding points (the closest points) of both entities to be registered are first defined; the average distance between the points pairs previously defined is then brought to minimum through a rigid transformation. Let P and X be two sets. The both point clouds to be registered.

Two subsets of these entities, respectively  $\{p_i\}$  and  $\{y_i\}$ , are put into correspondence through a similarity criterion. The matter, then, at issue is to find a rigid transformation marked  $(R,t)$  which will minimize the criterion to the meaning/in the direction of the following minimal squares:

$$e(R,t) = \frac{1}{N} \sum_{i=0}^N \left\| (Rp_i + t) - \psi(p_i) \right\|^2 \quad (5)$$

where the  $\psi$  function associates to each point  $p_i$  a point from the set X and N corresponds to the number of point pairs/couples. To an iteration k of the calculation procedure, the global

transformation is updated incrementally as follows.  $R = R_k R$  and  $t = t_k + t$

Finally, the criterion to be minimized becomes:

$$e(R_k, t_k) = \frac{1}{N} \left\| R_k (Rp_i + t) + t_k - \psi(p_i) \right\|^2 \quad (6)$$

The correspondence function  $\psi$  is defined by:

$$\psi(p_i) = x \left| \min_{x \in X} \left| D((Rp_i + t), x) \right. \right. \quad (7)$$

Once the flunking has been performed / completed, the comparison of the two facial surfaces will be based on the appropriate point pairs in both point clouds. This allows detecting the areas of the face mass containing malformations by locating the places where the crest lines of both subjects present striking discrepancy

### 3. TESTS AND RESULTS

We tested the suggested method on two sets of data: the first one contains malformations were simulated by translation and rotation, the second one contains real data.

#### 3.1 Detection of Malformations of the Simulated Data

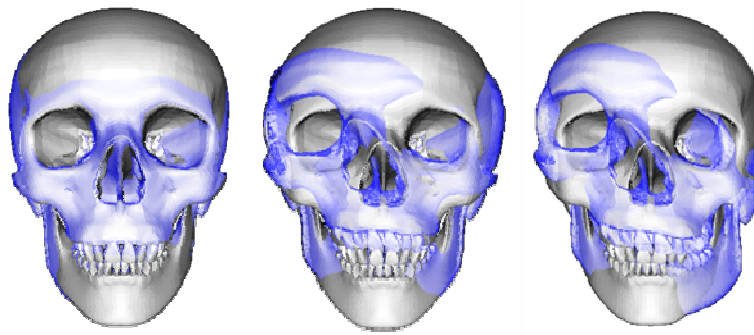
We devote this section to the study of possible malformations which the subjects of the simulated database have. To do this, we registrate (put in correspondence) each one of these subjects to the reference image. This will enable us to define the deformed zones for each cranium. The results obtained are presented by the figure 3-a where we present the image of reference by opaque surface and those of the patients by a transparent color. To define the deformed zones, we refer to a preexistent image of Atlas. The analysis of the figure enables us to deduce that:

- The first subject, on which we did not make many modifications, does not have any malformation except for some light shifts which we cannot interpret them like malformations
- The second subject is shifted compared to the image of reference, in its frontal, zygomatic and nasal bone,
- The third subject presents even more deformations than the second since a shift in the left vertical branch of its mandible was detected.

In order to validate the performance of the results of the method, we put in correspondence the ridge lines extracted from each subject compared to those of the image of reference. The result is given by the figure 3-b, highlighting that:

- The lines of the first subject, considered as healthy, are very close even coincided with those of the reference,
- The ridge lines of the second and third subjects are shifted on certain levels.

Referring to the figure 3-a, we can confirm that we succeeded in detecting some malformations as those identified by using the whole images. Indeed, we could detect the presence of malformations, for the second and the third subjects, each one in their frontal, nasal and zygomatic bones and a malformation in the left vertical branch of the mandible of the third subject.



(a) Registration of images

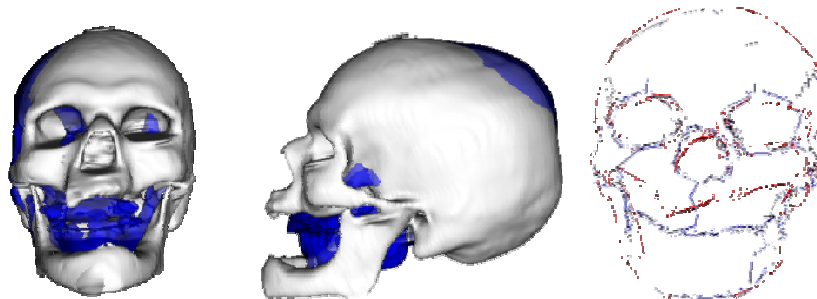


(b) Registration of crest lines

**FIGURE 5:** Results on simulated data

### 3.2 Detection of Malformations of the Real Data

Since the images of database correspond to only healthy cases, we will registrate two images by choosing one of them like reference. The result is given by the figure 4-a. As already mentioned, the two subjects are healthy, which justifies that the registration of two craniums did not give great deformations except for those due to the essential difference between the two patients. This result also is obtained, as illustrated by the figure 4-b, by registering the ridge lines (represented by different colors), of the two subjects. This figure highlights that the two whole of lines is very close. This shows that any of the subjects does not have malformations compared to the other subject.



(a) Registration of real craniums

(b) Registration of crest lines

**FIGURE 6:** Results on real data

### 3. CONCLUSION

This work presents a contribution to the analysis of deformations of the scanographic images for the detection and the localization of the malformed areas of the maxillo-facial mass. The study of the normality of a subject compared to a model of reference constituted the main axis of our work. In order to manipulate heavy volume data, we extracted from the cranial images a whole of linear characteristics called "ridge lines". This choice is justified by the simplicity of these lines and their anatomical relevance. Then, we tried to detect possible malformations by registering (putting in correspondence) the ridge lines of a subject in those of a model of reference. The results obtained by applying this method, succeed to identify the various pathological areas of the maxillo-facial mass while manipulating data of more condensed sizes and to show consequently the validity of the suggested approach in this work.

### REFERENCES

- [1] C. Zhang, T. Chen, "Efficient feature extraction for 2D/3D objects in mesh representation", Proc. of the International Conference on Image Processing (ICIP), vol.3. 2001.
- [2] C.Harris and M.J.Stephens. A combined corner and edge detector. In Alvey Vision Conference, pages147–152,1988.
- [3] J.A. Sethian et R. Kimmel, "Computing Geodesic Paths on Manifolds", Proc. Natl. Acad. Sci., 95(15):8431–8435, 1998
- [4] J-P.Thirion and A.Gourdon. The 3d marching lines algorithm. Graphical Models and Image Processing, 58(6):503509,November1996.
- [5] H Blum. A transformation for extracting new descriptors of shape . Models for the perception of speech and visual form, 1967
- [6] Hahn, M., Krüger, N.: Junction detection and semantic interpretation using hough lines. In: DAGM 2000
- [7] C. Clouchoux, O. Coulon, J-L. Anton, J-F. Mangin, J. Regis, "A new cortical surface parcellation model and its automatic implementation", in: LNCS - Medical Image Computing and Computer-Assisted Intervention - MICCAI 2006: 9th International Conference, R. Larsen and M. Nielsen and J. Sparring, n° 4191, pp. 193-200, octobre 2006
- [8] Thirion J.P., 1996b The Extremal Mesh and the Understanding of 3D Surfaces, in International Journal of Computer Vision, 19(2):115-128.
- [9] Max Wardetzky, Saurabh Mathur, Felix Kälberer, Eitan Grinspun, Discrete Laplace operators: No free lunch. Symposium on Geometry Processing, 2007, pp. 33-37. Laplacien
- [10] SORKINE O.:Laplacian mesh processing. State of The Art Report. Eurographics 2005),53–70
- [11] Olga Sorkine, "Differential Representations for Mesh Processing". Computer Graphics Forum, Vol. 25(4), 2006.
- [12] D. P. Luebke.: A Developer's Survey of Polygonal Simplification Algorithms. IEEE Computer Graphics and Applications'01, 24-35, (2001)
- [13] Jerry O. Talton. A Short Survey of Mesh Simplification Algorithms. University of Illinois at Urbana-Champaign, October 2004.

- [14] Sajid Hussain, Håkan Grahn, and Jan Persson, "Feature-preserving Mesh Simplification: A Vertex Cover Approach," Proc. of the IADIS International Conference on Computer Graphics and Visualization 2008 (CGV 2008), pages 270-275.
- [15] W. Schroeder, J. Zarge, and W. Lorensen, "Decimation of Triangle Meshes," Computer Graphics (Proc. Siggraph 92, vol. 26, ACM Press, New York, 1992, pp. 65-70.
- [16] R. V. Garimella and B. K. Swartz, "Curvature Estimation for Unstructured Triangulations of Surfaces", Los Alamos National Laboratory, 2003
- [17] Y.Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm", IEEE Transaction on Medical Imaging, vol. 20, no. 1, pp. 45–47, 2001.

# Unsupervised Multispectral Image Classification By Fuzzy Hidden Markov Chains Model For SPOTHRV Images

**Faiza DAKKA**

*Faculty of Sciences, Department of Physics,  
Mohamed V-Agdal University  
Rabat B.P. 1014, Morocco*

dakka\_f@yahoo.fr

**Ahmed HAMMOUCH**

*GIT-LGE Laboratory, ENSET,  
Mohamed V-Souissi University  
Rabat, B.P. 6207, Morocco*

**Driss ABOUTAJDINE**

*Faculty of Sciences, Department of Physics,  
Mohamed V-Agdal University  
Rabat B.P. 1014, Morocco*

---

## Abstract

This paper deals with unsupervised classification of multi-spectral images, we propose to use a new vectorial fuzzy version of Hidden Markov Chains (HMC).

The main characteristic of the proposed model is to allow the coexistence of crisp pixels (obtained with the uncertainty measure of the model) and fuzzy pixels (obtained with the fuzzy measure of the model) in the same image. Crisp and fuzzy multi-dimensional densities can then be estimated in the classification process, according to the assumption considered to model the statistical links between the layers of the multi-band image. The efficiency of the proposed method is illustrated with a Synthetic and real SPOTHRV images in the region of Rabat.

The comparisons of two methods: fuzzy HMC and HMC are also provided. The classification results show the interest of the fuzzy HMC method.

**Keywords:** Bayesian Image Classification, Markov Chains, Fuzzy Hidden Markov, Unsupervised Classification.

---

## 1 INTRODUCTION

Hidden Markov models are used in many forms and with different spatial structures like strings, fields or trees [3]. The popularity of these models is illustrated by the multitude and diversity of applications that have been proposed, in signal processing in speech recognition and also in image processing. In particular, the model of hidden Markov chains (CMC) has been successfully used for image classification [3].

In this case, 2D images are first transformed into a 1D vector path through Hilbert-Peano [5] to fit the one-dimensional structure of a chain Fig. 1 (a). The interest for this model comes from the fact that when the hidden process  $X = (X_1, \dots, X_N)$  can be represented by a finite Markov chain and when the structure of the noise is not too complex, then  $X$  can be reconstructed from the only observed process  $Y = (Y_1, \dots, Y_N)$ ,  $Y_n \in R$ , using different criteria for classification like MAP "Maximum A Posteriori" or MPM Maximal Posterior Mode. It is sometimes interesting to take into account not only the uncertainty in the observations (often due to noise), but also their vagueness (fuzzy). This imprecision may come from a motion sensor at the time of the acquisition of a photo, or the effect of surface / volume part found in

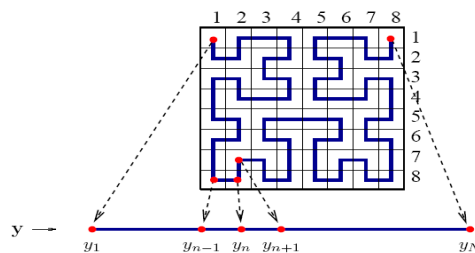


medical imaging and satellite. The latter effect reflects the fact that a pixel of the image is the result of integration over a surface not necessarily homogeneous, where several components are mixed in different proportions and unknown. In terms of classification, these pixels have an ambiguous and sometimes it is better not to get a "harsh response" from classifier (eg, Class 0 or Class 1) but a "vague response" measured by an interval value  $]0,1[$ . This response reflects the ambiguity of the pixels can be observed and interesting information. Indeed, the level of degrees or can be interpreted as the mixing ratio of class 0 versus class 1 in the pixel, or as the degree of belonging to class 0.

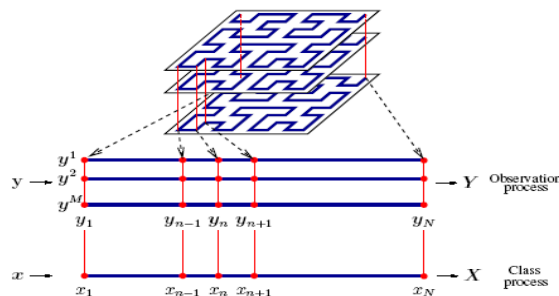
This paper is organized as follows: The CMC model is briefly presented in Section 2. In Section 3 the CMC model for fuzzy image classification is detailed, and also The CMC Parameter estimation, carried out with an extension of the algorithm Iterative Conditional Estimation (ICE) [1, 2]. Sections 4 illustrate the comparative results of Classification Unsupervised. Finally, Section 5 reviews the theoretical and experimental work and offers some perspectives.

## 2 UNSUPERVISED CLASSIFICATION USING HMC

This section is intended to give some recalls about the HMC model [6,7] and its use for unsupervised image classification [8]. The HMC model can be adapted to a 2D analysis through a Hilbert-Peano scan of the image, see Fig. 1. Hence all estimation and classification processings are applied on the 1D sequence, and the segmented 2D image is reconstructed by using a reverse Hilbert-Peano scan from the 1D classified sequences.



**FIGURE 1:** Hilbert-Peano scan construction for an 8\*8 image. This scan is used to transform a 2D image into a 1D signal ( $y$ ), and conversely.



**FIGURE 2:** Hilbert-Peano scan for an 8\*8 multi-component image ( $M$  = number of layers in the image,  $N$  = number of pixels in each layer).

The Hilbert-Peano scan presents the ability to take into account the neighborhood of the pixel of interest [6]. Fig. 2 shows the scan for a 8 \* 8 multi-component image with  $M$  layers.

### A. The HMC model

An image  $y = \{y_1, \dots, y_N\}$ ,  $N$  being the total number of pixels, is considered as a realization of the 1D observed process  $Y = \{Y_1, \dots, Y_N\}$ , each  $Y_n$  is a real-valued random variable. The segmented image  $x = \{x_1, \dots, x_N\}$  is considered as a realization of a hidden process  $X = \{X_1, \dots, X_N\}$ , each  $X_n \in \Omega = \{1, \dots, K\}$  is a discrete random variable. In classical HMC modeling,  $X$  is assumed to be a Markov chain, i.e.  $p(x_{n+1} | x_n, \dots, x_1) = p(x_{n+1} | x_n)$ .

The distribution of  $X$  is consequently determined by the distribution of  $X_1$ , denoted by  $\pi_k = p(X_1 = k)$ , and the set of transition matrices  $(A_n)_{1 \leq n \leq N}$  whose entries are  $a_{ij}^n = p(X_{n+1} = j | X_n = i)$ .

We further assume that  $X$  is a stationary Markov chain, i.e. entry  $a_{ij}^n = a_{ij}$  does not depend on  $n$ . With the following additional properties:

- (i)  $Y_n$  are independent conditionally to  $X$ , i.e.  $p(y | x) = \prod_{n=1}^N p(y_n | x)$ , and
- (ii)  $p(y_n | x) = p(y_n | x_n)$ , the distribution of the pairwise process  $(X, Y)$  can be written as

$$p(x, y) = \pi_{x_1} f_{x_1}(y_1) \prod_{n=2}^N a_{x_{n-1}, x_n} f_{x_n}(y_n)$$

with  $f_{x_n}(y_n) = p(y_n | x_n)$  the data-driven densities, which are assumed to be Gaussian in this work [8].

**B. Classification:**

The estimation of  $X$  from  $Y$  can be done by applying the MPM Bayesian criterion:

$$\forall n \in [1, \dots, N], x_n^{MPM}(y) = \arg \max_k \xi_n(k)$$

(1)

with  $\xi_n(k) = p(X_n = k | y)$  the marginal a posteriori probabilities.

The HMC model allows explicit computation of the MPM solution using the well-known Baum's "forward"  $\alpha_n(k)$  and "backward"  $\beta_n(k)$  probabilities [10], modified by Devijver [9] for computational reasons:

$$\alpha_n(k) \approx P(X_n = k, Y_1 = y_1, \dots, Y_n = y_n) \tag{2}$$

$$\beta_n(k) \approx \frac{P(Y_{n+1} = y_{n+1}, \dots, Y_N = y_N | X_n = k)}{P(Y_{n+1} = y_{n+1}, \dots, Y_N = y_N | Y_1 = y_1, \dots, Y_n = y)}$$

(3)

In the following, we use the numerically stable forward-backward recursions resulting from these approximations:

Forward initialization:

for  $1 \leq i \leq K$ .

$$\alpha_n(i) = \frac{\pi_i f_i(y_1)}{\sum_{1 \leq j \leq k} \pi_j f_j(y_1)}$$

(4)

Forward induction:

for  $n = 2, \dots, N$  and  $1 \leq i \leq K$ .

$$\alpha_n(i) = \frac{f_i(y_n) \sum_{1 \leq j \leq k} \alpha_{n-1}(j) a_{ij}}{\sum_{1 \leq l \leq k} f_l(y_n) \sum_{1 \leq j \leq k} \alpha_{n-1}(j) a_{lj}}$$

(5)

for  $1 \leq i \leq K$

$$\beta_N(i) = 1$$

Backward induction:

for  $n = N-1, \dots, 1$  et  $1 \leq i \leq K$

$$\beta_n(i) = \frac{\sum_{1 \leq j \leq k} a_{ij} f_j(y_{n+1}) \beta_{n+1}(j)}{\sum_{1 \leq l \leq k} f_l(y_{n+1}) \sum_{1 \leq j \leq k} a_{lj} \beta_{n+1}(j)}$$

(6)

It can be shown that marginal a posteriori probabilities involved in MPM classification can be written

$$\xi_n(i) = \alpha_n(i) \beta_n(i), \tag{7}$$

(7)

and joint a posteriori probabilities  $\Psi_n(i, j) = P(X_n = i, X_{n+1} = j | Y = y)$  as:

$$\Psi_n(i, j) = \frac{\alpha_n(i) a_{ij} f_j(y_{n+1}) \beta_{n+1}(j)}{\sum_{1 \leq m \leq k} f_m(y_{n+1}) \sum_{1 \leq l \leq k} \alpha_n(l) a_{lj}}$$

(8)

**C. Estimation:**

Before classification, all parameters involved in the CMC model have to be estimated.

$$\theta = \{\pi_k, a_{kl}, f_k\} \tag{9}$$

(9)

One well-known solution is to use the EM iterative procedure [11] aiming at optimizing the log-likelihood of data, according to the steps described in Algorithm 1, with following update equations

$$\forall k \in \Omega, \Pi_k^{[q]} = \frac{1}{N} \sum_{n=1}^N \xi_n^q(k) \tag{10}$$

$$\forall k, l \in \Omega, \hat{a}_{kl}^{[q]} = \frac{\sum_{n=1}^{N-1} \Psi_n^q(k, l)}{N \hat{\Pi}_k^q} \tag{11}$$

$$\forall k \in \Omega, \hat{\mu}_k^{[q]} = \frac{\sum_{n=1}^N \xi_n^q(k) y_n}{N \hat{\Pi}_k^q} \tag{12}$$

$$\forall k \in \Omega, \hat{\sigma}_k^{2 [q]} = \frac{\sum_{n=1}^N \xi_n^q(k) (y_n - \hat{\mu}_k^{[q]})^2}{N \hat{\Pi}_k^q} \tag{13}$$

**Algorithm 1 :** HMC parameters estimation using EM.

q ← 0: Initialize parameters  $\theta^{[0]}$

repeat

q ← q + 1

- Compute “Forward”  $\alpha_n^{[q]}$  and “Backward”  $\beta_n^{[q]}$  probabilities using equations (5) and (6).

- Compute a posteriori probabilities  $\xi_n^{[q]}(k)$  using eq. (6) and  $\psi_n^{[q]}(k, l)$  using eq. (7) and (8).

- Estimate CMC parameters  $\theta[q]$  using equations (10) and (11) for Markov parameters and equations (12) and (13) for Gaussian data-driven parameters.

until  $|\theta[q] - \theta[q-1]| < \text{Threshold}$ .

Iterative estimation of parameters is stopped when parameters do not vary much.

### 3 UNSUPERVISED CLASSIFICATION USING FUSSY HMC

#### A. Fuzzy Markov chains model

We consider a multi-component image of M layers. According to the Hilbert-Peano scan, we get N series of M data, denoted by  $y = \{y_1, \dots, y_N\}$ , where  $y_n = \{y_n^{[1]}, \dots, y_n^{[M]}\}^t$ ,  $1 \leq n \leq N$ .

In classical HMC approach, the aim is to classify each  $y_n \in \mathbb{R}^M$  into a set of K classes, the state space  $\Omega = \{w_1, \dots, w_K\}$ , in order to obtain the segmented chain  $x = \{x_1, \dots, x_N\}$  ( Fig. 2). The segmented image is then reconstructed from x using an inverse Hilbert-Peano scan.

For the sake of simplicity, we confine our study to the K = 2 case, i.e.  $\Omega = \{0, 1\}$ .

In fuzzy HMC context, the range of  $x_n$  is now the interval  $\Omega = [0, 1]$ . In the following,  $\varepsilon_n$  will denote a realization of random variable  $X_n$  and we will adopt the notation:

$\varepsilon_n = 0$  if the pixel is from class 0,

$\varepsilon_n \in ]0, 1[$  if the pixel is a fuzzy one,

$\varepsilon_n = 1$  if the pixel is from class 1.

#### B. Probabilities in fuzzy Markov chains context

As stated previously, each  $x_n$  takes its value in two types of sets: a hard one  $\{0, 1\}$ , and a fuzzy one defined over the range  $]0, 1[$ .

Let  $\delta_0$  and  $\delta_1$  be Dirac weights on 0 and 1, and  $\square_1$  the Lebesgue measure on  $]0, 1[$ .

By taking  $\nu = \delta_0 + \delta_1 + \square_1$  as a measure on  $\Omega$ , the distribution of  $X_n$  can be defined by a density h on  $\Omega$  with respect to  $\nu$ .

If we assume that X is homogeneous and the distribution of each  $X_n$  is uniform on the fuzzy class,  $P(X_n = \varepsilon_n) = h(\varepsilon_n) = \pi_{\varepsilon_n}$  can be written:

$h(\varepsilon_n = 0) = \pi_0;$

$h(\varepsilon_n = 1) = \pi_1;$   
 $h(\varepsilon_n) = \pi_{]0,1[}; \quad \varepsilon_n \in ]0, 1[;$   
 with  $\pi_0 + \pi_1 + \pi_{]0,1[} = 1.$

We can now detail the new expression for the transition probabilities of the Markov chain  $t_{\varepsilon_{n-1}, \varepsilon_n}$  :

$$\begin{aligned}
 t_{\varepsilon_{n-1}, \varepsilon_n} &= P(X_n = \varepsilon_n / X_{n-1} = \varepsilon_{n-1}) = \\
 &P(X_n = 0 / X_{n-1} = \varepsilon_{n-1})\delta_0(\varepsilon_n) \\
 &+ P(X_n = \varepsilon_n / X_{n-1} = \varepsilon_{n-1})t_{]0,1[}(\varepsilon_n) \\
 &+ P(X_n = 1 / X_{n-1} = \varepsilon_{n-1})\delta_1(\varepsilon_n) \\
 &\forall \varepsilon_{n-1}, \varepsilon_n \in \Omega \text{ and } \forall n \in \{2, \dots, N\}
 \end{aligned}$$

With  $t_{\varepsilon_{n-1}, \varepsilon_n} \geq 0$  and  $\int_{\Omega} t_{\varepsilon_{n-1}, \varepsilon_n} d\varepsilon_{n-1} = 1$

$$P(X = x) = \prod_{\varepsilon_1} \prod_{n=2}^N t_{\varepsilon_{n-1}, \varepsilon_n}$$

### C. Multi-component fuzzy HMC implementation.

$\varepsilon_n$  is a realization of a random variable  $X_n$ , and each  $y_n$  is a realization of a random vector  $y_n = \{y_n^{[1]}, \dots, y_n^{[M]}\}^t$ . Thus the problem is to estimate the unobserved realization  $x$  of a random process  $X$  from the observed realization  $y$  of a random process  $y = \{y_1, \dots, y_N\}$ ;

Similarly to classical HMC, multi-component fuzzy HMC based image classification methods consider the two following assumptions:

H1: the random variables  $Y_1, \dots, Y_N$  are independent conditionally on  $X$ ;

H2: the distribution of each  $Y_n$  conditionally on  $X$  is equal to its distribution conditionally on  $X_n$ .

It is important to note that the random variables  $(y_n^{[m]})_{1 \leq m \leq M}$  are not assumed to be mutually independent conditionally on  $X_n$ .

Assuming that distributions of  $(X_n, Y_n, X_{n+1}, Y_{n+1})$  are independent of  $n$ , each state  $\varepsilon_n$  of the state space (i.e. hard classes  $\{0,1\}$ , as well as the fuzzy class  $]0,1[$ ) is associated to a distribution characterizing the  $M$ -dimensional observations  $y_n$ :

$$f_{\varepsilon_n}(y_n) = P(Y_n = y_n \mid X_n = \varepsilon_n), \tag{14}$$

Given an observed sequence  $y = \{y_1, \dots, y_N\}$ , the joint state-observation probability is given by:

$$P(X = x, Y = y) = \prod_{\varepsilon_1} f_{\varepsilon_1}(y_1) \prod_{n=2}^N t_{\varepsilon_{n-1}, \varepsilon_n} f_{\varepsilon_n}(y_n) \tag{15}$$

In unsupervised classification, the distribution  $P(X = x, Y = y)$  is unknown and must first be estimated in order to apply a Bayesian classification technique (MAP or MPM). Therefore the following sets of parameters need to be estimated:

- 1) The set  $\Gamma$  characterizing the fuzzy Markov chain parameters, i.e. the initial probability vector  $\pi = \{\pi_\varepsilon\}_{\varepsilon \in \Omega}$  and the transition probabilities  $t_{\varepsilon_{n-1}, \varepsilon_n} \forall \varepsilon_{n-1}, \varepsilon_n \in \Omega$
- 2) The set  $\Delta$  regrouping the parameters of the  $M$ -dimensional distributions presented in (14), i.e. the distributions associated with the hard classes and the fuzzy one.

### D. Fuzzy Markov Chain parameters estimation

For the estimation of the parameters in  $\Gamma$ , we propose to use an adaptation of the general ICE algorithm [12] which is an alternative to the well-known Estimation-Maximization (EM) algorithm. In fact, ICE does not refer to the likelihood, but it is based on the conditional expectation of some estimators from the complete data  $(x, y)$ . It is an iterative method which produces a sequence of estimations  $\theta_q$  of parameter  $\theta$  as follows:

- 1) initialization  $\theta_0$ , obtained with an initial classification algorithm (k-means algorithm).

2) computation of  $\theta^{q+1} = [\hat{\theta}(x, y) / Y = y]$  where  $\hat{\theta}(x, y)$  is an estimator of  $\theta$ .

3) Stop the algorithm when  $\theta_{Q-1} \square \theta_Q$ .

This section is not intended to give a complete description of the ICE algorithm in the HMC context, interested readers may consult [5]. Similarly to the classical case, parameters in  $\Gamma$  can be calculated analytically by using the Baum-Welch algorithm [13]:

for the hard classes, the classical normalized Baum-Welch probabilities [14] can be used directly.

for the fuzzy class, the forward and backward probabilities are defined by:

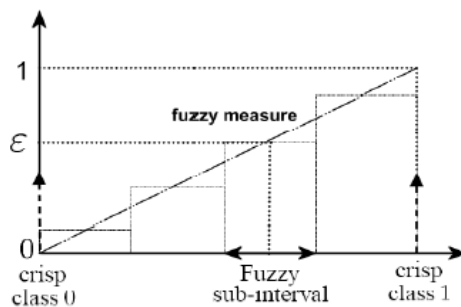
$$\alpha_{n+1}(\xi) \propto f_{\xi}(y_{n+1}) \int_{[0,1]} \alpha_n(\zeta) t_{\zeta, \xi} d\zeta \tag{16}$$

$$\beta_n(\xi) \propto \int_{[0,1]} \beta_{n+1}(\zeta) t_{\zeta, \xi} f_{\xi}(y_{n+1}) d\zeta \tag{17}$$

The integrals above can not be calculated analytically and numerical integration must be performed. Hence, the continuous interval ]0,1[ is partitioned into a given number F of subintervals. We thus reduce the domain of fuzzy membership degree ]0,1[ to F attributed values, corresponding to the medium value of the sub-interval of interest (see Fig. 3), thus resulting in a quantization of the fuzzy measure ]0; 1[.

We obtained F fuzzy classes "discrete, whose value corresponds to the fuzzy median subinterval considered.

For example, F = 2 implies  $\epsilon \square \{0.25, 0.75\}$ , F = 3 implies  $x \square \{0.165, 0.5, 0.825\}$ . More F is large, over the parameter estimation is accurate.



**FIGURE 3:** Partition of the continuous interval ]0,1[ into F = 4 sub-intervals. The attributed values  $\epsilon$  correspond to the medium value of the sub-interval of interest.

