

International Journal of Image Processing (IJIP)



ISSN : 1985-2304



VOLUME 4, ISSUE 4

PUBLICATION FREQUENCY: 6 ISSUES PER YEAR

International Journal of Image Processing (IJIP)

Volume 4, Issue 4, 2010

Edited By
Computer Science Journals
www.cscjournals.org

Editor in Chief Professor Hu, Yu-Chen

International Journal of Image Processing (IJIP)

Book: 2010 Volume 4, Issue 4

Publishing Date: 30-10-2010

Proceedings

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. Violations are liable to prosecution under the copyright law.

IJIPJournal 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

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.

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)

Dr. Jane(Jia) You
The Hong Kong Polytechnic University (China)

Professor. Davide La Torre
University of Milan (Italy)

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

Dr. Huiyu Zhou
Queen's University Belfast (United Kindom)

Editorial Board Members (EBMs)

Professor. Herb Kunze
University of Guelph (Canada)

Assistant Professor. Yufang Tracy Bao
Fayetteville State University (North Carolina)

Dr. C. Saravanan
National Institute of Technology, Durgapur West Benga (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. Gerald Schaefer
Loughborough University (United Kingdom)

Dr. Chin-Feng Lee
Chaoyang University of Technology (Taiwan)

Associate Professor. Wang, Xiao-Nian
Tong Ji University (China)

Professor. Yongping Zhang
Ningbo University of Technology (China)

Table of Content

Volume 4, Issue 4, October 2010

Pages

- 287 – 292 Analysis of an Image Secret Sharing Scheme to Identify Cheaters
Jung-San LEE, Pei-Yu Lin, Chin-Chen Chang
- 293 – 306 2D Shape Reconstruction Based on Combined Skeleton-Boundary Features
j.komala lakshmi
- 307 – 319 A Quantitative Comparative Study of Analytical and Iterative Reconstruction Techniques
Shrinivas D Desai, Linganagouda Kulkarni
- 320 – 328 Color Image Segmentation Technique Using “Natural Grouping” of Pixels
Nirmalya Chowdhury, Biplab Banerjee , Tanusree Bhattacharjee
- 329 – 341 An Interactive Content Based Image Retrieval Technique and Evaluation of its Performance in High Dimensional and Low Dimensional Space
Nirmalya Chowdhury, Biplab Banerjee , Tanusree Bhattacharjee
- 342 – 353 Improving the Accuracy of Object Based Supervised Image Classification using Cloud Basis Functions Neural Network for High Resolution Satellite Images
Imdad Ali Rizvi, B.Krishna Mohan
- 354 – 367 Interferogram Filtering Using Gaussians Scale Mixtures in Steerable Wavelet Domain
Gh.S.El-Taweel , Ashraf. K. Helmy
- 368 – 376 New Approach: Dominant and Additional Features Selection Based on Two Dimensional-Discrete Cosine Transform for Face Sketch Recognition
Arif Muntasa

- 377 – 389 Performance Comparison of Face Recognition Using DCT Against Face Recognition Using Vector Quantization Algorithms LBG, KPE, KMCG, KFCC
Shachi J. Natu, Prachi J. Natu, Tanuja K. Sarode, H. B. Kekre
- 390 – 400 A Novel Method for De-warping in Persian document images captured by cameras
Hadi dehbovid, Farbod razzazi, Shapor alirezaee
- 401 – 420 A Novel and Robust Wavelet based Super Resolution Reconstruction of Low Resolution Images using Efficient Denoising and Adaptive Interpolation
Liyakathunisa, C.N.Ravi Kuamr
- 421 – 435 Separation of mixed Document Images in Farsi Scanned Documents Using Blind Source Separation
Hossein Ghanbarloo, Farbod Razzazi, Shahpur Alirezaei
- 436 – 443 Anisotropic Diffusion for Medical Image Enhancement
Nezamoddin N. Kachouie
- 444 – 455 Algorithm to Generate Wavelet Transform from an Orthogonal Transform
H. B. Kekre, Archana Athawale, Dipali Sadavarti

Analysis of an Image Secret Sharing Scheme to Identify Cheaters

Jung-San Lee

*Department of Information Engineering and
Computer Science Feng Chia
University Taichung, 40724,
Taiwan*

leejs@fcu.edu.tw

Pei-Yu Lin

*Department of Information Communication
Yuan-Ze University Chung-li, 32003,
Taiwan*

pagelin3@gmail.com

Chin-Chen Chang

*Department of Information Engineering and
Computer Science Feng Chia University
Taichung, 40724, Taiwan*

alan3c@gmail.com

Abstract

Secret image sharing mechanisms have been widely applied to the military, e-commerce, and communications fields. Zhao et al. introduced the concept of cheater detection into image sharing schemes recently. This functionality enables the image owner and authorized members to identify the cheater in reconstructing the secret image. Here, we provide an analysis of Zhao et al.'s method: an authorized participant is able to restore the secret image by him/herself. This contradicts the requirement of secret image sharing schemes. The authorized participant utilizes an exhaustive search to achieve the attempt, though, simulation results show that it can be done within a reasonable time period.

Keywords: Analysis, Secret image sharing, (t, n) -threshold, Cheater detection

1. INTRODUCTION

Shamir first introduced the concept of secret sharing in 1979 [10]. Given a set of participants $P = \{P_1, P_2, \dots, P_n\}$, each of them possesses a secret shadow generated from the secret S . Hereafter, any t out of n members can reveal S by collecting t secret shadows, i.e. (t, n) -threshold mechanism. In such a system, participants with fewer than t shadows have no more knowledge of the secret than the one with nothing. This can effectively enhance the security of communications in an insecure network.

Engineers extend this concept to protect confidential images. Due to its practicability, secret image sharing mechanisms have been widely applied to the military, e-commerce, and communications fields [2, 3, 4, 5, 6, 8, 12, 13]. As illustrated in Fig. 1, in a $(2, 5)$ -threshold scheme, while delivering a secret image $F-14$ to five authorized members, an image owner constructs several shadows from the original image in advance. Then, the owner issues each

member a distinct shadow. No one who possesses fewer than two shadows can learn anything about the secret image. Only when two authorized members provide their shadows can the secret image be restored.

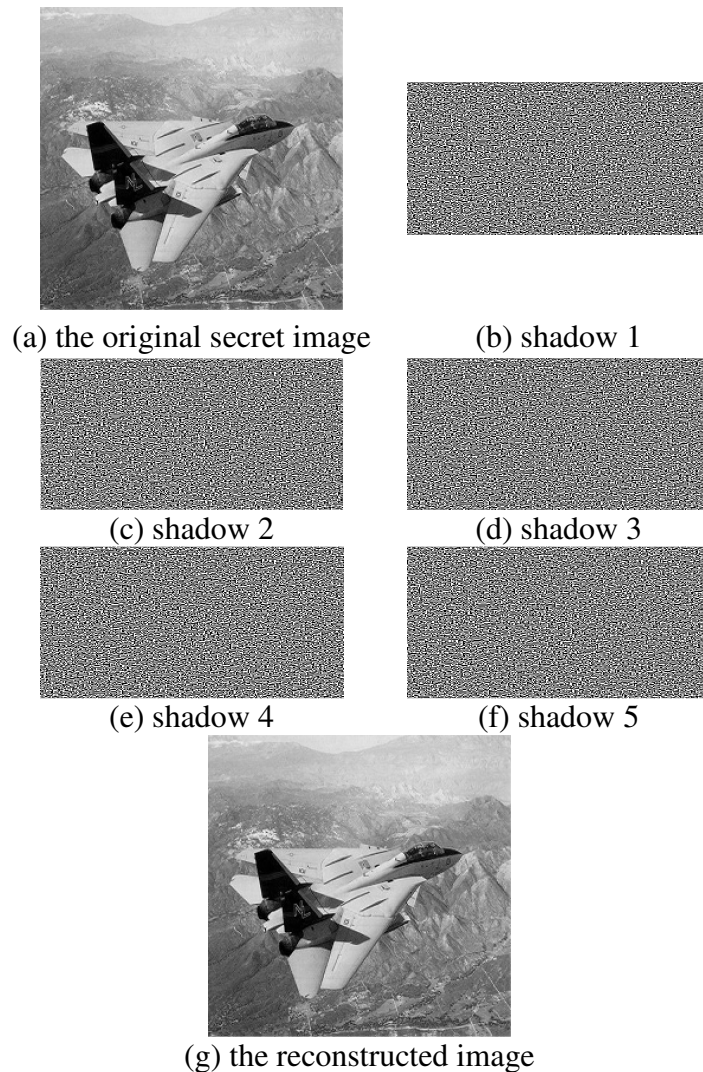


FIGURE 1: Secret image sharing: F-14

Recently, based on Thien and Lin's method, Zhao et al. proposed a novel secret image sharing scheme that introduces the concept of cheater identification [5, 7, 14]. This enables the image owner and authorized members to detect the cheater while reconstructing the secret image. It is claimed in [14] that their (t, n) - threshold method can confirm the following properties:

- i. Involved participants can detect cheaters no matter who they are;
- ii. At least t authorized participants can cooperate to reveal the secret image;
- iii. Participants can join in recovering different original secret image as long as they possess a secret shadow;
- iv. No secure channel is needed between the image owner and authorized participants;
- v. The size of shadow image is smaller than that of original secret image.

Unfortunately, we find that there exists a design weakness in Zhao et al.'s method: an authorized participant is able to figure out a congruent number of the private key of the image owner using the exhaustive search. Later, the participant can utilize this number to restore secret images

simply; this contradicts Property ii. Employing the exhaustive search, though, experimental results show that this can be done within a reasonable time period.

The rest of this article is organized as follows. We briefly introduce Zhao et al.'s mechanism in Section 2. The design weakness of the method is proven in Section 3. We make conclusions in Section 4.

2. REVIEW OF AN IMAGE SECRET SHARING SCHEME TO IDENTIFY CHEATERS

Zhao et al.'s method consists of three phases: initialization phase, construction phase, and verification phase. Assume that $P = \{P_1, P_2, \dots, P_n\}$ is the set of participants and any t out of n participants can cooperate to recover the secret image. Details of these phases are described as follows [14].

2.1 Initialization phase:

To begin with, the gray values of the secret image from 251 to 255 shall be truncated to 250 since 251 is the greatest prime not larger than 255. The image owner O selects two large primes (p, q) and computes $N = p \times q$. O picks a generator $g \in [N^{1/2}, N]$ and constructs an RSA-based public and private key pair (e_0, d_0) satisfying $e_0 \times d_0 = 1 \pmod{\varphi(N)}$. O publishes (e_0, g, N) [9,11].

Each participant $P_i \in P$ chooses a random number s_i ranged within $[2, N]$ as its secret shadow.

P_i computes $\alpha_i = g^{s_i} \pmod N$ and proves it to O . O shall ensure $\alpha_i \neq \alpha_j$ for $P_i \neq P_j$.

2.2 Construction phase:

Step 1: O computes $\alpha_0 = g^{d_0} \pmod N$ and $\beta_i = \alpha_i^{d_0} \pmod N, i = 1, 2, \dots, n$. O publishes α_0 .

Step 2: According to lexicography order, O divides the secret image I into several sections. For each section k containing t pixels, O constructs a $(t-1)$ th-degree polynomial as follows,

$$f_k(x) = a_0 + a_1x + \dots + a_{t-1}x^{t-1} \pmod{251}, \quad (1)$$

where a_0, a_1, \dots, a_{t-1} are the t pixels of section k .

Step 3: O computes $y_i = f_k(\beta_i)$, (2)

for $i = 1, 2, \dots, n$, and publishes y_1, y_2, \dots, y_n .

2.3 Verification phase:

Step 1: P_i utilizes its own secret shadow s_i to generate sub-secret $\beta_i = \alpha_0^{s_i} \pmod N$.

Step 2: Anyone can verify β_i by checking whether $\alpha_i = \beta_i^{e_0} \pmod N$ holds or not. If it holds, β_i is valid; otherwise, P_i may be a cheater.

Step 3: By collecting t pairs of (β_i, y_i) 's and the Lagrange interpolating polynomial, the participants can determine a $(t-1)$ th-degree polynomial as follows,

$$\begin{aligned} f_k(x) &= \sum_{j=1}^t y_j \prod_{i=1, i \neq j}^t \left(\frac{x - \beta_i}{\beta_j - \beta_i} \right) \pmod{251} \\ &= a_0 + a_1x + \dots + a_{t-1}x^{t-1} \pmod{251}. \end{aligned}$$

Hence, the secret image I is restored.

3. SECURITY ANALYSIS

In this section, we demonstrate that Zhao et al.'s (t, n) -threshold secret image sharing mechanism does not comply with Property ii. We first describe the system scenario. O possesses two secret images I_1 and I_2 . P_i keeps a secret shadow s_i and publishes $\alpha_i = g^{s_i} \bmod N$. For I_1 , according to Equation (1), O has to construct the polynomial

$$f_{k1}(x) = a_0 + a_1x + \dots + a_{t-1}x^{t-1} \bmod 251,$$

where a_0, a_1, \dots, a_{t-1} are the t pixels of section k . Furthermore, O must compute $\beta_i = \alpha_i^{d_0} \bmod N$ and publish $y_{i1} = f_{k1}(\beta_i)$, for $i = 1, 2, \dots, n$.

By the same manner, O computes the following polynomial for I_2 :

$$f_{k2}(x) = b_0 + b_1x + \dots + b_{t-1}x^{t-1} \bmod 251,$$

where b_0, b_1, \dots, b_{t-1} are the t pixels of section k . Moreover, O computes and publishes $y_{i2} = f_{k2}(\beta_i)$, for $i = 1, 2, \dots, n$, according to Equation (2).

Assume that P_i has joined in recovering the secret image I_1 and obtained the unique polynomial $f_{k1}(x) = a_0 + a_1x + \dots + a_{t-1}x^{t-1} \bmod 251$. We employ the following proposition to show the design weakness in [14].

Proposition: P_i can restore the secret image I_2 by itself.

Proof: Applying β_i to $f_{k1}(x)$, P_i yields the following:

$$\begin{aligned} y_{i1} &= a_0 + a_1(\beta_i \bmod 251) + \dots + a_{t-1}(\beta_i^{t-1} \bmod 251) \bmod 251 \\ &= a_0 + a_1(\alpha_i^{s_i} \bmod 251) + \dots + a_{t-1}((\alpha_i^{s_i})^{t-1} \bmod 251) \bmod 251 \\ &= a_0 + a_1(\alpha_i^{d_0} \bmod 251) + \dots + a_{t-1}((\alpha_i^{d_0})^{t-1} \bmod 251) \bmod 251. \end{aligned}$$

According to Fermat's Theorem [11] and the equation $\alpha_i^{d_0} \bmod 251$, P_i can fabricate d'_0 satisfying

$$d'_0 = d_0 \bmod 250.$$

That is, $0 \leq d'_0 \leq 249$. Using the exhaustive search, P_i can find an d'_0 answering to

$$\begin{aligned} \alpha_i^{d'_0} \bmod 251 &= \beta_i \bmod 251, \\ \alpha_i^{2d'_0} \bmod 251 &= \beta_i^2 \bmod 251, \\ &\vdots \\ \alpha_i^{(t-1)d'_0} \bmod 251 &= \beta_i^{(t-1)} \bmod 251. \end{aligned}$$

For the section k in I_2 , P_i collects $y_{12}, y_{22}, \dots, y_{t2}$, and constructs

$$\begin{aligned} y_{12} &= b'_0 + b'_1(\alpha_i^{d'_0} \bmod 251) + \dots + b'_{t-1}((\alpha_i^{d'_0})^{t-1} \bmod 251) \bmod 251, \\ y_{22} &= b'_0 + b'_1(\alpha_i^{2d'_0} \bmod 251) + \dots + b'_{t-1}((\alpha_i^{2d'_0})^{t-1} \bmod 251) \bmod 251, \\ &\vdots \\ y_{t2} &= b'_0 + b'_1(\alpha_i^{(t-1)d'_0} \bmod 251) + \dots + b'_{t-1}((\alpha_i^{(t-1)d'_0})^{t-1} \bmod 251) \bmod 251. \end{aligned}$$

Since $d'_0, \alpha_1, \alpha_2, \dots$, and α_i are known to P_i , P_i can obtain the following system of equations.

$$y_{12} = b'_0 + b'_1\delta_1 + \dots + b'_{t-1}\delta_1^{t-1} \bmod 251,$$

$$y_{22} = b'_0 + b'_1\delta_2 + \dots + b'_{t-1}\delta_2^{t-1} \pmod{251},$$

⋮

$$y_{t2} = b'_0 + b'_1\delta_t + \dots + b'_{t-1}\delta_t^{t-1} \pmod{251},$$

where $\delta_1 = \alpha_1^{d'_0} \pmod{251}, \delta_2 = \alpha_2^{d'_0} \pmod{251}, \dots, \delta_t = \alpha_t^{d'_0} \pmod{251}$. Hence, P_i has the $(t-1)$ th degree Vandermonde matrix:

$$A = \begin{bmatrix} 1 & \delta_1 & \delta_1^2 & \dots & \delta_1^{t-1} \\ 1 & \delta_2 & \delta_2^2 & \dots & \delta_2^{t-1} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ 1 & \delta_t & \delta_t^2 & \dots & \delta_t^{t-1} \end{bmatrix}.$$

As $\alpha_i \neq \alpha_j$, it implies $\delta_i \neq \delta_j$, for $i = 1, 2, \dots, n$, and

$$\det(A) = \prod_{1 \leq i < j \leq t} (\delta_j - \delta_i) \neq 0.$$

That is, A is a non-singular matrix. Thus, P_i can obtain a unique solution $\{b'_0, b'_1, \dots, b'_{t-1}\} = \{b_0, b_1, \dots, b_{t-1}\}$ from the system of equations. Eventually, P_i is able to restore the pixels of the secret image by itself. □

The proposition shows that Zhao et al.'s method violates Property ii. Even though P_i applies the exhaustive search, this attempt can be completed in 250 tries at most. We conduct experiments in the VC 6.0 language to confirm the feasibility of figuring out d'_0 . Simulators were performed on a PC with Intel L2300 CPU; the RSA algorithm was implemented according to the public OpenSSL library [1]. Since $g \in [N^{1/2}, N]$, the input size is set to 1024 bits. The length of module N is set to 256, 512, and 1024 bits, respectively. The running time of 250 tries is illustrated in Table 1. It is clear that P_i is able to imitate d'_0 within a very short time period under these cases.

	Module length (bit)		
	256	512	1024
Running time (second)	0.268151	0.411303	1.0416215

TABLE 1: Running Time under Different Module Length.

4. CONCLUSIONS

In this article, we have proven that an authorized participant can restore the secret image without the help of others in Zhao et al.'s (t, n) -threshold. Even if the compromise has been done by the exhaustive search, the simulation shows that it is completed within a rather short time interval.

ACKNOWLEDGMENT

This research is partially supported by the National Science Council, Taiwan, ROC, under contract No. NSC 98-2218-E-035-001-MY3. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

5. REFERENCES

- [1] The openssl project, <http://www.openssl.org>
- [2] T. S. Chen and C. C. Chang. "New method of secret image sharing based on vector quantization". *Journal of Electronic Imaging*, 10(4): 988-997, 2001
- [3] C. C. Chang and R. J. Hwang. "Sharing secret images using shadow codebooks". *Information Sciences*, 335-345, 1998
- [4] C. C. Chang, C. Y. Lin and C. S. Tseng. "Secret image hiding and sharing based on the (t, n) -threshold". *Fundamenta Informaticae*, 76(4): 399-411, 2007
- [5] C. Thien and J. Lin. "Secret image sharing". *Computer & Graphics*, 26(1): 765-770, 2002
- [6] J. B. Feng, H. C. Wu, C. S. Tsai and Y. P. Chu. "A new multi-secret images sharing scheme using Lagrange's Interpolation". *The Journal of Systems and Software*, 76: 327-339, 2005
- [7] R. J. Hwang, W. B. Lee and C. C. Chang. "A concept of designing cheater identification methods for secret sharing". *The Journal of Systems and Software*, 46: 7-11, 1999
- [8] R. Lukac and K. Plataniotis. "Colour image secret sharing". *Electronics Letters*, 40(9): 529-531, 2004
- [9] R. Rivest, A. Shamir and L. Adleman. "A method for obtaining digital signatures and public key cryptosystem". *Communications of the ACM*, 21(2): 120-126, 1978
- [10] A. Shamir. "How to share a secret". *Communications of the ACM*, 22(11): 612-613, 1979
- [11] W. Stallings. "Cryptography and Network Security – Principles and Practices", Pearson Education Inc., Fourth Edition, pp. 238-241 (2006)
- [12] C. S. Tsai, C. C. Chang and T. S. Chen. "Sharing multiple secrets in digital images". *The Journal of Systems and Software*, 64(2): 163-170, 2002
- [13] R. Wang and C. Su. "Secret image sharing with smaller shadow images". *Pattern Recognition Letters*, 27: 551-555, 2006
- [14] R. Zhao, J. J. Zhao, F. Dai and F. Q. Zhao. "A new image secret sharing scheme to identify cheaters". *Computer Standards & Interfaces*, 31(1): 252-257, 2009

2D Shape Reconstruction Based on Combined Skeleton-Boundary Features

J.komala Lakshmi

*Assistant Professor Department of Computer Science
SNR SONS college Research scholar,
Bharathiar University Coimbatore,
Tamil Nadu ,641006 India*

ashwathraju@yahoo.com

Dr. M. Punithavalli

*Director, Department Of Computer Science
Sri Ramakrishna College of Arts and
Sciences for woman, Coimbatore, 641
044, Tamil Nadu, India*

mpunitha_srcw@yahoo.co.in

Abstract

Reconstructing a shape into meaningful representation plays a strong role in shape-related applications. It is motivated by recent studies in visual human perception discussing the importance of certain shape boundary features as well as features of the shape area; it utilizes certain properties of the shape skeleton based on symmetry axes combined with boundary features based on curvature to determine protrusion strength. The main contribution of this paper is the combination of skeleton and boundary information by deploying the symmetry –curvature duality method to simulate human perception based on results of research in visual perception. The experiments directly compare our algorithm with experiments on human subjects. They show that the proposed approach meets the human perceptual intuition. In comparison to existing methods, our method gives a perceptually more reasonable and stable result. Furthermore, the noisy shape reconstruction demonstrates the robustness of our method ,experiments of different data sets prove the invariant representation of the combined skeleton-boundary approach.

Keywords: Protrusion, Symmetry-curvature duality, Contour, Bamboo boundary, Merging point.

1. INTRODUCTION

In all the research areas such as image retrieval and computer graphics, character recognition, image processing, and the analysis of biomedical images [1], the skeleton plays a major role for object representation and recognition. Skeleton-based representations [2] are the abstraction of objects, which contain both shape features and topological structures of original objects. Because of the skeleton's importance, many skeletonization algorithms [3] have been developed to represent and measure different shapes. A recent study on human perception [4] demonstrated that curvature properties of the boundary and the area properties of the enclosed regions affect the observers' identification of contour segments. Similar to previous results from [5], their experiments showed that negative minima of contour curvature depict segment boundaries. Additionally segment identification was determined by contour length, the turning angle at part boundaries and the width at the part's base. Motivated by these results, the proposed approach mimics the procedures of human perception assumed in [4] by combining concepts of skeleton and boundary in shape analysis. Our method shows that 'junction points' of the skeleton can provide the possibility of important protrusions. Boundary information as mentioned in [4], including contour length, the turning angle at part boundaries and the width at the part's base,

can determine the probability of important protrusions. The proposed reconstruction is then based on a definition of protrusion strength, based on both curvature of boundary points and their structural correspondence. Recently Bai et al. [6][7] presented robust approaches to compute a perceptually reasonably pruned discrete 2D skeleton of a single boundary. The author utilizes [8] this skeleton approach. It establishes a correspondence between structural information and boundary. Due to its robustness, we use this skeleton approach. Of course any skeleton algorithm offering 'junction points' is applicable.

2. PRELIMINARIES

Shape Reconstruction is an emerging field. It mainly splits into two classes of approaches: boundary and region based. Generally segmentation is best suited for uniform boundary images. Hence Segmentation splits the shapes into regions and the boundary of that particular region or contour is taken for processing. In [6], the author, segments the image into many number of contour based skeletal representation. In [7] Discrete skeletal evolution he identifies the skeleton evolving algorithm to identify the complete stable skeletal representation of the input image. In[9],The author utilizes the boundary feature and skeleton structure to decompose the input image. In the following sections, it will be seen that 'junction points' of the skeleton play a crucial role in finding a correct decomposition. The motivational connection between junction points and protrusions is the observation that a junction point emanates from the merging of two protrusions. Hence, if the junction points are known, the positions of protrusions can be predicted and the protrusions can be classified as parts or non-parts based on their strength. The experiments directly compare our algorithm with experiments on human subjects. This is mainly achieved through the crucial junction point information with the important boundary features.

3.Methodology of the proposed work

The image is seen into two classes. One from the outer shape boundary. The other is from the segment of the shape and its corresponding segment boundary or contour. The shape boundary is computed from the high positive maximum of the curvature and the segment boundary or contour is computed from the negative minima of curvature. In[10],the author explains between the axes that track the curvature maxima (positive axes and the curvature minima (negative axes) using a second derivative. In fig 3.1 it is depicted as follows.



Fig3.1 Positive (shown in blue) and negative (shown in red) axes

A. Reconstruction.

Deforming a given shape to a target shape has been a topic of interest in shape analysis. In a skeleton based representation, [11],[12],[13],[14],[15] one can obtain new shapes simply by modifying a skeletal representation. Note that even though small changes in the shape boundary may lead to significant changes in the shape skeleton, the opposite is not true [16]. Suppose we edit our skeletal representation by changing attributes or deleting/inserting a primitive. Since we have lost information as a result of excessive regularization, we cannot construct the new shape boundary simply by propagating the edited skeleton, as in [17],[18]. Therefore, we suggest an alternative solution by considering the transformations or diffeomorphisms, [19],[20], that transform one disconnection point set into another. Note that when such a transformation is found, it can be used to form a dense correspondence between two shape domains. The new shape boundary after the skeletal edit operations can be formed by applying the computed transformation to the original boundary points. In contrast to skeleton scale space methods which smooth the boundary first and then compute the skeleton, we, in a dynamic way, continuously smooth the shape boundary and record its interested points using visual data exploration and pixel connectivity till the whole boundary becomes simple enough. Hence we get the boundary as well as the skeleton representation collectively called as bamboo skeleton as in fig 4.2.1. As such, in [21][22],[23], the proposed representation is an unlabeled attributed point set and forms a trade-off between unstructured point sets, and in [24],[14],[25],[26] skeletal graphs. Skeletal points moving with a faster speed than the non-skeletal points can be detected by the minima of the gradient which are indirectly related to curvature maxima. We include the parameters symmetry, curvature duality, protrusion strength, width and the segment length. The relevant measure (high positive curvature maximum) for computing the boundary end point. Skeletal points moving with a faster speed than the non-skeletal points can be detected by the minima of the gradient which are indirectly related to curvature maxima. Hence it is worth exploring the algorithms that jointly determine point correspondences and estimate local deformation, [11], In [27] the author represents the unique, consistent and stable bamboo skeleton for all number of contour produced in [7].

B.Exploring point correspondences and local deformation.

The construction of skeleton is via maximal circles that are inscribable inside the shape and touch the shape boundary at more than one point. In this construction, circle radius plays the role of time of arrival in the symmetry axis function. Considering the envelope of maximal circles, one can reconstruct the shape boundary.

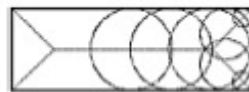


Fig3.b.1 maximal circles

In [28], [29],[30], While all of these constructions lead to the same representation, they inspired many others having different properties and being computed with different algorithms. No matter which major branch is chosen as a reference axis, the same axis must be chosen for similar shapes. Since there are two major axes of the same type, there is an ambiguity in the process. If the descriptions of two similar shapes depend on different coordinate frames, the matching algorithm will be unable to determine the similarities of shapes. This situation may necessitate creating at least two descriptions. To overcome this problem, we use each negative major symmetry branch as a reference axis for one half of the shape. Hence, the shape is described as two halves with each half having its own coordinate frame. This approach decreases the computation time of the matching algorithm drastically.

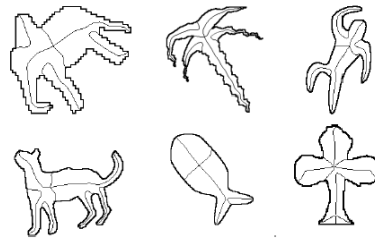


Fig 3.b.2 Symmetry axes for sample shapes after pruning.

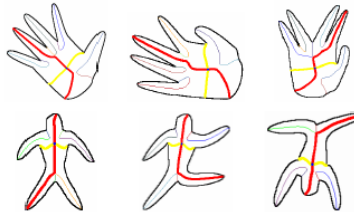


Fig 3.b.3

The major branches of the hand and human shapes. Positive major symmetry branches are shown in bold red and negative major symmetry branches are shown in bold yellow.

The *skeleton* S of a boundary B is the locus of the centers of maximal disks in correspondence to [10]. We use the discrete definition of a skeleton as given in [5]: using 8-neighborhood, a skeleton is a connected, thin point set $S = \{s_1 \dots s_n\}$ describing a geometric graph embedded in R^2 .

A shape *boundary* is a vector of points $B = \{b_1 \dots b_m\}$.

An *endpoint* $E_i \in S$ is a skeleton point having only one neighbor.

A *junction point* $J_k \in S$ is a skeleton point with three or more neighbors.

Given a junction point J_k of a skeleton, there is a set of corresponding boundary points intersecting with the maximal disk centered at J_k . We call these boundary points *tangent points* $t_i \in B$ of the junction J_k .

The shortest path between a pair of endpoints on a skeleton graph is called a *skeleton path* P ([31]). We call a partial path between junction or endpoints a *branch*.

With these definitions, we will now describe our decomposition approach. The goal is to find the end points of the part *lines* ([5]), i.e. straight connections between a pair of boundary points entirely inside of the shape boundary.

We first compute the skeleton using the Discrete Skeleton Evolution method as proposed in [5]. This approach first computes the skeleton based on Blum's medial axis definition. Then iterative skeleton pruning (removing end branches) is performed based on a relevance measure which describes the importance of the respective branch for shape reconstruction. The pruning process ends when only major relevant branches remain. We follow [10] and utilize a single threshold as stop criterion. Our experiments showed that this leads to visually correct skeletons in all examples. After the pruned skeleton is obtained, we choose 'relevant' (high positive curvature) pairs of tangent points as candidates to merge the shape boundary. The visual significance, and therewith the final decision for merging computed as protrusion strength and symmetry curvature duality is as follows.

There is a connection between positive curvature maxima of the shape boundary and the skeleton in the sense that each curvature maximum gives rise to a skeletal boundary [32]. This connection is often used to extract the skeleton. When the skeleton is extended to include branches that arise from curvature minima, a richer set commonly referred to as local symmetry axes is obtained [33]. The skeletal arc formed out of curvature minima contains the end points one at the boundary having high positive curvature and the other at the symmetry axes on the base

skeleton with high negative curvature. The protrusion strength is used to identify the junction point with curvature minima.

Following [9], our merging is guided by the following two rules:

- (1) Merging focuses on the turning angle of the part boundary, where the tangent points have positive maxima of curvature.
- (2) Segment identification performance (i.e. strength of protrusion) is related to the segment length and the width at the part's base

Protrusion strength:

With $c(.)$ denoting the curvature in a boundary point (concavities having negative curvature), compute the sum of curvature $C(i, j) = (c(l_i) + c(r_j))$ for all $(l_i, r_j) \in T L \times T R$. Also, for each (l_i, r_j) compute the *protrusion strength*, which is defined as follows:

$$P(l_i, r_j, b_k, r) = \frac{|b_k - r|}{|l_i - r_j|}$$

Where $|l_i - r_j|$ is the Euclidian distance of l_i and r_j , r is the radius of the maximal disk, see fig.3. b_k is the length of branch with junction point J_k , i.e. the length between J_k and the respective endpoint E . However, if another junction point \hat{J} is between J_k and E , which led to a cut before, b_k is defined as the (shorter) branch length between J_k and \hat{J} . With a given threshold T , the pair (l_i, r_j) , which minimizes the curvature value, is selected amongst all pairs of points with $P(l_i, r_j, b_k, r) > T$ $C(\hat{l}_i, \hat{r}_j) = \min\{C(i, j)\}$

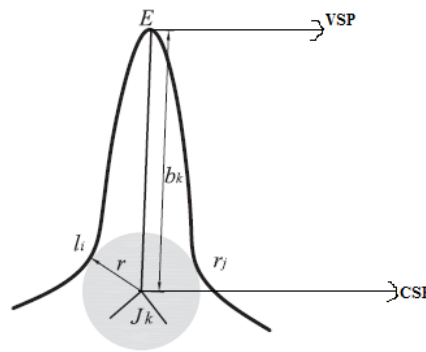


Fig 3.b.4 Computation of protrusion strength.

The skeleton branch is a line segment connecting the end points one at the boundary having curvature maximum and the other at the junction point of the symmetry axes/base skeleton having curvature minimum. The segment terminates at the curvature extremum of the opposite type.

C. Combining symmetry axes and curvature minima

According to symmetry-curvature duality theorem,[34]the tangent point $t_i \in B$ and the junction point $J_k \in S$ are the curvature extremum of the opposite type. According to the author in [9], $J_k \in S$ is a point with curvature minimum. $t_i \in B$ is a point with curvature maximum. According to symmetry axis the curvature maximum point gives raise to the boundary. Hence using pixel connectivity we connect all those $t_i \in B$, to obtain the boundary of the ununiform shape. This $E_i \in S$ is called as valance skeleton point (vsp)and $J_k \in S$ is called as core skeleton point (csp).This vsp is a point that can be used for merging and modifying the input image into another skeletal representation.

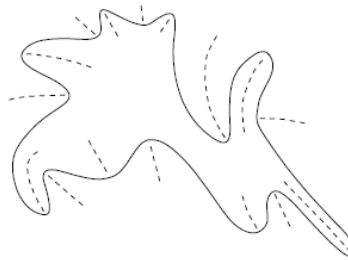


Fig 3.c.1

An illustration of the Symmetry-Curvature Duality Theorem

In [9] the protrusion strength is used to identify the inner most point as junction point having high negative curvature whereas our proposed method utilizes the by the symmetry –curvature duality theorem to compute the outer most point as valance skeletal point having high positive curvature and that can be used for further processing .