E. Multi-dimensional density estimation

At each ICE iteration, we need to estimate the multidimensional densities  $f_{\epsilon_n}(y_n)$ . Several strategies from multivariate data analysis are available, depending on the assumptions made on the statistical links between the layers, and on the choice of the shape of the one-dimensional densities. If independence between the layers is assumed,  $f_{\epsilon_n}(y_n)$  is the product of M densities  $g_{\epsilon_n}^1, \dots, g_{\epsilon_n}^M$  defined on R:

$$f_{\epsilon_n}(y_n) = \prod_{m=1}^M g_{\epsilon_n}^m(y_n^m)$$

For example, if we consider that  $f_{\epsilon_n}(y_n)$  are multidimensional

Gaussian densities, parameters estimation can be easily achieved from the first and second moments of a M-dimensional sample. Denoting by  $N(m, \sigma^2)$  the normal distribution with mean m and variance  $\sigma^2$ , the pdf of the hard classes are then expressed according to:

$$\epsilon_n = 0 : N(m_0; \sigma_0^2),$$

$$\epsilon_n = 1 : N(m_1; \sigma_1^2);$$

The parameters  $\Delta = \{m_0, m_1, \sigma_0, \sigma_1\}$  can be estimated by computing the empirical mean of

several estimates according to  $\theta^{q+1} = \frac{1}{L} \sum_{l=1}^L \hat{\theta}(x^l, y)$ , where  $x^l$  is an a posteriori realization

of  $X$  conditionally on  $Y$ . It can be shown that  $X | Y$  is a non homogeneous Markov chain whose parameters can be computed from the forward and backward probabilities: (16) and (17).

The definition of the fuzzy measure  $\bar{A}$ : “the pixel belongs to class 1” corresponding to the fuzzy class  $]0,1[$ , and its fuzzy membership function  $\mu_A$ , allows us to estimate the fuzzy parameters of the set  $\Delta$  in this new context. The proposed fuzzy membership function  $\mu_A$  is defined by:

$$\mu_A(m) = \begin{cases} 1 - \frac{m_1 - m}{m_1 - m_0} & \forall m \in [m_0, m_1] \\ 0 & \text{elsewhere} \end{cases}$$

Accordingly, the parameters of the Gaussian pdf for the fuzzy class can then be estimated by:

$$m_{\varepsilon_n} = (1 - \varepsilon_n)m_0 + \varepsilon_n m_1$$

$$\sigma_{\varepsilon_n} = (1 - \varepsilon_n)^2 \sigma_0 + \varepsilon_n^2 \sigma_1^2$$

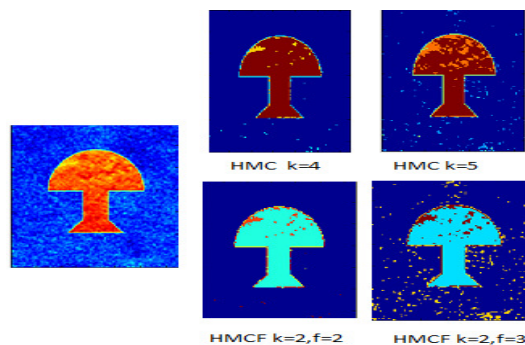
#### 4 MULTI-SPECTRALES IMAGES CLASSIFICATION

This section is intended to evaluate the two local classification approaches, and the benefit of relaxing the HMC model.

##### A. Validation

To validate the proposed approach we consider the synthetic image in Figure 4, we have considered a hard class (the opaque background and the umbrella), and the gradient between this class as the fuzzy zone (representing the border). Fig.4 present the results of two methods of classification based on HMC model and fuzzy HMC model.

We can be seeing that the image of Figure 4 appears to be classified much more precise with fuzzy HMC. The Taking into account a higher number of classes in the classical model of HMC did not have much interest because the additional classes tend to specialize on individual pixels.



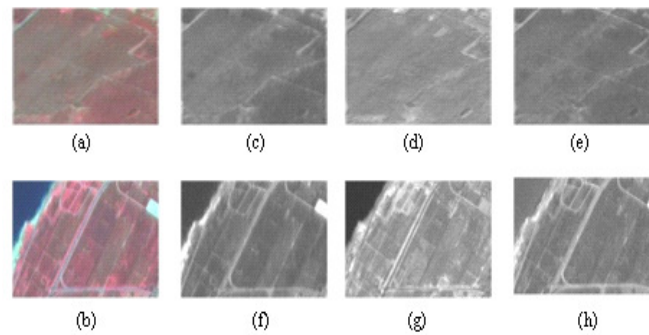
**FIGURE 4:** The synthetic image and their results of classification with CMC model ( 4 class and 5 class) and fussy CMC model (F=2, and F=3).

##### B. SPOTHRV Images Classification

###### 1. The studied images

The image chosen for the study is that of the forest region of Rabat Morocco. This forest is an ecological unit consisting of a typical natural vegetation and reforestation.

Figure 5 shows the satellite images studied, and its multi-spectral images (green, red and near infrared) of the forest Rabat.

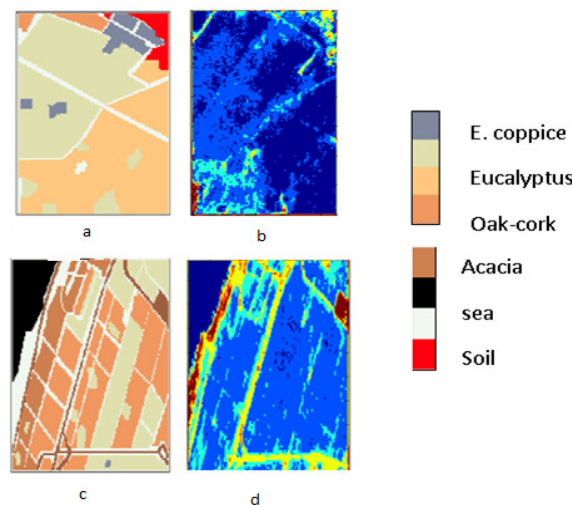


**FIGURE 5:** The two images studied (a and b) and their multi-spectral images (c, d and e) for the image (a) and (f, g and h) for the image (b).

HRV image - SPOT XS of this region, was acquired in December 1999 with a spatial resolution of 20m. in this image we extracted two test areas of size 140x90 pixels occupied mainly by a natural forest of orkoaks and eucalyptus planting, pine and acacia. The reality of the ground was obtained from the official plans of the forest, followed by field verification. These areas are characterized by heterogeneity in the distribution of forest Stands. Discrimination between different textures is not obvious to human observer.

2. Experimental protocol

In all experiments, parameters initialization was done with a k-means classifier. The ICE algorithm was stopped after 10 iterations, assuming it has converged.

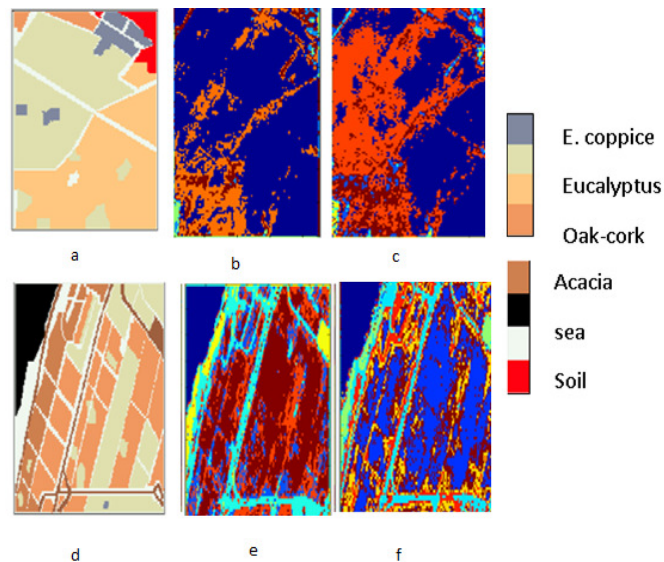


**FIGURE 6:** (a) and (c) Maps of the reality of the ground. (b) and (d) Results of classification with CMC (6 class) .

Figure 6 shows images of the realities ground (a and c) and the results of classification with CMC on both multi-spectral images of the forest in Rabat.(b and d). Computation times are 339.246seconds.

In this application, the “Sea” or the “Acacia” were considered as crisp class 0 and “Eucalyptus” or the “Oak-cork” crisp class 1. The fuzzy measure A thus corresponds to: “the pixel belongs to the free sea class”. The classification results depend on the partition of ]0,1[, the choice of the F sub-intervals implies different values of fuzzy measure  $\epsilon$ , e.g.  $F = 2$  implies  $\epsilon \in \{0.25, 0.75\}$ ,  $F = 3$  implies  $\epsilon \in \{0.165, 0.5, 0.825\}$ ,  $F = 4$  implies  $\epsilon \in \{0.125, 0.375, 0.625, 0.875\}$

Figure 7 shows images of the realities ground (a and c) and a classification results obtained with fussy vectorial CMC models for  $F=2$  sub-intervals(b and e) and  $F=3$  sub-intervals (c and f). Computation times are 545.636 seconds.



**FIGURE 7:** (a) and (d) Maps of the reality of the ground. Results of classification with fussy CMC for 4 crisp class and 2 fuzzy class ( $K=4, F=2$ ) (b) and (e) and for 3 crisp class and 3 fuzzy class ( $k=3, F=3$ ) (c) and (f).

The computational complexity involved in the fuzzy vectorial HMC model is quite higher than the one involved in the classical one. One can observe that a big number of fuzzy sub-intervals provide a fine characterization of the observed scene. Furthermore, the classification results produced by the fuzzy vectorial HMC model are more homogeneous. The global shape and frontiers of the class seem to be better defined. Nevertheless, this model clearly characterizes the frontiers between class and produces a reliable classification of class, and then illustrates the interest of the vectorial fuzzy HMC model.

## 5 CONCLUSION

In this work, we propose to extend the fuzzy HMC model to unsupervised multi-spectral image classification. In multi-component case, it is interesting not only to take into account the uncertainty measure of the noisy observation (characteristic of probabilistic approach in classical HMC), but also the imprecision measure of this observation (characteristic of fuzzy approaches [6]) By adding a fuzzy measure in a statistical model, we obtain an original model, different from classical and models. Indeed, it preserves the robustness of the statistical classification (based on measures of uncertainty), and enriches it with the fuzzy characteristic (measure of imprecision). The Experimental results confirm the validity of the proposed approach. The vectorial fuzzy HMC model seems to be promising in the field of multi-component image classification, due to the imprecision measure ability to take into account the multivariate information. We believe extending the fuzzy model to another model strictly more general than the CMC, called Markov chain couple [15] could also benefit from the fuzzy extension proposed in this paper.

## REFERENCES

- [1] W. Pieczynski. Statistical image segmentation. *Mach. Graph. and Vis.*, 1 : 261-268, 1992.
- [2] H. Caillol, W. Pieczynski, A. Hillion. Estimation of fuzzy Gaussian mixture and unsupervised statistical image segmentation. *IEEE Trans. on Im. Proc.*, 6(3) : 425-440, 1997.
- [3] W. Pieczynski. Modèles de Markov en traitement d'images. *Traitement du Signal*, 20(3) :255{278, 2003.
- [4] B. Benmiloud and W. Pieczynski. Estimation des paramètres dans les CMC et segmentation d'images. *Traitement du Signal*, 12(5) :433{454, 1995.



- [5] W. Skarbek. Generalized Hilbert scan in image printing. In R. Klette and W. G. Kropatsch, editors, *Theoretical Foundations of Computer Vision*. Akademik Verlag, Berlin, 1992.
- [6] S. Derrode, G. Mercier, and W. Pieczynski, "Unsupervised change detection in SAR images using a multicomponent HMC model," in *MultiTemp'03*, (Ispra, Italy) (2003).
- [7] L. R. Rabiner, "A tutorial on HMMs and selected applications in speech recognition," *Proc. IEEE* 77(2), 257–286 (1989). [doi:10.1109/5.18626].
- [8] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.* 19(5), 465–475 (1997). [doi:10.1109/34.589206].
- [9] P.-A. Devijver, "Baum's Forward-Backward algorithm revisited," *Pattern Recogn. Lett.* 3,369–373 (1985).
- [10] L. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains," *Ann. Math. Stat.* 41(1), 164–171 (1970).
- [11] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. Roy. Stat. Soc.* 39(1), 1–38 (1977).
- [12] W. Pieczynski, "Statistical image segmentation," *Mach. Graph. and Vis.*, vol. 1, pp. 261–268, 1992.
- [13] L. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," in *Proc. IEEE*, vol. 77, no. 2, Feb. 1989, pp. 257–286.
- [14] P. A. Devijver, "Baum's forward-backward algorithm revisited," *Pattern Recognition Letters*, vol. 39, no. 3, pp. 369–373, december 1985.
- [15] S. Derrode, C. Carincotte, and J.M. Boucher. Unsupervised image segmentation based on high-order hidden Markov chains. In *IEEE ICASSP*, Montreal (Ca), 17-21 mai 2004.

## Fusion Based Gaussian noise Removal in the Images Using Curvelets and Wavelets With Gaussian Filter

**NAGA SRAVANTHI KOTA**

*Department of Electronics and Communication Engineering  
Sri Venkateswara University  
Tirupati - 517 502, India.*

*nagasravanthi.kota@gmail.com*

**G. UMAMAHESWARA REDDY**

*Department of Electronics and Communication Engineering  
Sri Venkateswara University  
Tirupati - 517 502, India*

*umaskit@gmail.com*

---

### Abstract

Denoising images using Curvelet transform approach has been widely used in many fields for its ability to obtain high quality images. Curvelet transform is superior to wavelet in the expression of image edge, such as geometry characteristic of curve, which has been already obtained good results in image denoising. However artifacts those appear in the result images of Curvelets approach prevent its application in some fields such as medical image. This paper puts forward a fusion based method using both Wavelets and Curvelet transforms because certain regions of the image have the ringing and radial stripe after Curvelets transform. The experimental results indicate that fusion method has an abroad future for eliminating the noise of images. The results of the algorithm applied to ultrasonic medical images indicate that the algorithm can be used efficiently in medical image fields also.

**Keywords:** Gaussian Filtering, Wavelet Transform, Curvelets Transform, Image Fusion, Denoising.

---

### 1. INTRODUCTION

Sparse representation of image data, where most of the information is packed into small number of data, is very important in many image processing applications. Image denoising using wavelets has been widely used in recent years and provides better accuracy in denoising different types of images. Wavelets are suitable for dealing with objects with point singularities. Wavelets can only capture limited directional information due to its poor orientation selectivity. By decomposing the image into a series of high-pass and low-pass filter bands, the wavelet transform extracts directional details that capture horizontal, vertical, and diagonal activity. However, these three linear directions are limiting and might not capture enough directional information in noisy images, such as medical Magnetic Resonance Images, which do not have strong horizontal, vertical, or diagonal directional elements. Wavelets provide a very sparse and efficient representation for piecewise smooth signals, but it cannot efficiently represent discontinuities along edges or curves in images or objects.

Ridgelet improves denoising; however, they capture structural information of an image based on multiple radial directions in the frequency domain. Line singularities in ridgelet transform provides better edge detection than its wavelet counterpart. One limitation to use ridgelet in image denoising is that ridgelet is most effective in detecting linear radial structures, which are not dominant in medical images. Due to the above-mentioned shortcoming of wavelet transform, Donoho and others proposed Curvelet transform theory and their anisotropy character is very useful for the efficient expression of image edges and gets good results in image denoising. The curvelet transform is a recent extension of ridgelet transform that overcome ridgelet weaknesses in medical image denoising. Curvelet is proven to be particularly effective at detecting image

activity along curves instead of radial directions which are the most comprising objects of medical images. However, Curvelet transform for denoising has also brought some negative impacts shows that it appears slightly “scratches” and “ringing” phenomenon in the reconstructed image, due to this situation, we have proposed fusion based method using images reconstructed from wavelet transform, curvelets transform and Gaussian filter for image denoising.

## 2. DENOISING METHODS

### 2.1 Gaussian Filtering

2D Gaussian filters are useful to provide image smoothing with minimal computations. Smoothing can reduce high frequency noise in an image while creating an image where a pixel and its neighbors are correlated with each other. Gaussian filters are able to smooth images with minimal computations because they are separable. This means that instead of using a 2D filter you can for example apply a 1D filter along the x-axis of the image and another 1D filter along the image's y-axis.

From a linear algebra perspective separable filters exploit the fact that the 2D Gaussian filter is really a rank one outer product of the two 1D filters. A 2D filter is only separable if it is aligned with the image axis and centered at the origin. In two-dimensions, one can vary a Gaussian in more parameters: not only may one vary a single width, but one may vary two separate widths, and rotate: one thus obtains both circular Gaussians and elliptical Gaussians, accordingly as the level sets are circles or ellipses. A particular example of Gaussian function is

$$f(x, y) = A e^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)} \quad (1)$$

Here the coefficient  $A$  is the amplitude,  $x_0, y_0$  is the center and  $\sigma_x, \sigma_y$  are the  $x$  and  $y$  spreads of the blob.

### 2.2 Wavelet Thresholding Using BayesShrink

Chang et al. proposed the BayesShrink scheme. In BayesShrink we determine the threshold for each Subband assuming a **Generalized Gaussian distribution (GGD)** [1]. The GGD is given by

$$GG_{\sigma_x, \beta}(x) = C(\sigma_x, \beta) \exp[-\alpha(\sigma_x, \beta)|x|^\beta] \quad (2)$$

$-\infty < x < \infty, \sigma_x > 0, \beta > 0$  Where

$$\alpha(\sigma_x, \beta) = \sigma_x^{-1} \left[ \frac{\Gamma(3/\beta)}{\Gamma(1/\beta)} \right]^{1/2} \quad (3)$$

$$C(\sigma_x, \beta) = \frac{\beta \cdot \alpha(\sigma_x, \beta)}{2\Gamma(1/\beta)} \text{ and}$$

$$\Gamma(t) = \int_0^\infty e^{-u} u^{t-1} du$$

The parameter  $\sigma_x$  is the standard deviation and  $\beta$  is the shape parameter.  $\beta$  ranges from 0.5 to 1. Assuming such a distribution for the wavelet coefficients, we empirically estimate  $\beta$  and  $\sigma_x$  for each subband and try to find the threshold  $T$  which minimizes the Bayesian Risk, i.e. the expected value of the mean square error.

$$\tau(T) = E(\hat{X} - X)^2 = E_X E_{Y/X}(\hat{X} - X)^2 \tag{4}$$

Where  $\hat{X} = \eta_T(Y), Y/X \sim N(x, \sigma^2)$  and  $X \sim GG_{\sigma_X, \beta}$  the optimal threshold  $T^*$  is then given by

$$T^*(\sigma_X, \beta) = \arg \min_T \tau(T) \tag{5}$$

This is a function of the parameters  $\sigma_X$  and  $\beta$ . since there is no closed form solution for  $T^*$ , numerical calculation is used to find its value. It is observed that the threshold value set by

$$T_B(\sigma_X) = \frac{\sigma^2}{\sigma_X} \tag{6}$$

is very close to  $T^*$

The estimated threshold  $T_B = \sigma^2 / \sigma_X$  is not only nearly optimal but also has an intuitive appeal.

When  $\sigma / \sigma_X \ll 1$ , the signal is much stronger than the noise,  $T_B / \sigma$  is chosen to be small in order to preserve most of the signal and remove some of the noise, when  $\sigma / \sigma_X \gg 1$ , the noise dominates and the normalized threshold is chosen to be large to remove the noise which has overwhelmed the signal. Thus, this threshold choice adapts to both the signal and the noise characteristics as reflected in the parameters  $\sigma$  and  $\sigma_X$

Parameter Estimation to determine the threshold: The GGD parameters,  $\sigma_X$  and  $\beta$  need to be estimated to compute  $T_B(\sigma_X)$ . The noise variance  $\sigma^2$  is estimated from the subband  $HH_1$  by the robust median estimator.

$$\hat{\sigma} = \frac{\text{Median}(|Y_{ij}|)}{0.6745}, Y_{ij} \in \text{subband HH1.}$$

The observation model is  $Y = X + V$ , with  $X$  and  $V$  independent of each other, hence

$$\sigma_Y^2 = \sigma_X^2 + \sigma^2 \tag{7}$$

With  $\sigma_Y^2$  is the variance of  $Y$ . since  $Y$  is modeled as zero-mean,  $\sigma_Y^2$  can be found empirically by

$$\hat{\sigma}_Y^2 = \frac{1}{n^2} \sum_{i,j=1}^n Y_{ij}^2 \tag{8}$$

Where  $(n \times n)$  is the size of the subband under consideration. Thus

$$\hat{T}_B(\hat{\sigma}_X) = \frac{\hat{\sigma}^2}{\hat{\sigma}_X} \tag{9}$$

Where

$$\hat{\sigma}_X = \sqrt{\max(\hat{\sigma}_Y^2 - \hat{\sigma}^2, 0)} \tag{10}$$

In the case that  $\hat{\sigma}^2 \geq \hat{\sigma}_Y^2$ ,  $\hat{\sigma}_X$  is taken to be zero i.e.  $\hat{T}_B(\hat{\sigma}_X)$  is  $\infty$  or in practice,  $\hat{T}_B(\hat{\sigma}_X) = \max(|Y_{ij}|)$  and all coefficients are set to zero.

### 2.3 Curvelet Transform

Curvelet transform is proposed by Candes and Donoho in 1999, its essence is derived from the ridge-wave theory [2]. In the foundation of single ridge-wave or local ridge-wave transform, we can construct Curvelet to express the objects which have curved singular boundary, Curvelet combines the advantages of ridge-wave which is suitable for expressing the lines' character and wavelet which is suitable for expressing the points' character and take full advantage of multi-scale analysis, it is suitable for a large class of image processing problems and has got quite good results in practical application [3]. Curvelet transform develops from 1999 until now, has transformed from the first generation (J.L.Starck, 2002) to the second generation theory (E.J.Candes, 2005) [4]. Based on this, research scholars realize the conversion process more simply and quickly, they develop another new algorithm. The first method is based on unequally-spaced fast Fourier transforms (USFFT) and the second is based on the wrapping of specially selected Fourier samples. The two implementations essentially differ by the choice of spatial grid used to translate Curvelets at each scale and angle. The 'wrapping' approach used in this paper assumes the same digital coronization as digital curvelet transform theory, but makes a different, somewhat simpler choice of spatial grid to translate Curvelets at each scale and angle[6]. Instead of a tilted grid, we assume a regular rectangular grid and define 'Cartesian' Curvelets in essentially the same way as before,

$$c(j, l, k) = \int \hat{f}(\omega) \tilde{U}_j(S_{\theta_1}^{-1} \omega) e^{i\langle b, \omega \rangle} d\omega \tag{11}$$

Discrete curvelet transform formula

$$c(j, l, k) = \int \hat{f}(\omega) \tilde{U}_j(S_{\theta_1}^{-1} \omega) e^{i\langle S_{\theta_1}^{-T} b, \omega \rangle} d\omega \tag{12}$$

Notice that  $S_{\theta_1}^{-T} b$  has been replaced by  $b \simeq (k_1 2^{-l}, k_2 2^{-\frac{l}{2}})$  taking on values on a rectangular grid. As before, this formula for b is understood When  $\theta \in \left(-\frac{\pi}{4}, \frac{\pi}{4}\right)$  or  $\theta \in \left(\frac{3\pi}{4}, \frac{5\pi}{4}\right)$ .

Wrapping round origin point is the core of Wrapping based curvelet. It realizes one to one through the periodization technology in the affine region. As shown in figure 1, Curvelet's essence is multi-scale localized ridge-wave, Curvelet transform is localized multi-scale ridge-wave transform, in fact. Curvelet's decomposition includes the following steps:

- 1) Sub-band decomposition: use wavelet transforms to the image and decomposes it to multiple sub-band components;
- 2) Smooth segmentation:" smooth segment" every subband to some sub-blocks, the size of the sub-blocks after each scale's division can be determined according to specific needs and can be different each other.
- 3) Ridge-wave analysis: make localized ridge-wave transform to each sub-blocks after segmentation.

Practice has proved that traditional Curvelet transform method has the phenomenon that it would appear slightly "scratches" and "ringing" in the image which is dialed with by denoising and reconstruction.

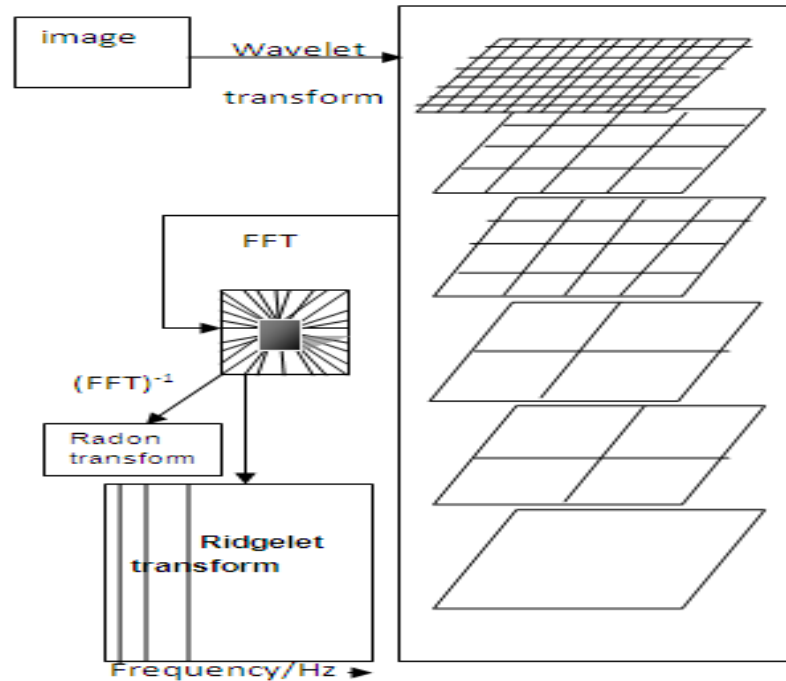


FIGURE1 .Curvelet transform

## 2.4 Proposed Method

The image contains a variety of areas, such as texture region, smooth region and so on; these different areas have different tolerances to the noise. If we use the same processing method with no difference, the consequences are filtering too much detailed information and damaging the detailed information of the edge or the edges are protected but to retain too much noise or it may cause a variety of distortions. The traditional Curvelet transform has the problem, has the phenomenon that it would appear slightly “scratches” and “ringing” in the image which is dialed with by denoising and reconstruction. To solve this problem, we can make the improvement, the steps are as follows:

- 1) Use Gaussian filtering on the noisy image and obtain the reconstructed image.
- 2) Use two-dimensional wavelet to denoise the original image which concludes noise.
- 3) Use Curvelet transform to denoise the original image which concludes noise.
- 4) Fusion the images which are processed by step (1),step(2) and step(3) , we can obtain better denoised image.

## 3. RESULTS AND DISCUSSIONS

This Denoising is carried out for the three standard gray scale images like Lena, house and boat of size 512x512, peppers colour image and also one MRI image of size 512x512 taken from the Visible Human Male data set of visible human project, released in November 1994. Then denoising is performed on these images corrupted with white Gaussian noise. To evaluate our method, we use the Peak Signal to Noise Ratio (*PSNR*), which is defined as:

$$PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right) \quad (13)$$

Where *MSE* is the Mean Square Error:

$$MSE = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} |x(m,n) - \hat{x}(m,n)|^2 \tag{14}$$

Where M and N are the number of lines and columns of the image,  $x(m, n)$  and  $\hat{x}(m, n)$  are the pixels of the original and the processed image. However, the PSNR do not correlate well with subjective quality evaluation. In fact, the HVS is sensitive to the noise on the uniform zones. Its perception on the textured zones is more difficult. To take in account this characteristic of the HVS, we use the weighted PSNR (*WPSNR*) that use the local variance of the image to ponder the error:

$$WPSNR = 10 \log_{10} \left( \frac{255^2}{wMSE} \right) \tag{15}$$

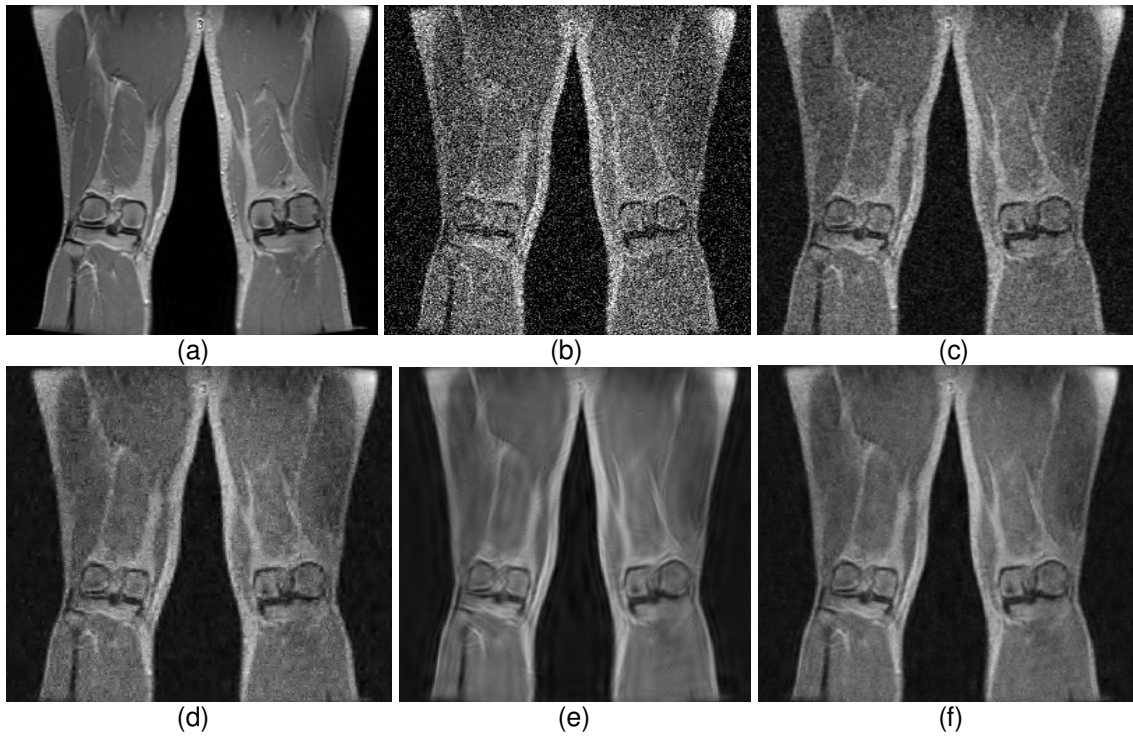
Where

$$WMSE = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \left| \frac{x(m,n) - \hat{x}(m,n)}{I + \text{var}(m,n)} \right|^2 \tag{16}$$

image	sigma(dB)	Gaussian filtering	Wavelet Transform denoising	Curvelet Transform denoising	fusion based denoising
MRI	30	34.2372	34.6475	34.7834	35.2639
"	50	30.0025	30.7335	30.7855	31.1393
LENA	30	36.0373	36.8384	37.5526	38.093
"	50	32.0341	33.5867	34.0649	34.6634
BOAT	30	35.488	35.8198	35.8387	36.6309
"	50	31.7195	32.4621	32.414	33.2787
HOUSE	30	36.3317	37.7527	38.6751	39.1076
"	50	32.2772	34.9115	35.4168	35.8162

**TABLE 1:** List of WPSNR values for images denoised from four denoising methods for four standard images corrupted by Gaussian noise of standard deviations  $\sigma = 30$  and  $\sigma = 50$





**FIGURE 2: MRI test image:** (a) Original image, (b) Noisy image with  $\sigma = 50$  dB, (c) Image after denoising with Gaussian filter, (d) Denoised image from BayesShrink Wavelet Thresholding, (e) Denoised image from Curvelet transform, (f) Denoised image from Proposed method.