4 Experiments

4.1 Experiment on Abstract Shapes

Figure 4.1.1 gives the original data set with darkened boundary parts for test. The data set has different levels of part significance with respect to the part base line and the elongation: the base width for the test segment increases from top to bottom, the segment length for the test segment increases from left to right. The least significant segment is therefore located in the bottom row, left column. The result in [4] shows the trend that it was harder for observers to identify the segments of shapes shown at the bottom and left side compared to those at the top and right side of figure 4.1.2 For example, only about 40–50 percent of the observers identified the defined segment in (row/column) 3,1 and 4,1 as ‘significant’.

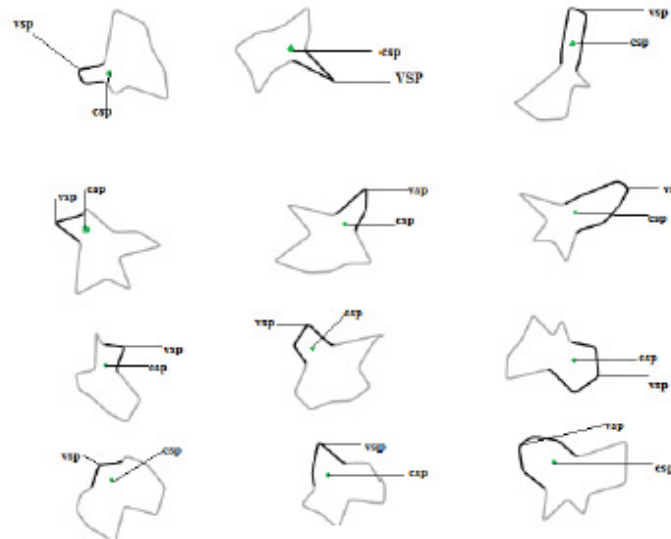


FIGURE 4.1.1

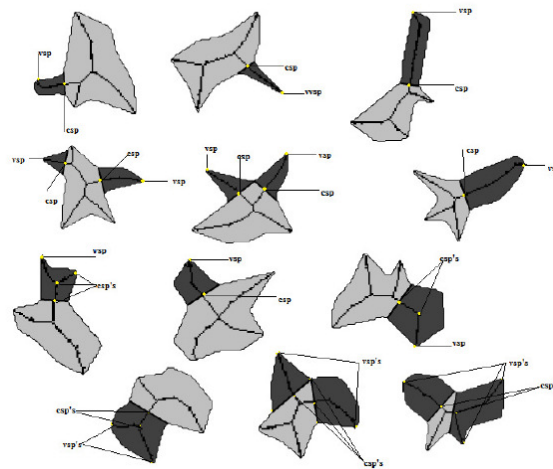


FIGURE 4.1.2

4.2 Experiments on Different Shapes.

A. Existing Method

The first and second row has seven shapes from the existing approach. It is different for any number contours.

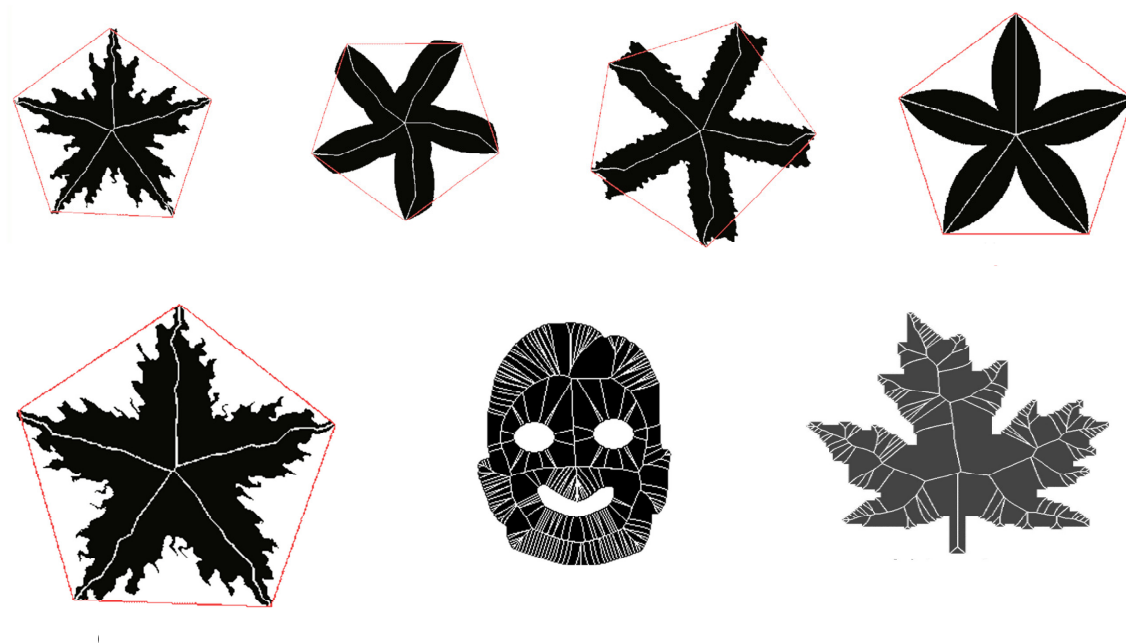


FIGURE 4.2.A.1

B.Proposed Bamboo skeleton Representation

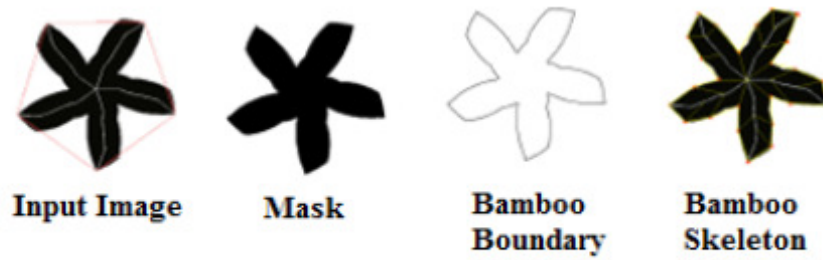


FIGURE 4.2.B.1

In [33]The first and second row has seven shapes from our approach. It is common for any number of contours.

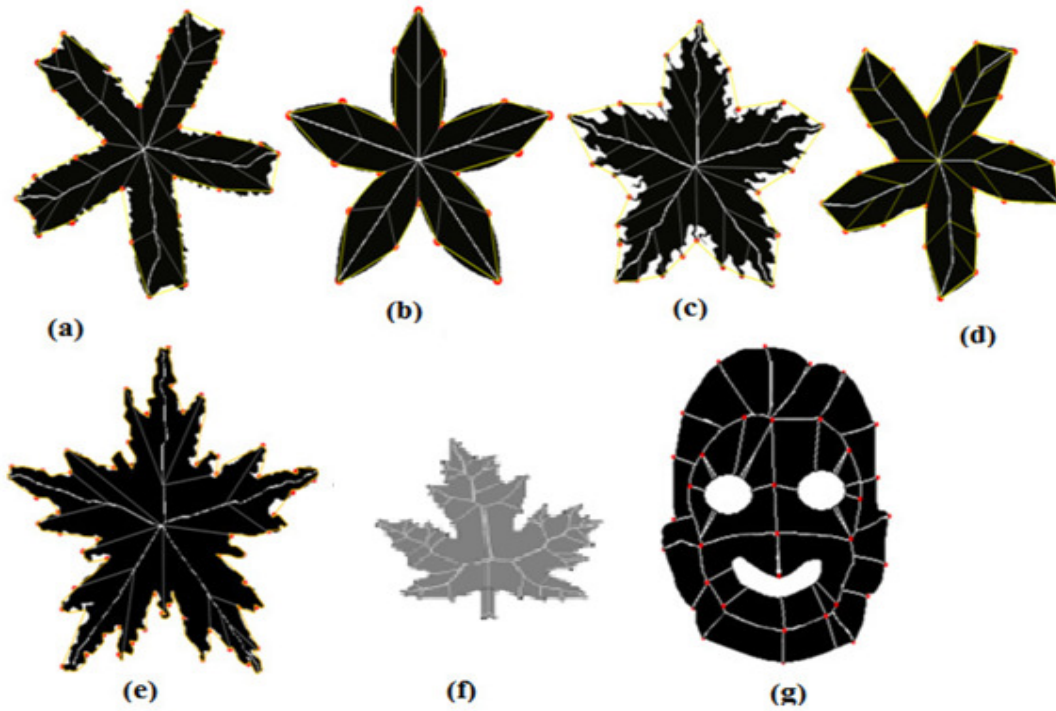


FIGURE 4.2.B.2

C.Proposed Merging Point Computation

In [35] the merging points are computed and labeled.

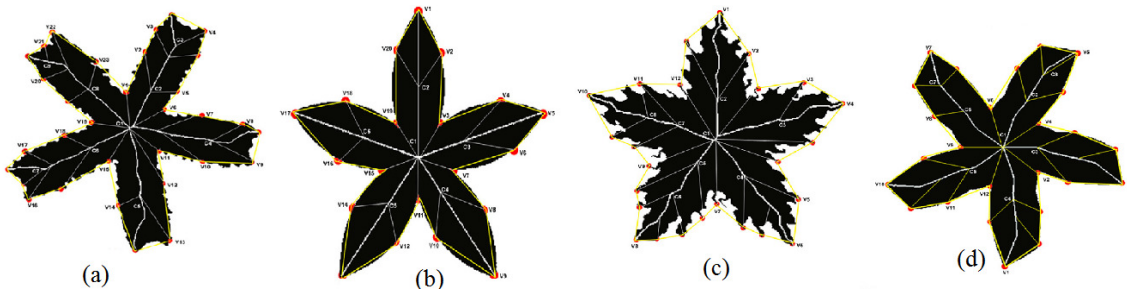


FIGURE 4.2.C.1 Merging points computed by our approach. The results is in accord with the experiment in[4]

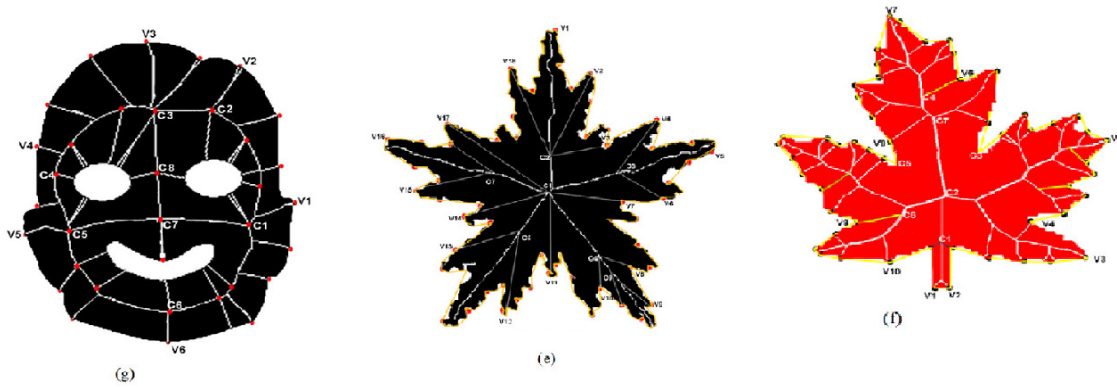


FIGURE 4.2.C.2: Merging points computed by our approach. The results is in accord with the experiment in [4]

4.3. Reconstruction results of our approach on different shapes.

This experiment shows reconstruction of different shapes, taken from [7], Figure 4.3.1 shows some results of the proposed algorithm. The consistent reconstruction of the object in row1 is especially remarkable, although heavily distorted by noise.

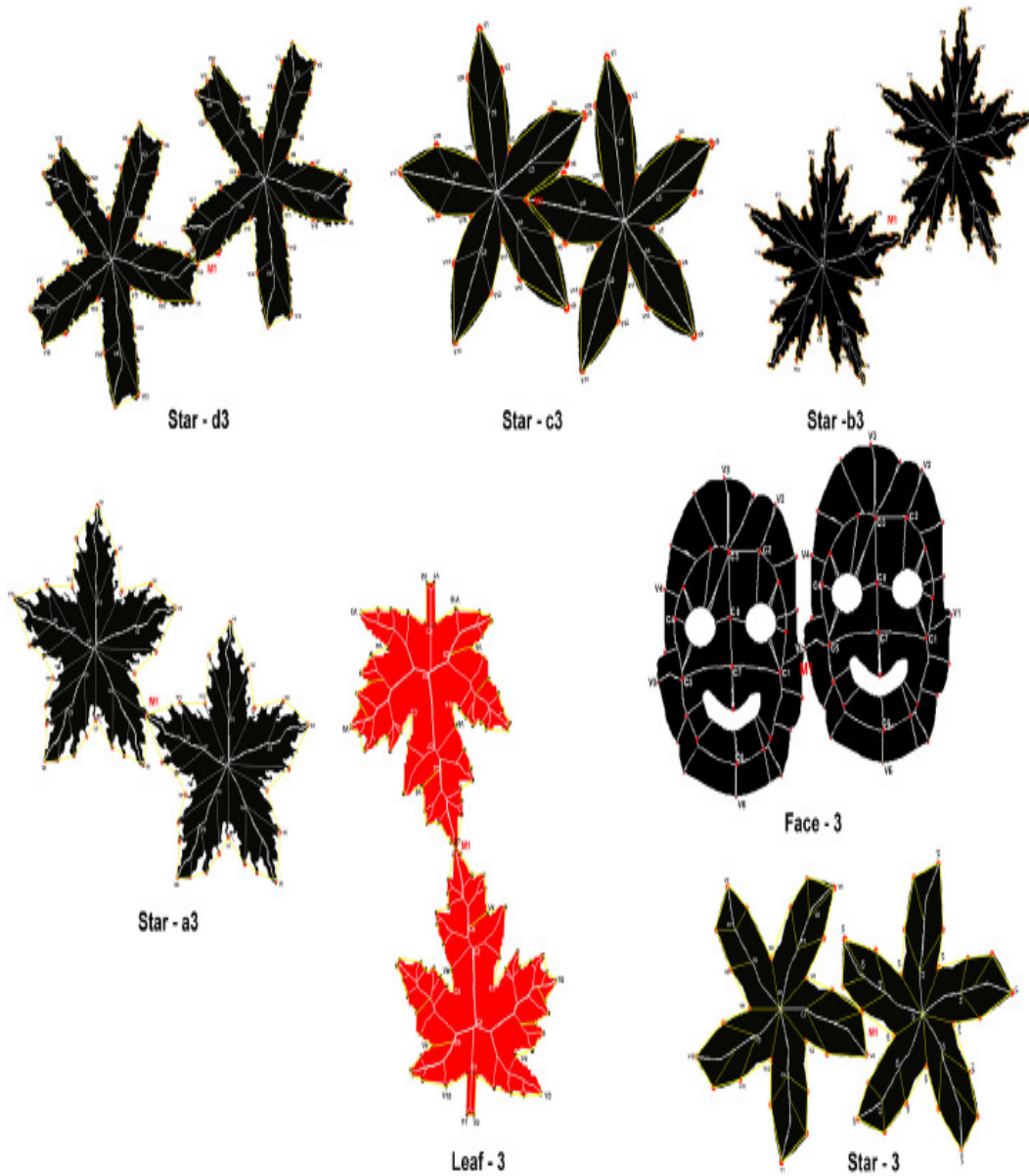


FIGURE 4.3.1 Transformed skeletal representation

4.4 Comparison

In the first row the existing pruning method with DCE produces different contour representation for the input image. The second row depicts the proposed skeleton representation.

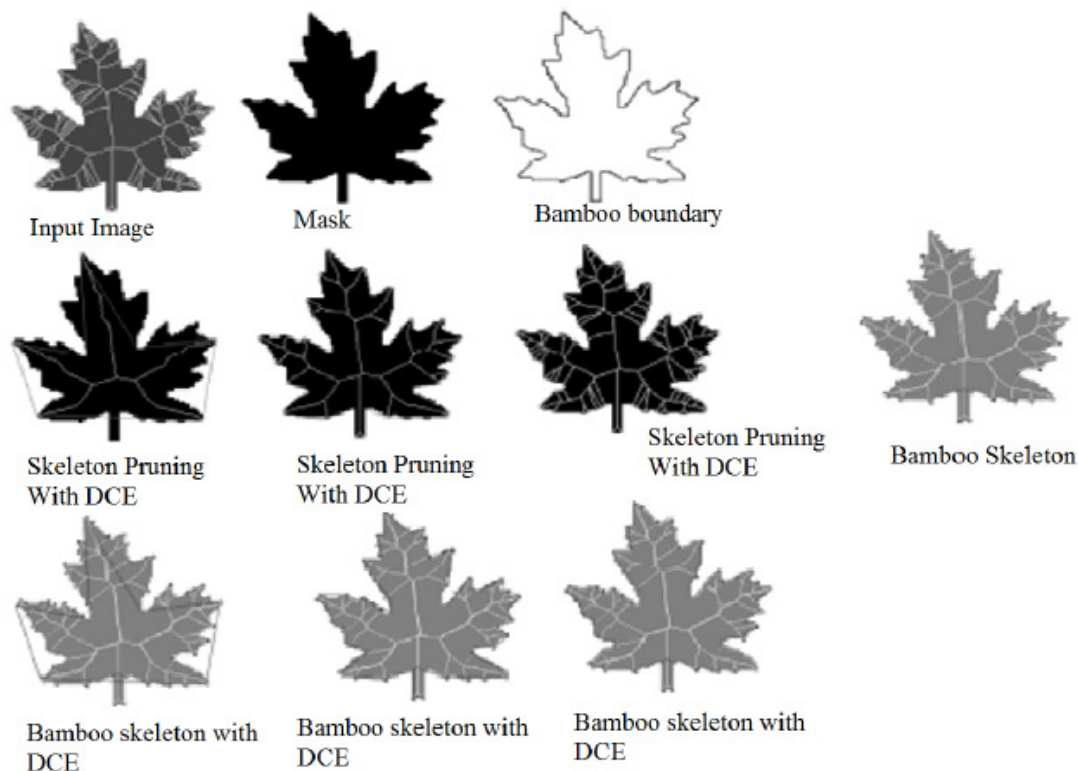


FIGURE 4.4.1.