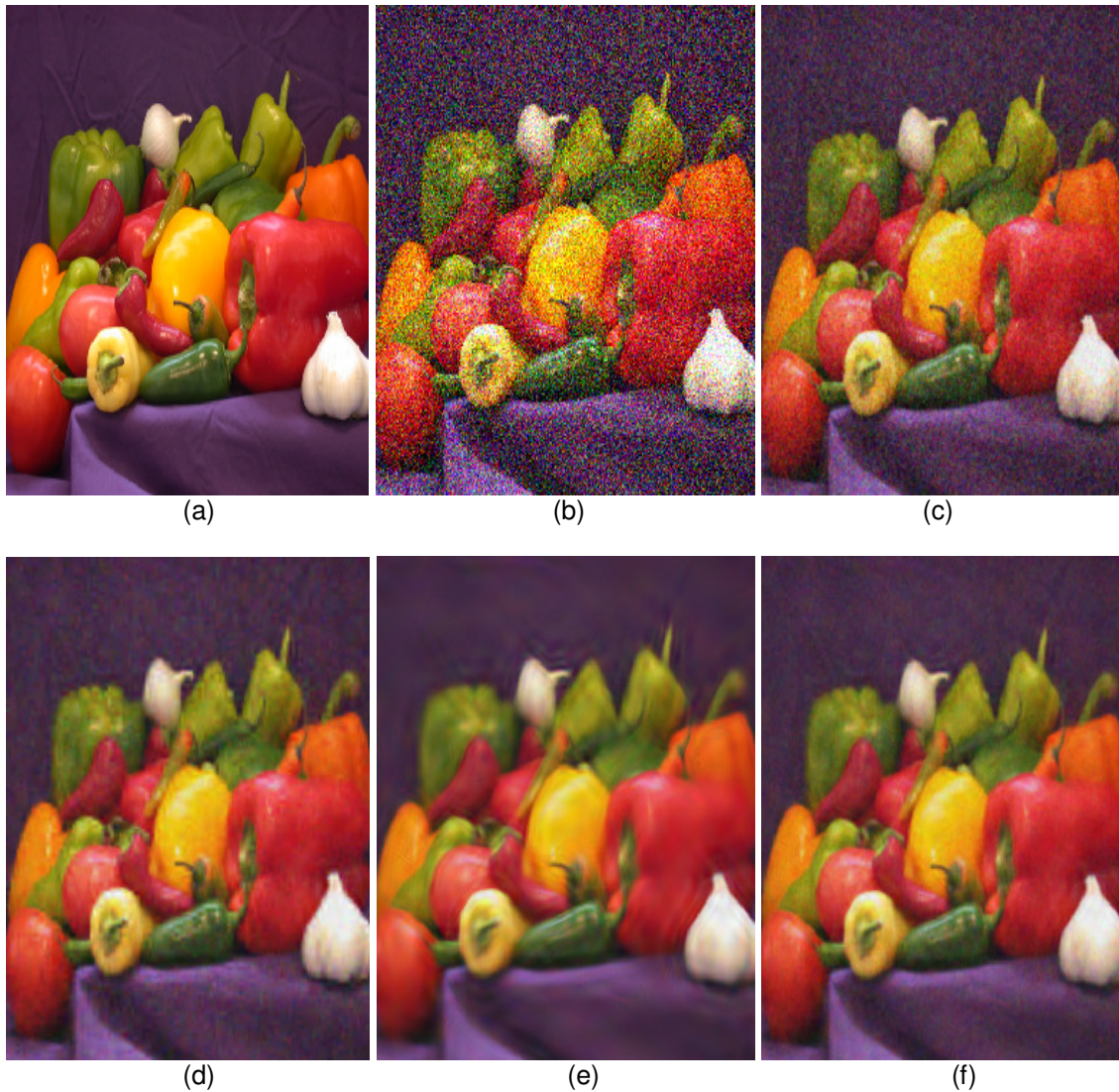


**FIGURE 3: GIRL test image:** (a) Original image, (b) Noisy image with  $\sigma = 50$  dB, (c) Image after denoising with Gaussian filter, (d) Denoised image from BayesShrink Wavelet Thresholding, (e) Denoised image from Curvelet transform, (f) Denoised image from Proposed method.





**FIGURE 4:** Lena colour test image: (a) Original image, (b) Noisy image with  $\sigma = 50$  dB, (c) Image after denoising with Gaussian filter , (d) Denoised image from BayesShrink Wavelet Thresholding, (e)Denoised image from Curvelet transform, (f)Denoised image from Proposed method.



**FIGURE 5: Peppers** colour test image: (a) Original image, (b) Noisy image with  $\sigma = 50$  dB, (c) Image after denoising with Gaussian filter , (d) Denoised image from BayesShrink Wavelet Thresholding, (e)Denoised image from Curvelet transform, (f)Denoised image from Proposed method.

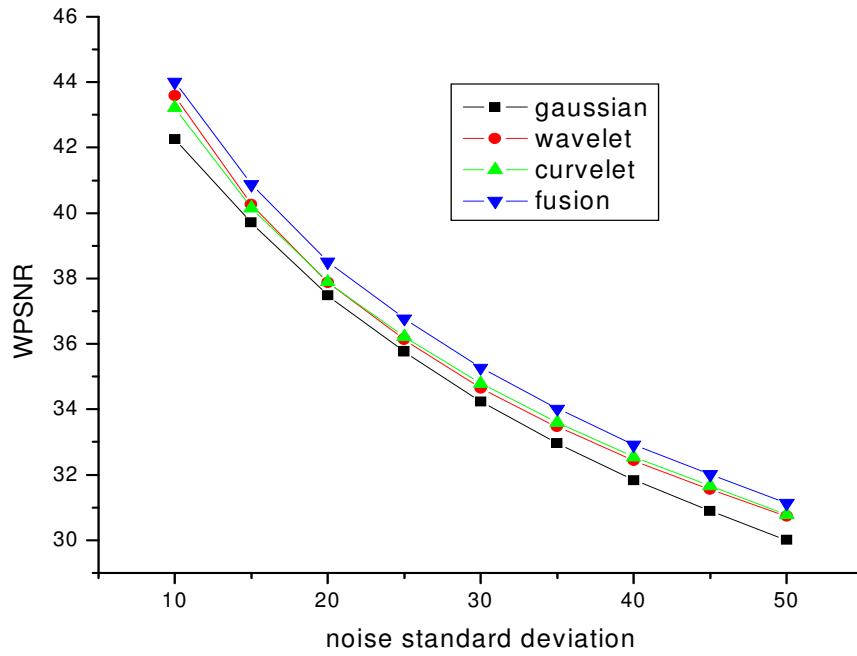


FIGURE 6: WPSNR versus noise standard deviation for four denoising methods on MRI test image

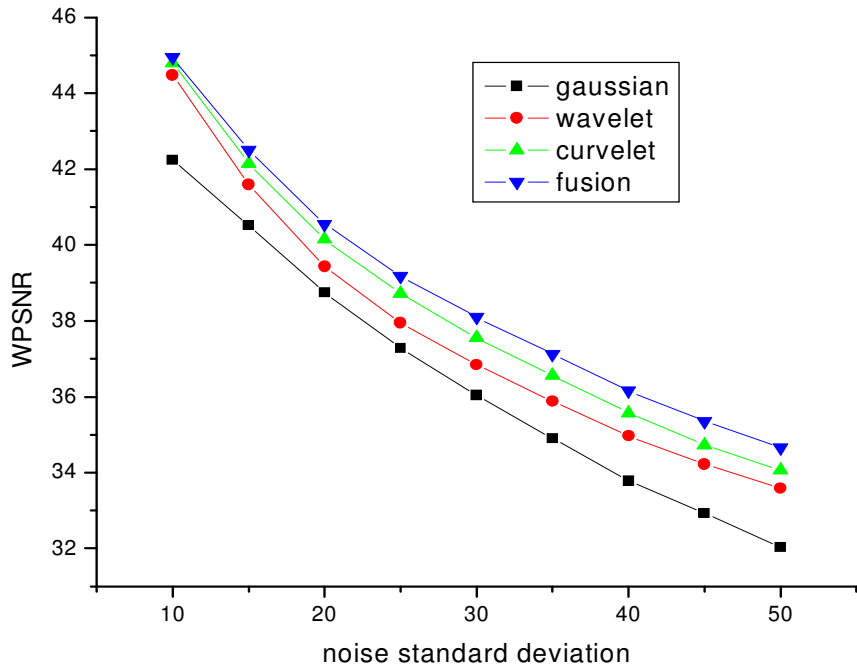
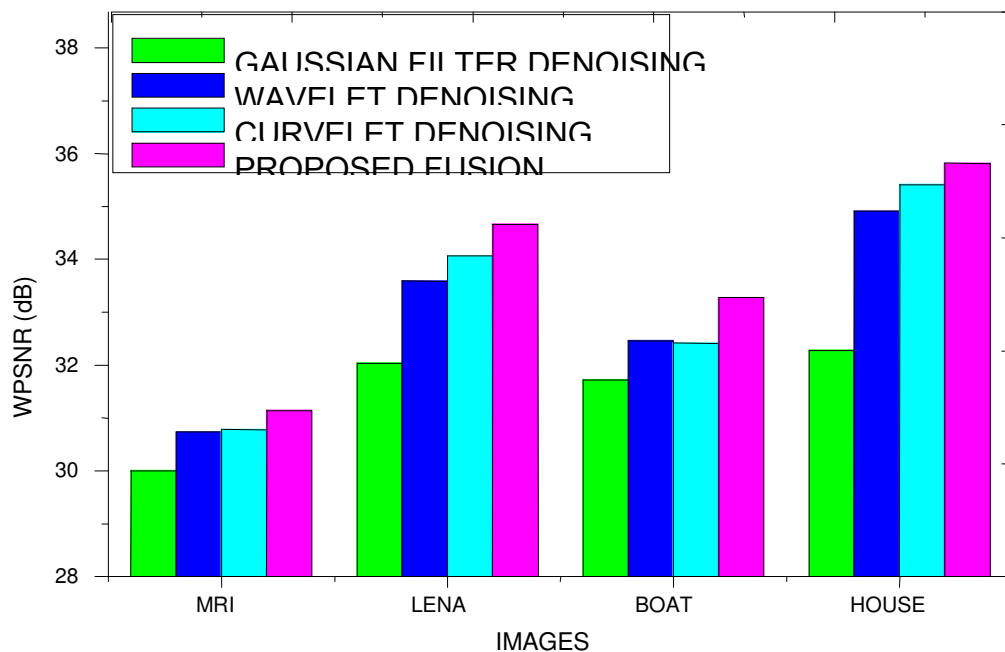


FIGURE 7: WPSNR versus noise standard deviation for four denoising methods on Lena test image



For the four standard images shown below, the comparison of denoising performance in terms of WPSNR is given in the Figure 8



**FIGURE 8:** Comparison of denoising performance (WPSNR) for four standard images using four denoising methods discussed in this paper.

Though wavelets are well suited to point singularities, they have limitations with orientation selectivity hence do not represent changing geometric features along edges effectively. Curvelet transform exhibits good reconstruction of the edge data by incorporating a directional component to the traditional wavelet transform.

From figure 2, figure 3 and figure 4, we can find out that image (c) is the smoothed version of noisy image after Gaussian filtering which retains most of the noise whereas image (d) lost partial edge information after wavelet denoising, the hat's outer edge and ribbon edge and soon become blurred, on the contrary, every edge of image (e) which is denoised by Curvelet transform maintains very well, but appears slight "ringing" phenomenon, appears slight "scratches" phenomenon in the face and hat body and some smooth area. Regardless of losing edge information and "scratches" and "ringing" phenomenon, we do not hope them to appear. Comparatively, image (f) denoised by the fusion of images (c), (d) and (e) we can not only maintain the edge information but also reduce "scratches" and "ringing" phenomenon. On the

whole, the fusion method obtained better denoising results than curvelet transform and two-dimensional wavelet transform individually.

#### 4. CONCLUSION

We tested four denoising methods such as Bayes thresholding based wavelet denoising, curvelet transform denoising, Gaussian filtering, and fusion based denoising. Though “scratches” and “ringing” phenomenon are appearing by using traditional Curvelet transform for denoising, we proposed the fusion based method using images after curvelet, wavelet transforms and Gaussian filtering. Our method applies two-dimensional wavelet transform, curvelet transform and Gaussian filtering. After two-dimensional wavelet transform, edge and detail information lost mostly. After curvelet transform, detail information retained better than wavelet transform, meanwhile in denoising process emerged the ringing and radial stripe. After Gaussian filtering image is smoothed to a great extent reducing high frequency noise. So thus with the fusion of three images obtained from the methods discussed we obtained better denoised image by preserving edges and removing Gaussian noise to a great extent and the experimental results prove that the proposed Fusion algorithm is also well suits for medical images as well.

#### 5. REFERENCES

- [1].M V S Grace Chang, Bin Yu. “Adaptive wavelet thresholding for image denoising and compression”, *IEEE Transactions on Image Processing*, Vol.9 (9), Sep 2000, pp.1532-1546.
- [2]. E.J.Candes, “Ridgelets: Theory and applications, Department of Statistics”, Stanford University, pp.1998, 23-38.
- [3].E.J.Candes,” Monoscale Ridgelets for the Representation of Image with Edges”, Department of Statistics, Stanford University, Stanford,CA,1999, pp.1-26.
- [4]. J.L. Startck, E.J. Candes, D.L. Donoho, “The curvelet transform for image denoising,” *IEEE Transactions on Image Processing*, Vol.11 (6), 2002, pp.670-684.
- [5]. E.J.Candes and D.Donoho, “Curvelet: a surprisingly effective non adaptive representation for object with edges”, Proceeding of Curves and Surfaces IV, France, 1999, pp.105-121.
- [6]. David L.Donoho & Mark R.Duncan,”Digital Curvelet Transform: Strategy, Implementation and Experiments”, Department of statistics, Stanford University, Nov 1999, pp.1-20.
- [7]. J.Fadili, J.L.Starck, “Curvelets and Ridgelets”, France, Oct 24, 2007, pp.1-30.
- [8]. E.Candes, L.Demanet, D.Donoho, L.Ying,”Applied and Computational Mathematics”, Department of Stanford university, Stanford, Mar 2006, pp.1-44.
- [9].Anil A.Patil, J.Singhai,” Image Denoising Using Curvelet Transform: an approach for edge preservation”, *Journal of scientific and Industrial Research*, vol.69, Jan 2010, pp.34-38.
- [10].S.G.Mallat,”A theory for Multiresolution signal decomposition: The wavelet representation”, *IEEE transactions on Pattern recognition*, vol.11, pp. 674-693, 1989.
- [11].J.M.Johnstone, D.L.Donoho,”Adapting to smoothness via wavelet shrinkage”, *Journal of the Statistical association*, 1995, vol.90, pp.1200-1224.
- [12].D.L.Donoho,”Denoising by soft thresholding”, *IEEE transactions on Information theory*, 1995, vol.41, pp.613-627.

- [13].D.Donoho and I.Johnstone,"Minimax estimation via wavelet shrinkage", *Annals of statistics*, 1998, vol.26, pp.879-921.
- [14].B.Vidakovic,"Statistical modeling by wavelets", John Wiley & Sons, NY, 1999.
- [15].I. Daubechies, "Wavelet transforms and orthonormal wavelet bases, different perspectives on wavelets," in *Proceedings of the Symposia in Applied Mathematics*, Vol. 47, pp. 1–33, American Mathematical Society,1993, San Antonio, Tex, USA.
- [16].D. L. Donoho, "Ridge functions and orthonormal ridgelets," *Journal of Approximation Theory*, 2001, Vol. 111, no. 2, pp. 143–179.
- [17].W. Fourati, F. Kammoun, and M. S. Bouhlel, "Medical image denoising using wavelet thresholding," *Journal of Testing and Evaluation*,2005, Vol. 33, no. 5, pp. 364–369.
- [18].M. Do and M. Vetterli, "Image denoising using orthonormal finite ridgelet transform," in *Wavelet Applications in Signal and Image Processing*,2003,vol. 4119 of *Proceedings of SPIE*, pp. 831–842.
- [19].L. Demanet, "The curvelet Organization," <http://www.curvelet.org/software.html>.

# An Efficient Multiplierless Transform Algorithm for Video Coding

## **Geetha. K. S**

*Associate Professor, Dept of E&CE  
R.V.College of Engineering,  
Bangalore, 560 059, India*

*geethakomandur@gmail.com*

## **M K Pushpa**

*Associate Professor, Dept of IT  
M.S.Ramaiah Institute of Technology,  
Bangalore,560 054, India*

*pushpachandan@rediffmail.com*

## **Dr.M.Uttarakumari**

*Professor, Dept of E&C  
R.V.College of Engineering,  
Bangalore,560 059, India*

*dr.uttarakumari@gmail.com*

## **Dr. S.Sethu Selvi**

*Professor & Head, Dept of E&C  
M.S.Ramaiah Institute of Technology,  
Bangalore,560 054, India*

*selvi@msrit.edu*

---

### **Abstract**

This paper presents an efficient algorithm to accelerate software video encoders/decoders by reducing the number of arithmetic operations for Discrete Cosine Transform (DCT). A multiplierless Ramanujan Ordered Number DCT (RDCT) is presented which computes the coefficients using shifts and addition operations only. The reduction in computational complexity has improved the performance of the video codec by almost 58% compared with the commonly used integer DCT. The results show that significant computation reduction can be achieved with negligible average peak signal-to-noise ratio (PSNR) degradation. The average structural similarity index matrix (SSIM) also ensures that the degradation due to the approximation is minimal.

**Keywords:** Ramanujan Ordered Number DCT, Multiplierless DCT, Video Coding.

---

## **1. INTRODUCTION**

Digital video applications have become more and more popular in our everyday life. Currently, there are several video standards, such as H.261 [1], H.263 [2], and MPEG [3][4], established for different applications. All these standards use motion compensated prediction, Discrete Cosine Transform (DCT), quantization, zigzag scan, and Variable Length Coding (VLC) as their basic functional blocks. Among these building blocks, Motion estimation (ME) in the motion compensated (MC) prediction is the most computationally intensive part, and then the DCT and the Inverse DCT (IDCT). Many fast algorithms have been developed to speed up the computation for Motion estimation. In this paper, an efficient technique is investigated to accelerate software video encoders by reducing the number of operations for DCT and quantization. The DCT and the quantization processes require a lot of multiplications, which are computationally expensive. A modification is proposed by replacing the 2-D DCT block in the standard MPEG-2 video codec with the 2-D Multiplierless Recursive DCT block. The performance is then compared with the existing DCT algorithms.

The organization of the paper is as follows. In Section 2, different blocks of video coding as in MPEG coder/decoder are explained, Section 3, explains the use of the multiplierless DCT coefficient computation that reduces the computation in the video encoder. In Section 4, the methodology of the proposed technique with the simulation results is discussed.

## 2. MPEG CODER/DECODER

The international standard [5, 6] describe a system, MPEG-2, for encoding and decoding digital video data. The standard allows for the encoding of video over a wide range of resolutions, including higher resolutions commonly known as HDTV.

In this system, encoded pictures are made up of pixels. If each  $8 \times 8$  array of pixels is known as a block, then an  $2 \times 2$  array of blocks is termed a macroblock. In this paper, an  $8 \times 8$  array of pixels is used as macroblock. Compression is achieved using the well known techniques of prediction (motion estimation in the encoder, motion compensation in the decoder), 2-D DCT, quantization of DCT coefficients, and Huffman/run(remove space) length coding. Pictures called I pictures are encoded without prediction and maintained as reference frames. Pictures termed P pictures may be encoded with prediction from previous pictures. B pictures may be encoded using prediction from both previous and subsequent pictures. A simplified MPEG-2 encoder and decoder is shown in Figure 1.

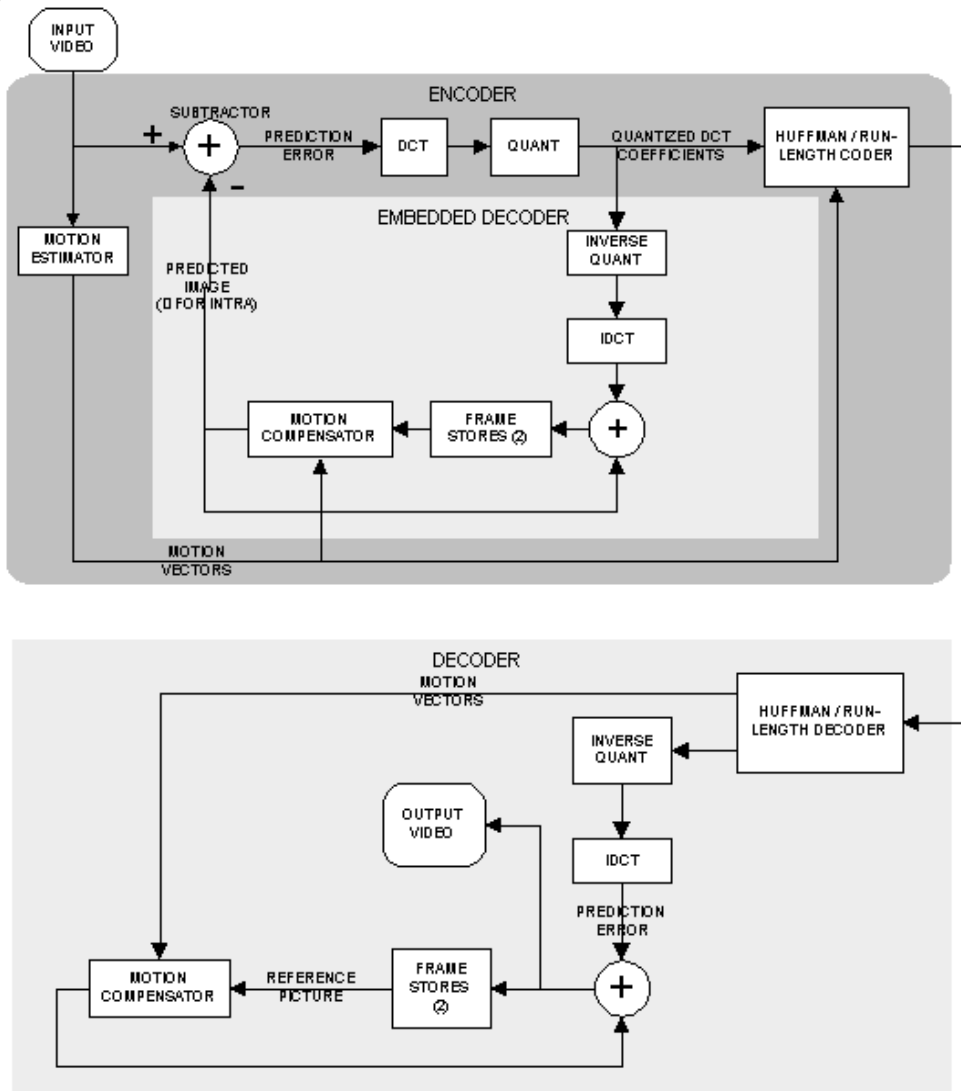


FIGURE1. MPEG-2 Encoder and Decoder

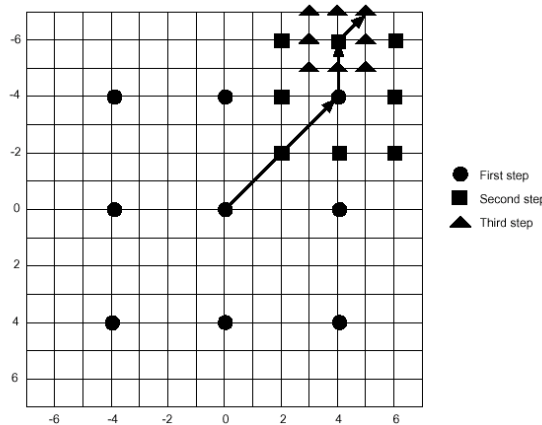


Before DCT is performed, motion compensated prediction is done for every macro block ( $8 \times 8$  pixels) on inter-coded frames. The objective of motion estimation is to find the best match of the current macro block within the search region in the reference frame. The common matching criterion used for finding the best match in the search region is the Mean Absolute Difference (MAD).

$$MAD = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} |C_{ij} - R_{ij}| \tag{1}$$

Where N is the size of the macro block,  $C_{ij}$  and  $R_{ij}$  are the pixels being compared in current macro block and the reference macro block respectively.

In motion compensated predictive coding, before performing the DCT computation, the Three Step Search algorithm [7, 8] is used to find the motion vectors. The best macroblock is found by using the MAD as a measure. The search algorithm is started with the search location at the centre of the macroblock as (0, 0). The step size is then fixed as  $S=4$ , and the search parameter as 7 for a macroblock of size  $8 \times 8$ . So, the search continues for the eight neighborhood pixels around location (0, 0). Out of these 9 locations, the pixel with the least cost function is then reinitiated as the new search origin and the step size is then reduced by half. So,  $S=S/2$ . The procedure is repeated until  $S=1$ . The pixel with the least cost function would then be the best match. The vector that represents the best match is saved.



**FIGURE 2:** Three Step Search Procedure. (Motion Vector is (5,7))

Each motion compensated macro block consists of four  $8 \times 8$  luminance and two  $8 \times 8$  chrominance prediction error blocks (difference blocks). These  $8 \times 8$  blocks are transformed to generate  $8 \times 8$  DCT coefficients and these coefficients are quantized for compression.

### 3. PROPOSED VIDEO CODEC

The DCT and the quantization processes require a lot of multiplications, which are computationally expensive. The standardized DCT block requires floating-point multipliers and for an  $8 \times 8$  block, evaluation of coefficients require 12 floating-point multipliers. The implementation of such a codec is more expensive as the complexity is concentrated towards the floating-point multipliers. This disadvantage is overcome by replacing the floating-point DCT block with a multiplierless DCT block where the coefficients are evaluated using Ramanujan ordered numbers. Computation of DCT coefficients involves evaluation of cosine angles of multiples of  $2\pi/N$ . Evaluation of these angles is accomplished by using a 4<sup>th</sup> degree polynomial that approximates the cosine function with error of approximation in the order of  $10^{-3}$  [13]. If N is

chosen such that it could be represented as  $2^{-l}+2^{-m}$ , where  $l$  and  $m$  are integers, then the trigonometric functions can be evaluated recursively by simple shift and addition operations. Such integers are called Ramanujan ordered numbers. Use of Ramanujan ordered Number for computing DCT was outlined by the author in [11,12]. Matrix factorization of the transformation matrix reduced the complexity to  $\frac{N}{2} \log_2 N$  shifts and  $\frac{3}{2}N \log_2 N - N + 1$  additions [12] thereby eliminating the use of multipliers.

**3.1 Multiplierless Ramanujan Ordered Number DCT(RDCT)[11,12]**

The 2-D Discrete Cosine Transform (DCT) can be defined as follows:

$$C(k_1, k_2) = \frac{4}{N_1 N_2} \sum_{n_1=0}^{N_1-1} \sum_{n_2=0}^{N_2-1} x(n_1, n_2) \cos\left(\pi k_1 \frac{(2n_1+1)}{2N_1}\right) \cos\left(\pi k_2 \frac{(2n_2+1)}{2N_2}\right) \quad (2)$$

Neglecting the scaling factors and using the property of Seperability, the DCT equation can be written as:

$$C(k_1, k_2) = \sum_{n_1=0}^{N_1-1} \left( \sum_{n_2=0}^{N_2-1} x(n_1, n_2) \cos\left(\pi k_2 \frac{(2n_2+1)}{2N_2}\right) \right) \cos\left(\pi k_1 \frac{(2n_1+1)}{2N_1}\right) \quad (3)$$

Thus, 2-D  $N_1 \times N_2$  DCT can be implemented by computing the row transformation followed by the column transformation. Hence, a 1-D transformation can be considered as a process of evaluating the sequences in the form as follows:

$$c_n = \sum_{n=0}^{N-1} x(n) \cos\left(\frac{2\pi}{N} 2^{-2} (2n+1) k\right) \quad (4)$$

**3.1.1 Evaluation of Transform Coefficients Using Chebyshev Recursion**

Computation of DCT coefficients requires evaluation of sequences of type

$$\{c_n \mid c_n = p \cos\left(\frac{2\pi n}{N}\right) n = 0, 1, 2 \dots N-1, p \in \Re\} \quad (5)$$

where  $\Re$  is the set of real numbers. These computations are done via a Chebyshev-type of recursion.

Let us define

$$W(M, p) = \{w_n \mid w_n = p \cos(2\pi n / M)\} \\ n = 0, 1, \dots, \Psi, \quad p \in \Re \quad (6) \\ \Psi = \left(\frac{M}{4} - 1\right), \quad M = \beta N$$

where,  $\beta$  is equal to 1, if N is divisible by 4. It is equal to 2, if N is divisible by 2, but not by 4. Otherwise, it is equal to 4(N is not divisible by 2). The use of  $\beta$  facilitates the computation of  $w(M, p)$  by considering cosine values from the first quadrant of the circle.