1. It gives the same common skeletal representation for three different contour against that of the existing DCE method. It reduces the unused storage space for the particular shape representation as in contour1.
2. The boundary deformation is reduced ,In the contour 2 in which the boundary points that are essential for shape representation are not included in contour2 whereas in the proposed representation the boundary points are computed using symmetry –curvature duality based protrusion method and the skeletal arc have been included in contour2 of the proposed method, and hence the boundary deformation is reduces to the great extent.
3. The number of skeletal arc is discrete, unique and standard for all the contour representation. The proposed skeletal representation is unique and stable and satisfies the skeleton topology.
4. Merging of two input images could be achieved to produce new recombinant skeletal representation. The following fig represents merging of two input images to produce transformed or merged images that can be used for producing the new real time objects. The computed boundary points can be used in merging two images so as to transform the existing skeletal representation into a modified recombinant skeletal representation that can be used for generating real time objects.
5. The bamboo skeletal representation is suited for ununiform contour based shape representation against the existing method which is finds difficulties in ununiform contour based images.

4.5 Robustness to Noise

In this experiment, noise was added to data extracted from [7] to show the robustness of the decomposition method. Figure 8 presents 4 groups of shapes, each containing a basic shape in different poses. These poses show non-rigid deformations. Since the approach is based on robust skeleton detection and important boundary features, the reconstruction result is consistent. Figure 9 gives the decomposition result for two shapes with evolved noise. It indicates that the increasing noise does not influence the merging result.

4.6 Invariant Representation

A shape centered coordinate frame allows one to describe the organization of the primitives in a way that is robust to changes in scale, rotation and articulation. This coordinate frame can be formed by designating the prominent branches of the shape as the reference axes.

5. Conclusion and future scope

The proposed method is best suited for 2D object representation. It can be extended for 3D images also. Any modifications developed in the generated real time objects can also be studied and accordingly the work can be extended from objects point of view and its representation.

Acknowledgement.

The author would like to express her gratitude to Mr. Dr V.Sengoden, Principal and secretary, SNR SONS COLLEGE, COIMBATORE for providing necessary infrastructure and for his constant encouragement that led to improvise the presentation of the quality of the paper.

References

- [1] H. Blum. "A transformation for extracting new descriptors of shape, in Models for the Perception of Speech and Visual Form (W.Wathen-Dunn, ed.)", Cambridge MA: MIT Press, pp. 362-380, (1967)
- [2] J. K. Lakshmi and M. P. Valli. "A Survey on skeletons in digital image processing" . International conference proceedings of IEEE computer Society,2009
- [3] J. K. Lakshmi and M. P. Valli. "A Survey on skeletonization in digital image processing". International conference proceedings of Managing Next Generations Software Applications08,Sponsored by CSIR, New Delhi, 2008
- [4]Cohen, E.H., Singh, M. " Geometric determinants of shape segmentation: Tests using segment identification". Vision Research 47: 2825–2840, 2007
- [5] Siddiqi, K., Kimia, B.B. "Parts of visual form: Computational aspects". IEEE Transactions on Pattern Analysis and Machine Intelligence, 17: 239–251, 1995
- [6]Bai, X., Latecki, L.J., Liu, W. "Skeleton pruning by contour partitioning with discrete curve evolution". IEEE Transactions on Pattern Analysis and Machine Intelligence, 29: 449– 462, 2007
- [7] Bai, X., Latecki, L.J.. "Discrete skeleton evolution". 6th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition,2007
- [8]J. Komalalakshmi1, Dr. M. Punithavalli. "Impact of Boundary Points in Skeleton Based Images". International journal of advanced engineering and applications,(IJAEA)Italy, 2010
- [9] J. Zeng, R. Lakaemper, X. Yang, Xin Li, Temple University, Philadelphia, PA. 2D "Shape Decomposition Based on Combined Skeleton-Boundary Features", Lecture Notes in Computer Science.
- [10]Aslan, C.; Erdem, A.; Erdem, E.; Tari, S.; Microsoft Corp., Redmond, WA "Disconnected Skeleton: Shape at Its Absolute Scale,pattern analysis and machine intelligence". Ieee transaction on, IEEE computer society, 30(12): 2008
- [11] H. Chui, A. Rangarajan. "A new point matching algorithm for non-rigid registration". *CVIU*, 89(2-3):114–141, 2003.
- [12] W. Mio, A. Srivastava, and S. Joshi. "On shape of plane elastic curves". Submitted to IJCV.

- [13] A. Peter and A. Rangarajan. "Shape matching using the Fisher-Rao Riemannian metric: Unifying shape representation and deformation". *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, 1164–1167, 2006
- [14] T. B. Sebastian, P. N. Klein, B. B. Kimia. "Recognition of shapes by editing their shock graphs". *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(5):550–571, 2004
- [15] J. Shah. "An H^2 type Riemannian metric on the space of planar curves". *Workshop on the Mathematical Foundations of Computational Anatomy (in conjunction with MICCAI'06)*, 2006
- [17] K. Leonard. "Classification of 2D shapes via efficiency measures". Unpublished work, 2006.
- [18] H. Blum. "Biological shape and visual science". *Journal of Theoretical Biology*, 38:205–287, 1973
- [19] F. L. Bookstein. "*Morphometric Tools for Landmark Data—Geometry and Biology*". Cambridge Univ. Press, 1991
- [20] V. Camion and L. Younes. "Geodesic interpolating splines". In *EMMCVPR*, pp. 513–527, 2001
- [21] S. Belongie, J. Malik, J. Puzicha. "Shape matching and object recognition using shape contexts". *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(4):509–522, 2002
- [22] M. C. Burl and P. Perona. "Recognition of planar object classes". In *CVPR*, pp. 223–230, 1996
- [23] H. Chui, A. Rangarajan. "A new point matching algorithm for non-rigid registration". *CVIU*, 89(2-3):114–141, 2003
- [24] D. Geiger, T. L. Liu, R. V. Kohn. "Representation and self-similarity of shapes". *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(1):86–99, 2003
- [25] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, S. W. Zucker. "Shock graphs and shape matching". *Int. J. Comput. Vision*, 35(1):13–32, 1999
- [26] J. Glaunes, A. Trounev, and L. Younes. "Diffeomorphic matching of distributions: A new approach for unlabelled point-sets and sub-manifolds matching". In *CVPR*, pp. 712–718, 2004
- [27] J. komalalakshmi, Dr .M. Punithavalli. "Discrete Skelton Reconstruction using bamboo skeleton communicated to the IEEE Transaction on Image processing".
- [28] M. Brady, H. Asada. "Smoothed local symmetries and their implementation". *The International Journal of Robotics Research*, 3(3):36–61, 1984
- [29] J.W. Bruce, P.J. Giblin, C.G. Gibson. "Symmetry sets". *Proc. R. Soc. Edinb. Sect. A-Math*, 101:163–186, 1985
- [30] D. Geiger, T. L. Liu, R. V. Kohn. "Representation and self-similarity of shapes". *IEEE Trans. Pattern Anal. Mach. Intell.*, 25(1):86–99, 2003
- [31] Bai, X., Latecki, L. "Path similarity skeleton graph matching". *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30 (2008)
- [32] M. Leyton. "A process-grammar for shape". *Artificial Intelligence*, 34(2):213–247, 1988

[33] M. Brady, H. Asada. "Smoothed local symmetries and their implementation". *The International Journal of Robotics Research*, 3(3):36–61, 1984

[34] Michael Leyton. "Symmetry-Curvature Duality, computer vision ,graphics and image processing". 38:327-341, 1987.

[35] J.k. lakshmi, Dr .M.Punithavalli. "Computation of merging points in skeleton based digital images".GJCST, global journal of computer science and technology, 9(5):164-171, 2009.

A Quantitative Comparative Study of Analytical and Iterative Reconstruction Techniques

Shrinivas D Desai

*Assistant Professor, Dept of ISE
B V B College of Engineering & Technology
HUBLI – 580031
Karnataka, India*

sd_desai@bvb.edu

Dr Linganagouda Kulkarni

*Director JPNES Group of Institutions – School of
Engineering / School of Pharmacy
MEHABUBNAGAR
Andra Pradesh, India*

linganagouda@yahoo.co.uk

Abstract

Measurement of visual quality is of great importance in the field of medical image applications such as X-ray tomography. In applications such as CT and MRI scanning, most of the time quality of reconstructed image is assessed qualitatively by radiologist or domain experts which are purely subjective evaluation [1]. However quantitative analysis or machine evaluation of reconstructed image appeals more to medical practitioners and normally aids the diagnosing of disease. Hence there is immense need for parameters which quantitatively measures the quality of reconstructed images. This paper proposes six image quality measurement parameters (IQM) and comparison of both qualitative ratings and proposed quantitative method for images reconstructed in CT by analytical and iterative techniques [2] is presented here.

Projections (parallel beam type) for the image reconstruction are calculated analytically by defining Shepp logan phantom head model with coverage angle ranging from 0 to $\pm 180^\circ$ with rotational increment of 2° to 10° . For iterative reconstruction coverage angle of $\pm 90^\circ$, iteration up to 10 is used. The original image is grayscale image of size 128 X 128.

Experiment results reveal very close similarity among assessment done qualitatively with those of assessment done quantitatively. Hence the proposed six quality measurement parameters appear to be effective for assessing the quality of reconstructed image in CT applications.

Keywords: Reconstruction algorithm, Simple-Back projection algorithm (SBP), Filter-Back projection algorithm (FBP), Algebraic Reconstruction Technique algorithm (ART), Image quality, coverage angle, Computed tomography (CT).

1. INTRODUCTION

For applications such as CT and MRI in which images are to be viewed by human beings, the ultimate method of quantifying visual image quality is through subjective evaluation [4]. In practice, however, subjective evaluation is usually too inconvenient; require expertise, time-consuming and expensive. The majority of the proposed perceptual quality assessment models have followed a strategy of modifying the MSE measure so that errors are penalized in accordance with their visibility. The most reported objective method for image quality measurement parameters are Peak Signal to Noise Ratio, Human Visual System [3], Picture Quality Scale [5, 6], Noise Quality Measure [7], Fuzzy [8], Multi-scale Structural Similarity Index Metric [9], Information fidelity criterion [10], and Visual information fidelity [11] which works with luminance only. While IQM which works on color images are Sarnoff model [12, 13], BSDM [14], Universal image quality index [15] and DC tune [16].

The resented reported IQM for structural similarity is MSSIM index, which is combination of three comparisons: luminance, contrast and structure [4]. Quantifying differential qualities using various dissimilar metric is yet another IQM proposed to discriminate reconstructed image. Literature presents six dissimilarity metrics namely Euclidean distance, Manhattan distance, Canberra distance, Bray-Curtis distance, Squared Chord distance and Chi-Squared distance to identify the metric that provide better classification between the reconstructed and original image [17].

In CT applications, where images are grayscale, the image quality measurement doesn't necessarily depend on luminance or color. In this regard we propose six IQM parameters to objectively measure gray scale reconstructed image which are generated in CT applications using analytical and iterative techniques.

2. METHOD

The following section deals with different methods available for reconstruction of image from projections.

2.1 Reconstruction:

Given the sinogram $p(r, \theta)$ we want to recover the object described in (x, y) coordinates. Basically it works by subsequently "smearing" the acquired $p(r, \theta)$ across a film plate. This is simple back projection [18]

$$b(x, y) = B\{p(r, \theta)\} = \int_0^{\Pi} p(x.\cos\theta + y.\sin\theta, \theta)d\theta \quad \text{-- 2.1}$$

2.2 Filtered Back projection:

We need a way to equalize the contributions of all frequencies in the FT's polar grid, this can be done by multiplying each $P(k, \theta)$ by a ramp function. This way the magnitudes of the existing higher-frequency samples in each projection are scaled up to compensate for their lower amount. The ramp is the appropriate scaling function since the sample density decreases linearly towards the FT's periphery [18].

Filtered back projection is given as

$$f(x, y) = \int_0^{\Pi} \left(\int_{-\infty}^{\infty} P(k, \theta) \cdot |k| \cdot e^{i2\Pi kr} dk \right) d\theta \quad \text{-- 2.2}$$

where $P(k, \theta)$ is 1 D Fourier transform of $p(r, \theta)$, multiplying with $|k|$ gives ramp filtering, integrating the expression from ∞ to $-\infty$ gives inverse 1D Fourier transform, and integrating from 0 to Π gives filtered back projection for all angles.

2.3 Algebraic Reconstruction:

An entirely different approach for Tomographic imaging consists of assuming that the cross section consists of an array of unknowns and the setting up algebraic equations for the unknowns in terms of the measured projection data. All algebraic reconstruction technique methods are iterative procedures. There are two approaches for reconstructing the image using iterative technique [19].

In first approach the process starts with an initial estimate and tries to push the estimate closer to the true solution. Instead of back-projecting the average ray value, the projections corresponding to the current estimate are compared with the measured projections. The result of the comparison is used to modify the current estimate, thereby creating a new estimate. In the second approach ART pose the reconstruction problem as a set of simultaneous equations. Here a square grid is assumed to superimpose over the unknown cross-section image. Image values are assumed to be constant within each cell of the grid. The set of linear equation for each ray is then formed.

$$\sum_{j=1}^N w_{ij} f_j = p_i \quad , \quad i = 1, 2, \dots, M \quad \text{-- 2.3}$$

where p_i are projection data along the i^{th} ray, w_{ij} the fraction area of j image cell intercepted by i^{th} ray, f_j are the values of the reconstruction grid element (pixel) , M is the total number of rays (in all the projection) and N is the number of grid cell.

For a large values of M and N there exist an iterative methods for solving (2.3), which is One iteration step the Kaczmarz method.

The **Kaczmarz method**, based on the work of the Polish mathematician Stefan Kaczmarz [20], is a method for solving linear systems of equations $Ax = b$. It was rediscovered in the field of image reconstruction from projections where it is called the Algebraic Reconstruction Technique (ART) [21].

$$f_j^{(i)} = f_j^{(i-1)} + \lambda \cdot \frac{p_k - \overrightarrow{w_k} \cdot \overrightarrow{f^{i-1}}}{\sum_{i=1}^N W_{kl}^2} \cdot W_{kj} \quad \text{-- 2.4}$$

Where $f_j^{(i)}$ = New pixel value, $f_j^{(i-1)}$ = Old pixel value, λ = Relaxation parameter,

$p_k - \overrightarrow{w_k} \cdot \overrightarrow{f^{i-1}}$ = The difference, $\sum_{i=1}^N W_{kl}^2$ = Normalizing the difference, W_{kj} = Weighting the correction

Equation (2.4) is the well-known algebraic reconstruction technique (ART) where the second term on the left is the correction factor for ray i^{th} . Apparently, equation 2.4 is used to update the j^{th} pixel on every ray equation.

3. EXPERIMENTS

Figure. 3.1 shows the 2D shepp logan phantom head model used in our study. It is constructed from basic ellipsoids as per standard [24], which allows us to calculate their projections analytically [22]. The size of the model defined here is 128 X 128, grayscale image.



Figure 3.1 2D Shepp Logan phantom head model

Figures 3.2 a, b, c are the standard medical images of brain, abdomen and breast are used in our study [23]. Projections are calculated mathematically and reconstructed using SBP, FBP and ART techniques.

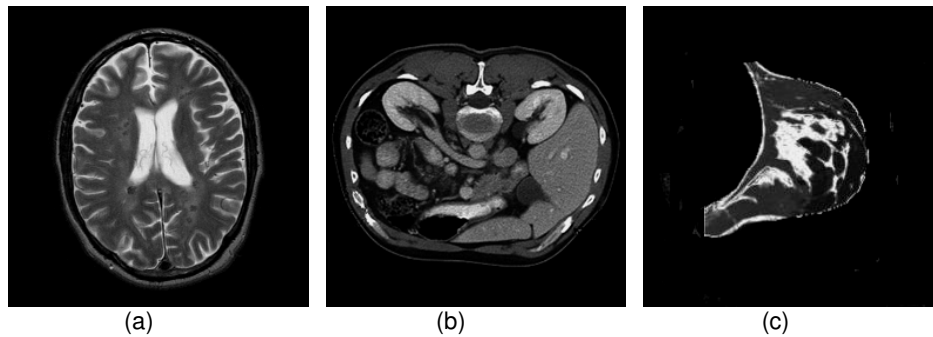


Figure 3.2 Standard medical image (a) Brain (b) Abdomen (c) Breast

3.1 Quality measurement parameters: The quality of reconstructed image is quantified by six quality measurement parameters as listed below.

1) Mean Square Error (MSE): Mean square error is the sum over all squared value differences divided by image size. It's a measure between the original image and the reconstructed image

$$MSE = \frac{1}{MN} \sum_{j=1}^M \sum_{k=1}^N [(I(x, y) - I'(x, y))]^2 \quad \text{--3.1}$$

Where $I(x,y)$ is the original image, $I'(x,y)$ is the reconstructed image, and M,N are the dimensions of the images.