Let us then define

$$x = \frac{2\pi}{N} 2^{-2}$$

$$\therefore w_n = \cos((2n+1)x) \tag{7}$$

$x$  is then represented using Ramanujan ordered number of degree 2 as  $\hat{x} = 2^{-l} + 2^{-m}$  where  $l$  and  $m$  are non-negative integers.

For ex: If  $N=8$ , then

$$x = \frac{2\pi}{8} 2^{-2} \cong (2^{-1} + 2^{-2}) 2^{-2}$$

$$\hat{x} = 2^{-3} + 2^{-4} \tag{8}$$

$$\therefore w_n = \cos(n'\hat{x})$$

where  $n'$  is the scaled and shifted time samples and  $\hat{x}$  being the Ramanujan ordered number. Evaluation of these cosine values is by cosine approximation using 2<sup>nd</sup> order polynomial. Let the polynomial be defined as

$$\alpha = \frac{\hat{x}^2}{2!}$$

$$\therefore t_n(\alpha) = \cos(n\alpha) \tag{9}$$

$t_n(\alpha)$  are then computed using the recursive equations as

$$t_0(\alpha) = 1$$

$$t_1(\alpha) = (1 - \alpha)$$

$$\vdots$$

$$t_{n+1}(\alpha) = 2(1 - \alpha)t_n(\alpha) - t_{n-1}(\alpha)$$

$$n = 1, 2, \dots, (\Psi - 1) \tag{10}$$

It is observed that the above recursive equations are closely related to Chebyshev polynomial of the first kind. Since the evaluation of the recursive equations involve only numbers of powers of two,  $t_n(\alpha)$ 's and therefore  $c_n(\alpha)$ 's can be computed by simple shift and addition operations. RDCT kernel needs samples only at  $(2n+1)$ , and thus all the samples of  $t_n(\alpha)$  need not be stored.

TABLE I. COMPARISON OF COMPUTATIONAL COMPLEXITY

Operations	Floating-point DCT $N \times M$ [9]	Integer DCT $N \times M$ [10]	RDCT $N \times M$ [11]
Multiplications	$(NM/2) \log_2 M$ (Floating-point)	$NM$ (Integer)	Nil
Additions	$(3NM/2) \log_2 M$ $+ (-2NM + N + M)$	$2(\log_2 NM - 1)$ $+ NM + 2$	$(3NM/2) \log_2 M$ $- 2NM + N + M$
Lifting Steps	Nil	$(3N/2) \log_2 N$ $- 3N + 3$	Nil
Shifts	Nil	Nil	$(NM/2) \log_2 M$

Table I gives the comparison of the reduction of the computational complexity of the proposed algorithm. To compute  $N \times M$  DCT the proposed algorithm takes  $(3NM/2)\log_2 M - 2NM + N + M$  additions and  $NM/2\log_2 M$  shift operations. Thus for  $N=M$ , the proposed algorithm for a  $8 \times 8$  block DCT evaluation, requires 96 shift operation, and 176 addition operations. The proposed algorithm being recursive ensures that the storage of the trigonometric values is not required.

**4. SIMULATION RESULTS**

To demonstrate the efficacy of the proposed algorithm on MPEG Video codec, the results were compared with the existing algorithm of the standard MPEG-2 video codec and the results are tabulated. The proposed RDCT is tested by replacing the two-dimensional DCT block in the MPEG-2 standard algorithm with the 2-D RDCT block. The performance is then compared by using commonly used multiplierless 2-D Integer DCT. --. DC coefficient is quantized and coded separately and transmitted. The AC coefficients are encoded with very few coefficients removing the completely zero coefficients block.

Table.II gives the average PSNR of the original frame with decoded frame, using 60 frames of input video sequence (video grabbed at 30fps), with a GOP (group of pictures) as 10 and the encoding format as  $I_1P_4B_2B_3P_7B_5B_6I_{10}B_7B_8$ . The step size is considered as 10 to decode all 10 frames in the display format as  $I_1B_2B_3P_4B_5B_6P_7B_8B_9I_{10}$ . The simulation has been evaluated for both forward and bidirectional prediction and the results shows that the motion estimation in both the formats gives better results for the proposed RDCT when compared with the floating-point DCT and the integer DCT. From Table II it is clear that the proposed RDCT offers same accuracy in average PSNR as that of the floating-point DCT with a deviation of 0.01%, and the deviation with Integer DCT is by 0.01% for standard test sequence like Alex.avi. The deviation in PSNR of the RDCT with floating-point DCT is 0.005% and with integer DCT is 0.08% for real time data sequence. This clearly shows that the proposed technique of using RDCT for the video codec is providing better reconstructed picture quality.

TABLE II AVERAGE PSNR IN dB OF THE DECODED FRAMES

Test Sequence	Frame Format	Floating-point DCT[9]	Integer DCT [10]	Multiplierless RDCT
Real time Data (Frames grabbed at 30 fps)	IBBPBBPBBP	35.7010	35.6694	35.6991
	IPPPPPPPPP	33.6581	33.6132	33.6525
San_Fran_Traffic	IBBPBBPBBP	34.8076	34.7776	34.7996
	IPPPPPPPPP	31.7476	31.7176	31.7462
Alex	IBBPBBPBBP	36.0928	36.0809	36.0876
	IPPPPPPPPP	31.583	31.5756	31.579

The Structural Similarity Index Matrix (SSIM) index seeks to separately discover differences in local image luminance  $l(x,y)$ , contrast  $c(x,y)$  and structure  $s(x,y)$  between the original and compensated images. Given the pixel points  $(x,y)$ , the SSIM is defined as

$$SSIM(x, y) = l(x, y) \cdot c(x, y) \cdot S(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \cdot \frac{\sigma_{xy} + C_3}{\sigma_x + \sigma_y + C_3} \tag{11}$$

where  $\mu_x, \mu_y, \sigma_x, \sigma_y$  and  $\sigma_{xy}$  are the local sample means, variances, and cross-covariance of  $x$  and  $y$ . The constants  $C_1, C_2, C_3$  stabilize SSIM when the means and variances become small.

SSIM index varies between 0(worst) and 1(best). Table III shows the average SSIM for decoded frames with original frames.

Table III AVERAGE SSIM BETWEEN THE DECODED AND ORIGINAL FRAMES

Test Sequence	Frame Format	Floating-point DCT[9]	Integer DCT [10]	Multiplierless RDCT
Real time Data (Frames grabbed at 30 fps)	IBBPBBPBBP	0.9223	0.9218	0.9223
	IPPPPPPPPP	0.921645	0.921628	0.921635
San_Fran_Traffic	IBBPBBPBBP	0.85028	0.85020	0.85026
	IPPPPPPPPP	0.85701	0.85014	0.85693
Alex	IBBPBBPBBP	0.8689	0.8684	0.8690
	IPPPPPPPPP	0.8678	0.8668	0.8682

From Table III, it is clear that the quality of decoding is very good with RDCT and achieves the same performance as that of the floating-point DCT. This is ensured by taking the difference frame between the reference frame and the decoded frame. The difference frame is as shown in the Figure 3a and 3b. The difference between the RDCT and the floating-point DCT in terms of SSIM is 0.01% for standard test sequence like Alex.avi and the difference between the RDCT and the Integer DCT in terms of SSIM is 0.07% for the same test sequence. For the real time data the difference between RDCT and floating-point DCT is 0 in terms of SSIM and between RDCT and integer DCT is 0.05% in terms of SSIM. These values clearly indicate that the reconstructed frame with proposed RDCT is very good in subjective quality when compared with the reconstructed frame with Integer DCT.

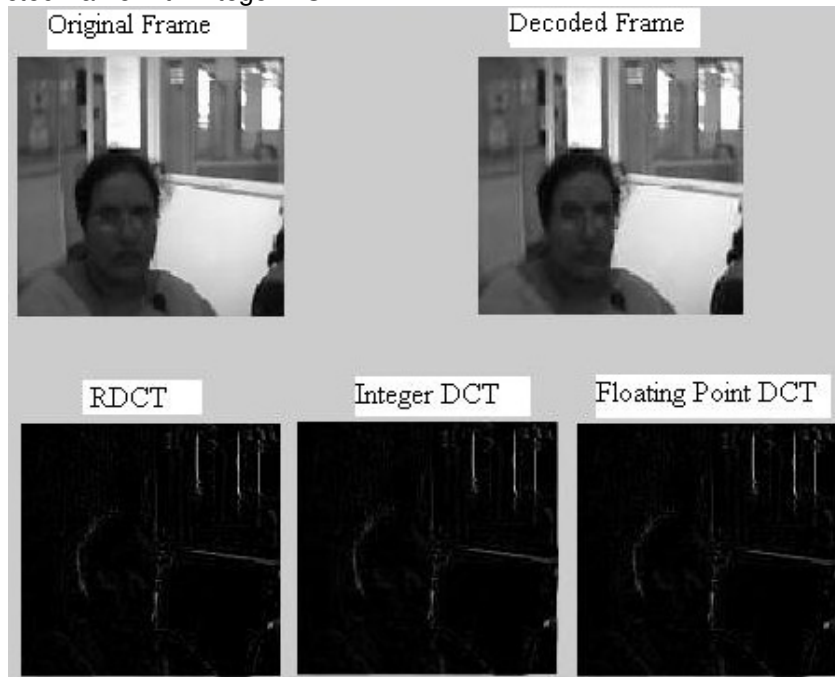
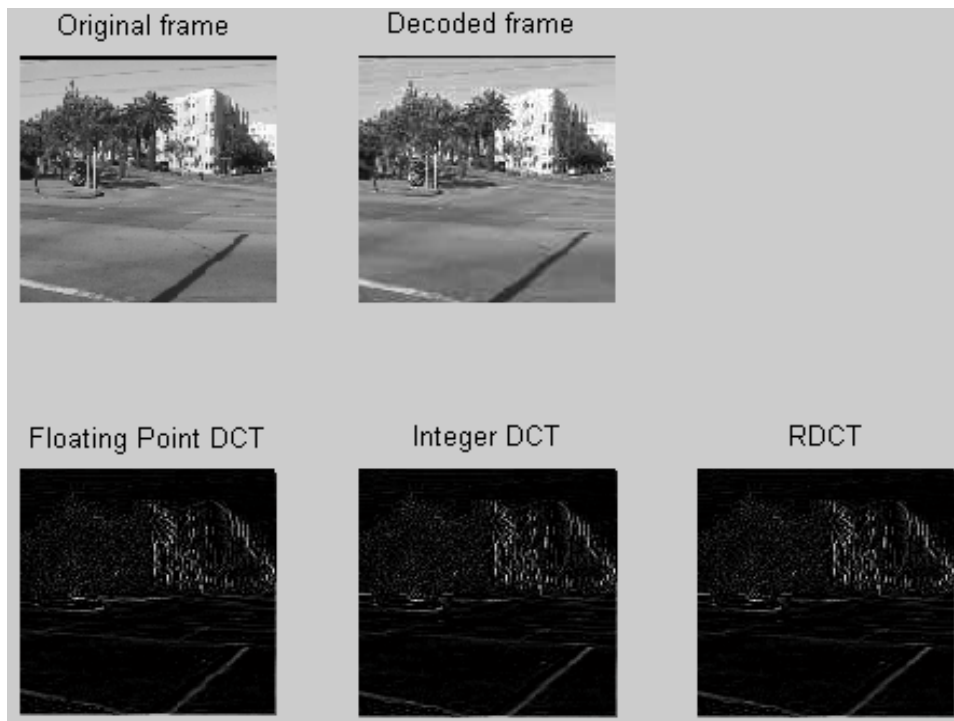


FIGURE 3a Difference between original and decoded frame (real time sequence)



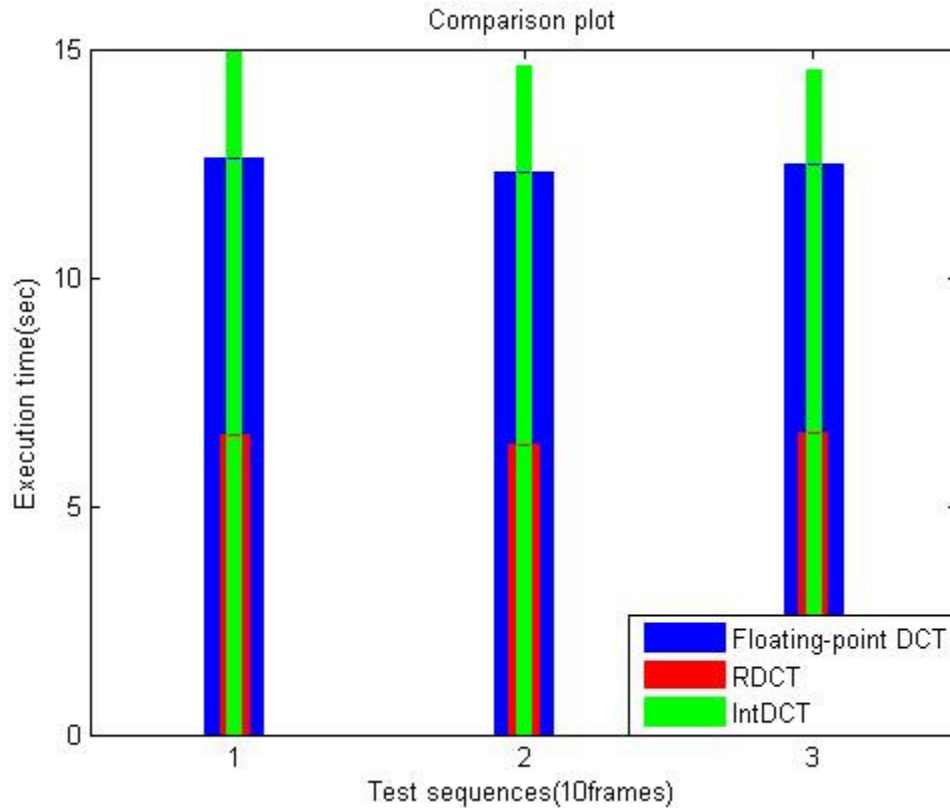
**FIGURE 3B** Difference between original and decoded frame (San\_Fran\_Traffic.avi)

Table IV shows the comparison of the computation time for decoding I reference frame and decoding 60 frames, with different algorithms with a GOP of 10 frames. The computation was performed on a Intel Core 2 Duo Processor, @ 1.80 GHz.

TABLE IV DECODING TIME IN SECONDS

Test Sequence	Decoding frame	Floating-point DCT	Integer DCT	Multiplierless RDCT
Real time Data (Frames grabbed at 30 fps)	Reference frame	0.191	0.313	0.125
	10 frames	12.609	14.953	6.562
San_Fran_Traffic	Reference frame	0.296	0.308	0.109
	10 frames	12.322	14.641	6.335
Alex	Reference frame	0.245	0.325	0.125
	10 frames	12.484	14.547	6.593

Table IV shows that the proposed RDCT has reduction in decoding time for 10 frames by 47.9578% when compared with the floating-point DCT whereas it has an improvement of 56.1158% over the commonly used integer DCT for a real time data sequence. However, the reduction in the time is 47.1884% when compared with the floating-point DCT whereas it has an improvement of 54.6779% over integer DCT for a standard data sequence like Alex.avi.



**FIGURE 4.** Comparison plot for sequences 'Real-time sequence', 'San\_Fran\_Traffic' & 'Alex'

Fig 4 gives us better comparison in terms of the execution times for decoding 10 frames using different algorithms namely RDCT, floating-point DCT and the IntDCT. The plot clearly shows the RDCT outperforms the floating-point DCT and the IntDCT. This improvement in the decoding time is due to the improvement in the computational complexity of the DCT algorithm.

## 5. CONCLUSION

The computationally less complex video coding technique is presented in this paper using multiplierless Ramanujan ordered DCT. This method allows us to evaluate the cosine function using only integers which are powers of 2 thereby replaces the complex floating-point multiplications by shifters and adders. This algorithm takes  $N/2 \log_2 N$  shifts and  $(3N/2 \log_2 N) - N + 1$  addition operations to evaluate an N-point DCT. The cosine approximation increases the overhead on the number of adders by 13.6% but totally avoids floating point multiplications. The reduction in complexity is reflected in the time required for the decoding of video frames. There is an improvement of 58% from the existing commonly used Integer DCT video codec. The average SSIM and average PSNR values indicate that the quality of decoding using the RDCT is same as that of the Integer DCT. Hence, the proposed algorithm is an efficient multiplierless transform for video coding that offers less computationally complexity but assures the same quality as that of the existing algorithms.

## 6. REFERENCES

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of ITU-T Recommendation H.261, "Video codecs for audiovisual services at p x 64 kb/s," Mar. 1993.
- [2] ITU-T Recommendation H.263, "Video coding for low bitrate communication," Mar. 1996.

- [3] ISO/IEC 11 172-2, "Information technology - coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s: Part 2 Video," Aug. 1993.
- [4] ITU-T Recommendation H.262 | ISO/IEC 13818-2, "Information technology - generic coding of moving pictures and associated audio information: video," 1995.
- [5] ISO/IEC 13818-2 "Generic Coding of Moving Pictures and Associated Audio Information: Video",
- [6] ATSC document A/54 "Guide to the Use of the ATSC Digital Television Standard"
- [7] Renxiang Li, Bing Zeng and Ming I.Liou, "A New Three-Step Search Algorithm for Block Motion Estimation", IEEE Trans. Circuits And Systems For Video Technology, Vol.4, No.4, pp. 438-442, Aug 1994.
- [8] Aroh Barjatya, "Block Matching Algorithms For Motion Estimation" , DIP 6620 Spring 2004 Final Project Paper.
- [9] H.S. Hou, "A Fast Recursive Algorithms for Computing the Discrete Cosine Transform". IEEE Trans. Acoust., Speech, Signal Processing, Vol.35, pp 1455-1461, Oct 1987.
- [10] Yonghong Zeng, Lizhi Cheng, Guoan Bi, and Alex C. Kot, "Integer DCT's and Fast Algorithms", IEEE Signal Proc.141-14 (2000).
- [11] Geetha.K.S, V.K.Ananthashayana, "A Novel Recursive Multiplierless Algorithm for 2-D DCT",Proc. ICSPCN 2009,Aug 2009.
- [12] Geetha.K.S, M.Uttarakumari, "Multiplierless Recursive algorithm using Ramanujan ordered Numbers," in IETE Journal of Research, vol. 56, Issue 4, JUL-AUG 2010.
- [13] Geetha.K.S, M.Uttarakumari, "A Novel Cosine approximation for high-speed evaluation of DCT" International Journal of Image Processing, CSC Journals Volume: 4 Issue: 6 Pg 539 – 548 Jan-Feb 2011.



# Video Audio Interface for Recognizing Gestures of Indian Sign Language

**P.V.V.Kishore**

*Research Scholar/ECE  
Andhra University College of Engineering  
Visakhapatnam, 530003, INDIA*

*pvvkishore@gmail.com*

**P.Rajesh Kumar**

*Associate Professor/ECE  
Andhra University College of Engineering  
Visakhapatnam, 530003, INDIA*

*rajeshauce@gmail.com*

**E.Kiran Kumar**

*Postgraduate Student/ECE  
Dadi Institute of Engineering and Technology  
Visakhapatnam, 530017, INDIA*

*kiraneepuri@gmail.com*

**S.R.C.Kishore**

*Post Graduate Student/ECE  
Pydah College of Engineering  
Visakhapatnam, 530037, INDIA*

*ravickishore@gmail.com*

---

## Abstract

We proposed a system to automatically recognize gestures of sign language from a video stream of the signer. The developed system converts words and sentences of Indian sign language into voice and text in English. We have used the power of image processing techniques and artificial intelligence techniques to achieve the objective. To accomplish the task we used powerful image processing techniques such as frame differencing based tracking, edge detection, wavelet transform, image fusion techniques to segment shapes in our videos. It also uses Elliptical Fourier descriptors for shape feature extraction and principal component analysis for feature set optimization and reduction. Database of extracted features are compared with input video of the signer using a trained fuzzy inference system. The proposed system converts gestures into a text and voice message with 91 percent accuracy. The training and testing of the system is done using gestures from Indian Sign Language (INSL). Around 80 gestures from 10 different signers are used. The entire system was developed in a user friendly environment by creating a graphical user interface in MATLAB. The system is robust and can be trained for new gestures using GUI.

**Keywords:** Sign language, Edge Detection, Wavelet Transform, Image Fusion, Elliptical Fourier Descriptors, Principle Component Analysis, Fuzzy Inference System.

---

## 1. INTRODUCTION

Sign language is seen as the core distinctive feature that defines a deaf community. The role of sign language recognition systems in the society is to ensure that deaf people have equality of opportunity and full participation in society. Sign language is represented basically by continuously varying different hand shapes and movements by a signer. Sign language recognition is decoding and understanding the information embedded in the hand shapes and converting them to meaning full words. Sign language is a natural language of the deaf community. By developing sign language recognition system a hearing impaired person can easily interact with a normal person at different levels in the society. We have attempted to design a basic system of sign language pertaining to Indian Sign Language.

Going back to early days of sign language Stokoe et.al [1] showed that signs are made of basic articulatory units initially referred to as cheremes commonly referred in these times as phonemes in resemblance with that to words in spoken language.

Logically sign language understanding consists of linguistic analysis of hands tracking, hands shapes, hands orientations, sign verbalization and also with important linguistic information communicated with head movements and facial expressions. Sign language is in many ways different from spoken language such as facial and hand terminology, references in virtual signing space, and grammatical differences as explained.

The major difficulty in sign language recognition compared to speech recognition is to recognize simultaneously [2] different communication attributes of a signer such as hands and head movement, facial expressions and body pose. All these attributes has to be considered simultaneously for a good recognition system.

The second major problem faced by sign language recognition system designers is tracking the signer in the clutter of other information available in the video. This is addressed by many researchers as signing space [3]. A sign language space can be created with entities such as humans or objects stored in it around a 3D body centered space of the signer. The entities are executed at a certain location and later referenced by pointing to the space. A major challenge faced by researchers to define a model for spatial information containing the entities created during the sign language dialogue.

Additional difficulties arise in the form of background in which signer is located. Most of the methods developed so far use simple backgrounds in controlled set-up such as simple backgrounds, special hardware like data gloves, restricted sets of actions, restricted number of signers, resulting different problems in sign language feature extraction.

For adapting the system for signer independent as in case of automatic speech recognitions systems which are robust being able to cope with different dialects. Speaker adaptation techniques known for speech recognition can be used to make the system more robust. While for recognizing signs of few signers only the intrapersonal variabilities in appearance and velocity of their hands needed to be modeled. As the number of signers increases, the quantity and diversity of variabilities is extremely increased.

In continuous speech recognition as well as continuous sign language recognition, co articulation effects have to be considered. Owing to location changes in virtual signing space, we have to encounter movement epenthesis problem [3,4]. Movement epenthesis refers to movements which occur regularly in natural sign language in order to change the location in signing space.

In automatic speech recognition the energy of the audio signal is used to detect silence in the sentences. New features and models have to be defined for silence detection in sign language recognition. The simplest way to do this is by motion analysis in the video with no reliability, because words are signed by just holding a particular posture in the signing space. In order to automate the sign language recognition system there should be a consistent silence detection and sentence boundaries adding to speed and recognition performance.

Sign language similar to spoken language is not confined to a particular region or territory. It is practiced differently around the world. In USA it is known as American Sign Language (ASL) [5, 6] and in Europe [7, 8], Africa [9] and in Asia [10, 11]. Unlike America and Europe sign language, India does not have a standard sign language with necessary variations, though. But recent past Ramakrishna missions Vivekananda University, Coimbatore came up with an Indian sign language dictionary. There are around 2037 signs [12] currently available in sign language.

Accordingly sign language recognition systems are classified in to two broad categories: sensor glove based [13] and vision based systems [14, 15]. The first category requires signers to wear a sensor glove or a colored glove. The wearing of the glove simplifies the task of segmentation

during processing. Glove based methods suffer from drawbacks such as the signer has to wear the sensor hardware along with the glove during the operation of the system. In comparison, vision based systems use image processing algorithms to detect and track hand signs as well as facial expressions of the signer, which is easier to the signer without wearing gloves. However, there are accuracy problems related to image processing algorithms which are a dynamic research area.

Thad sterner proposed a real time American Sign Language recognition system using wearable computer based video [14] which uses hidden markov model (HMM) for recognizing continuous American Sign Language system. Signs are modeled with four states of HMMs which have good recognition accuracies. Their system works well but it is not signer independent. M.K.Bhuyan [16] used hand shapes and hand trajectories to recognize static and dynamic hand signs from Indian sign language. The used the concept of object based video abstraction technique for segmenting the frames into video object planes where hand is considered as a video object. Their experimental results show that their system can classify and recognize static, dynamic gestures along with sentences with superior consistency.

Rini Akmelia et.al [17] proposed a real time Malaysian sign language translation using color segmentation technique which has a recognition rate of 90%. Nariman Habili et.al [18] proposed a hand and face segmentation technique using color and motion cues for the content based representation of sign language video sequences. Yu Zhou, Xilin chen [19] proposed a signer adaptation method which combines maximum a posteriori and iterative vector field smoothing to reduce the amount of data and they have achieved good recognition rates.

Gesture recognition systems that are implemented statistical approaches [20], example based approaches [21], finite state transducers [22] have registered higher translation rate. In the last decade there were more efficient algorithms for training [23] and more efficient algorithms for generation [24] with the advancement of powerful computers and bigger parallel corpora.

The Essential Sign Language Information on Government Networks (eSIGN Project) [25] is a European project that has contributed in a large way in developing tools for automatic creation of sign language contents. The important breakthrough in this project has been 3D avatar (VGuido) to represent signs in the sign language resourcefully. The project also consists of visual settings for creating sign animations in rapid and easy way translating web content into sign language. This translation of web content in to sign language helps deaf people significantly in using the internet. Currently this project is translating websites in Germany, UK and the Netherlands.

Recently researchers are developing systems that can translate speech to sign language which require a parallel corpus to be able to train the language and translation models, and to evaluate the systems. The European Cultural Heritage Online Organization (ECHO) presents a corpus in British and Netherlands sign languages [26]. Another was made by The American Sign Language Linguistic Research Group at Boston University [5].

In this paper various image processing and artificial intelligence techniques are combined to generate a fully automated sign language recognition system. This system converts videos of signs by different signers into text and voice messages. The system can be trained to handle new signs at any point during operation. A graphical user interface (GUI) has been developed to make the system more user friendly. We named our system Indian Sign Language Recognition system (INSLR).



















The paper is organized as follows: in Section 2, creation of database representation of Indian Sign language. Section 3, presents an overview of the system along with different stages of design. Section 4, analysis of results and performance of the system under various conditions and finally section 5 summarizes the main conclusions.

## 2. DATABASE REPRESENTATIONS OF INDIAN SIGN LANGUAGE

In order to approach the problem of translating signs in to speech it is necessary to create a database of videos of different signs by multiple signers. Unlike American Sign Language or British sign language, Indian sign language does not have a standard database that is available for use. Hence we have created our own video database of Indian signs in collaboration with Indian Deaf Society [12] and shanthi ashram school for deaf children. Initially we have created 80 signs of alphabets, numbers, words and sentences by multiple signers of Indian sings.

The experimental setup consists of a dark background and the signer wearing a dark jacket with full sleeves. This controlled environment reduces tracking and segmentation problems. The RGB videos are acquired using a Sony cyber shot H7 digital cam coder having a resolution of 640×480 pixels capturing 25 frames per second. Higher resolution incorporates longer delays in the video acquisition process and higher execution times. The videos are acquired under normal lighting conditions to simulate real-time environment. Natural lightning of around 15-20 lux was present during the video acquisition process.

A total of nine different signers volunteered where each signer is asked to repeat the sign twice under different conditions with a total number of 1440 gesture videos for a total of 80 signs. A video acquisition process is subjected to many environmental conditions such as position of the camera, illumination and noise. Some of the signs in our data base are presented in TABLE 1. The only constraint of our system is that the signer must wear a dark colored full sleeves jacket with a dark background.

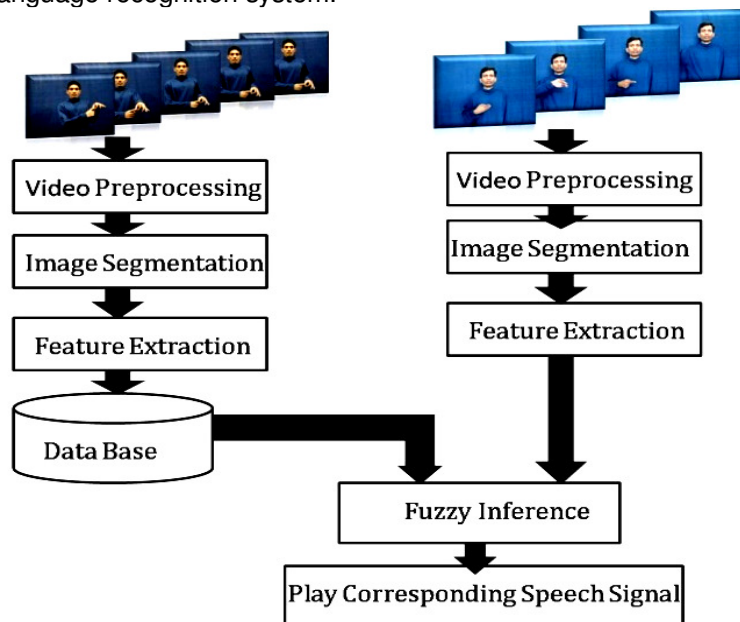
Sl.No	Frame#10	Frame#20	Frame#30	English Representation of the Sign
1				I DON'T WANT TO GO
2				COW
3				UPWARDS
4				SMALL
5				CROW
6				DUCK

7				TEN
8				PEACOCK
9				FAT

**TABLE 1:** Shows different signs in our database displaying three Frames per video

### 3. DESIGN OF PROPOSED SYSTEM

The system design is based on four broad issues related as video preprocessing, image segmentation, feature extraction and pattern classification. Figure 1 shows the framework of our proposed sign language recognition system.