2) Peak signal to noise ratio (PSNR): is a measure of the peak error

$$PSNR = 10 \log \left(\frac{255}{\sqrt{MSE}} \right) \quad \text{-- 3.2}$$

A lower value for MSE means lesser error, and as seen from the inverse relation between the MSE and PSNR, this translates to a high value of PSNR. Logically, a higher value of PSNR is good because it means that the ratio of Signal to Noise is higher. Here, the 'signal' is the original image, and the 'noise' is the error in reconstruction

3) Normalized Cross-Correlation (NCC) : Normalized Cross-Correlation is one of the methods used for template matching, a process used for finding incidences of a pattern or object within an image

$$NCC = \frac{\sum_{j=1}^M \sum_{k=1}^N (I(x, y) - I'(x, y))^2}{\sum_{j=1}^M \sum_{k=1}^N (I(x, y))^2} \quad \text{-- 3.3}$$

4) Structural Content (SC) : Structural content establishes the degree to which an image in the collection matches. It's the measure of image similarity based on small regions of the images containing significant low level structural information. The more the number of such regions common to both images, the more similar they are considered.

$$SC = \frac{\sum_{j=1}^M \sum_{k=1}^N (I(x, y))^2}{\sum_{j=1}^M \sum_{k=1}^N I'(x, y)^2} \quad \text{-- 3.4}$$

5) Maximum Difference (MD): It's the variation of the method of paired comparisons.

$$MD = \text{Max} \left(\left| x_{j,k} - x'_{j,k} \right| \right) \quad \text{-- 3.5}$$

6) Normalized Absolute Error (NAE): It's the numerical difference between the original and reconstructed image.

$$NAE = \frac{\sum_{j=1}^M \sum_{k=1}^N \left| x_{j,k} - x'_{j,k} \right|}{\sum_{j=1}^M \sum_{k=1}^N \left| x_{j,k} \right|} \quad \text{-- 3.6}$$

4. RESULTS & DISCUSSION

Comparison of reconstruction techniques such as SBP, FBP (analytical), & ART (iterative) with respect to quality of reconstructed image is presented in this section. Projections calculated analytically are assumed for parallel projection data and the center of rotation as the center point of the projections.

Number of projections and angular range: The reconstruction discrepancy such as MSE was calculated for 18, 24, 36, and 72 projections, taken over angular ranges of 0 to $\pm 180^\circ$ with incremental ranging from 2° to 10° . The original image is of size 128 X 128. It was found that, for a given phantom and angular range, the discrepancy decreased as the number of projections increased, regardless of initial guess. Experiment is done using Matlab (The MathWorks, Inc., Natick, MA), version 7.0.4, on WindowsXP platform (Microsoft Corp., Redmond, WA). The reconstructed images are best seen at VGA output, 1024 X 768 resolutions.

4.1 Simple back projection

In SBP an individual sample of projection is back projected by setting all the image pixels along the ray pointing to the sample to the same value. The following Figure 4.1 (a) shows the reconstruction of 2D shepp logan phantom head model (Figure. 3.1) with different projections. The experimental result as well as Figure 4.1(b) shows that minimum 72 projections, with coverage angle ranging from 0 to 180° with an incremental value of 2° is necessary to reconstruct the image with acceptable quality.

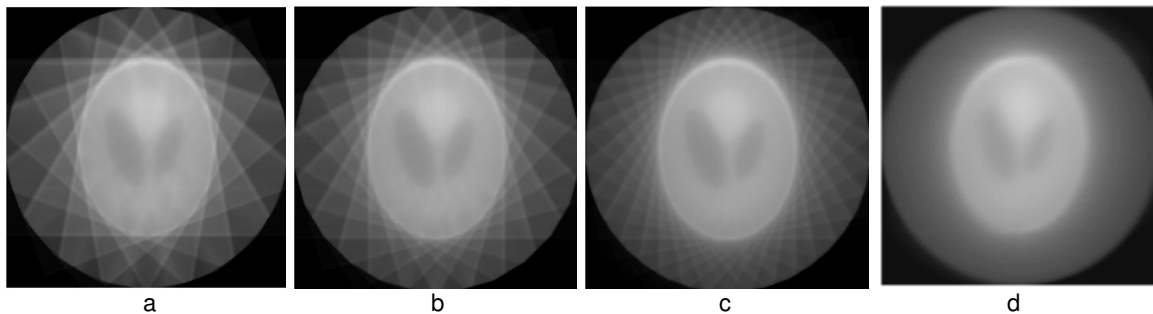


Figure. 4.1 (a) Reconstructed image using SBP with a) 18, b) 24, c) 36, d) 72 projections.

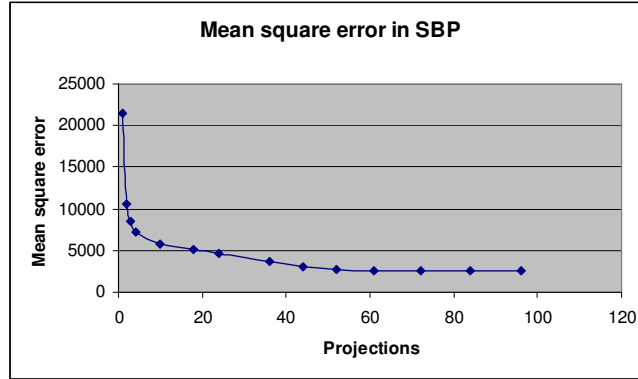


Figure 4.1 (b) Mean square error v/s projections for SBP

SBP can be referred to as a “brute force” technique. It’s a simple and non effective method of obtaining an approximate reconstruction from multiple projections, and this technique generates star or spoke artifact. Image padding is necessary so that pixel values at borders are retained even after rotating the P-H model. Figure 4.1 (b) clearly reveals that the quality of reconstructed image increases as number of projection increases, MSE settles down from 72 projections. However the reconstructed image appears to be very blurry.

4.2 Filtered Back projection:

In FBP each view is filtered before the back projection to counteract the blurring effect. That is, each of the one-dimensional views is convolved with a one-dimensional filter kernel to create a set of filtered views. These filtered views are then back projected to provide the reconstructed image, a close approximation to the correct image. In fact, the image produced by filtered back projection is identical to the correct image when there are an infinite number of views and an infinite number of points per view. The Figure 4.2 (a) shows reconstruction of phantom head model by FBP with coverage angle ranging from 0 to 180° with an incremental value of 10° to 2°.

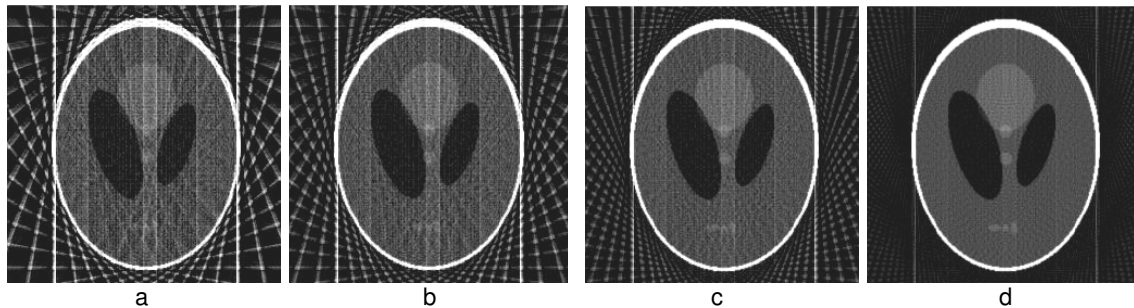


Figure 4.2 (a) Reconstructed image using FBP with a) 18, b) 24, c) 36, d) 72 projections.

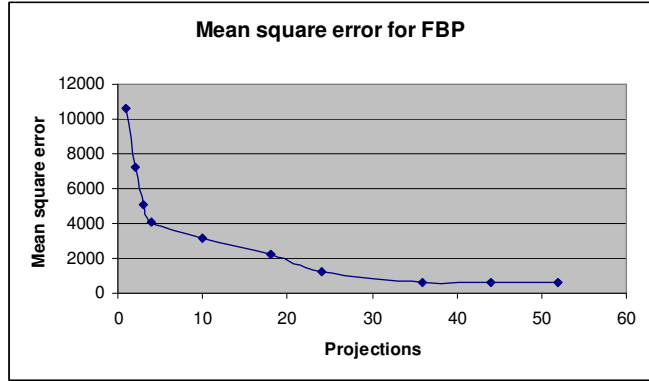


Figure 4.2 (b) Mean square error v/s projections for FBP

Figure 4.2 (b) clearly reveals that the quality of reconstructed image increases as number of projection increases, the MSE of the reconstructed image remains constant after 36 projections. This algorithm requires minimum of 36 projections with rotational increment of 5° to display acceptable reconstructed image. It is the most common method of removing the star artifact which is normally generated in SBP by using various filters. FBP algorithm used for reconstruction of image from its projections is found to be fast and efficient with large number of projections. It is fast and efficient with large number of projections. We can say that it is direct inversion of the projection formula. But here quantitative imaging is difficult. FBP has become more popular and commercialized in nuclear medicine.

4.3 Algebraic Reconstruction technique:

In ART all the pixels in the image array is set to some arbitrary value. In experiment it is set to average value of the phantom head model. An iterative procedure is then used to gradually change the image array to correspond to the profiles. Using first approach after the first complete iteration cycle, there will still be an error between the ray sums and the measured values. This is because the changes made for any one measurement disrupts all the previous corrections made. The idea is that the errors become smaller with repeated iterations until the reconstructed image converges to the proper solution. The Figure 4.3 shows reconstruction of phantom head model with “smart” initial guess, coverage angle 0 to 90° , increment of 5° , iterations from 3 to 8.

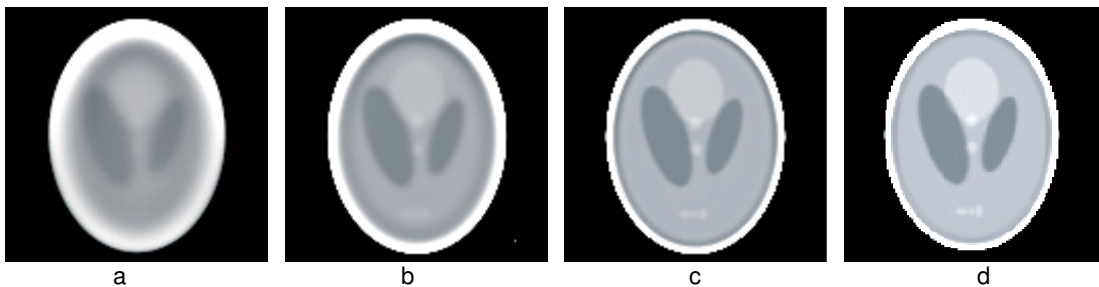


Figure 4.3 Reconstructed image with ART: a) $\theta = 5$, Iteration=3 b) $\theta = 5$, Iteration=4 c) $\theta = 5$, Iteration=6 d) $\theta = 5$, Iteration=8

The principle of ART is to find a solution by successive estimates. The projections corresponding to the current estimate are compared with the measured projections. The result of the comparison is used to modify the current estimate, thereby creating a new estimate. Corrections are carried out either as addition of differences or multiplication by quotients between measured and estimated projections. Figure 4.4 (a) and (b) shows the difference between reconstructed and original image by drawing NAE versus iterations for difference correction and multiplicative corrections respectively.

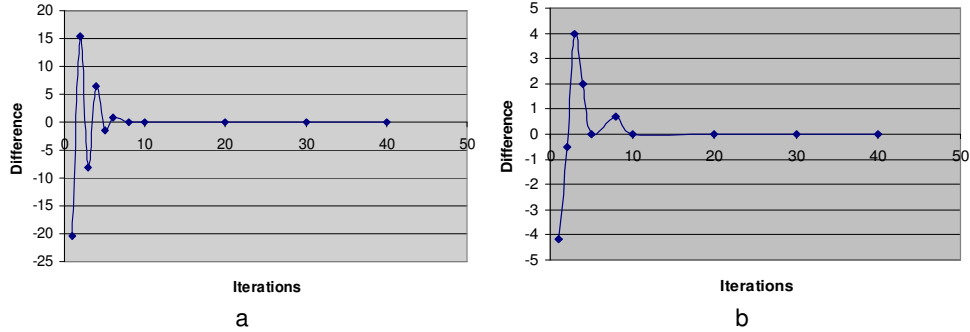


Figure 4.4 NAE versus iteration a) difference correction b) multiplicative corrections.

In ART obviously, the errors of phantom obtained from the algebraic formulations are less than that of the filtered back projection; the star artifact appearing in image reconstructed from the filtered back projection is disappeared with the method of algebraic formulation. However, the algebraic formulations take more time to complete the process than does the filtered back projection.

The following table shows quality of reconstructed image measured by six quality measurement parameters. Table 4.1(a) shows the quality measurement of reconstructed by varying number of projections, & keeping common coverage angle, while Table 4.1(b) shows the quality measurement by varying coverage angle, & keeping common number of projections.

Image Quality Measurement	Coverage angle: $\pm 90^\circ$ Number of projections: 72			Coverage angle: $\pm 90^\circ$ Number of projections: 36		
	SBP	FBP	ART	SBP	FBP	ART
MSE	4938.6	572.7858	2586.4	2532.4	1275.6	286.25
PSNR	11.1947	20.5509	14.0039	5.689	7.56	10.57
NCC	1.0113	0.9949	1.0039	0.58	0.57	0.4612
SC	0.8773	0.9984	0.9356	0.48	0.47	0.4824
MD	115	255	143	56	75	129
NAE	0.2119	0.0262	0.1534	0.125	0.078	0.0112

Table 4.1 (a) : Image quality measurement by varying number of projections keeping fixed coverage angle.

The experiments revealed major observations; as the number of projections within a given angular range was increased, the quality of reconstructed image appeared better. Regarding the angular range, it was found that the discrepancy decreased with increasing angular range.

Image Quality Measurement	Coverage angle: $\pm 90^\circ$ Number of projections: 36			Coverage angle: $\pm 180^\circ$ Number of projections: 36		
	SBP	FBP	ART	SBP	FBP	ART
MSE	2532.4	1275.6	286.25	1526.5	624.25	135.2
PSNR	5.689	7.56	10.57	2.589	3.251	5.26
NCC	0.58	0.57	0.4612	0.24	0.245	0.218
SC	0.48	0.47	0.4824	0.18	0.214	0.26
MD	56	75	129	29	32	72
NAE	0.125	0.078	0.0112	0.08	0.026	0.059

Table 4.1 (b) : Image quality measurement by varying coverage angle keeping fixed number of projections

Table 4.1 (c) shows Image quality measurement for Figure 3.2 a, b, c reconstructed from SBP, FBP, ART. The experiment reveals the fact that FBP effectively eliminates Star artifacts created by SBP. ART performs better even at limited views, and has better noise tolerance.

		MSE	PSNR	NCC	SC	MD	NAE
Model	Technique						
Brain	SBP	1646.2	3.731567	0.3371	0.292433	38.33333	0.070633
	FBP	190.9286	6.8503	0.331633	0.3328	85	0.008733
	ART	862.1333	4.667967	0.334633	0.311867	47.66667	0.051133
Abdomen	SBP	1975.44	4.47788	0.40452	0.35092	46	0.08476
	FBP	229.1143	8.22036	0.39796	0.39936	102	0.01048
	ART	1034.56	5.60156	0.40156	0.37424	57.2	0.06136
Breast	SBP	3292.4	7.463133	0.6742	0.584867	76.66667	0.141267
	FBP	381.8572	13.7006	0.663267	0.6656	170	0.017467
	ART	1724.267	9.335933	0.669267	0.623733	95.33333	0.102267

Table 4.1 (c) Image quality measurement for Figure 3.2 a, b, c reconstructed from SBP, FBP, ART.

Table 4.2 shows the reconstruction time for Figure 3.1 Shepp logan phantom head model using analytical & iterative techniques.

Technique	Number of projections			
	18	24	36	72
SBP	5.24	10.26	15.24	20.72
FBP	0.75	1.23	1.95	2.38
ART	5.42	11.24	16.58	21.24

Table 4.2: Reconstruction time* for Figure 3.1 for different number of projections. Time in seconds.

ART requires more time compared to others for reconstruction, while FBP requires least. However, the SBP take more time to complete the process than does the FBP, but less as compared to ART. Table 4.3 shows the reconstruction time for three standard medical images brain, abdomen, and breast with different number of projections

Technique	Brain		Abdomen		Breast	
	36	72	36	72	36	72
SBP	10.34	20.35	10.45	20.36	9.96	19.94
FBP	0.95	2.16	0.98	2.15	0.86	2.03
ART	11.24	23.15	11.28	24.16	10.56	20.56

Table 4.3: Reconstruction time* for Figure 3.2 a, b, c for different number of projections. Time in seconds

* Results obtained using Matlab (The MathWorks, Inc., Natick, MA), version 7.0.4, under the Window XP operating system (Microsoft Corp., Redmond, WA), on Intel Pentium Core2 Duo, CPU 2.80GHz, 1GB RAM.

Figure 4.5 (a) to (c) shows the best quality image that can be reconstructed for different standard medical images brain, abdomen, and breast from projections using SBP (1st column), FBP (2nd column), ART (3rd column) the coverage angle is from 0 to 180°, with incremental of 5° to 2°. For ART the reconstruction is done with “smart” initial guess, coverage angle ranging from 0 to 90°, increment of 5°, and iterations from 5 to 8.