**FIGURE 1:** Framework of sign language recognition system

From the figure 1 we can understand the overall working process of the system. Video of signer is preprocessed using different techniques to avoid any false segmentation in the next stage. Image segmentation stage divides the entire frames of images of the signers into hands and head segments of the signer. Shape features are extracted and are optimized before saving to the database. This process is repeated for all the available signs using a video of a signer. The sign language database can be upgraded to add new signs into the database.

The system can now be tested with the signs present in the database using fuzzy inference system based on rules. The rule based system is a very powerful tool used by many researchers for pattern classification tasks. Once trained the fuzzy inference system produces a voice and text response corresponding to the sign irrespective of the signer and changing lighting conditions.

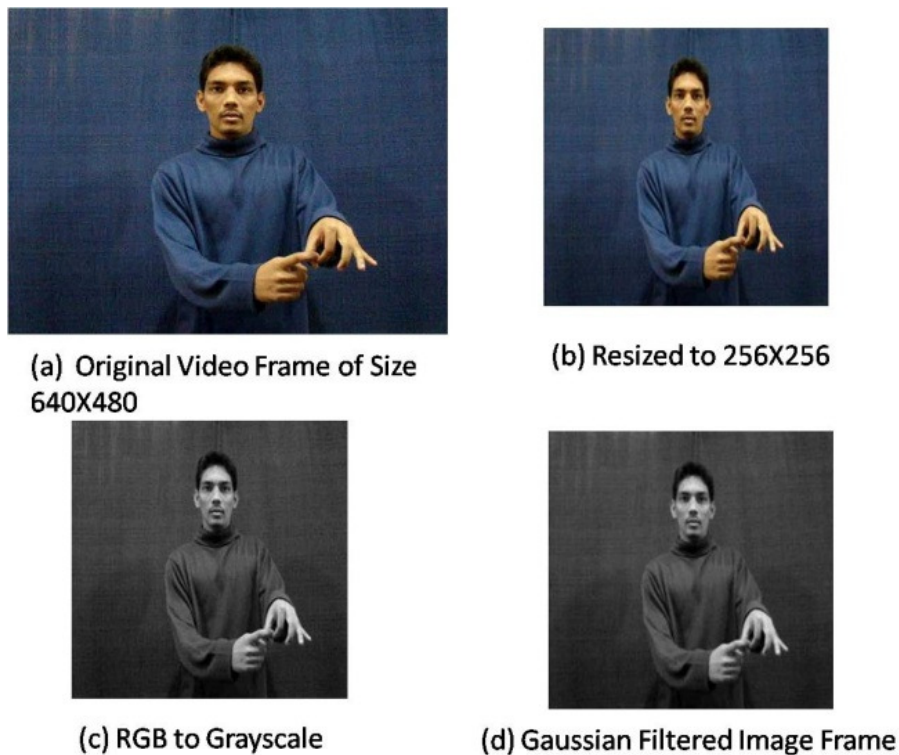


### 3.1 Video Preprocessing

Video preprocessing comprise of three operations on the input video sequence of the signer to make the segmentation process efficient. The first in the process is scaling the video form 640×480 pixels to 256×256 for reducing data and thereby saving the processing time. Scaling is done using bilinear interpolation algorithm, where the output pixel value is a weighted average of pixels in the nearest 2×2 neighbourhood.

The resized RGB (Red, Green, and Blue) video has 3 color planes per frame. Any minor alteration to the color video involves modification in all three R, G and B plane. Hence operating in RGB plane can decelerate the recognition process. This problem is solved by converting RGB color video into Grayscale video by eliminating the hue and saturation information and retaining the luminance. The grayscale image frame is formed by a weighted sum of R, G and B components.

During video acquisition the video is subjected to high frequency periodic noise naturally caused by the presence of electrical or electromechanical interference. Also moving objects in the video causes motion blur. These types of noise and blur can be eliminated by applying a Gaussian low pass filter. Convolving noisy blurred video frame with Gaussian filter point spread function gives a output filtered video frame. All the preprocessing steps are shown in figure2.



**FIGURE 2:** Video Preprocessing

### 3.2 Segmentation

One of the most serious challenges we faced is to automatically segment objects from a video sequence of the signer. Canny fused Wavelet based video object segmentation have turn out to be an indispensable part in achieving better segmentation, because it combines the necessary qualities of canny operator and two dimensional wavelet transform. Segmentation of hands and head portion is made perfect by applying a series of image processing techniques such as

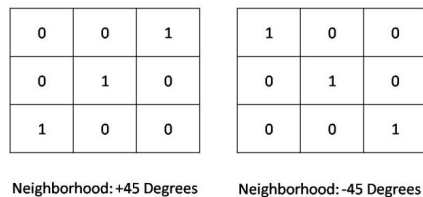
morphology, canny operator, and wavelet transform and image fusion. These entire operations guarantee that hand shape in any orientation is perfect in each and every frame of the video.

A straightforward and proficient edge detection algorithm is proposed based on morphology and multilevel discrete wavelet transform (DWT). The edge detection process is a combination of gradient calculation, applying two dimensional discrete wavelet transform and finally calculating inverse wavelet transform for image reconstruction. Gradient image of canny edge detector and the approximations of wavelet transform are fused together. For the non-uniform illumination problem, a global thresholding algorithm (Otsu's Method), can be constructed to minimize the inter class variance of black and white pixels. By applying the algorithm on signer videos show that it not only detects edges correctly but also suppresses the noise to a great extent.

Edge is a result of gray level discontinuities. The edge detection of images is an important research area in the field of image processing. For long it has been an objective of the edge detection algorithms to emphasize problems to detect edges in the noisy environments, problem of false edge detection and problem of un-detection [27]. The most common approach for detection is by using 2D first and second order derivatives. The process is convolving 2D gradient with the image which returns a one for non-uniform regions and a zero for uniform regions. For high-quality segmentation, the first step is selection of suitable edge detection method.

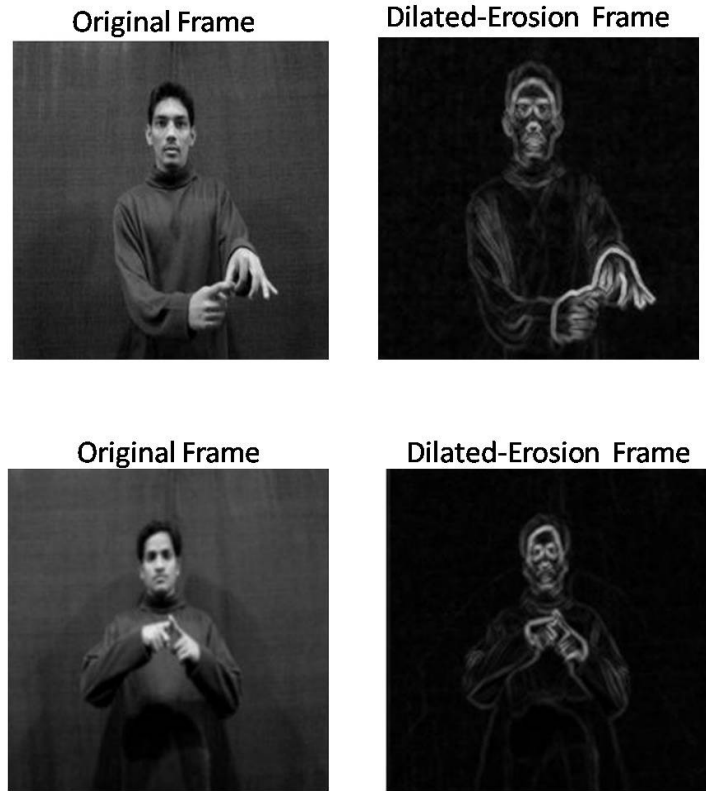
Traditional edge detection techniques are based on the gradient operators, such as Sobel, Prewitt and Roberts. The most popular edge detectors such as laplacian of Gaussian and canny operators have the ability to detect true weak edges. Due to edge complexity, they are very sensitive to noise and brightness variations. Baris Sumengen, B. S. Manjunath [28] proposed a simple and efficient edge detection algorithm showing that using multiscale edge detection techniques eliminates the need for explicit scale selection and edge tracking. There are many different methods for edge detection; the traditional methods use first-order derivative operators to detect edges, such as Robert operator, Prewitt operator [29]. Wavelet has well local performance in time-frequency field and the characteristic of multi-resolution analysis, which is very suitable for image processing [30, 31].

Dilation and Erosion are two fundamental morphological operations. Dilation adds pixels to the borders of items in an image while erosion removes pixels on item borders. The number of pixels added or removed from the items in an image depends on the size and shape of the structuring element used to process the image. An essential part of the dilation and erosion operations is the structuring element used to probe the input image. A structuring element is a matrix consisting of only 0's and 1's that can have any arbitrary shape and size. The pixels with values of 1 define the neighbourhood. In this paper we have chosen a two line structuring element of length three which are oriented diagonally at +45 degrees and -45 degrees. The two structuring elements are shown in the figure 3.



**FIGURE 3:** Flat Structuring elements containing 3 elements

In segmenting videos dilation and erosion are used in combination to produce a binary gradient image before applying discrete wavelet transform. Our algorithms dilate and erode the original image by using the structuring elements defined in the figure1 and subtracts the dilated image from the eroded image. The result of above operation applied to image frames is shown in figure 4.



**FIGURE 4:** Result of dilation-erosion

An easy way to comply with the conference paper formatting requirements is to use this document as a template and simply type your text into it. Wavelet transform is defined as decomposing a signal into a family of functions that are the translation and dilation of a unique function  $\theta(x)$  called the basic wavelet. A function  $\theta(x)$  is a basic wavelet if its average is zero.  $L^2(\mathbb{R})$  denotes the Hilbert space of measurable, square-integrable 1-D function  $f(x)$ . For  $f \in L^2(\mathbb{R})$ , the wavelet transform of  $f(x)$  at the scale  $a$  and position  $\tau$ , is defined by

$$wf(a, t) = f * \psi_a(t) = \frac{1}{\sqrt{a}} \int f(x) \psi\left(\frac{x-t}{a}\right) dt \quad [1]$$

Where  $*$  denotes the convolution operator, and

$$\psi_a(x) = \frac{1}{s} \psi\left(\frac{x}{s}\right) \quad [2]$$

The dynamic wavelet transform means the scale varies only along the dyadic sequence  $(2^j)_{j \in \mathbb{Z}}$ , and it has a fast algorithm. The first or second-order derivatives of a smoothing function  $\theta(x)$  can be used to detect the sharp variation points. The extrema of the first derivative correspond to the edge points.  $\theta(x, y)$  denotes a two dimension smoothing function. The first derivative respectively along the horizontal and vertical orientation is defined as basic wavelet function

$$\psi^1(x, y) = \frac{\partial \theta(x, y)}{\partial x} \quad \psi^2(x, y) = \frac{\partial \theta(x, y)}{\partial y} \quad [3]$$

The wavelet transform of  $f(x, y)$  at scale  $a$  has two components defined by



$$W_a^1 f(x, y) = f * \psi_a^1(x, y) \quad \text{and}$$

$$W_a^2 f(x, y) = f * \psi_a^2(x, y)$$

Let  $a$  equal  $2^j$ . At each scale, the modulus of the gradient vector is proportional to

$$M_{2^j} f(x, y) = \sqrt{\left|W_{2^j}^1 f(x, y)\right|^2 + \left|W_{2^j}^2 f(x, y)\right|^2}$$

The angle of the gradient vector with the horizontal orientation is given by

$$\text{Arg}[WTf(2^j, x, y)] = \text{tg}^{-1} \left[ \frac{WT^2 f(2^j, x, y)}{WT^1 f(2^j, x, y)} \right]$$

So the edge points of  $f * \theta_{2^j}(x, y)$  are the points  $(x, y)$ , where the modulus  $M_{2^j} f(x, y)$  is local maxima in the direction of the gradient given by  $A_{2^j} f(x, y)$ . Many small local maxima are due to noise, so the maxima smaller than a given threshold value are eliminated. The basic wavelet in our algorithm is the quadratic spline wavelet, which is subject to the Battle-Lemarie spline wavelets family and is orthogonal with compact support and integral shift.

### 3.2.1 Fusion Algorithm of Multiscale Wavelet Transform

The multiscale edge images are obtained by first performing 2D multiscale Daubechies wavelet transform up to level two producing approximated, vertical, diagonal and horizontal details, which are important characteristics of an image. The 2D wavelet transform divides the image in to low frequency (L) and high frequency components (H) at level1. Further in level2 decomposition the low frequency information can be divided in to LL and high frequency information LH. The high frequency component in level1 is decomposed in to low frequency information HL and high frequency HH. The wavelet decomposition process is shown in the figure.5

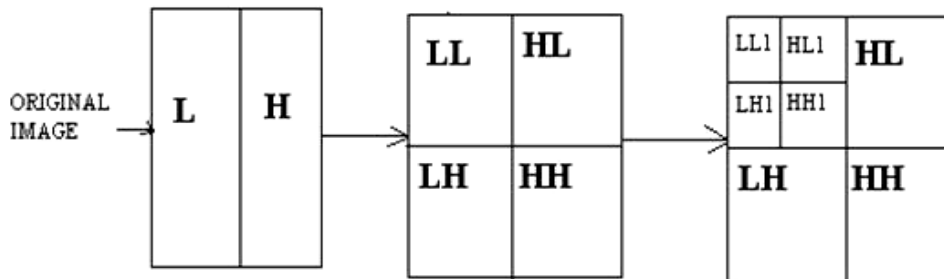
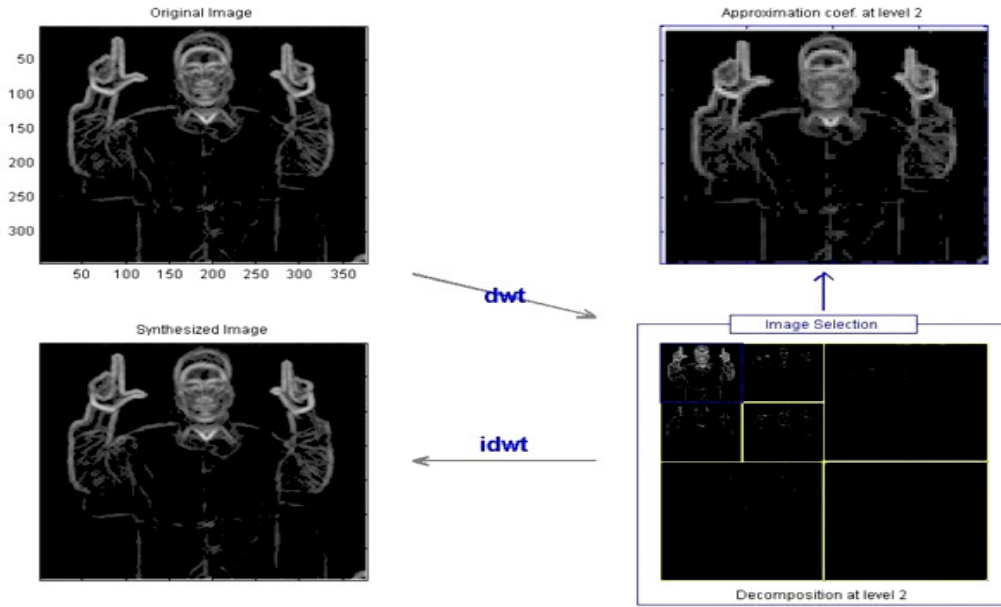


FIGURE 5: Wavelet Decomposition Based on Wavelet Transform

The results of wavelet decomposition using 2D Daubechies two wavelets for level2 on ‘Cow’ Video image frame is shown in figure6.

Though the wavelet transform applied on the image eliminates most of the noise in the image, the authentic edges in the image are also mixed with the noise, particularly the HH areas are greatly affected by noise. Thus many methods of edge detection discard these HH areas. This paper presents a new technique for edge detection by fusion algorithm based on wavelet transform and canny edge detector.



**FIGURE 6:** Daubechies2 wavelet transform of level2.

The notion  $L^2(\mathbb{R}^2)$ , where  $\mathbb{R}$  is a set of real numbers, denote the finite energy function  $f(x,y)$  in  $\mathbb{R}^2$ ; and  $x,y$  in  $\mathbb{R}$ . In two dimension wavelet transform, a 2D scaling function  $\phi(x,y)$ , and three two dimensional wavelets,  $\phi^H(x,y)$ ,  $\phi^V(x,y)$  and  $\phi^D(x,y)$  are produced as shown in figure.5. The above functions represent gray level variations along different directions such as horizontal variations, vertical variations and diagonal variations. The DWT of  $f(x,y)$  of size  $M \times N$  is

$$W_\phi(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f \phi_{j_0, m, n} \quad [4]$$

Where  $j, m, n, M, N$  are integers,  $i = \{H, V, D\}$ ,  $j_0$  is an arbitrary starting scale and the coefficients  $W_\phi$  define an approximation of  $f$  at scale  $j_0$ .

$$W_\phi^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f \psi_{j, m, n}^i \quad [5]$$

The coefficients in the above equation add horizontal, vertical and diagonal details as shown in figure.5 for scales  $j \leq j_0$ . The  $\phi_{j, m, n}$  and  $\psi_{j, m, n}^i$  denote scaled and translated basis functions as shown below:

$$\begin{aligned} \phi_{j, m, n}(x, y) &= 2^{j/2} \phi(2^j x - m, 2^j y - n) \\ \psi_{j, m, n}^i(x, y) &= 2^{j/2} \psi^i(2^j x - m, 2^j y - n) \end{aligned}$$

Given  $W_\phi$  and  $W_\psi^i$ ,  $f$  is obtained via inverse DWT as:

$$f = \frac{1}{\sqrt{MN}} \sum_m \sum_n \left( W_\phi^{j_0} \phi_{j_0} + \sum_i \sum_{j=j_0}^{\infty} W_\psi^j \psi_j^i \right) \quad [6]$$

Figure.7 shows the block diagram and figure 8 the application of image fusion process. Both the images undergo wavelet decomposition and reconstruction again in this stage using the above

equation. The fusion parameter selection rule is based on absolute maximum of horizontal coefficients and vertical coefficients:

$$W_{\varphi} = \begin{cases} W_{\psi}^H, & |W_{\psi}^H| \geq |W_{\psi}^V| \\ W_{\psi}^V, & \text{otherwise} \end{cases} \quad [7]$$

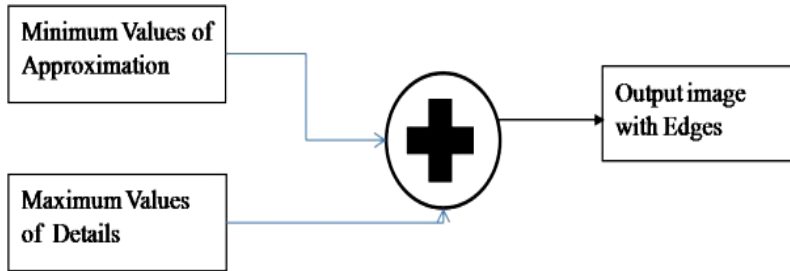


FIGURE.7. Image Fusion Process

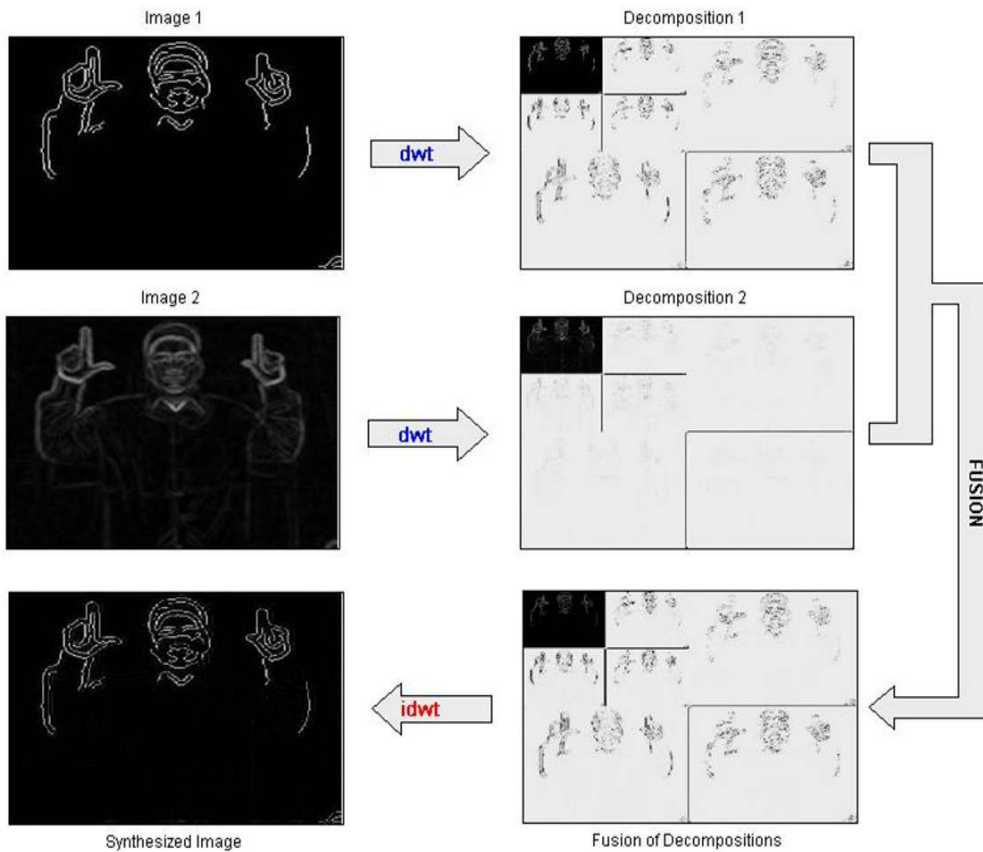


FIGURE 8: Image Fusion Applied to a video frame

To be acknowledged as a correct edge line, thresholding is desirable. Thresholding is performed on the wavelet reconstructed image by using a global thresholding algorithm known as Otsu's method. Converting a greyscale image to monochrome is a common image processing task. Otsu's method, named after its inventor Nobuyuki Otsu, is one of many binarization algorithms. Otsu's thresholding method involves iterating through all the possible threshold values

and calculating a measure of spread for the pixel levels each side of the threshold, i.e. the pixels that either falls in foreground or background. The aim is to find the threshold value where the sum of foreground and background spreads is at its minimum. In Otsu's method we exhaustively search for the threshold that minimizes the intra-class variance, defined as a weighted sum of variances of the two classes given by

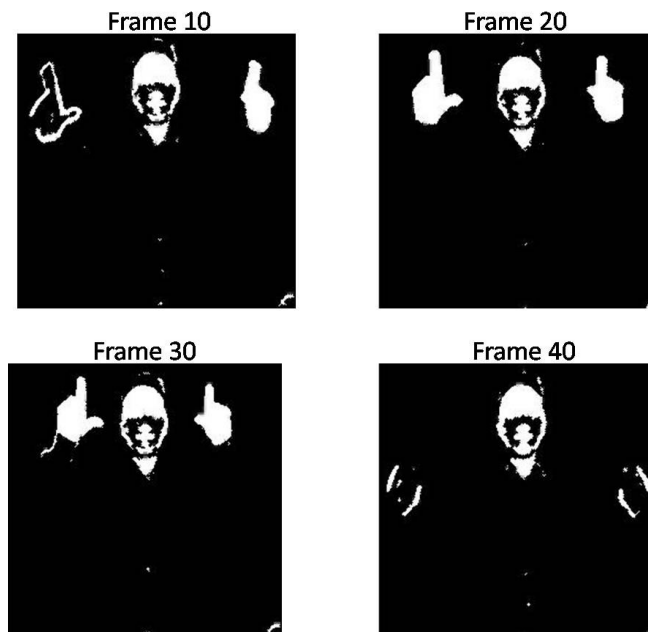
$$\sigma_w^2(t) = W_1(t)\sigma_1^2(t) + W_2(t)\sigma_2^2(t) \quad [8]$$

Where  $W_i$  are probabilities of 'i' classes separated by a threshold 't' and  $\sigma^2$  are variances of these classes. In this method image histogram is considered having a two Gaussian distribution representing the object and the background. This proves an effective algorithm for threshold calculation. After thresholding the video frames produced brighter edges compared to traditional algorithms for edge detection as shown in figure 9 for some frames of a sign.

### 3.3 Feature Extraction

High level feature extraction primary concern is finding shapes in computer images [32]. The feature extraction process is bound to shape our world view. During feature extraction process we seek invariance parameters so that the extraction process does not vary according to specified circumstances. That is, techniques used for feature extraction should find shapes reliably and robustly irrespective of changes in illumination levels, position, orientation and size of the object in a video. Objects in an image are represented as collection of pixels. For object recognition we need to describe the properties of these groups of pixels. The description of an object is a set of numbers called as object's descriptors.

Recognition is simply matching a set of shape descriptors from a set of known descriptors. A usable descriptor should possess four valuable properties. The descriptors should form a complete set, they should be congruent, rotation invariant and form a compact set. Objects in an image are characterized by two forms of descriptors region and shape descriptors. Region descriptors describe the arrangement of pixels within the object area whereas shape descriptors are the arrangement of pixels in the boundary of the object.



**FIGURE 9:** Result of Segmentation obtained by fusion of canny and DWT showing 4 frames in the video

#### 3.3.1 Elliptical Fourier Descriptors

The segmentation yields a low data in the form of pixels along a boundary or pixels contained in a region. We are mainly interested in the boundary pixels that are hand shapes of the signer. To

extract shape outline with minimum number of pixels for an image frame without losing shape information we choose Fourier Descriptors [32, 33, 34, 35]. Fourier descriptors often attribute to early work by cosgriff (1960), allows us to bring the power of Fourier theory to shape description. The basic idea of Fourier descriptors is to characterize a curve by a set of numbers that represent the frequency content of a whole shape. The Fourier descriptors allow us to select a small set of numbers that describe a shape for an image frame. This property Fourier descriptors is helpful because these Fourier coefficients carry shape information which is not insensitive to translation, rotation and scale changes. But the changes in these parameters can be related to transformations on descriptors.

Fourier representations are articulated in terms of orthogonal basis functions, causing the parameters to be distinct and hence avoid redundancy. The boundary of hands and head or its contour can be represented by closed curves. As a result of periodicity of curves projection on the vertical and horizontal axis we used elliptical Fourier representations to model hand and head contours. The elliptical Fourier descriptors of a closed contour describe a curve in 2D space. Each pixel in the image plane is represented by a complex number. Thus any shape in an image is defined as

$$s(t) = x(t) + jy(t) \tag{9}$$

Where the parameter 't' is given by arc-length parameterization.

To obtain elliptical Fourier descriptors of a curve we need to obtain Fourier expansion of the shape in the above equation. The Trigonometric fourier expansion can be obtained using the expression

$$s(t) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos(k\omega t) + b_k \sin(k\omega t)) \tag{10}$$

Accordingly the general expression for elliptical fourier coefficients for any closed contours can be defined as

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} a_0 \\ c_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix} \tag{11}$$

Where

$$\begin{aligned} a_0 &= \frac{1}{2\pi} \int_0^{2\pi} x(t) dt & c_0 &= \frac{1}{2\pi} \int_0^{2\pi} y(t) dt \\ a_k &= \frac{1}{2\pi} \int_0^{2\pi} x(t) \cos(kt) dt & b_k &= \frac{1}{2\pi} \int_0^{2\pi} x(t) \sin(kt) dt \\ c_k &= \frac{1}{2\pi} \int_0^{2\pi} y(t) \cos(kt) dt & d_k &= \frac{1}{2\pi} \int_0^{2\pi} y(t) \sin(kt) dt \end{aligned}$$

Thus any closed contour can be represented by its fourier coefficients  $a_0, c_0, a_k, b_k, c_k$  and  $d_k$ . Where 'k' represents the rank of the ellipse and for  $k=1$  corresponds to fundamental component of closed curve. For different values of k, the trigonometric summation defines the locus of an ellipse in a complex plane. As we change the parameter 't' the point traces ellipses moving at a speed proportional to the harmonic number 'k'.

From geometric point of view a simple ellipse is modeled by the equation

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \times \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \tag{12}$$

Where A is semi major axis oriented in horizontal axis direction and B is semi minor axis oriented in vertical axis direction. The model starting point lies with semi major axis. Considering rational angle and phase shift from major axis we obtain a more comprehensive representation of the ellipse as

$$\begin{bmatrix} x(t) \\ y(t) \end{bmatrix} = E \times \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \quad [13]$$

Where

$$E = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix} \times \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix} \times \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix}$$

$\theta$  = Rotational Angle

$\Phi$  = Phase shift of the ellipse respectively.

This geometric interpretation of ellipse gives us better visualization. For  $K^{th}$  ellipse the defined parameters are as shown in the figure 10.

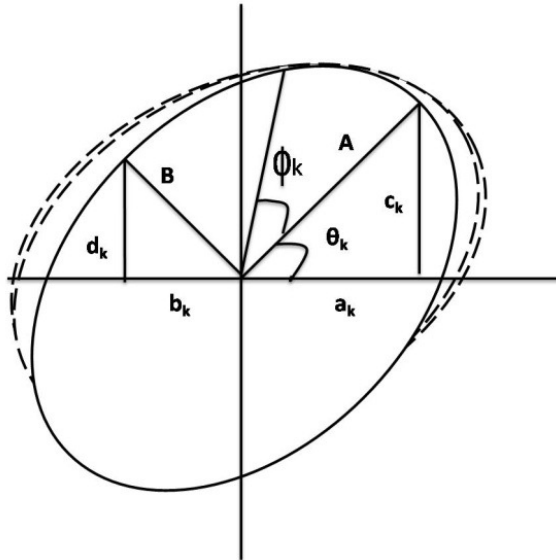


FIGURE 10:  $K^{th}$  Ellipse with Parameters.