Figure 4.5 (a): Reconstruction of Brain image with projections: coverage angle: iteration (for ART) : .
 Column 1: SBP, Column 2: FBP, Column 3: ART

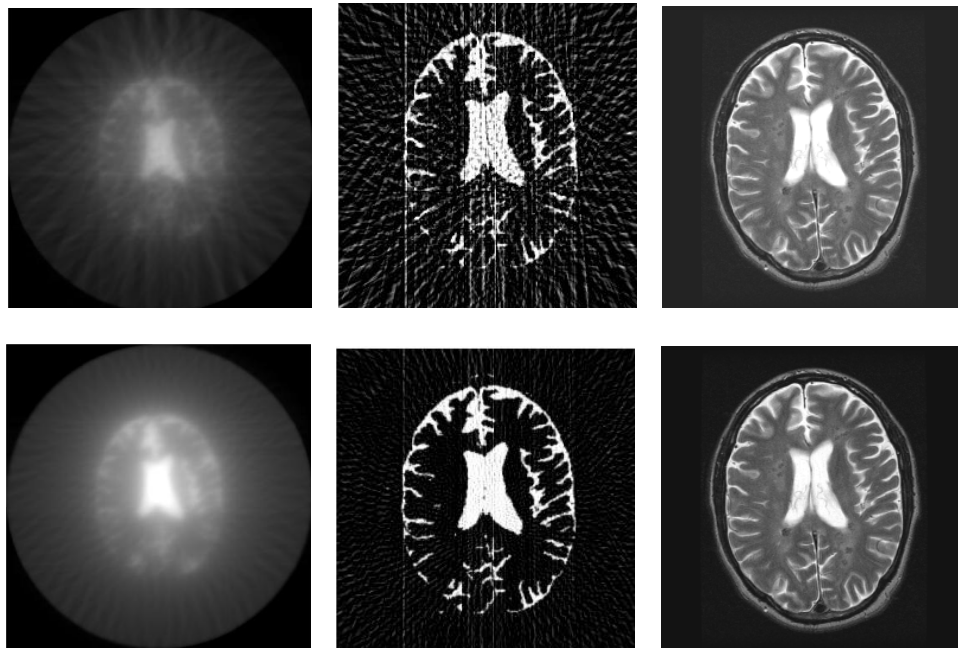


Figure 4.5 (b): Reconstruction of Abdomen image with projections: coverage angle: iteration (for ART) : .
 Column 1: SBP, Column 2: FBP, Column 3: ART

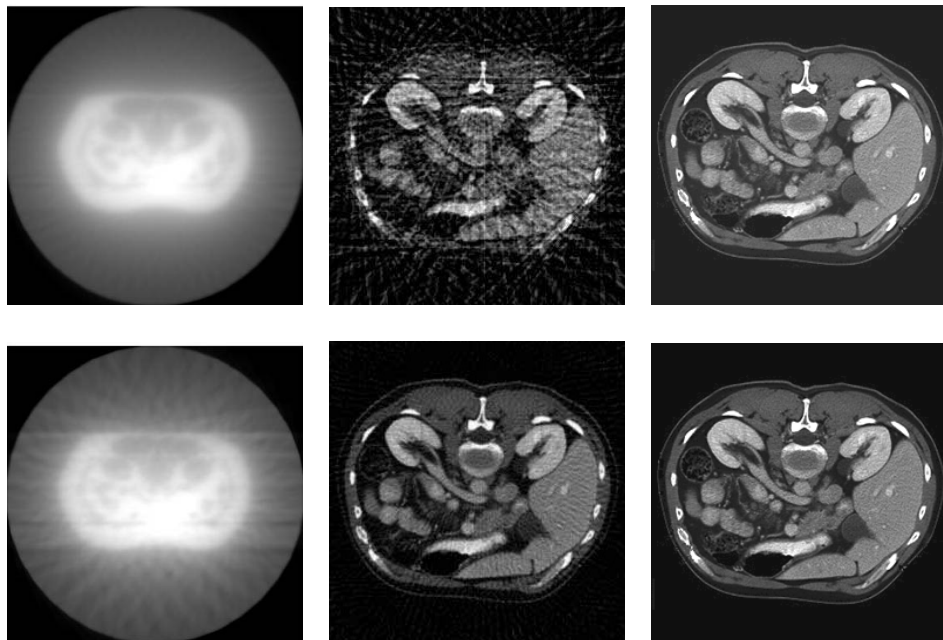
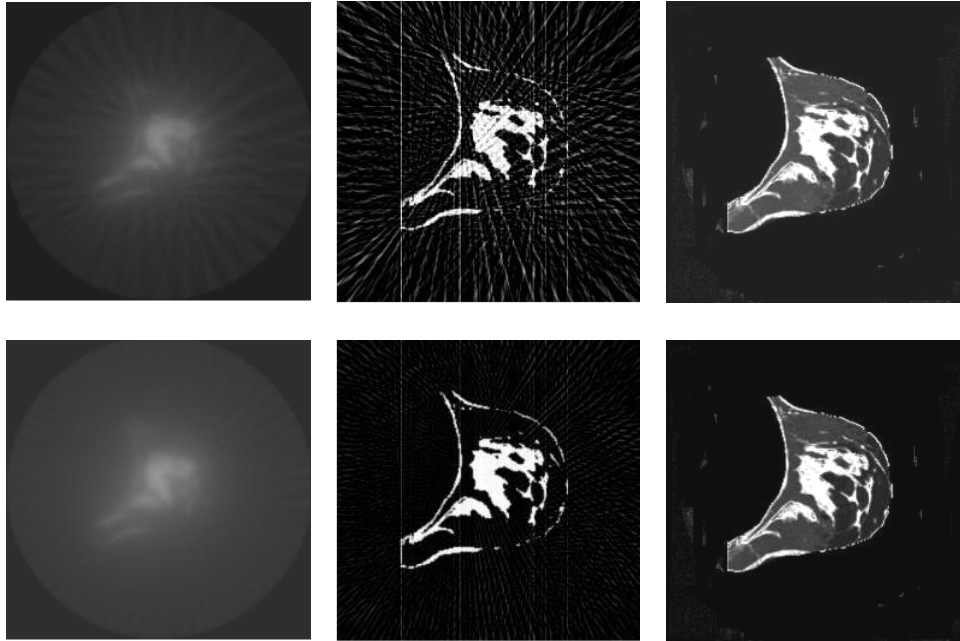


Figure 4.5 (c): Reconstruction of Breast image with projections: coverage angle: iteration (for ART) : .
Column 1: SBP, Column 2: FBP, Column 3: ART



5.CONCLUSION

The quantitative comparative study of analytical and iterative techniques for image reconstruction from projections for computed tomography revealed a case of diminishing returns, which are concluded as below.

In this work, objective measurement by Mean Square Error, Peak signal to noise ratio, Normalized Cross-Correlation, Structural Content, Maximum Difference and Normalized Absolute Error led to an ability to subjectively judge the reconstructed image quality.

The major problem with SBP reconstruction is that it leaves “extra” counts on the image because of which reconstructed image appears severe blur nature or has “Star” like pattern. In case of FBP; it removes the star artefact produced in SBP, and by using Ram-Lak with Hamming / Butterworth filter we can simultaneously reduce high-frequency components (containing much noise) and low-frequency component (containing blur). The image produced by the ART technique has poor density resolution, acceptable spatial resolution, but requires large computational time. The spoke artifact appearing in image reconstructed from the filtered back projection is disappeared in algebraic formulation; it possesses better noise tolerance. These subjective evaluations are in close with objective evaluation done using six IQM. From the experiments, we shall conclude that proposed six IQM all together is reliable and practical to measure the quality of reconstructed images for CT applications.

6. REFERENCES

1. H. R. Sheikh, M. F. Sabir, A. C. Bovik. "A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algorithms". IEEE Transactions on Image Processing, 15(11): 2006
2. P. P. Bruyant. "Analytic and Iterative Reconstruction Algorithms in SPECT". Journal of Nuclear Medicine, by Society of Nuclear Medicine, 43(10): 1343-1358, 2002
3. N. Yamsang and S. Udomhunsakul. "Image Quality Scale (IQS) for Compressed Images Quality Measurement". Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, Hong Kong, 2009
4. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli. "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Transactions On Image Processing, 13(4): 2004
5. M. Miyahara, K. Kotani, V. R. Algazi,. "Objective Picture Quality Scale (PQS) for image coding," IEEE Trans. Commun., 46(9), 1215–1225,1998
6. CIPIC PQS ver. 1.0, [Online]. Available at: <http://msp.cipic.ucdavis.edu/estates/ftp/cipic/code/pqs>.
7. N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, A.C. Bovik. "Image quality assessment based on a degradation model". IEEE Trans. Image Process., 4(4): 636–650, 2000
8. D. V. Weken, M. Nachtegael, E. E. Kerre. "Using similarity measures and homogeneity for the comparison of images". Image Vis. Comput., 22: 695–702, 2004
9. Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," presented at the IEEE Asilomar Conf. Signals, Systems, and Computers, 2003
10. H. R. Sheikh, A. C. Bovik, G. de Veciana. "An information fidelity criterion for image quality assessment using natural scene statistics". IEEE Trans. Image Process., 14(12): 2117–2128, 2005
11. H. R. Sheikh, A. C. Bovik. "Image information and visual quality". IEEE Trans. Image Process., 15(2), 430–444, 2006
12. JNDmetrix Technology Sarnoff Corp., evaluation Version, 2003 [Online]. Available at : <http://www.sarnoff.com/products-services/video-vision/jndmetrix/downloads.asp>
13. J. Lubin. "A visual discrimination mode for image system design and evaluation". in Visual Models for Target Detection and Recognition (E. Peli, ed.), Singapore: World Scientific Publishers pp. 245–283, (1995)
14. I. Avcibas, B. Sankur, K. Sayood. "Statistical evaluation of image quality measures". J. Electron. Imag., 11(2): 206–23, 2002.
15. Z. Wang, A. C. Bovik. "A universal image quality index". IEEE Signal Processing Letters, 9: 81–84, 2002.
16. A. B. Watson. "DCTune: A technique for visual optimization of DCT quantization matrices for individual images". Soc. Inf. Display Dig. Tech. Papers, vol. XXIV, pp. 946–949, (1993)
17. K. B. Raja, M. Madheswaran, K. Thyagarajah. "Quantitative and Qualitative Evaluation of US Kidney Images for Disorder Classification using Multi-Scale Differential Features". ICGST-BIME Journal, 7(1): 2007

18. K. Mueller and R. Yagel. "On the Use of Graphics Hardware to Accelerate Algebraic Reconstruction Methods". Presented at the *SPIE Medical Imaging Conference Physics of Medical Imaging*, San Diego, 1999
19. A.M. Ali, Z. Melegy , M. Morsy , R.M. Megahid , T. Bucherl , E.H. Lehmann. "Image reconstruction techniques using projection data from transmission method". *Annals of Nuclear Energy*, 31: 1415–1428, Elsevier Ltd, 2004
20. Kaczmarz, S. „Angenäherte Auflösung von Systemen linearer Gleichungen“. *Bulletin International de l'Academie Polonaise des Sciences et des Lettres*, series A, 35 :335-357, 1937
21. Gordon, R., R. Bender, G.T. Herman. "Algebraic reconstruction techniques (ART) for threedimensional electron microscopy and x- ray photography". *Journal of Theoretical Biology*, 29:471-481, 1970.
22. A. K. Jain. "Fundamentals of Digital Image Processing", Prentice – Hall of India, 1 - Edition, pp.431 – 470, (1989)
23. Medical Image Samples Sebastien Barre, Online image database, "Archive of DICOM images" [online] Available at:<http://www.barre.nom.fr/medical/samples/>
24. S. D. Desai. "Reconstruction of image from projections-an application to MRI & CT Scanning". In proceedings of International conference ICSCI-2005

Color Image Segmentation Technique Using “Natural Grouping” of Pixels

Biplab Banerjee

*Computer Science and Engg. Dept
Jadavpur University Kolkata –
700032, India*

gettbiplab@gmail.com

Tanusree Bhattacharjee

*Computer Science and Engg. Dept
Jadavpur University Kolkata –
700032, India*

tanusree_bhattacharjee85@yahoo.co.in

Nirmalya Chowdhury

*Faculty, Computer Science and Engg. Dept
Jadavpur University Kolkata –
700032, India*

nirmalya_chowdhury@yahoo.com

Abstract

This paper focuses on the problem Image Segmentation which aims at sub dividing a given image into its constituent objects. Here an unsupervised method for color image segmentation is proposed where we first perform a Minimum Spanning Tree (MST) based “natural grouping” of the image pixels to find out the clusters of the pixels having RGB values within a certain range present in the image. Then the pixels nearest to the centers of those clusters are found out and marked as the seeds. They are then used for region growing based image segmentation purpose. After that a region merging based segmentation method having a suitable threshold is performed to eliminate the effect of over segmentation that may still persist after the region growing method. This proposed method is unsupervised as it does not require any prior information about the number of regions present in a given image. The experimental results show that the proposed method can find homogeneous regions present in a given image efficiently.

Keywords: Segmentation, Region Growing, Natural Grouping.

1. INTRODUCTION

Segmentation of an image entails the division or separation of the image into regions of similar attributes. In another way, image segmentation is nothing but pixel classification. The level to which the segmentation process is to be carried out depends on the particular problem being solved. It is considered as an important operation for meaningful analysis and interpretation of the images acquired. It is one of the most critical components of an image analysis and/or pattern recognition system and still is considered as one of the most challenging tasks in the field of

image processing. It has applications in several domains like Medical Science, Analysis of Remotely Sensed Image, Fingerprint Recognition, and Traffic System Monitoring and so on.

Image segmentation algorithms are generally based on one of two basic properties of intensity values of the image pixels: discontinuity and similarity. In the first category, the approach is to partition an image based on abrupt changes in intensity values. Edge detection techniques fall in this category which is similar to boundary extraction. On the other hand, in the second category, the idea is to partition the image into different regions such that pixels belonging to a given region are similar with respect to a set of predefined criteria's. Researchers have been working on these two directions for years and have proposed various methods keeping those region based properties in mind. It should be noted that there is no fixed approach for segmentation. Based on the similarity or discontinuity criteria, many image segmentation methods have been proposed which can be broadly classified into six groups: (1) Edge Detection, (2) Histogram based method, (3) Clustering (K-Means clustering, Fuzzy C-means clustering etc.), (4) Region based methods (Region growing, Region splitting & merging), (5) Physical Model based approach, (6) Neural Network based segmentation methods [20].

The edge detection method is one of the widely used approaches to the problem of image segmentation. It is based on the detection of points with abrupt changes at gray levels. The main disadvantages of the edge detection technique are that it does not work well when images have many edges because in that case the segmentation method produces an over segmented result, and it cannot easily identify a closed curve or boundary. For an edge based method to be efficient, it should detect the global edges and the edges have to be continuous.

Histogram-based methods are very efficient in terms of time complexity when compared to other image segmentation methods because they typically require only one pass through the pixels. In this technique, a histogram is computed from all of the pixels in the image, and the peaks and valleys in the histogram are detected. Now the pixels in between two consecutive valleys can be considered to belong to a single cluster. One of the disadvantages of this method is that it is unable to work well when the image has no apparent gray level histogram peak. The other disadvantage is that the continuity of the segmented regions cannot be ensured. For the histogram based method to be efficient, we should focus on global peaks which are likely to correspond to the dominant regions in the image.

Clustering methods are also used for image segmentation purpose [4] [5]. The segmentation method incorporating clustering techniques encounters great difficulties when determining the number of clusters actually present in the corresponding feature space or extracting the appropriate feature.

The region based methods are based on the similarity of pixels within a region. Sometimes a hybrid method incorporating the edge based and region based methods have been proved to be useful for certain applications. The first region growing method was the seeded region growing method. This method takes a set of seeds as input along with the image. The seeds mark each of the objects to be segmented. The regions are iteratively grown by comparing all unallocated neighboring pixels to the regions. The difference between a pixel's intensity value and the region's mean, δ , can be used as a measure of similarity. The pixel with the smallest difference measured this way is allocated to the respective region. This process continues until all pixels are allocated to a region. Seeded region growing requires seeds as additional input. The segmentation results are dependent on the choice of seeds. Noise in the image can cause the seeds to be poorly placed. Region splitting & merging is a modified algorithm that doesn't require explicit seeds. It starts off with a single region first, which is the image as a whole. Then the region is split into four different sub region based on some dissimilarity measures to construct a quad tree structure of regions. The regions are split to the extreme level so that no more regions splitting can be done now. Then the split regions are merged iteratively to have the final segmentation.

The model based image segmentation assumes that given an image, its individual regions follow a repetitive form of geometrical structure. This is segmentation using texture feature. Neural Network segmentation relies on processing small areas of an image using an artificial neural network or a set of different neural networks. After such processing the decision-making mechanism marks the areas of an image accordingly to the category recognized by the neural network. A type of network designed especially for this is the Kohonen self-organizing map.

The proposed technique is based on the region based method for image segmentation. In general, the region based methods suffer from the problem of selection of initial seeds. We have formalized a method to sort out this problem. The method proposed in this paper uses the concept of “natural grouping” [10] based on the three color channels to find the seed pixels. Hence each seed pixel acts as a pivot for a region of the given image. Then we start the process of region growing from each such seed pixel. After the completion of this region growing process, we obtain as many regions as the number of seed pixels. The region growing stage is followed by a region merging method which reduces the effect of over segmentation to a great extent.

The rest of the paper is organized as follows, section 2 discusses about some existing works on image segmentation, then in section 3 a problem definition is given and in section 4, our method is proposed then in section 5, some experimental results are shown and a comparison is done with some other well-known segmentation techniques to show the effectiveness of our method. Section 6 deals with the concluding discussion and future improvements that can be done.

2. A BRIEF OVERVIEW OF THE RELATED WORK

Color image segmentation has always been a challenging task for the researchers over the years. Pal & Pal [12] provided a detail review on various segmentation techniques. Among myriads of existing image segmentation techniques, many can be considered as unsupervised clustering methods. For example, region merging is the analogue of agglomerative clustering [1]. Graph cut methods such as minimal cut and normalized cut represent the clustering problem in a graph theoretic manner [2]. A major problem known as the problem of validity for this kind of methods is how to determine the number of clusters in an image. Since the problem is largely unresolved, most techniques need the user to provide a terminating criterion. B.Sowmya and B.Sheelarani segmented color image using soft computing techniques [3]. The soft computing techniques they used were Possibilistic C means algorithm (PCM) and competitive neural network. Fuzzy set and Fussy logic techniques are also used by researchers for solving segmentation problem. A. Borji et al. described a CLPSO-based Fuzzy Color Image Segmentation [6]. H. D. Cheng et al. segmented color image on the basis of fuzzy homogeneity approach [7]. Besides this, artificial neural network (ANN) and Genetic algorithm (GA) techniques also have been used for image segmentation [8]. There are numerous segmentation techniques in medical imaging depending on the region of interest. There are atlas-guided techniques and region growing segmentation methods. Some of them use a semi-automatic approach and still need some operator interaction. Others are fully automatic and the operator has only a verification role. Fan et al. proposed a method of automatic image segmentation by integrating color edge detection and seeded region growing [9]. They used fast Entropy thresholding for edge extraction. After they obtained color edges, which provided the major geometric structures in an image, the centroids between these adjacent edge regions were taken as the initial seeds for seeded region growing. These seeds were then replaced by the centroids of the generated homogeneous image regions by incorporating the required additional pixels step by step. Adams and Bischof proposed another method using seeded region growing [13]. Franc Y Shis and Shouxian Cheng proposed another region based image segmentation method [14] where initial seeds are selected based on the idea of calculating standard deviation in a neighbor and to check whether the value is under a threshold and then assigning each pixel in that region as seeds. This seed selection is followed by a region growing and region merging.

“Natural grouping” of dataset is a method of identifying the hidden clusters of data points present in the dataset. For a set of data points, we can draw the scatter diagram of the set. What one perceives to be the groups present in the set is termed as the natural groups present in the set. These sorts of algorithms are better than the general clustering algorithms in the fact that, they don't need any prior information about the number of clusters present actually in the dataset. Nirmalya Chowdhury and C. A. Murthy proposed a method of obtaining natural groups present in a dataset using minimum spanning tree and bayes classifier [10]. Sugar et al. proposed an information theoretic approach for finding natural groups present in a dataset [11]. Premananda Jana and Nirmalya Chowdhury proposed another method of natural grouping of a given dataset using MST of dataset [15].

3. STATEMENT OF THE PROBLEM

Given an image I and homogeneity predicate P , the segmentation of I is a partition of I into K regions $\{R_1, R_2, \dots, R_K\}$ satisfying the following conditions,

Each pixel in the image should be assigned to a region.

$$\text{i.e. } \bigcup_{p=1}^k R_p = I \quad (1)$$

Each pixel is assigned one and only one region.

$$\text{i.e. } R_k \cap R_{kk} = \emptyset \quad \forall k, kk = 1, 2, \dots, K, k \neq kk \quad (2)$$

Each region satisfies the predicate P , which can be similarity measure in color space or similarity based on texture or any similarity criteria combining both of them etc.

$$\text{i.e. } P(R_k) = \text{True for } k=1, 2 \dots K \quad (3)$$

Two different regions cannot satisfy P at the same time.

$$\text{i.e. } P(R_k \cap R_{kk}) = \text{False where } k \neq kk \quad (4)$$

4. THE PROPOSED METHOD IN THE FORM OF AN ALGORITHM

Fig 1 presents an overview of our algorithm. Here after finding the unique RGB tuples of pixels of a given image, we perform a MST based natural grouping of the pixels. The pixels closest to the centers of the groups thus obtained are found out. These pixels serve as the pivot for region growing image segmentation purpose. This step is immediately followed by a region merging process. We can divide the algorithm into three parts, namely, seed selection, region growing and region merging. These steps are discussed in detail in the next section,

4.1 Selection of seed pixels

Seed pixel selection is the initial step for image segmentation based on region growing. The pixels selected as seeds should satisfy the following properties,

- Every seed should be similar to its neighbors with respect to some criteria P .
- Any pair of seeds should be distinct with respect to the criteria P .

For seed pixel selection purpose, we first perform the natural grouping on the image pixels in the RGB feature space as RGB space is suitable for color display. In image processing and analysis, we often transform the $\langle R G B \rangle$ components into other color spaces. Every color space has its own advantages and disadvantages. Cheng et. al. compared several color spaces including RGB , YIQ , YUV , CIE etc. for color image segmentation purpose [16]. For performing natural grouping, we use the natural grouping method using minimum spanning tree (MST) proposed by Premananda Jana and Nirmalya Chowdhury. Using pixels as the vertices of the MST, it has been used for obtaining the natural groups of pixels based on their RGB values. Euclidean distance (Equation 5) between the data points has been taken to be the edge weight of the said MST.

$$D_e = \sqrt{(R_i - R_j)^2 + (G_i - G_j)^2 + (B_i - B_j)^2} \quad (5)$$

Then we calculate the sum of the edge weight (l_n) of the MST. The threshold for cluster separation is taken to be,

$$Th_{NG} = \frac{\beta \times l_n}{n-1} \quad (6)$$

where n is the no of nodes in the MST [17]. It has been experimentally found that $\beta=2$ has provided consistently good results [15].

Now the edges of the MST are removed for which edge weights exceeds the threshold (6) for cluster separation. Nodes (data points or pixels here) in each path constitute different clusters of data points (pixels). The centers of those clusters are found out. Now the image pixels having intensity values nearest to those cluster centers values are selected as seeds.

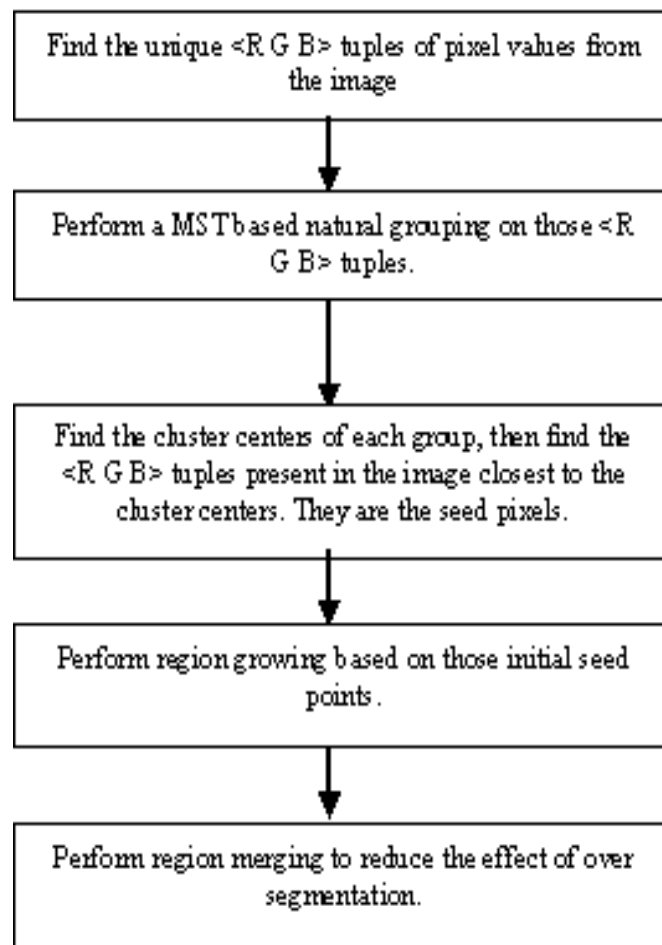


FIGURE 1: Outline of the Proposed Method.

4.2 Region growing step

Seed pixels selected by the above method acts as the pivot for each possible region. For efficiency purpose, every region in the image should contain a seed pixel. But in our proposed method of seed selection, it may so happen that no seed falls in any of the Region of Interest. We eliminate this problem effectively in this region growing stage. In the beginning, we are performing region growing from those initially selected seed pixels. Steps for region growing process are stated below,

- For each natural group we compute the Euclidian distance of all data points from the remaining data points of that cluster using the formula (5). The maximum distance so obtained is recorded. Then we sort those recorded distances in the increasing order and chooses the minimum value to be the threshold (Th_{RG}) for region growing purpose.
- For each seed pixel, we compute the Euclidian distance between the seed pixel and four of its neighboring pixels.
- If the distances thus found are less than the threshold (Th_{RG}), then they are included in the region marked by that particular seed.
- If one or more of the neighboring pixels do not satisfy the condition for region growing as stated above, then they remain unassigned.
- The above methods are repeated for all the seeds.
- For a given seed, the pixels included by region growing method narrated above are labeled by the color (RGB) of the seed pixel.
- Label all the unassigned pixels by the color of the seed pixel to which they are nearest.

4.3 Region merging step

It is possible that the above mentioned method for region growing may result in generation of more number of segments than that is actually present in a given image. Thus at this stage, one needs to use a suitable region merging criteria to merge those regions which should not be given the identity of separate regions. So to get rid of this over segmentation problem a method of region merging is employed which is described below,



FIGURE 2: Some Sample Images from the Database Used for Experimentation.

Two regions (R and R') are merged if,

- They are adjacent.
- If the Euclidian difference between their mean color intensity values is within the threshold value (Th_{RM}):

$$Th_{RM} = \sqrt{b^2(R) + b^2(R')} \quad (7)$$

where $b(R)$ is defined as: -

$$b(R) = g \times \sqrt{\frac{1}{2 \times Q} \left(\frac{1}{|R|} \right) \ln \frac{R_{|R|}}{\partial}} \quad (8)$$

where g is the maximum of particular color channel (R or G or B) of the mean color value of region R. Q is total no of elements in the set of random variable that is used to represent each color channel. Basically q denotes the level of merging. For our purpose we make use q from 1 to 32 and observe the changes in result. |R| represent the cardinality (no of pixels) of region R. ∂ is a parameter which is defined as $1/|I|^2 * 6$ where |I| is the cardinality of the image I.

The merging threshold is based on the work done by Nock and Nielsen [18] where they segment images using statistical region merging technique. The above mentioned threshold (Th_{RM}) is based on the statistical measure having the idea that pixels within a region have the same expectation value and pixels in different regions have different expectation value.

5. EXPERIMENTAL RESULTS AND DISCUSSIONS

We implemented the algorithm in matlab and tested it on a Pentium Core-2 Duo system with 3 GB RAM. We have performed our experiment on about 100 images collected from the web. Fig. 2 shows some ample images from the database created. fig 3 shows some of the samples from the collected image database for showing experimental results. The images are color images of different resolutions. We resized the images into 100 x 100 after starting the segmentation process. The images are basically scenery images. We have also performed comparison with some well-known segmentation techniques.

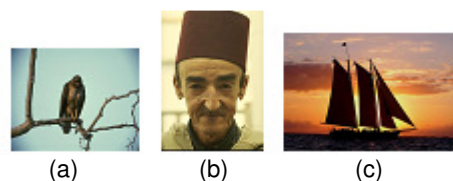


FIGURE 3: Some Sample Images Collected for Experiment.

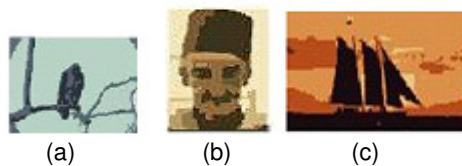


FIGURE 4: Results of the Proposed Method.



FIGURE 5: Results of the Statistical Region Merging Based Method.

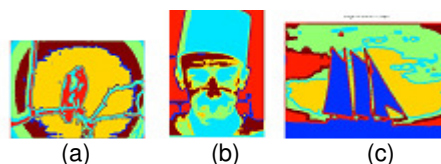


FIGURE 6: Results of the K-Means Based Method.

Fig 4, Fig 5 and Fig 6 show the comparison of the results obtained by our method with the results obtained by [18] and [19]. We obtain both the implementation versions of those algorithms from the web. In fig 4, the results of the segmentation process obtained by the proposed method on

some sample images collected are depicted. Then the results of [18] are shown in fig 5. The results show that our method is relatively free from over segmentation problem. This problem is there in most of the results obtained by the method proposed in [18]. For example, in the result shown in Fig 5(a), the sky is segmented in three different segments while our method shows in Fig 4(a) that the sky is segmented into two parts. Besides, the algorithm proposed in [18] sometimes over merged some regions which make the result difficult to analyze. For example, in the picture Fig 5(b), the right eye of the man is merged with the eye brow and the region covering them. But in our method in Fig 4(b), the eye and eye brow can be identified separately. The sky in Fig 4(c) is easy to understand in comparison with the result in Fig .5(c).

We have made another comparative study with the segmentation algorithm using k-means clustering proposed by Tse-Wei Chen Yi-Ling Chen Shao-Yi Chien [19]. Fig 6 shows the results. This algorithm label a segmented region by a randomly selected color value. The results in 6(a) and 6(c) show that, the sky is over segmented in more number of segments. Whereas the result in 6(b) shows that, most of the parts are over merged.

Our algorithm lacks in two aspects. It requires an additional scanning of the source image after the region growing stage. The preprocessing step needed for the natural grouping step, i.e. drawing the MST requires some addition storage as it requires building the adjacency matrix of the pixel points. Except these two shortcomings, the algorithm works fine for most of the images.

The time complexity for our algorithm spans three different phases: natural grouping for seed selection, region growing and region merging. The natural grouping is MST based and we use Prim's algorithm for finding the MST. Time complexity of prim's algorithm is $O(E+V) \log V$ where E is the no of edges and V is the no of vertices present in the MST. It requires $O(E)$ to find the groups. The time complexity of seed selection is $O(c \times m \times n)$ where c is the number of clusters obtained and $m \times n$ is the image resolution which is 100×100 . In region growing, each unclassified pixel is inserted into the sorted list exactly once. Checking neighboring regions and calculating distances can be done in constant time. Putting a pixel into the sorted list requires $\log(m \times n)$. Therefore, it takes $O((m \times n) \log(m \times n))$ for region growing. After the completion of the region growing, we again scan the whole image to label the unassigned pixels. It takes $O(m \times n)$ time.

In region-merging, calculating the differences between regions takes $O(r^2)$, where r is the number of regions. To calculate sizes for all the regions, it takes $O(m \times n)$. Usually, r is much less than $m \times n$. To merge two regions, we need to label the pixels in the two regions, calculate the mean for the new merged region, and calculate the distances between this and other regions. Therefore, it takes $O(m \times n)$ to merge two regions.

6. CONCLUSION

We have presented an efficient algorithm for color image segmentation with automated seed selection. We performed natural grouping on the image pixels to find out different color bands present in the image. Experimental results show that our method can produce good results. This work can be carried on in future by incorporating texture feature along with color feature for region growing purpose as texture is an important region descriptor.