Hence Eq. (13) becomes

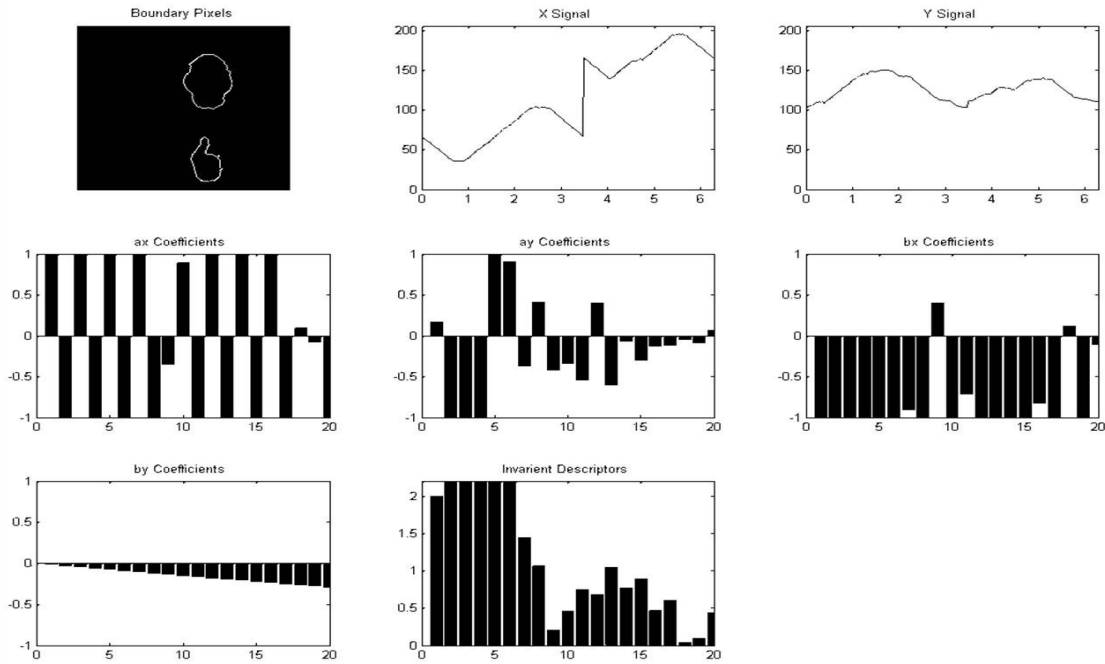
$$\begin{aligned} \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} &= \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} + \sum_{k=1}^{\infty} \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \times \begin{bmatrix} \cos(kt) \\ \sin(kt) \end{bmatrix} \\ \begin{bmatrix} x(t) \\ y(t) \end{bmatrix} &= \begin{bmatrix} A \cos(\theta) & -B \sin(\theta) \\ A \sin(\theta) & B \cos(\theta) \end{bmatrix} \times \begin{bmatrix} \cos(\phi) & -\sin(\phi) \\ \sin(\phi) & \cos(\phi) \end{bmatrix} \\ &= E * \begin{bmatrix} \cos(t) \\ \sin(t) \end{bmatrix} \end{aligned} \quad [14]$$

Where 
$$E_k = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix}$$

From the above equation we can write the fourier coefficients of  $k^{\text{th}}$  ( $a_k, b_k, c_k, d_k$ ) as

$$\begin{aligned} a_k &= A_k \cos \theta_k \cos \phi_k - B_k \sin \theta_k \sin \phi_k \\ b_k &= -A_k \cos \theta_k \sin \phi_k - B_k \sin \theta_k \cos \phi_k \\ c_k &= A_k \sin \theta_k \cos \phi_k + B_k \cos \theta_k \sin \phi_k \\ d_k &= -A_k \sin \theta_k \sin \phi_k + B_k \cos \theta_k \cos \phi_k \end{aligned}$$

Where  $A_k, B_k, \theta_k, \phi_k$  are more understandable parameters of the same ellipse and the relationship between these set parameters could be computed. The plot of different fourier coefficients along with the view of head and hand contours is shown in figure 11.



**FIGURE 11:** Plot gives the boundary pixels of a signers head and hand shapes represented using 20 coefficients and Fourier descriptors computed using direct implementation of Eq.11. The last figure shows the invariant descriptors.

The number of coefficients can be specified by the user by using the parameter 'k'. Details in the characterization are given by the number of coefficients used. For our application the maximum number of coefficients has been limited to 20. Figure 11 shows the original curve,  $x(t)$  and  $y(t)$  coordinate functions and the fourier descriptors described by invariant descriptors. The coefficients remain the same after changing the location, orientation and scale clearly indicating elliptical fourier descriptors truly characterize the shape of an human head and hands.

For a typical video of the signer with 45 frames we obtain a matrix of  $45 \times 20$  values representing the shape information of hands and head. For 80 signs, our feature vector size becomes very large and consumes more processing time in the next stages.

### 3.3.2 Principal Component Analysis

Here we treat the current feature vector using principle component analysis (PCA) also known as karhunen Loeve(KL) transform which uses factorization to transform data according to its statistical properties. This data transformation is particularly useful for classification problems when data set is very large. PCA reduces the dimensions of highly correlated input vectors. This technique has three effects: it orthogenesis the vector components obtained using fourier descriptors, secondly it orders the resulting orthogonal components in ascending order from largest variations to least variations and finally it discards the least contributing components of the data set.

The fourier descriptors matrix for a single video is 45x20 and we have 9 signers with around 80 signs each which makes our feature matrix huge. Now we reduce the dimension of our feature vector matrix by using PCA. Always start with normalizing the input vectors so they have zero mean and unity variance. The mathematics of PCA can be summarized in the following steps. Obtain the feature matrix  $\mathbf{c}_x$  from the data. Each column of the matrix defines a feature vector. Compute the covariance matrix  $\Sigma_x$ . This matrix gives information about the linear independence between the features. Obtain the Eigen values by solving the characteristic equation.

$$\det(\lambda_i I - \Sigma_x) = 0$$

These values form the diagonal covariance matrix  $\Sigma_y$ . Since the matrix is diagonal, each element is the variance of the transformed data.

Obtain the eigenvectors by solving for  $\mathbf{w}_i$  in

$$(\lambda_i I - \Sigma_x) \mathbf{w}_i = 0$$

for each Eigen value. Eigen- vectors should be normalized and linear independent. The transformation  $\mathbf{W}$  is obtained by considering the eigenvectors as their columns. Obtain the transform features by computing  $\mathbf{c}_y = \mathbf{c}_x \mathbf{W}^T$ . The new features are linearly independent. For classification applications, select the features with large values of  $\lambda_i$ . Remember that  $\lambda_i$  measures the variance and features that have large range of values will have large variance.

For example, two classification classes can be obtained by finding the mean value of the feature with largest  $\lambda_i$ .

For compression, reduce the dimensionality of the new feature vectors by setting to zero components with low  $\lambda_i$  values. Features in the original data space can be obtained by

$$\mathbf{C}_X^T = \mathbf{W}^T \mathbf{C}_Y^T$$

The result of elliptical Fourier descriptors and principle component analysis give a matrix consisting of feature vectors. Figure 12 shows feature vector matrix and figure 13 shows their uniqueness. Each row in the feature vector matrix corresponds to a video sign in the data base.



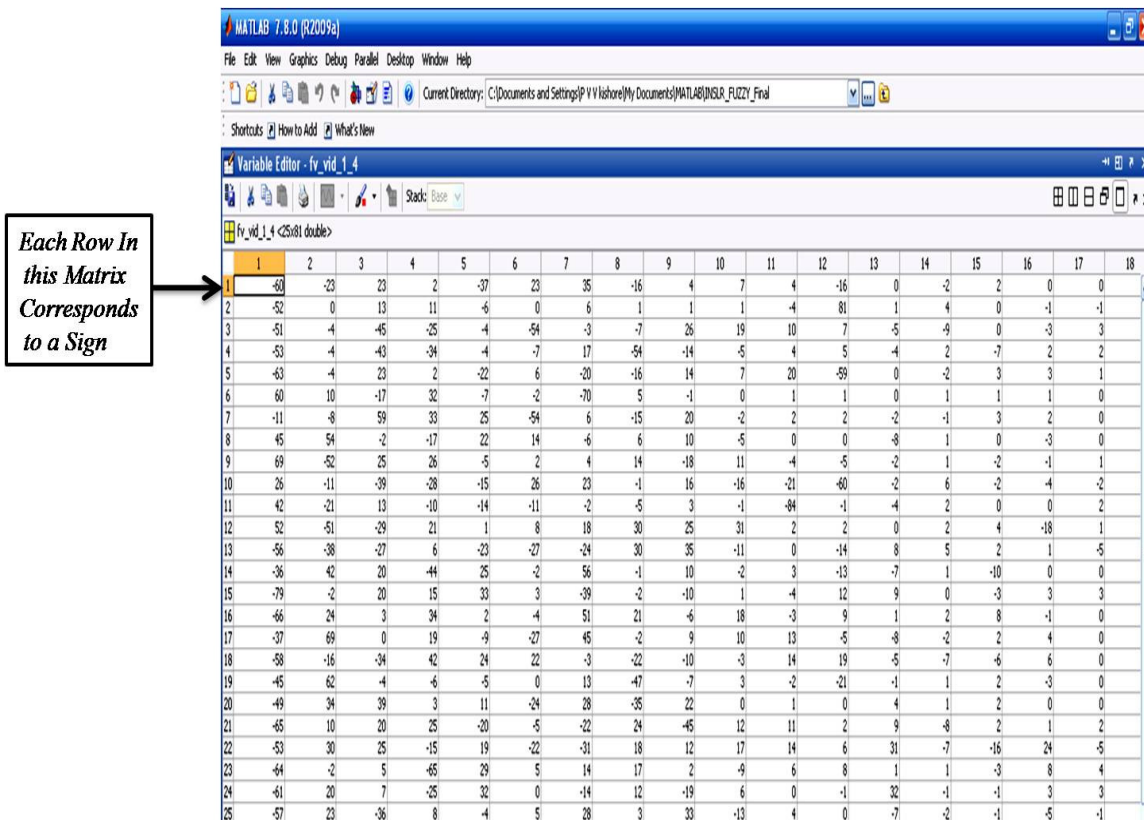


FIGURE 12: Feature Vector Matrix.

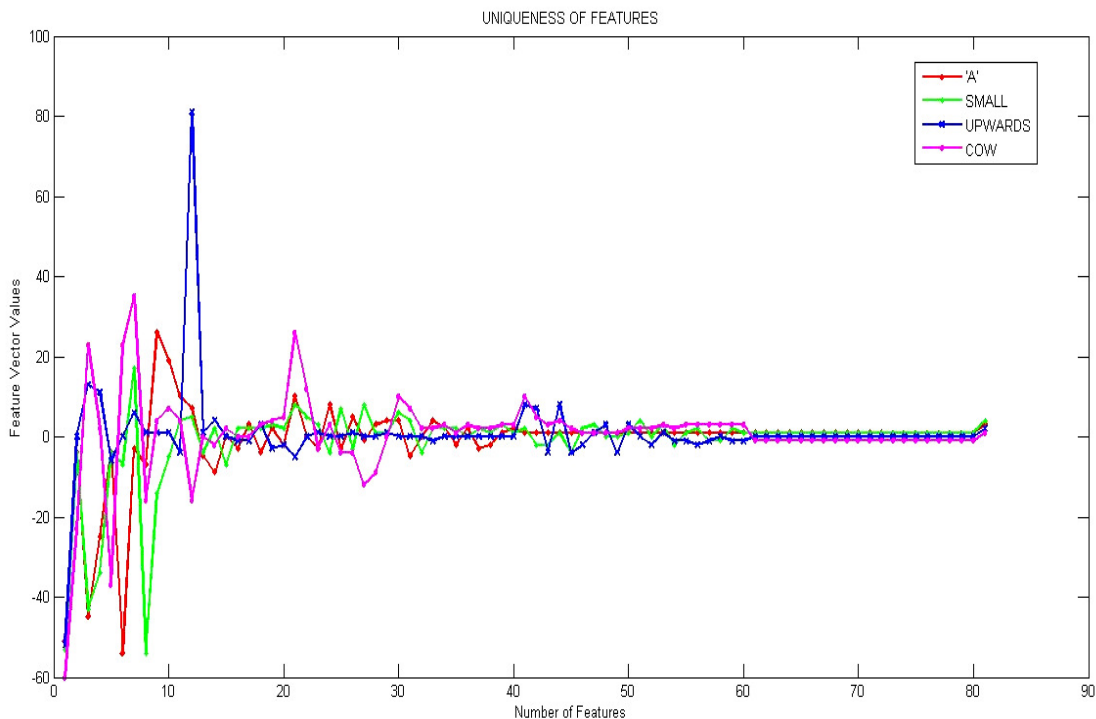


FIGURE 13: Plot showing uniqueness of feature vector for four different signs

#### 4. PATTERN RECOGNITION

The final stage of the system is classification of different signs and generating voice messages corresponding to the correctly classified sign. In the past this part of the system was implemented using hidden markov models (HMM), Neural networks, Baye’s theorem and to some extent fuzzy logic. For our sign language recognition system we have deployed a fuzzy inference system.

##### 4.1 Fuzzy Inference Systems (FIS)

Fuzzy logic is being applied to pattern recognition and pattern classification problems with if-then rules [36-37]. Numerous techniques have been proposed in the past to generate fuzzy if-then rules automatically [38-40]. There are various methods for generating fuzzy if-then rules for pattern classification problems. Generating fuzzy if-then rules from numerical data is considered for our system. The method was first proposed by Ishibuchi [41]. Generation of fuzzy if-then rules from numeric data is done in two sets of processes. One is mapping a pattern into fuzzy subspace using fuzzy partition and two is determination of fuzzy if-then rules each fuzzy subspace. An example of fuzzy partition system is given by a simple grid where a 2D pattern space is divided into fuzzy subspace.

The performance indicators in a if-then rule based fuzzy classification systems depends on the selection of fuzzy partition. If the fuzzy classification is too large, the performance may be low with many patterns being wrongly classified. If fuzzy partition is too small, many fuzzy if-then rules cannot be generated because lack of training patterns in the corresponding fuzzy subspaces. This problem was eliminated by the use of distributed fuzzy if-then rules [41] where all fuzzy if-then rules corresponding to several fuzzy subsets were simultaneously employed in fuzzy inference system. The main drawback of this system is that the number of fuzzy if-then rules becomes enormous for pattern classification problems.

##### 4.2 Fuzzy Classification Method

For pattern classification we considered Takagi-sugeno-kang (TSK) or simply sugeno type Fuzzy inference system because the output membership functions are linear or constant. Sugeno fuzzy inference system consists of five steps, fuzzification of input variables, applying fuzzy ‘and’ or ‘or’ operator, calculating the rule weights, calculating the output level and finally defuzzification. Many methods are proposed to generate fuzzy rule base. The basic idea is to study and generate the optimum rules needed to control the input without compromising the quality of control. In this paper the generation of fuzzy rule base is done by subtractive clustering technique in sugeno fuzzy method for classification video. Cluster center found in the training data are points in the feature space whose neighborhood maps into the given class. Each cluster center is translated into a fuzzy rule for identifying the class. A generalized type-I TSK model can be described by fuzzy IF-THEN rules which represent input output relations of a system. For multi input single output first order type-I TSK can be expressed as

IF  $x_1$  is  $Q_{1k}$  and  $x_2$  is  $Q_{2k}$  and ... and  $x_n$  is  $Q_{nk}$ ,  
THEN  $Z$  is

$$w = p_0 + p_1X_1 + p_2X_2 + \dots + p_nX_n \tag{15}$$

Where  $x_1, x_2, \dots, x_n$  and  $Z$  are linguistic variables;  $Q_{1k}, Q_{2k}, \dots$  and  $Q_{nk}$  are the fuzzy sets on universe of discourses  $U, V, \dots$  and  $W$ , and  $p_0, p_1, \dots, p_n$  are regression parameters.

With subtractive clustering method,  $X_j$  is the  $j^{\text{th}}$  input feature  $x_j$  ( $j \in [1, n]$ ), and  $Q_{jk}$  is the MF in the  $k^{\text{th}}$  rule associated with  $j^{\text{th}}$  input feature. The MF  $Q_{jk}$  can be obtained as

$$Q_{jk} = \exp \left[ -\frac{1}{2} \left( \frac{X_j - \chi_{jk}}{\sigma} \right)^2 \right] \tag{16}$$

Where  $x_{jk}$  is the  $j^{\text{th}}$  input feature of  $x_k$ , the standard deviation  $\sigma$  of Gaussian MF given as

$$\sigma = \frac{1}{\sqrt{2a}} \tag{17}$$

### 5. RESULTS AND ANALYSIS

In this section we will analyze the performance of our system by its capability to recognize gestures from videos. We also discuss the difficulties faced while designing the system. Table 2 summarizes different gestures used for the analysis of our proposed system.

All English Alphabets, Numbers, cow, donkey, duck, peacock, fat, feather, foolish, funny, nest, what, crow, young, upwards, towards, come, good, love, mother, father, where are you going, do your home work etc.

**TABLE 2:** Summary of Sign language used

Classification of different gestures is done using Fuzzy inference system for 80 different gestures of Indian sign language by 9 different signers. Table 3 gives details of the fuzzy inference system used for gesture classification. Figure 14 shows the input membership functions. The performance of the system is evaluated based on its ability to correctly classify signs to their corresponding speech class. The recognition rate is defined as the ratio of the number of correctly classified signs to the total number of signs:

$$\text{Recognition Rate (\%)} = \frac{\text{Number of Correctly Classified Signs}}{\text{Total Number of Signs}} \times 100$$

Name	'fis_inslr'
Type	'sugeno'
And Method	'min'
Or Method	'max'
Defuzz method	'wtaver'
Imp Method	'prod'
Agg Method	'sum'
Input	[1x80 struct]
Output	[1x1 struct]
Rule	[1x25 struct]

**TABLE 3:** Details of fuzzy inference system used for gesture classification.

A graphical User Interface (GUI) has been created to automatically train and recognize the gestures as shown in the figure 15. The GUI has two parts, a training part and a recognition part. Training is used to train new signs and store them in the data base of sign language system. Recognition tries to recognize the trained videos along with unknown videos of different signers. Training consists of four processes that can be activated by the use of buttons Load Video, Extract Features, Save Features along with the text in the text box and finally train button to finish the training. View FIS shows the user fuzzy inference system used. Once the load video button is pressed the video is played in the top portion of the training side which is interfaced to Microsoft windows media player.

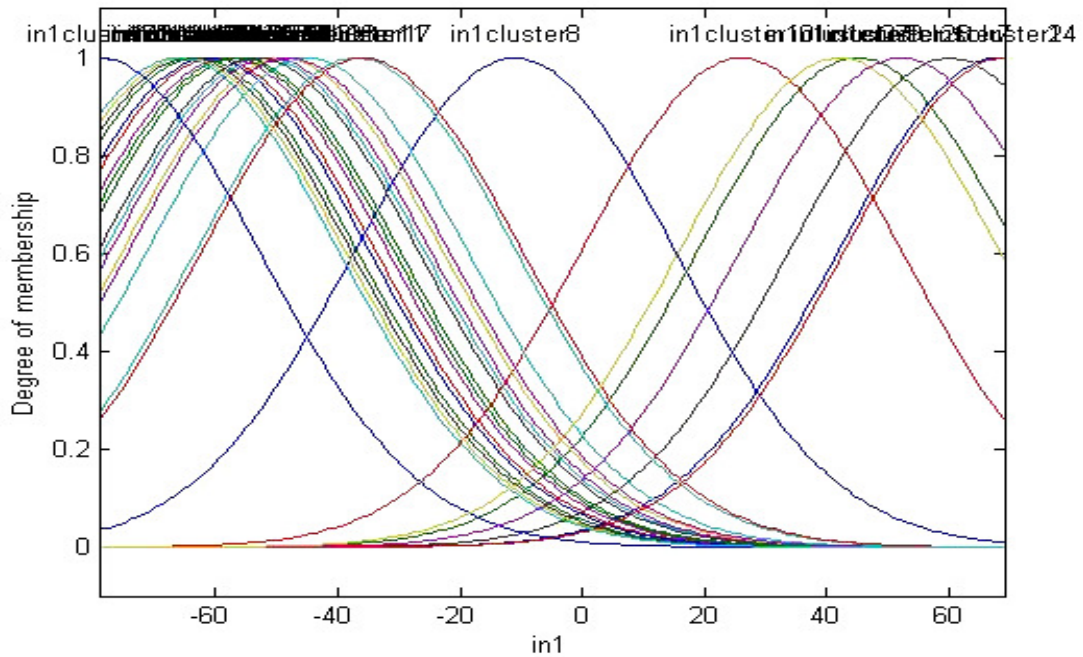


FIGURE 14: Input Membership Functions

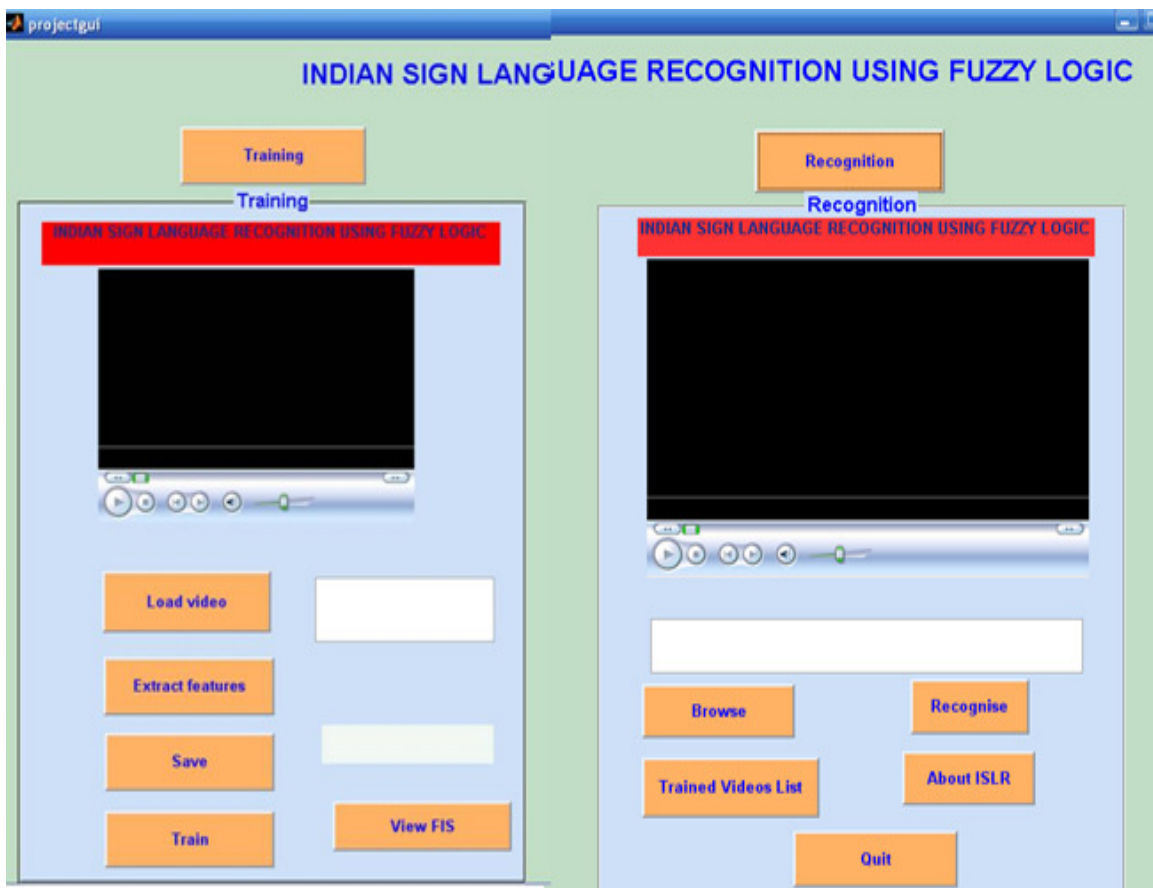


FIGURE 15: GUI of Sign Language Recognition System

The recognition half consists of five buttons and a text box to display the word or sentence recognized by the system. The browse option is used to load video of the signer into the system. Once loaded and the recognize button is pressed, a voice message is heard along with the text message that is displayed in the text box provided on top of all buttons corresponding the gesture loaded into the system. The user can also see trained video list by using the trained videos list button. About ISLR button gives all the help information regarding our system.

Figure 16 gives a demo of training the fuzzy inference system using a video of signer representing a gesture of number 'six'. Figure 17 shows the recognition of sign 'six'. Table 4 shows the results obtained when training 10 samples for each gesture with different signers. Table shows recognition rates of some signs used for classification. The total number of signs used for testing is 80 from 10 different signers and the system recognition rate is close to 96%. The system was implemented with MATLAB version 7.0.1.

## **6. CONCLUSION AND FUTURE WORK**

In this paper we developed a system for recognizing a subset of the Indian sign language. The work was accomplished by training a fuzzy inference system by using features obtained using DWT and Elliptical fourier descriptors by 10 different signer videos for 80 signs with a recognition rate of 96%. Most of the research in sign language is mainly the Extraction of Head and Hand Gesture Features for Recognition of Sign Language by the use of Sobel edge operator, DCT (Discrete Cosine Transform) for feature extraction and neural network for classification. In his framework he segmented head and hands separately and found the areas of the segmented head and hand portions. Then he applied DCT to extract features and finally a neural network is used to classify the gesturers.

When implemented the same process we observed that the sobel operator is unable to mark true edges. Hence we used canny operator with double thresholding to mark true edges on our videos. Though DCT based feature extraction gives minimum energy features but it fails to extract shapes of the segmented head and hand portions which was again done separately for all three areas increasing considerably the processing time of their algorithm. As we used Elliptical Fourier Descriptors which are proved shape extractors with minimum number of descriptors in our case it is 20 per frame. This greatly reduced processing time and extracted finite value feature vector.

Finally the video acquisition process was under controlled environment with good lighting and definite camera displacement. Our videos are shot under various lightning conditions with different signers as can be observed from Table 1.

Some researchers used hough transform to model hand shapes to extract features along with neural network to classify signs. Hough transform is good at identifying shapes but they should be almost straight lines. Hence the algorithm using hough transform was unable to give us good results for video of gesturers. It works well for images of gestures with only one hand at a time as we have identified. Hence we used Elliptical Fourier Descriptors which are proved shape extractors with minimum number of descriptors in our case it is 20 per frame. This greatly reduced processing time and extracted finite value feature vector.

Most authors used fuzzy logic for recognition of gestures as we did, but we observed that our algorithm gives better recognition rates for different gesturers under different conditions over a wide data base of signers. We also implemented the same process using error back propagation neural network giving fairly the same recognition rate.

In future we are looking at developing a system for Indian sign language that works in real-time. To accomplish this task we are building 3D models of hands and head of signer.

Sign	Correctly Recognized Signs	False Recogniti-on	Recognition Rate (%)
A	9	0	100
B	9	0	100
C	9	0	100
D	9	0	100
X	9	0	100
M	7	2	77.77
N	8	1	88.88
Y	6	3	66.67
Cow	8	1	88.89
Duck	9	0	100
Crow	5	4	55.56
Fat	8	1	88.89
Feather	7	2	77.77
Love	9	0	100
Together	9	0	100
Come here	6	3	66.67
Do your Home Work	5	4	55.56
Numbers 1-10	90	0	100
Upwards	9	0	100
Total	231	21	90.90%

**TABLE 4:** Results obtained when training signs with 10 different signers.



**FIGURE 16:** Training Demo of Gesture sign 'SIX'.



FIGURE 17: Recognition of sign 'SIX'

## 7. REFERENCES

- [1] W. Stokoe, D. Casterline, and C. Croneberg. "A Dictionary of American Sign Language on Linguistic Principles." Gallaudet College Press, Washington D.C., USA, 1965.
- [2] U. R. Wrobel. Referenz in eb"ardensprachen: Raum und Person. Forschungsberichte des Instituts f"ur Phonetik und Sprachliche Kommunikation der Universit"at M"unchen, 37:25–50, 2001.
- [3] C. Vogler and D. Metaxas. "A framework for recognizing the simultaneous aspects of american sign language". Computer Vision & Image Understanding, 81(3):358–384, March 2001.
- [4] Ruiduo Yang, Sudeep Sarkar, and Barbara Loeding. Enhanced level building algorithm to the movement epenthesis problem in sign language. In "CVPR", MN, USA, June 2007.
- [5] ASL corpus: <<http://www.bu.edu/asllrp/>>.
- [6] Christopoulos, C., Bonvillian, J., 1985. "Sign language". Journal of Communication Disorders 18, 1–20.
- [7] Atherton, M., 1999. Welsh today BSL tomorrow. In: Deaf Worlds 15 (1),pp. 11–15.
- [8] Engberg-Pedersen, E., 2003. From pointing to reference and predication: pointing signs, eyegaze, and head and body orientation in Danish Sign Language. In: Kita, Sotaro (Ed.),

Pointing: Where Language, Culture, and Cognition Meet. Erlbaum, Mahwah, NJ, pp. 269–292.

- [9] Nyst, V., 2004. Verbs of motion in Adamorobe Sign Language. Poster. In: TISLR 8 Barcelona, September 30–October 2. Programme and Abstracts. (Internat. Conf. on Theoretical Issues in Sign Language Research; 8), pp. 127–129.
- [10] Abdel-Fattah, M.A., 2005. Arabic sign language: a perspective. *Journal of Deaf Studies and Deaf Education* 10 (2), 212–221.
- [11] Masataka, N. et al., 2006. Neural correlates for numerical processing in the manual mode. *Journal of Deaf Studies and Deaf Education* 11 (2), 144–152.
- [12] Indian Sign language, empowering the deaf, Indian Sign Language Database <<http://www.deafsigns.org>>.
- [13] Gaolin Fang and Wen Gao, “Large Vocabulary Continuous Sign language Recognition Based on Transition-Movement Models”, *IEEE Transaction on Systems, MAN, and Cybernetics-Vol.37, No.1, January 2007*, pp 1-9.
- [14] T. Starner and A. Pentland “Real-Time American Sign Language Recognition from video using Hidden Markov Models”, Technical Report, MIT Media laboratory Perceptual computing section, Technical Report number.375, 1995.
- [15] Ming-Hsuan Yang and Narendra Ahuja, “Extraction of 2D Motion Trajectories and its Application to Hand Gesture Recognition”, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, Vol.24, No.8, August 2002, pp1061-1074.
- [16] M.K. Bhuyan and P.K. Bora, “A Frame Work of Hand Gesture Recognition with Applications to Sign Language”, *Annual India Conference, IEEE*, pp1-6.
- [17] Rini Akmelawati, Melanie Po-Leen Ooi and Ye Chow Kuang, ‘Real-Time Malaysian Sign Language Translation Using Colour Segmentation and Neural Network’, *IEEE on Instrumentation and Measurement Technology Conference Proceeding, Warsaw, Poland 2006*, pp. 1-6.
- [18] Nariman Habibi, Cheng Chew Lim and Alireza Moini, ‘Segmentation Of The Face And Hands In Sign Language Video Sequences Using Color And Motion Cues’, *IEEE Transactions on Circuits and Systems For Video Technology 2004*, Vol. 14, No. 8, , pp.1086 – 1097
- [19] Yu Zhou and Xilin Chen, “Adaptive sign language recognition with Exemplar extraction and MAP/IVFS”, *IEEE signal processing letters*, Vol 17, No-3, March 2010, pp297-300.
- [20] Och J., Ney. H., 2002. Discriminative training and maximum entropy models for statistical machine translation. In: Annual Meeting of the Ass. For Computational Linguistics (ACL), Philadelphia, PA, pp. 295–302.
- [21] Sumita, E., Akiba, Y., Doi, T., et al., 2003. A Corpus-Centered Approach to Spoken Language Translation. Conf. of the Europ. Chapter of the Ass. For Computational Linguistics (EACL), Budapest, Hungary, pp. 171–174.
- [22] Casacuberta, F., Vidal, E., 2004. Machine translation with inferred stochastic finite-state transducers. *Computational Linguistics* 30 (2), 205–225.



- [23] Och, J., Ney, H., 2003. A systematic comparison of various alignment models. *Computational Linguistics* 29 (1), 19–51.
- [24] Koehn, P., 2004. Pharaoh: a beam search decoder for phrase-based statistical machine translation models. AMTA.
- [25] ESIGN project: <<http://www.sign-lang.uni-hamburg.de/eSIGN/>>.
- [26] ECHO corpus: <<http://www.let.kun.nl/sign-lang/echo/>>.
- [27] Liu Cai, “a kind of advanced Sobel image edge detection algorithm”, *Guizhou Industrial College Transaction (Natural Science Edition)*,2004, 33(5):77-79.
- [28] Baris Sumengen, B. S. Manjunath “ multi-scale edge detection and image segmentation”, ece dept. UC, santa Barbara, USA,(National Conference)2005.
- [29] Lily Rui, Liang, Carl G., Looney, “Competitive Fuzzy Edge Detection”, *Applied Soft Computing*, Vol.3, no.4, pp. 123-137, 2003.
- [30] G. Pajares, “A Wavelet-based Image Fusion Tutorial”, *Pattern Recognition*, Vol. 37, no.10, pp. 1855-1872, 2004.
- [31] J.N. Ellinas, M.S. Sangriotis, “Stereo Image Compression Using Wavelet Coefficients Morphology”, *Image and Vision Computing*, Vol.22, no.2, pp. 281-290, 2004.
- [32] Kuhl, F. P. and Giardina, C. R., *Elliptic Fourier Descriptors of a Closed Contour*, *CVGIP*, **18**,pp. 236–258, 1982
- [33] Lin C. C. and Chellappa, R., *Classification of Partial 2D Shapes using Fourier Descriptors*,*IEEE Trans. PAMI*, **9**(5), pp. 686–690, 1987.
- [34] Montiel, E., Aguado, A. S. and Zaluska, E., *Fourier Series Expansion of Irregular Curves*, *Fractals*, **5**(1), pp. 105–199, 1997.
- [35] Persoon, E. and Fu, K.-S., *Shape Description Using Fourier Descriptors*, *IEEE Trans. SMC*, **3**, pp. 170–179, 1977
- [36] [I] M. Sugeno, “An introductory survey of fuzzy control,” *Inform. Sci.*, vol. 36, pp. 59-83, 1985.
- [37] C. C. Lee, “Fuzzy logic in control systems: Fuzzy logic controller-Part-I and Part-II,”*IEEE Trans. Syst., Man, Cybem.*, vol. 20, no. 2, pp. 404-435, 1990.
- [38] T. Takagi and M. Sugeno, “Fuzzy identification of systems and its applications to modeling and control,” *IEEE Trans. Syst., Man, cybern.*, vol. 15, no. 1, pp. 116-132, 1985.
- [39] L. X. Wang and I. M. Mendel, “Generating fuzzy rules by learning from examples,” *IEEE Trans. Syst., Man, Cyben.*, vol. 22, no. 6, pp. 1414-1427, 1992.
- [40] M. Sugeno and T. Yasukawa, “A fuzzy-logic-based approach to qualitative modeling,” *IEEE Trans. Fuzzy Systems*, vol. 1, no. 1, pp. 7-31, 1993.
- [41] H. Ishibuchi, K. Nozaki, and H. Tanaka, “Distributed representation of fuzzy rules and its application to pattern classification,” *Fuuy Sets and syst.*, vol.52, pp 21-32, 1992.

# An Analysis and Comparison of Quality Index Using Clustering Techniques for Spot Detection in Noisy Microarray Images

## A. Sri Nagesh

Faculty, Computer Science & Engineering Department,  
R.V.R. & J.C. College of Engineering,  
Guntur -522019. India

asrinagesh@gmail.com

## Dr. G. P. Saradhi Varma

Faculty, Information Technology Department,  
S.R.K.R.Engineering. College,  
Bhimavaram -534204. India

gpsvarma@yahoo.com

## Dr.A.Govardhan

Faculty, Computer Science & Engineering Department,  
JNTUCEH, Jagtial,  
Hyderabad. India

govardhan\_cse@yahoo.co.in

## Dr. B. Raveendra Babu

Director, Operations,  
Delta Technologies,  
Hyderabad. India

rbhogapathi@yahoo.com

---

### Abstract

In this paper, the proposed approach consists of mainly three important steps: preprocessing, gridding and segmentation of micro array images. Initially, the microarray image is preprocessed using filtering and morphological operators and it is given for gridding to fit a grid on the images using hill-climbing algorithm. Subsequently, the segmentation is carried out using the fuzzy c-means clustering. Initially the enhanced fuzzy c-means clustering algorithm (EFCMC) is implemented to effectively clustering the image whether the image may be affected by the noises or not. Then, the EFCM method was employed the real microarray images and noisy microarray images in order to investigate the efficiency of the segmentation. Finally, the segmentation efficiency of the proposed approach was compared with the various algorithms in terms of quality index and the obtained results ensures that the performance efficiency of the proposed algorithm was improved in term of quality index rather than other algorithms.

**Keywords:** Microarray Image, Genes, Spot Segmentation, Morphological Operator, Fuzzy K-Means, Fuzzy C-means, Enhanced fuzzy C-means Clustering (EFCM).

---

## 1. INTRODUCTION

In this research, we have proposed an efficient approach for microarray image segmentation to quantify the intensity of each spot and locate differentially articulated genes. The proposed approach contains three important steps such as, preprocessing, gridding and segmentation. The preprocessing stage contains the following process such as, top-hat filtering, binarization and morphological operations. Subsequently, the preprocessed image is given to the gridding process to accurately fit a grid into a spot. Here, we make use of hill climbing algorithm to effectively spot the grids on microarray using objective functions. Then, the image is given to the proposed clustering algorithm for segmentation. The designed clustering algorithm improves the microarray image segmentation by taking the advantage of spatial information along with the gray level pixel values. Furthermore, we have introduced the neighborhood fuzzy factor in the proposed clustering algorithm to effectively handling the spatial and intensity values in order to find the

appropriate cluster. The neighborhood fuzzy factor can be able to accurately detect the absent spots as well as the noisy spots.

The organization of the paper is as follows: Section 2 presents a brief review of some recent significant researches in Microarray image segmentation. The properties of proposed methodology for microarray image segmentation utilizing the enhanced fuzzy c-means clustering algorithm are 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.

## **2. REVIEW OF RELATED WORKS**

Numerous researches based on gridding and clustering techniques have been proposed by researchers for the segmentation of microarray images. A brief review of some important contributions from the existing literature is presented in this section.

Luis Rueda and Juan Carlos Rojas [5] have proposed a pattern recognition technique based method for DNA micro array image segmentation. Using a clustering algorithm, the method has first performed an unsupervised classification of pixels, and the resulting regions have been subsequently subjected to supervised classification. Further fine tuning is achieved by discovering and merging region edges, and eliminating noise from the spots by morphological operators. The reasonable potential of the proposed technique for segmentation of DNA micro array images has been demonstrated by the very high accuracy obtained by the results on background and noise separation in various micro array images.

Volkan Uslan and Dhsan Omur Bucak [6] have performed a study in the microarray image processing to make a fine difference against the gene expressions. They have experimented and compared two methods for this. In particular, the segmentation phase of the microarray image has been analyzed. Clustering techniques have been used in addition to the segmentation methods utilized in commercial packages. They have examined the results of the application of fuzzy c-means and k-means techniques.

Maroulis D. and Zacharia E. [2] have presented a morphological modeling of spots based automatic micro array images segmenting approach. The proposed approach has been shown to be extremely effective even for noisy images and images with spots of diverse shapes and intensities by the carried out experiments.

## **3. PROPOSED METHODOLOGY FOR MICROARRAY SEGMENTATION USING CLUSTERING TECHNIQUES**

The proposed approach consists of three important steps: preprocessing, gridding and segmentation. Initially, the microarray image is preprocessed using filtering and morphological operators and it is given for gridding to fit a grid on the images using hill-climbing algorithm. Subsequently, the segmentation is carried out using the proposed clustering algorithm, which is developed utilizing the fuzzy c-means clustering. The enhanced fuzzy c-means clustering algorithm (EFCMC) proposed in this paper makes use of the neighborhood pixel information along with the gray level information to effectively clustering the image whether the image may be affected by the noises or not. Then, the proposed method was employed the real microarray images and noisy microarray images in order to investigate the efficiency of the segmentation.

### **3. (A) PROPERTIES: ABSENT SPOT DETECTION AND NOISE TOLERANCE**

#### **Property 1: Absent Spot Detection**

In general, the main challenge behind the microarray image segmentation is to accurately detect the absent spots (case 1) and in addition to accurately segment the high intensity spots (case 2). By looking into these challenges, case 2 can be easily achieved by the conventional clustering algorithms. But, the absent spots can be very difficult to find by the traditional clustering algorithms so in order to detect absent spots accurately, we make use of the neighborhood

dependent fuzzy factor to balance the image details whenever the membership values of the pixel values is calculated. The factor proposed in the clustering algorithm is adaptively changed in all iteration by considering the intensity values of neighborhood pixels and thus preserving the insensitiveness to boundary values by converging it to the central pixel's value.

**Property 2: Noise Tolerance**

The proposed algorithm can efficiently tackled the following two challenges even if the microarray image is corrupted by the noises.

**Case 1:** if the central pixel is not affected by the noise and some pixels within its neighbors may be corrupted by noise.

**Case 2:** if the central pixel is corrupted by noise and the other pixels within its neighbors may not be corrupted by noise. These two cases are easily dealt with the proposed clustering algorithm due to the introduction of the neighborhood dependent fuzzy factor which is easily ignoring the added noises. On the other hand, it can be adaptively adjusted their membership values according to its neighborhood pixel so that the segmentation accuracy of the proposed clustering algorithm can be improved even if the image is corrupted by the noises.

**3. (B) QUALITY ASSESSMENT ANALYSIS**

The input image taken for microarray image segmentation is given to the proposed algorithm to obtain the segmented results. Then, the quality index is computed based on the definition given below to assess the quality of the proposed approach. The quality index given in [7, 8] is used to evaluate the performance of the proposed approach in microarray image segmentation. The quality index is defined as follows,

$$q_{index}(IDSpot) = \frac{q_{com2R}(IDSpot) + q_{com2G}(IDSpot)}{2} \tag{1}$$

$$q_{com} = (q_{size} * q_{sig-noise} * q_{bkg1} * q_{bkg2})^{1/4} * q_{sat} \tag{2}$$

$$q_{size} = \exp(-|F_{pixel} - F_{pixel0}| / F_{pixel0}) \tag{3}$$

$$q_{sig-noise} = F_{mean} / (F_{mean} + B_{mean}) \tag{4}$$

$$q_{bkg1} = f1 / (BSD / B_{mean}), f1 = 1 / [\max(BSD / B_{mean})] \tag{5}$$

$$q_{bkg2} = f2 * (bkg0 / (bkg0 + B_{mean})), f2 = 1 / [\max(bkg0 / (bkg0 + B_{mean}))] \tag{6}$$

$$q_{sat} = \begin{cases} 1 & ; \% sat \leq 10 \\ 0 & ; else \end{cases} \tag{7}$$

Where,  $F_{pixel}$  → number of pixel per spot

$F_{pixel0}$  → Average number of pixel per spot

$F_{mean}$  → Mean of foreground pixel intensities per spot

- $B_{mean}$  → Mean of local background pixel intensities
- $BSD$  → Standard deviation of local background per spot
- $bkg0$  → Global average of background per array
- $\%sat$  → Percentage of saturated pixel per spot

Here,  $q_{size}$  assesses the irregularities of spot size,  $q_{sig-noise}$  is a measure for the signal-to-noise ratio,  $q_{bkg1}$  quantifies the variability in local background and  $q_{bkg2}$  scores the level of local background.

Here, the quality index is computed for each spots presented in the microarray image after applying the clustering algorithms such as, k-means clustering, FCM and the proposed clustering (EFCMC). Then, the quality index obtained is plotted as graph shown in figure 4 and figure 5 for both channels. Fig 8.a and 9.a shows the input microarray image from different channels and Fig 4.b and 5.b illustrates the comparative quality index graph of the three algorithms corresponding to the input image. By analyzing these graphs, the proposed algorithm exactly detects the absent spots, which has zero quality index compared with other algorithms and at the same time, the intensity spots are accurately segmented since their quality index is greater than the other algorithms.

#### 4.1 Experimental Dataset

The performance of the proposed approach is carried out in a set of real microarray images obtained from the publically available database [9]. The image taken from the database contains 24 blocks and each block contains 196 spots, i.e. 14×14 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 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 [48, 1]. The overall segmentation results of the proposed approach are given in figure 2.

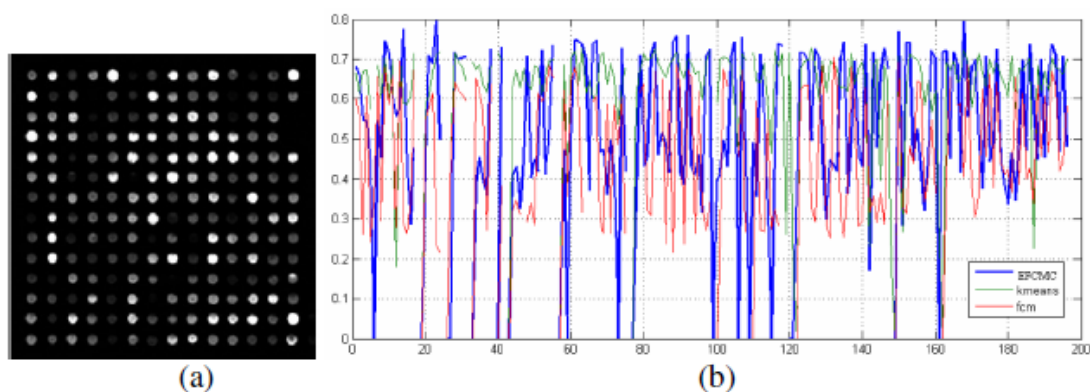
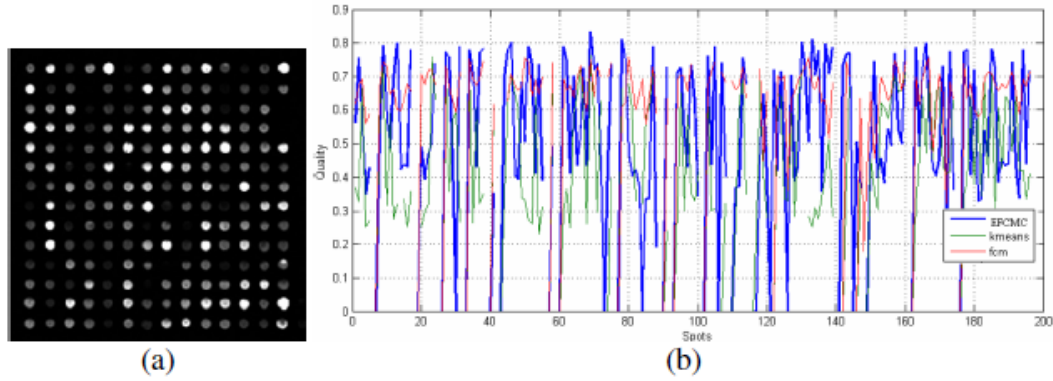


Fig. 1: (a) input microarray image-red channel (b) Comparative Quality index graph

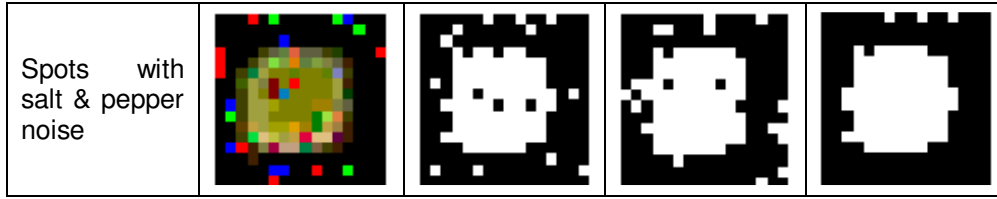


**Fig. 2:** (a) input microarray image-Green channel (b) Comparative Quality index graph

**4.3 Analysis: Absent Spot Detection and Noise Tolerance**

Here, we have analyzed the property of the proposed algorithm in identifying the low intensity spots and the detecting of absent spots. For analysis, we have taken typical spots, low intensity spots, absent spots and spots with various noises and then, different algorithms are applied on those spots to identify the efficiency of the algorithms. The obtained results are tabulated in the following figure 10. For a typical spot, three algorithms provide the identical results and for low intensity spots, FCM and EFCMC achieved better results compared with k-means clustering. The results obtained by the proposed algorithm for the absent spot is better compared with the k-means and FCM and those algorithms failed to identify the absent spots as per figure shown in below. For Gaussian and salt and pepper noise, the proposed approach accurately segments the spots and correctly removes the noisy pixels.

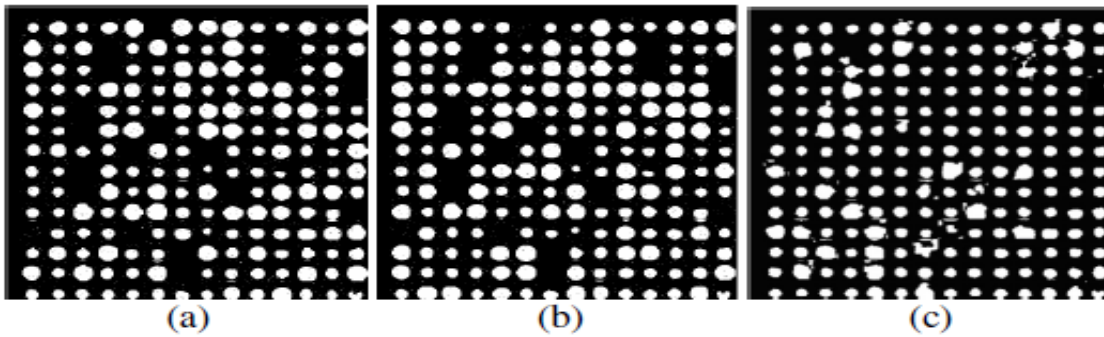
	Raw Image	K-means clustering	Fuzzy c-mean clustering	Proposed clustering
Typical spot				
Low intensity spot				
Absent spot				
Spots with Gaussian noise				



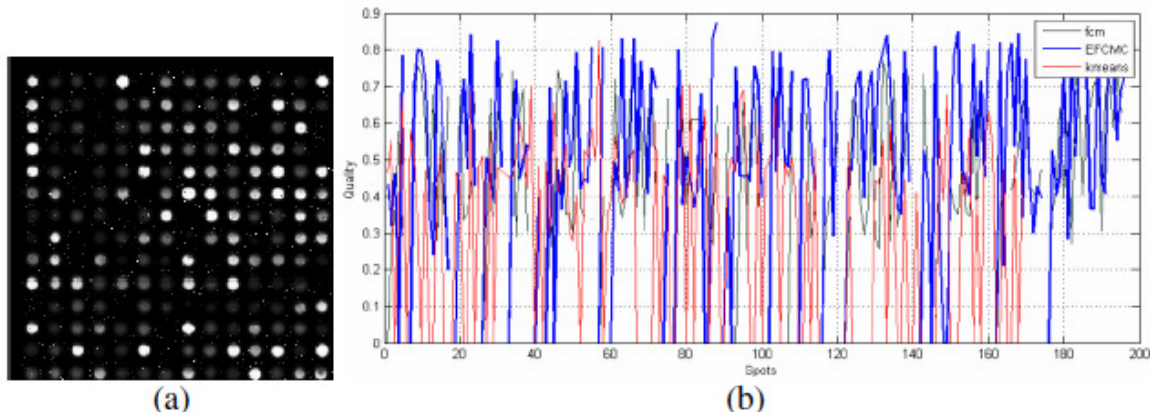
**Fig. 3:** Segmentation results in a typical spot, low intensity spot, absent spot and noisy spots using K-means, FCM and EFCMC

#### 4.4 Quality Assessment Analysis for the Noisy Images

The noise tolerance property of the proposed approach is analyzed by adding the salt & pepper noise in the microarray image. The input image is added with the salt & pepper noise and it is given to the proposed approach for segmentation. The results obtained by the proposed approach are used to compute the quality index so that the noise tolerance property is analyzed. The noisy input shown in fig.5.a and 7.a is given to the different algorithms, which provides the segmented results shown in figure 4 and 6. As per segmentation results of the proposed approach, the noisy pixels are exactly removed but in case of k-means and FCM, the noisy pixels still presented in the results. When we looking into the quality index graph shown in fig 5 and 7, the proposed approach provide the zero quality index for absent spots but other algorithms cant able to provide the accurate results for absent spot and at the same time, it falsely identify the absent spots.



**Fig. 4:** Segmentation results of salt & pepper -Green channel (a) k-means clustering (b) FCM clustering (c) EFCMC



**Fig. 5:** (a) Noisy input (salt & pepper)-Green channel (b) Quality index graph-Green channel



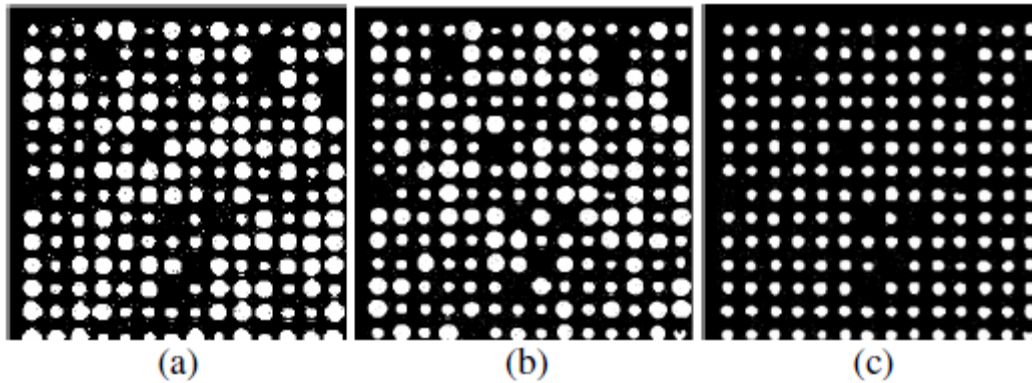


Fig. 6: Segmentation results of salt & pepper- Red channel (a) k-means clustering (b) FCM clustering (c) EFCMC

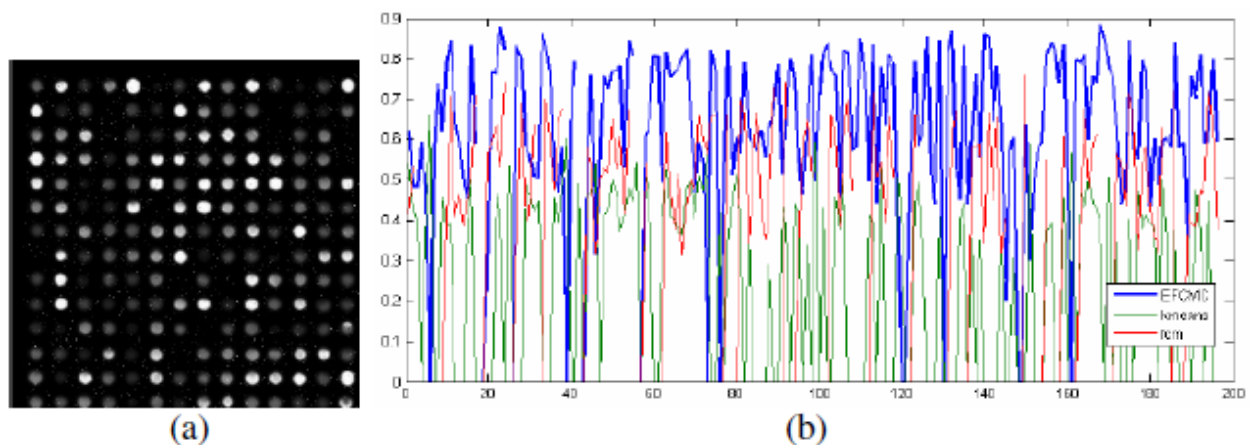


Fig. 7: (a) Noisy input (salt & pepper)-Red channel (b) Quality index graph-Red channel

## 5. CONCLUSION

In this paper, we developed and implemented and utilized an enhanced Fuzzy C-means clustering algorithm (EFCM) which was compared with the various algorithms in terms of quality index to investigate the performance efficiency in segmenting the microarray spot images. The comparative analysis proved that the proposed EFCM algorithm improved the quality index when compared with other algorithms.

## ACKNOWLEDGEMENTS

The Authors would like to thank all the authors and contributors for this outcome of the paper.

## 6. REFERENCES

- [1] Wu, H., Yan, H., "Microarray Image Processing Based on Clustering and Morphological Analysis", In First Asia Pacific Bioinformatics Conference, 111-118, 2003.
- [2] Maroulis D., Zacharia, E., "Microarray image segmentation using spot morphological model", in proceedings of the 9th International Conference on information Technology and Applications in Biomedicine, Larnaca, pp: 1-4, 2009.
- [3] A.Sri Nagesh, Dr.A.Govardhan,Dr G.P.S.Varma, Dr G.S.Prasad, "An Automated Histogram Equalized Fuzzy Clustering based Approach for the Segmentation of Microarray images."ANU Journal of Engineering and Technology, pp 42-48 volume 2, Issue 2 December 2010, ISSN: 0976-3414.



- [4] "Microarray Images", from <http://lmpp.nih.gov/lymphoma/data/rawdata/>
- [5] Luis Rueda, Juan Carlos Rojas, "A Pattern Classification Approach to DNA Microarray Image Segmentation", in Proceedings of the 4th IAPR International Conference on Pattern Recognition in Bioinformatics, 2009.
- [6] Kaushik Suresh, Debarati Kundu, Sayan Ghosh, Swagatam Das, Ajith Abraham and Sang Yong Han, "Multi-Objective Differential Evolution for Automatic Clustering with Application to Micro-Array Data Analysis", Sensors, Vol. 9, pp. 3981-4004, 2009.
- [7] Sebastiano Battiato, Gianpiero Di Blasi, Giovanni Maria Farinella, Giovanni Gallo and Giuseppe Claudio Guarnera, "Adaptive techniques for microarray image analysis with related quality assessment", vo. 16, no.4, 2007.
- [8] U. Sauer, C. Preininger, and S. R. Hany, "Quick & simple: quality control of microarray data", Bioinformatics, Advance Access, 2004.
- [9] Laurie Heyer, "MicroArray Genome Imaging & Clustering (MAGIC) Tool", Davidson College, Available: <http://www.bio.davidson.edu/projects/magic/magic.html>
- [10] 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, p. 116–21, 2003.
- [11] Volkan Uslan and Đhsan Ömür Bucak, "Microarray Image Segmentation Using Clustering Methods", Mathematical and Computational Applications, Vol. 15, No. 2, pp. 240-247, 2010.
- [12] A.Sri Nagesh, Dr G.P.S.Varma, Dr.A.Govardhan "An Improved Iterative Watershed and Morphological Transformation Techniques for Segmentation of Microarray Images" IJCA Special Issue on "Computer Aided Soft Computing Techniques for Imaging and Biomedical Applications" CASCT, 2010.

## Two Methods for Recognition of Hand Written Farsi Characters

**Mohammad Reza Jenabzade**  
*Department Of Computer  
Islamic Azad University, Yazd branch  
Yazd, Iran*

*jenabzadeh@gmail.com*

**Reza Azmi**  
*Department Of Computer  
Alzahra University  
Tehran, Iran*

*azmi@alzahra.ac.ir*

**Boshra Pishgoo**  
*Department Of Computer  
Alzahra University  
Tehran, Iran*

*boshra.pishgoo@gmail.com*

**Samanesadat Shirazi**  
*Department Of Computer  
Amir Kabir University Of Technology  
Tehran, Iran*

*sh.samaneh@gmail.com*

---

### Abstract

Optical character recognition (OCR) is one of the active bases of sample detection topics. The current study focuses on automatic detection and recognition of hand written Farsi characters. For this purpose; we proposed two different methods based on neural networks and a special post processing approach to improve recognition rate of Farsi uppercase letters. In the first method, we extracted wavelet features from borders of character images and learned a neural network based these patterns. In the second method, we divided input characters into five groups according to the number of their components and used a set of appropriate moment features in each group and classified characters by the Bayesian rule. In a post-processing stage, some structural and statistical features were employed by a decision tree classifier to reduce the misrecognition rate. Our experimental results show suitable recognition rate for both methods.

**Keywords:** Optical Character Recognition, Hand Written Farsi Characters, Neural Networks, Wavelet Transform, Decision Tree.

---

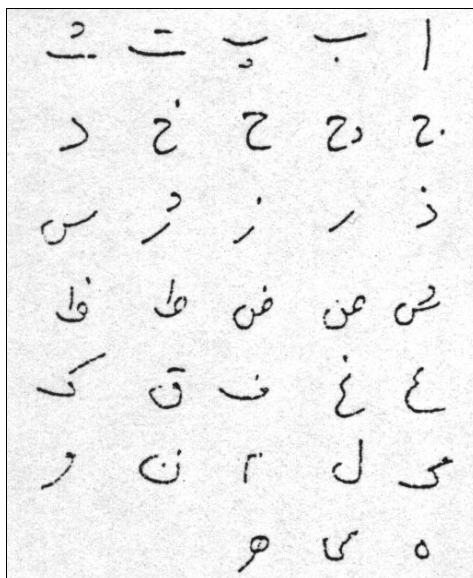
### 1. INTRODUCTION

In today world, almost all of the information is kept and processed by computers. However, using paper and paper forms are common in collecting information yet. The collected information in paper form must be changed to computer information anyway. A simple way of that is employment of many employees besides high costs, low speed and lots of mistakes to enter these forms into computers directly. The better way is an efficient method that a machine can extract and save the information of forms automatically. This automatic method is called Optical Character Recognition (OCR).

The application of systems based on OCR have been common in many commercial and industrial places like hospitals, banks, post offices, insurance offices and journal publishers [1-6]. In these applications, use of optical character recognition has two suitable properties [7]. i) the speed of information access is increased because there is possibility of search and edit in contrast to pictures and ii) the required space for information saving is decreased because texts files that extracted from pictures are usually take less space than picture files.

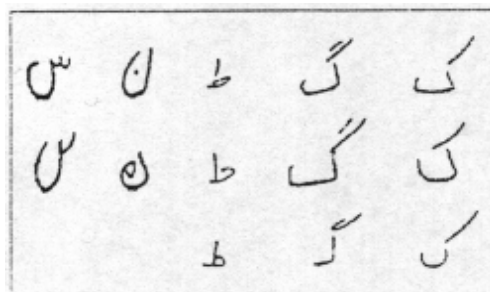
Totally, OCR systems are categorized into two domains [8]. One focuses on picture detection of letters after entrance to system (by scanner or digital camera) that is called offline recognition. In the other domain however, the writer enters the texts directly to system (by an optical pen) that is called online recognition. Online detection is easier than offline method because of some additional properties such as pen motion, the pen pressure and etc [9]. So in this study, we concentrate on hand written characters that are collected in offline mode.

Recognition of totally unconstrained hand-written characters is an attractive subject in the field of pattern recognition. Unlike English language, there has been only a few works on Arabic and Farsi characters recognition [10-14]. However, Farsi/Arabic texts have main specifications which make them difficult to recognize. Farsi/Arabic texts are cursive and are written from right to left. A Farsi/Arabic character might have several shape forms (1 to 4 shapes) depending on its relative position in the word. In addition, some Farsi/Arabic characters have the same shape and differ from each other only by existing of dots or zigzag bar. Figure 1 shows a sample set of upper case handwritten characters.



**FIGURE 1:** A sample set of Farsi Characters

The recognition algorithm, tolerates a high degree of style variation and distortion. Figure 2 shows this fact. However severely broken characters are mainly rejected by this algorithm.



**FIGURE 2:** Style variation and distortion

Totally, almost all of the OCR systems have four stages [15], that each of them has its own problems and effects on the system. These four stages are Pre-processing; feature extracting [16-18], character categorization and Post-processing [19-22]. In this paper, we concentrate on

these stages and present 2 methods for recognition of hand written Farsi characters. The remainder of this paper is organized as follows: sections 2 and 3 describe our proposed methods in details. Section 4 explains our dataset and illustrates experimental results for comparison our methods to each other. Finally section 5 concludes paper.

## 2. FIRST PROPOSED METHOD

In this method, first we describe pre-processing stage that consists of picture binary making and noise filtering in product scanning. There is property extraction of collected pictures next. In this part we give a picture body or a character to wavelet by border pixels as an input. The output is the properties that we want. The neural network is a simple and effective method of learning and training different samples that have efficient strength to noise. At the following subsections, we explain three stages pre-processing, feature extraction and recognition in this method in details. This method has no post-processing stage.

### 2.1 Pre Processing

In this step, binary making pictures and noise filtering are performed in scanning process. The saved pictures are in gray. Gray surfaces are necessary for defining the pen pressure and other information like that; but in suggested system only the main body of a structure is important and picture binary making causes the calculation volume decrease severely and program speed increases. Often the text optical character recognition systems are double characters pictures. In the proposed method the noises are cancelled first and then a picture is doubled. For noise canceling, total quantity of gray pixels is divided to the number of whole gray pixels and the resulting number is the grayness average of a picture. The threshold of grayness average is 10% of picture grayness and the pixels that are less than this limit are white pixels and the rest are black. While noise removing, we consider the dots as character's body.

### 2.2 Feature Extraction

In offline optical character recognition, the features are usually related to the items forms in pictures. The algorithm which is used for characters picture border pass is similar to the ones that described in [23] with small changes. The algorithm is designed for 8-neighbor connections and can detect the holes completely. After pass of letters picture border, the border following (cantor) is saved in  $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$  that number  $N$  shows the picture border pixels. The border pixels are re-sampled in 128 pixels for normalizing  $N$  (number of border pixels) in the collected pictures, close coefficients finding and features extraction in detail as a whole.

Different methods are used for feature extraction from sample picture border like chain code and transform of border follow to Fourier transform that the transform parameters are used for feature extraction [24]. The wavelet transform is also like Fourier transform for feature extraction from item picture cantor. The method is in a way of giving the picture cantor to a switch or changer and we use it coefficient as a feature set for detection. This formation can be either in continuous (integral) or none continuous (total) that we considered centralized continuous wavelet in the current study. For each function of  $f(t) \in L^2(\mathbb{R})$  in wavelet transform with several degrees of separation we have Equation (1).

$$f(t) = \sum_{k=-\infty}^{+\infty} c_{j_0}(k) \phi_{j_0,k}(t) + \sum_{k=-\infty}^{+\infty} \sum_{j=j_0}^{+\infty} d_j(k) \psi_{j,k}(t) \quad j_0 \in \mathbb{Z} \quad (1)$$

In above equation, first addition of function with low resolution or coarse approximation is shown and the second addition shows approximation function with high resolution which increases with  $j$  of function detail raise and the quality rises.  $c_{j_0}$  Is the approximation coefficient in  $j_0$  scale and the set of  $d_j$  are details of a signal in different resolutions. The wavelet transform used in the thesis is db1. For combining the coefficient in lower scale and gaining coefficient in higher scale, first the samples are taken up-sampling and then the resulting signal is filtered.

### 2.3 Character Recognition

As studied in previous subsection, our feature extraction method is wavelet transform. This method creates a feature vector for each character and then all of the vectors are normalized. In this method, our classifier is a neural network so at the next step, finalized vectors are taken to entrance layer of neural network as inputs. By consideration of the many examples and proper training, a neural network can recognize the samples that were not detected before which topic is useful for recognition of hand writings from a person to another. The applied neural network is Multi Layer Perceptron (MLP) that has 3 layers and each one is connected to the past layer. We train the network by an algorithm after mistakes derivation [25]. According to the point that we have 33 characters, the net output is made of 33 neurons that in system execution, the winner neuron shows the target character.

### 3. SECOND PROPOSED METHOD

In this method, first we divide input characters into five groups according to the number of their components and then use a set of appropriate moment features in each group and classify characters by the Bayesian rule. Finally, in a post-processing stage, some structural and statistical features are employed by a decision tree classifier to reduce the misrecognition rate. At the following subsections, we explain four stages pre-processing, feature extraction, recognition and post processing in this method in details.

#### 3.1 Pre Processing

The input character is scanned at a resolution of 200 dpi and its binary image is smoothed. The smoothing process makes the primary grouping less sensitive to broken characters. Using a component labeling algorithm based on 8-connectivity [26], each character is segmented into its parts. If the number of black pixels in a component is less than a threshold value, that component is filtered out as a noise. This threshold value is adjusted to the pen thickness. Input characters according to the number of their parts and dots are assigned to one of the five groups, shown in Figure 3.

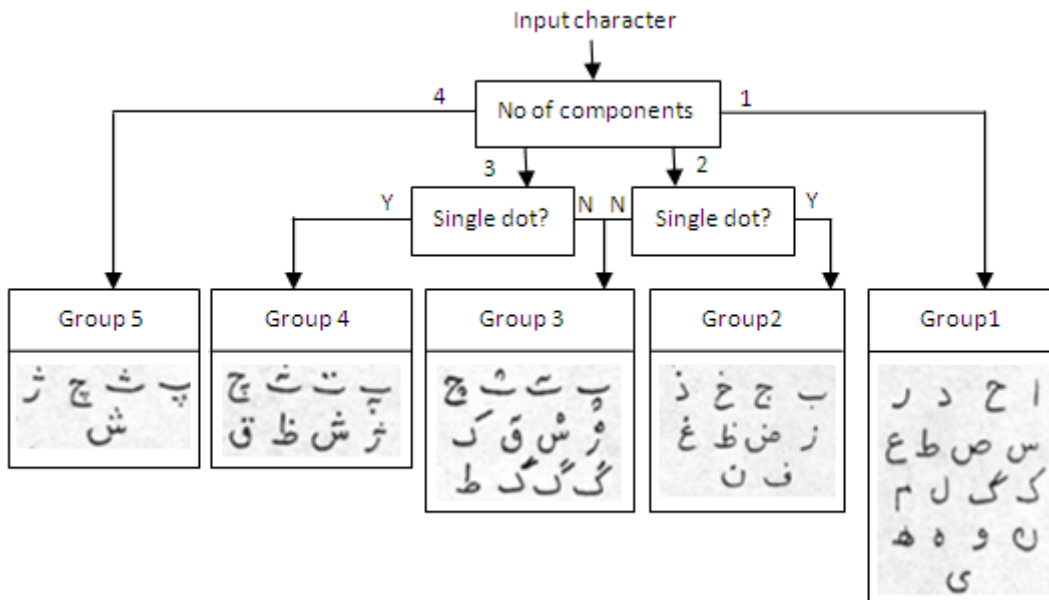


FIGURE 3: Primary grouping for input characters

#### 3.2 Feature Extraction

In this method, the central normalized moments which are defined as Equation (2) and (3) are chosen as main features [26].

$$\mu_{pq} = \sum \sum_{(m,n) \in R} (m - \bar{m})^p (n - \bar{n})^q f(m,n) \tag{2}$$

$$m_{pq} = \frac{\mu_{pq}}{\mu_{00} \binom{p+q-2}{2}} \tag{3}$$

If  $f(m,n)$  is a binary image of character,  $R$  is a region where  $f(m,n) \neq 0$ ,  $\mu_{pq}$  is central moment and  $m_{pq}$  is normalized central moment. To select an appropriate set of moments, 25 moments for  $p, q \leq 5$  were considered. Among these moments,  $m_{00}$ ,  $m_{01}$ ,  $m_{10}$  are always constant. From the remaining 22 moments, we select a set of moments for each group. Here an add-on feature selection method based on the total classification error was used [27].

### 3.3 Character Recognition

In this method, The Bayesian Rule was used for main classification. The distribution of features was considered to be multi dimensional normal one. The mean vectors and covariance matrices were estimated using the training samples for each class.

### 3.4 Post Processing

As shown in Table 1, some input characters are confused with other characters. For each output class we define a confusion set, consisting of the characters misrecognized in that class with an error rate more than 4%. For example the confusion set for the output classes “ح” and “ک” are “ع” and “زح اگ و ن ل ک ع ط ص س” respectively. To improve the recognition rate, in a post-processing stage, for any output character having non-empty confusion set, a heuristically decision tree is used. Decision is taken on the basis of some structural features and also other classes put forward by the Bayesian classifier as the lower priority options. For example, if the output class is “ح” the character loci features of the upper parts of the characters is used to recover those “ح” misrecognized as “ع”.

		Output class																
		ا	ح	د	ر	س	ص	ط	ع	ک	گ	ل	م	ن	و	ز	ه	ی
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Input character	ا	0	41	0	0	0	1	0	0	0	5	2	0	1	0	0	0	0
	ح	1	2	38	0	0	0	0	0	5	3	1	0	1	0	0	0	0
	د	2	0	0	47	0	0	0	2	0	0	0	0	0	0	0	0	1
	ر	3	0	0	1	40	0	0	0	0	7	1	0	0	0	0	0	1
	س	4	0	0	0	0	26	18	0	0	2	0	0	0	0	0	0	4
	ص	5	0	0	0	0	1	42	0	0	5	2	0	0	0	0	0	0
	ط	6	0	0	2	0	0	0	44	0	1	0	3	0	0	0	0	0
	ع	7	0	7	0	0	0	0	0	40	2	0	0	0	0	0	1	0
	ک	8	0	0	0	0	0	0	0	32	16	0	0	0	0	0	0	2
	گ	9	0	0	0	0	1	0	0	5	42	0	0	1	0	0	0	1
	ل	10	0	0	0	0	0	0	0	2	0	46	0	0	0	0	0	2
	م	11	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0
	ن	12	0	0	0	1	1	0	0	2	0	1	0	29	0	0	1	15
	و	13	0	0	0	2	0	0	1	0	1	3	0	0	42	0	0	1
	ز	14	0	0	1	0	0	0	1	0	0	0	0	0	0	32	11	5
	ه	15	0	0	0	0	0	0	1	0	0	0	0	0	0	1	48	0
	ی	16	0	1	1	0	9	1	0	0	0	0	0	2	0	1	0	35

TABLE 1: Confusion matrix for group 1 in a training stage (50 samples for training and 50 samples for test)

Table 2 shows the effect of post-processing stage on the character in Table 1. As one can see, the total recognition rate is increased using post processing.

		Output class																																
		ا	ب	پ	ت	ث	ج	چ	ح	خ	د	ذ	ر	ز	س	ش	ص	ط	ظ	ع	غ	ف	ق	ک	گ	ل	م	ن	و	ه	و	ی		
		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
Input character	ا	0	46	0	0	0	1	0	0	1	1	0	0	1	0	0	0	0	1	1	0	0	1	0	0	0	0	0	0	0	0	0		
	ب	1	2	45	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	1	1	0	1	0	0	0	0	0	0	0	0	0		
	پ	2	0	0	47	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0			
	ت	3	0	0	1	48	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		
	ث	4	0	0	0	0	46	1	0	0	0	2	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	1		
	ج	5	0	0	0	0	5	42	0	0	2	1	0	0	0	0	0	0	0	2	1	0	0	0	0	0	0	0	0	0	0	0		
	چ	6	0	0	0	0	0	0	49	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0		
	ح	7	1	1	0	0	0	0	0	0	46	1	0	0	0	0	0	0	46	1	0	0	0	0	0	0	0	0	1	0	0	0		
	خ	8	0	0	0	0	0	0	0	0	0	47	1	0	0	0	0	0	0	47	1	0	0	0	0	0	0	0	0	0	0	2		
	د	9	0	0	0	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	47	0	0	0	0	0	0	0	0	0	0	3		
	ذ	10	0	0	0	0	0	0	0	0	2	0	46	0	2	0	0	0	0	2	0	46	0	2	0	0	0	0	0	0	0	0		
	ر	11	0	0	0	0	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	50	0	0	0	0	0	0	0	0	0	0		
	ز	12	0	0	0	1	0	0	0	0	2	0	1	0	40	0	0	1	5	2	0	1	0	40	0	0	1	5	2	0	0	0		
	س	13	0	0	1	2	0	0	1	0	0	3	0	0	0	42	0	0	1	3	0	0	0	42	0	0	1	3	0	0	0	1		
	ش	14	0	0	0	0	0	0	2	0	0	0	0	0	1	0	45	2	0	0	0	0	1	0	45	2	0	0	0	0	0	2	0	
	ص	15	0	0	0	0	0	0	1	0	0	0	0	0	0	0	3	46	0	0	0	0	0	0	0	0	0	0	3	46	0	0		
	ط	16	0	1	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	46			

TABLE 2: Confusion matrix for group 1 after post-processing

#### 4. EXPERIMENTAL RESULTS

In this section we evaluate our proposed methods for recognition of hand written Farsi characters through two datasets. The first one consist of 6600 samples (200 samples for each character) written by different peoples and information of the other one is extracted from 420000 registration forms of national special talents high school and middle school tests [28].

For evaluation of the first method, wavelet function was considered Haar or db1. In property extraction for letters (33 categories),  $J_0$  was 3. In other words, we sampled to level 3 from beginning level of sampling downwards. Therefore, approximation coefficient of wavelet function  $cA_3$  was a vector with 17 entries and its details coefficients ( $cD_3, cD_2, cD_1$ ) were 3 vectors with 66, 33 and 17 entries in orders from up to down. According to coefficients vectors measure, all the coefficients together will cause slowing the recognition process and more mistakes probably occur. Therefore, all the detail coefficient sets must be considered from the coarsest (level 3) to level 2. Then, we will have 67 specifications (17, 17 and 33) for every border pixels set. Specifications measures are normalized and saved in a vector of 134\*1 for each picture.

The finalized vector is taken to entrance layer of neural network as an input which use reason is based on its actions. According to the point that we have 33 letters, the net output is made of 33 neurons that in system execution, the winner neuron shows the target letter. Therefore, the training and test sets are recognized with 95.17 % and 86.3 % accuracy.

We use confusion matrix for improving the results. Confusion matrix is usually used for training with an observer (and adaptation matrix without observer) [29]. The matrix finds similar classes to each other and decreases the categories numbers. In both cases of similarities, confusion matrix are gained that can be seen in Table 3.

Total Categories	Amount of similarity	Recognition of Train Samples (%)	Recognition of Test Samples (%)
33	0	95.17	86.3
26	Greater than 4 similarity	93.43	87.09
21	Equal and Greater than 4 similarity	97.24	91.51

TABLE 3: Letters Recognition with different categorizing numbers results

For the second method, Half of the samples were selected for the training and the other for the test. Table 4 shows the detailed results of the recognition process for each group in a training stage. As seen from this table, the post-processing stage improved the recognition rate for groups 1, 2 and 3 significantly.

Group	1	2	3	4	5
No Of Class	17	10	10	8	5
No of Moment	8	9	12	14	8
Recognition rate before Post-processing (%)	79.39	89.34	81.8	94	98.5
Recognition rate after Post-processing (%)	91.35	96.2	91.3	95	98.8

TABLE 4: No of features selected and recognition rate for each group in a training stage

Final result is shown as a confusion matrix in Table 5. The overall recognition rate in this experiment is 90.64%.

		Outputclass																																			
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33			
Input character	1	94	0	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0		
	2	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	3	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	4	0	0	1	95	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0		
	5	0	0	0	1	98	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	6	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0		
	7	0	0	0	2	2	0	76	0	0	0	0	0	0	8	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
	8	0	0	0	0	0	0	0	90	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	1	2	0	5	0	0	0	0	0		
	9	0	0	0	0	0	0	0	1	98	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	
	10	0	0	0	0	0	0	0	0	0	88	0	1	0	0	0	0	0	0	0	5	0	0	0	0	0	0	1	0	0	3	1	1	0	0		
	11	0	0	0	0	0	0	0	0	0	0	96	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0		
	12	3	0	0	0	0	0	1	0	1	0	88	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	0	4	0	0	0	0	0		
	13	0	0	0	0	0	0	0	0	1	0	5	0	89	0	0	0	0	0	0	1	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	
	14	0	0	0	0	0	0	1	0	0	0	0	0	0	98	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	
	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	89	0	1	0	0	0	0	0	0	0	8	1	0	0	1	0	0	0	0	0	0	
	16	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	95	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0	0	
	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	84	0	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	
	18	0	0	0	0	0	1	0	0	3	0	0	0	0	0	0	0	88	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0
	19	0	0	0	0	0	0	0	0	0	5	0	1	0	0	0	0	0	0	87	0	0	0	0	0	0	1	1	0	1	0	0	0	0	0	4	
	20	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	99	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
	21	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	90	0	0	0	0	1	0	1	0	2	0	0	0	0	0	
	22	0	0	0	0	0	0	1	0	6	0	0	0	0	0	0	0	0	0	0	0	0	93	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	23	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	96	0	0	0	0	0	0	0	0	0	0	0	0	
	24	0	0	1	1	7	0	2	0	0	0	0	0	9	0	6	0	1	0	0	0	0	0	0	72	0	1	0	0	0	0	0	0	0	0	0	
	25	0	0	1	1	0	0	0	0	0	0	0	2	0	2	1	3	0	0	0	0	0	0	0	2	82	2	0	0	1	0	0	3	0	0		
	26	0	0	0	1	0	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	5	88	0	0	0	0	0	1	0	0	0		
	27	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	8	0	0	0	0	5	86	0	0	0	0	0	0	0	0		
	28	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	2	95	0	0	0	0	0	0	0		
	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	1	0	0	90	0	0	0	0	0	1		
	30	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	2	0	0	0	0	0	0	2	0	0	1	90	1	0	0	0	0		
	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	1	0	7	0	82	2	2	0	0	0	2		
	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	2	0	6	0	2	83	1	0	0	0	0		
	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	2	1	93	0	0	

TABLE 5: Confusion matrix for second proposed method in recognition algorithm



## 5. CONCLUSION

In the current study we proposed a system for automatic recognition of hand written Farsi characters. For this purpose; we proposed two different methods based on neural networks and a special post processing approach to improve recognition rate of Farsi uppercase letters. In the first method, we extracted wavelet features from borders of character images and learned a neural network based these patterns. In the second method, we divided input characters into five groups according to the number of their components and used a set of appropriate moment features in each group and classified characters by the Bayesian rule. In a post-processing stage, some structural and statistical features were employed by a decision tree classifier to reduce the misrecognition rate. Our experimental results are shown suitable recognition rate for both methods.

## REFERENCES

- [1] Elmhurst, IL, "Optical Character Recognition and the Years Ahead", The Business Press, 1969.
- [2] Pas d'auteur, "Auerbach on Optical Character Recognition", Aurbach Publishers, Inc., Princeton, 1971.
- [3] G. Vamvakas, B. Gatos, I. Pratikakis, N. Stamatopoulos, A. Roniotis and S.J. Perantonis, "Hybrid Off-Line OCR for Isolated Handwritten Greek Characters", The Fourth IASTED International Conference on Signal Processing, Pattern Recognition, and Applications (SPPRA 2007), pp. 197-202, Austria, Feb. 2007.
- [4] W.R. Xu, H.G. Zhang, J. Guo, G. Chen, "Discrimination Between Printed and Handwritten Characters for Check OCR System", Proceedings of the first International Conference on Machine Learning and Cybernetics, Beijing, Nov. 2002.
- [5] M.D. Garris, D.L. Dimmick, "Form Design for High Accuracy Optical Character Recognition", IEEE Transactions PAMI, June 1996.
- [6] R. Fox, W. Hartmann, "An Abductive Approach to Hand-written Character Recognition for Multiple Domains", 2005.
- [7] B.Timsary, H.Fahimi, "Recognition Letters in Persian words typed using morphology", MS thesis, Department of Electrical Engineering, Isfahan University of Technology, 1992.
- [8] A. Aburas, S.M.A. Rehiel, "Off-line Omni-style Handwriting Arabic Character Recognition System Based on Wavelet Compression", ARISER, Vol. 3, No. 4, pp. 123-135, 2007.
- [9] M. Liana, G. Venu, "Offline Arabic Handwriting Recognition: A Survey", Transactions On Pattern Analysis and Machine Intelligence, IEEE, Vol. 28, No. 5, pp. 712-724, 2006.
- [10] Badie and M. Shimura, "Machine Recognition of Arabic Cursive Script ", Pattern recognition in Practice, E.S. Gelsema and L.N.Kanal (eds.), pp.315-323, North Holland publishing Company, 1980.
- [11] H. Al-Muallim and S. Yamaguchi, "A Method of Arabic Cursive Hand-Writing", IEEE trans. Patt. Annal. and Machine Intel., PAMI-9, No 5, pp. 715-722, sep. 1987.
- [12] T.S. Al-Sheikh and J.G. El-taweel, " Real Time Arabic-Handwritten Character Recognition", Proc. Int. Conf. Image Processing and its Application, pp. 212-216, Warwick, UK. July. 1989.

- [13] H.Al-Yousefi and S.S.Udpa, "Recognition of Arabic Characters", IEEE Trans. Patt. Analysis and Machine Intell. Vol. 14, No. 8, pp. 853-857, 1992.
- [14] K. Masruri and E. Kabir, "Recognition of Hand-Printed Farsi Characters by a Fuzzy classifier", Proc. Second Asia Conf. Computer Vision, ACCV'95, Singapore, pp. II.607-II.610, Dec. 1995.
- [15] S. Mori, H. Nishida, H. Yamada, "Optical Character Recognition", JohnWiley & Sons, NY, 1999.
- [16] H. Almuallim, S. Yamaguchi, "A method of recognition of Arabic cursive handwriting", Transactions on Pattern Analysis and Machine Intelligence, IEEE, Vol. PAMI-9, No. 5, Sept. 1987.
- [17] T. El-Sheikh, R. Guindi, "Computer recognition of Arabic scripts", Pattern Recognition, Vol. 21, No. 4, pp. 293-302, 1988.
- [18] M. El-Wakil, A. Shoukry, "On-line recognition of handwritten isolated Arabic characters", Pattern Recognition, Vol. 22, No. 2, pp. 97-105, 1989.
- [19] Ben Amor N., Essoukri N., "Combining a hybrid Approach for Features Selection and Hidden Markov Models in Multifont Arabic Characters Recognition", Conferences on Document Image Analysis for Libraries, IEEE, 2006.
- [20] Saeed M., Karim F, Hamidreza R., "Feature Comparison between Fractal Codes and Wavelet Transform in Handwritten Alphanumeric Recognition Using SVM Classifier", 7th International Conference on Pattern Recognition (ICPR'04), IEEE, 2004.
- [21] Saeed M., Karim F., Hamidreza R, "Recognition of Isolated Handwritten Farsi/Arabic Alphanumeric Using Fractal Codes", 7th International Conference on Pattern Recognition (ICPR'04), IEEE, pp. 104-108, 2004.
- [22] A. Mowlaei, K. Faez, A.T. Haghghat, "Feature Extraction with Wavelet Transform for Recognition of Isolated Handwritten Farsi/Arabic Characters and Numerals", IEEE, pp. 923-926, 2002.
- [23] <http://www.icaen.uiowa.edu/~dip/LECTURE/Segmentation2.html>
- [24] R.Azmi," Recognition printed Farsi texts", PhD thesis, Faculty of Engineering, Tarbiat Modarres University,1999.
- [25] Moller , Neural Networks, Vol. 6, 1993, pp. 525 to 533
- [26] A.K. Jain, " Fundamental of Digital Image Processing", Ch. 9, Prentice Hall, Englewood Cliffs, N.J. 1989.
- [27] A.N. Mucciardi and E.E. Gose, " A Comparison of Seven Techniques for Choosing Subsets of Pattern Recognition properties", IEEE Trans. Computers, Vol. C-20, No. 9,pp. 1023-1031, Sept. 1971.
- [28] S. Khosravi, F. Razzazi, H. Rezaei, M. R. Sadigh, "a Comprehensive Handwritten Image Corpus of Isolated Persian/Arabic Characters for OCR Development and Evaluation, Signal Processing and Its Applications, 9th International Symposium on Volume , Issue , 12-15, PP.1 – 4, 2007

## INSTRUCTIONS TO CONTRIBUTORS

The *International Journal of Image Processing (IJIP)* aims to be an effective forum for interchange of high quality theoretical and applied research in the Image Processing domain from basic research to application development. It emphasizes on efficient and effective image technologies, and provides a central forum for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing.

We welcome scientists, researchers, engineers and vendors from different disciplines to exchange ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 5, 2011, IJIP appears in more focused issues. Besides normal publications, IJIP intend to organized special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

### LIST OF TOPICS

The realm of International Journal of Image Processing (IJIP) extends, but not limited, to the following:

- Architecture of imaging and vision systems
- Character and handwritten text recognition
- Chemistry of photosensitive materials
- Coding and transmission
- Color imaging
- Data fusion from multiple sensor inputs
- Document image understanding
- Holography
- Image capturing, databases
- Image processing applications
- Image representation, sensing
- Implementation and architectures
- Materials for electro-photography
- New visual services over ATM/packet network
- Object modeling and knowledge acquisition
- Photographic emulsions
- Prepress and printing technologies
- Remote image sensing
- Autonomous vehicles
- Chemical and spectral sensitization
- Coating technologies
- Cognitive aspects of image understanding
- Communication of visual data
- Display and printing
- Generation and display
- Image analysis and interpretation
- Image generation, manipulation, permanence
- Image processing: coding analysis and recognition
- Imaging systems and image scanning
- Latent image
- Network architecture for real-time video transport
- Non-impact printing technologies
- Photoconductors
- Photopolymers
- Protocols for packet video
- Retrieval and multimedia

- Storage and transmission

- Video coding algorithms and technologies for ATM/p

### **CALL FOR PAPERS**

---

**Volume:** 6 - **Issue:** 1 - February 2012

**i. Paper Submission:** November 30, 2011

**ii. Author Notification:** January 01, 2012

**iii. Issue Publication:** January / February 2012

## **CONTACT INFORMATION**

### **Computer Science Journals Sdn Bhd**

B-5-8 Plaza Mont Kiara, Mont Kiara

50480, Kuala Lumpur, MALAYSIA

Phone: 006 03 6207 1607

006 03 2782 6991

Fax: 006 03 6207 1697

Email: [cscpress@cscjournals.org](mailto:cscpress@cscjournals.org)

CSC PUBLISHERS © 2011  
COMPUTER SCIENCE JOURNALS SDN BHD  
M-3-19, PLAZA DAMAS  
SRI HARTAMAS  
50480, KUALA LUMPUR  
MALAYSIA

PHONE: 006 03 6207 1607  
006 03 2782 6991

FAX: 006 03 6207 1697  
EMAIL: [cscpress@cscjournals.org](mailto:cscpress@cscjournals.org)