7. REFERENCES

1. H. Frigui, R. Krishnapuram "Clustering by competitive agglomeration". Pattern Recognition, 30(7):1109-1119, 1997
2. Y. Boykov. "Graph Cuts and Efficient N-D Image Segmentation". International Journal of Computer Vision, 70(2):109-131, 2006
3. B. Sowmya, B. Sheelarani. "Colour Image Segmentation Using Soft Computing Techniques". International Journal of Soft Computing Applications, 4:69-80, 2009

4. Li. Yanling and Yi. Shen. "Robust Image Segmentation Algorithm Using Fuzzy Clustering Based on Kernel-Induced Distance Measure". In Proceedings of International Conference on Computer Science and Software Engineering, 2008
5. E. Dana, and P. F. Whelan. "Color image segmentation using a spatial k-means clustering algorithm". In Proceedings of 10th International Machine Vision and Image Processing Conference, 2006
6. A. Borji, M. Hamidi and A. M. E. Moghadam. "CLPSO-based Fuzzy Color Image Segmentation". In Proceedings of North American Fuzzy Information Processing Society, 2007
7. H. D. Cheng, C. H. Chen, H. H. Chiu, H. Xu. "Fuzzy Homogeneity Approach to Multilevel Thresholding". IEEE Transaction On Image Processing, 7(7):1998
8. P. M. Birgani, M. Ashtiyani and S. Asadi. "MRI Segmentation Using Fuzzy C-means Clustering Algorithm Basis Neural Network". In Proceedings of 3rd International Conference on Information and Communication Technologies: From Theory to Applications, 2008
9. J. F. David, K. Y. Yau, A. K. Elmagarmid. "Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing". IEEE Transactions On Image Processing, 10(10):2001
10. N. Chowdhury, C. A. Murthy: "Minimal spanning tree based clustering technique: Relationship with Bayes Classifier". Pattern Recognition. 30(11): 1919-1929, 1997
11. A. Catherine A, Sugar, G. M. James (2003). "Finding the number of clusters in a data set: An information theoretic approach". Journal of the American Statistical Association. 98: 750-763, 2003
12. N.R. Pal, S.K. Pal. "A Review On Image Segmentation Techniques". Pattern Recognition, 26(9): 1227-1294, 1993
13. R Adams, L. Bischof. "Seeded Region Growing". IEEE Transaction on pattern analysis and machine intelligence, 16(6): 641-647, 1994
14. F. Y. Shih, S Cheng. "Automatic seeded region growing for color image segmentation". Image And Vision Computing, 23:877-886, 2005
15. P. Jana and N. Chowdhury. "Finding the Natural Grouping in a Data Set Using MST of the Data Set". In Proceedings of IICAI, 2005
16. H. D. Cheng, X. Jiang, Y. Sun, J. Wang. "Color image segmentation: advances and prospects". Pattern Recognition, 34(12): 2259-2281, 2001
17. J. Harvey and Greenberg. "Greedy algorithms for minimum spanning tree". University of Colorado, Denver, 1998
18. R. Nock, F. Nielsen. "Statistical Region Merging". IEEE Transaction on Pattern Analysis. Machine Intelligence, 26(11): 1452-1458, 2004
19. T. W. Chen, Y. L. Chen and S. Y. Chien. "Fast image segmentation based on K-Means clustering with histograms in HSV color space". In Proceedings of 10th IEEE Workshop on Multimedia Signal Processing, 2008
20. Image Segmentation Available at:<http://www.en.wikipedia.com>

An Interactive Content Based Image Retrieval Technique and Evaluation of its Performance in High Dimensional and Low Dimensional Space

Tanusree Bhattacharjee

*Computer Science and Engg. Dept
Jadavpur University Kolkata –
700032, India*

tanusree_bhattacharjee85@yahoo.co.in

Biplab Banerjee

*Computer Science and Engg. Dept
Jadavpur University Kolkata –
700032, India*

gettbiplab@gmail.com

Nirmalya Chowdhury

*Faculty, Computer Science and Engg. Dept
Jadavpur University Kolkata –
700032, India*

nirmalya_chowdhury@yahoo.com

Abstract

In this paper we have developed an Interactive Content Based Image Retrieval System which aims at selecting the most informative images with respect to the query image by ranking the retrieved images. The system uses relevance feedback to iteratively train the Histogram Intersection Kernel Based Support Vector Machine Classifier. At the end of the training phase of the classifier, the relevant set of images given by the final iteration of the relevance feedback is collected. In the retrieval phase, a ranking of the images in this relevant set is done on the basis of their Histogram Intersection based similarity measure with query image. We improved the method further by reducing dimensions of the feature vector of the images using Principle Component Analysis along with rejecting the zero components which are caused by sparseness of the pixels in the color bins of the histograms. The experiments have been done on a 6 category database created whose sample images are given in this paper. The dimensionality of the feature vectors of the images was initially 72. After feature reduction process, it becomes 59. The dimensionality reduction makes the system more robust and computationally efficient. The experimental results also agree with this fact.

Keywords: Relevance feedback, Similarity measures, Content Based Image Retrieval.

1. INTRODUCTION

With increase in size of digital image collections, there is a need of developing an efficient search engine for the images for browsing, searching and retrieval.

There are two types of image retrieval technique. The first one is the text based image retrieval [15],[16] where each image in the database is labeled by keywords describing the image. For searching for an image, user formulates the query using keywords which best describes his query image. In [17] a comprehensive survey is provided in this regard.

Since image databases are very large, so annotating such a huge collection is a very tedious task. Also subjective description of a query image may differ from person to person due to differences in human perception.

The second type of image retrieval technique is the content based image retrieval which has become an attractive research area in the last few decades [12],[13],[14] since it do not possess the difficulties of text based approach. The word content here means low level feature descriptors like color ([18],[19]), texture([20],[21]), shape([22],[23]) etc. The CBIR is the technique to map each of the images in the database to a feature space and then retrieve based on the feature of the query image. The main challenge of CBIR is to bridge the semantic gap between low level content descriptors with high level concepts (like faces, flowers, architectures etc).

For this reason relevance feedback has become a critical component of CBIR systems. Through its use the CBIR system interacts with user and tries to bridge the semantic gap. Relevance feedback has shown considerable improvement in the performance of the CBIR system ([24],[25]).

CBIR can be thought of as a classification task. It classifies the images in the image database into two sets: relevant and irrelevant. The relevant set contains all those images which are relevant to the user's query image and the irrelevant set contains the rest of the images in the database. This relevance with user's query image can be found out using relevance feedback given by user.

SVM_{Active} implemented by Tong et.al.[1] uses the relevance feedback component with the traditional SVM classifier for CBIR. It has given better results than existent Query Point Movement (QPM) and Query Expansion (QEX) scheme. But SVM_{Active} has been implemented with traditional kernels like RBF or polynomial. Recently S. Maji et.al. [5] has shown that classification using Intersection kernel based SVM (IKSVM) not only shows improvement in performance but can also be made computationally efficient by using an approximation technique.

In this work we have designed an interactive Content Based Image Retrieval Technique which uses Relevance Feedback. The Classifier used for this purpose is the Intersection Kernel based SVM which is integrated with a ranking based retrieval using Histogram Intersection based similarity measure.

The color feature subspace computed by dividing into bins resulted into non uniform distribution of the pixels. The pixel concentration is fully dependent on the particular image being considered. Hence due to such sparseness of pixel values there is a necessity to reduce the dimensionality of the feature space.

In this context of dimensionality reduction Sirovich et. al.[28] directly used eigen images. Faloutsos and Lin[29], Chandrasekaren *et al.* [30] and Brunelli and Mich [26] used principal component analysis (PCA) to perform the dimension reduction in feature spaces.

One problem with PCA is the unnecessary information loss which may prove to be disastrous. Unnecessary information loss due to feature reduction do not occur for our case as it is found that over 10 dimensions in the feature space became zero after reduction. So we are only eliminating the zero dimensions and considering all the rest of the dimensions.

The paper is organized in the following way. Section 2 contains an overview of Principle Component Analysis. Support Vector Machine and Histogram intersection kernel are discussed in chapter 3 and 4 respectively. Section 4 has the description of histogram intersection based similarity measure. Section 5 contains the feature extraction module, and finally our proposed algorithm is discussed in Section 6. Experimental results are given in section 7. The paper finishes with conclusion and discussions in Section 8.

2. PRINCIPAL COMPONENT ANALYSIS

Our task of content based image retrieval typically focuses on representing the images in the training database and the test database in a relatively high dimensional feature vector. We can apply some dimensionality reduction scheme to eliminate those dimensions that have low impact on the classification process, we are discussing here a method of dimensionality reduction using PCA[11].

PCA aims at highlighting the variability of data. Let us consider we have s_n number of data points in the two dimension space. If they are represented as $(x_i, y_i) \{i = 1, 2, \dots, s_n\}$. The mean of the data points are calculated by the equation,

$$x' = \sum_{i=1}^n x_i / s_n \tag{1}$$

$$y' = \sum_{i=1}^n y_i / s_n \tag{2}$$

These means are subtracted from the data points for the purpose of normalization,

$$Data\ Adjusted = \begin{bmatrix} (x_1 - x') & \dots & (y_1 - y') \\ | & & | \\ (x_{s_n} - x') & \dots & (y_{s_n} - y') \end{bmatrix} \tag{3}$$

We can compute the relative variability of the dimensions in terms of a co variance matrix,

$$\begin{bmatrix} Cov(x, x) & Cov(x, y) \\ Cov(y, x) & Cov(y, y) \end{bmatrix} \tag{4}$$

where the principle diagonal indicates the variances of the dimensions x and y respectively. The Covariance of x, y can be computed as follows,

$$cov(x, y) = \sum_{i=1}^n \{(x_i - x'), (y_i - y')\} / n - 1 \tag{5}$$

From the covariance matrix the eigen values are calculated. There will be two eigen values for 2 dimensional data points stated above. From these eigen values eigen vectors are calculated. The two eigen vectors corresponding to the two eigen values will be orthogonal to each other. From the point of view of PCA, the eigen vectors should be of unit length. The eigen vector with the highest eigen value is called the principal component of the data set as it contributes maximum information about the patterns in data. The pattern information given by the other eigen vectors decreases with the eigen values. Hence a vector of eigen vector is formed where the eigen vectors are arranged in decreasing order of corresponding eigen value.

$$\text{Feature Vector} = (\text{eig_vec1 eig_vec2}) \quad (6)$$

The lower order eigen vectors can be ignored as they contribute very little to the pattern information of the data set. The final data is got by the following equation,

$$\text{Final Data} = \text{Data Adjusted}^T \times \text{Feature Vector}^T \quad (7)$$

The original data set can be got back from this final data using reverse mathematical operations. PCA has been used successfully in many application areas like face recognition, and dimensionality reduction in many data mining applications

3. SUPPORT VECTOR MACHINE

Support Vector Machine, an important machine learning technique has been used efficiently for variety of classification purposes like object based image analysis[2], hand written digit recognition[3], image segmentation[cite reference] among others. SVM can be used efficiently as a binary classifier as well as multi class classifier [27].

3.1 Theoretical background of SVM

We have a set of data points (images) $\{x_i\}$ where $i=1, \dots, l$ (where l is the no. of training points) and each x_i is drawn from a d dimensional feature space (in our case $d = 72$). Our aim is to draw a hyperplane using SVM classifier [4] that will separate those training points into two separate classes i.e. +1 for positive classes and -1 for negative classes. The classes are denoted as y_i $[y_i \in \{-1, +1\}]$.

Linear SVM is the simplest type of SVM classifier which separates the points into two classes using a linear hyperplane that will maximize the margin between the positive and negative set. In our case as we are using SVM with relevance feedback, initially we are not labeling all the training points. We are labeling them iteratively by showing the user some images from the database, which the user marks as relevant (positive) or irrelevant (negative) classes.

The concept of non linearity comes when the data points can not be classified into two different classes using a simple hyperplane, rather a nonlinear curve is required. In this case, the data points are mapped non-linearly to a higher dimensional space so that they become linearly separable. For this purpose the kernel functions are used. Different types of kernel functions are already invented. A choice of kernel function for classification using non linear SVM depends upon the problem domain. Equations for some of them are stated below,

Polynomial :

$$K(x, z) = (w \cdot x + B)^p, \quad p = \text{deg ree} \quad (8)$$

Radial Basis:

$$K(x, z) = e^{-\gamma \|x - z\|^2} \quad (9)$$

In both the equations, x and z are the data points. $K(x,z)$ gives the kernel function for computation of non linear SVM. The main drawback of SVM is its time complexity which can be reduced reasonably as discusses in the next section.

4. INTERSECTION KERNEL BASED SVM

The histogram Intersection Kernel $K(H, H') = \sum_{i=1}^L \min(H_i, H'_i)$ [6] is used as measure of similarity between histograms H and H' with L number of bins. The Histogram Intersection Kernel based Support Vector Machine (IKSVM) has a complexity of $O(mn)$ where n is the dimension of the data points in the data set and m is the number of support vectors of the learned classifier. IKSVM algorithm proposed by S. Maji et. al [5] shows that time complexity is reduced to $O(n \log m)$ and space complexity is $O(nm)$. Finally they are successful in formulating an approximate method which has time and space complexity devoid of the number of support vectors that is $O(n)$. Hence they proved that classification using IKSVM is efficient in terms of resources required.

This result is particularly important because intersection kernel SVMs have been shown to be successful in detection and recognition (pyramid match kernel [7] and spatial pyramid matching [8]), but this is computationally inefficient compared to linear SVMs because non linear kernels require computation and space linearly proportional with number of support vectors (for classification) which also increases with increase in training data [9].

5. HISTOGRAM INTERSECTION BASED SIMILARITY MEASURE

It is shown in [10] Euclidean based similarity measure may not be able to capture user's query concept or user's perception. Inspired by the author [10] we have not used Euclidean based similarity measure.

If two histograms H and H' has L number of bins, then the histogram intersection distance between them is given by,

$$d(H, H') = \frac{\sum_{i=1}^L \min(H_i, H'_i)}{\min\left(\sum_{i=1}^L H_i, \sum_{i=1}^L H'_i\right)} \tag{10}$$

6. FEATURE EXTRACTION

We calculate both Color and Texture features for each image in the database. The shape feature is calculated as a property of color and texture features. Number of components calculated for feature vector for each image is 72. Each image represents a point in 72 dimensional feature spaces.

Characteristic Values	R	G	B
1	0	0	0
2	0	0	255
3	0	255	0
4	0	255	255
5	255	0	0
6	255	0	255
7	255	255	0
8	255	255	255

TABLE 1: Characteristic RGB Values.

6.1 Color Feature

We have used histogram based feature extraction by binning the pixel values based on RGB color space. We have defined 7 bins for which the ranges are defined in Table 1. Each of the bin corresponds to the pixel values between the Range x and Range $x+1$ where $x=1,2,\dots,7$.

The number of pixels in each of the bin varies from image to image. Many of the bins contain very little number of pixels in comparison to other bins.

Since RGB color space does not provide any information of image brightness and how each pixel is saturated with white color, RGB space is not an effective way of computing feature space. Hence we convert the pixels in each bin from RGB to HSV space where H corresponds to hue or the brightness, S corresponds to the amount by which each pixel is saturated with white color and V corresponds to the dominant color. In each bin, for each of H, S and V channel statistical moments like mean and variance are calculated. Hence they provide $7(\text{number of bins}) \times 3(\text{channels per bin}) \times 2(\text{number of features per channel per bin}) = 42$ features.

Using Gaussian distribution of average mean (average over three channels) and average variance, spreadness of bin and elongation of bin are calculated. These provides an additional $7(\text{number of bins}) \times 2(\text{number of features per bin}) = 14$ features.

6.2 Texture Feature

Texture is the tactile characteristic of a surface. For texture feature extraction purpose, we first obtain the discrete wavelet transformed version of a given image using haar wavelet mask. Wavelets are functions generated from a single function by its dilations and translations. The haar wavelet transform forms the simplest and oldest compression of this kind. The haar wavelet transform sub divides the image into 4 parts which denotes the image orientations in 4 different angular values, 0,45,90,180; the process corresponds to the following filtering operations,

- *Top left*: passing the image through 2-D lowpass filter (Lo-Lo).
- *Top right*: passing the image through horizontal highpass and vertical lowpass filter (Hi-Lo).
- *Lower left*: passing the image through horizontal lowpass and vertical highpass filter (Lo-Hi).
- *Lower right*: passing the image through 2-D highpass filter (Hi-Hi).

From an image of resolution $M \times N$, the haar wavelet transform obtains sub images in four different orientation each of resolution $M/2 \times N/2$. From these 4 sub images, we calculate the grey level co-occurrence matrix (GLCM). GLCM represents the occurrences of a given pair of intensity values of an image in a certain direction and hence helps in detecting the repeating patterns in the image. We calculate the energy co-efficient from each of the GLCM and sum them up. The inherent idea is that, the energy measure of a complex signal can be obtained by summing up the energy measures of its constituent signals. The energy measure obtained can be used as an useful image feature because it highlights on the local (pixel wise) as well as global (images in different orientations) descriptions of the image.

The energy measure for the 4 GLCMs are calculated as follows

$$Energy = \sum_i \sum_j P_d^2(i, j) \quad Energy = \sum_i \sum_j P_d^2(i, j) \quad (11)$$

where (i,j) represents the (row, col) pair of each GLCM represented by P_d where d represents displacement of the GLCM.

From each GLCM we obtain energy mean, energy variance, energy spreadness and energy elongation. The energy spreadness and energy elongation are calculated from Gaussian distribution of energy mean and energy variance. These provides another additional $4(\text{number of GLCM}) \times 4(\text{number of features per GLCM}) = 16$ features

7. PROPOSED METHOD

The proposed method shows some relevant images to the user based on the query image the user selects. This is done by interactively asking the user to label a few images from training set into relevant and irrelevant category. Then using those labeling the histogram intersection kernel based SVM is trained. These feedback rounds go on iteratively and each time a refinement of the images shown to the user is done. Finally those images which are judged by the SVM as positive are taken. The query image is plotted in the positive set. Now those images which are nearest to the query image based on a histogram intersection based similarity ranking are taken and shown to the user.

The performance of the proposed method is checked in two stages. The first and second stage is without and with dimensionality reduction respectively. The above paragraph described the first stage briefly. The second stage blindly follows the first stage with only one exception. After each feature extraction phase, there is a following dimensionality reduction phase with the PCA in the second stage. We are showing the second phase in the form of algorithm as the first phase could be easily derived from it by eliminating the dimensionality reduction phase.

Notations	Descriptions
C	Number of categories of images in the database
p	Number of images in each category
Tr	Training set
r	Number of images of each category in Tr
Ts	Test set
k	Number of images displayed to user for feedback
F	k images forms set F
Rr	Set of images marked as relevant by user
Ri	Set of images marked as irrelevant by user
Fr	Feature vector set of Rr
Fi	Feature vector set of Ri
F_{Tr}	Set of images classified as positive by SVM
F_{Ti}	Set of images classified as negative by SVM
g	Number of images to be displayed according to user requirement

TABLE 2: Notations for Algorithm.

We have collected a database with C categories of images, each category having p number of images. The training set Tr is formed from this parent database by selecting randomly r ($r < p$) images for each of the category. Hence Tr has total rxC images. The rest of the images in the database form the test set Ts . Hence Ts has total $(p-r)xC$ images.

7.1 Training by $IKSVM_{Active}$

Phase 1: Querying:

User selects an image from Ts .

Phase 2: Relevance Feedback:

Step 1: A fixed number (k) of images are selected randomly without replacements from Tr and kept in set F .

Step 2: User is shown the set F containing those images.

Step 3: User marks them as relevant or irrelevant.

Step 4: Relevant images form set Rr and irrelevant images form set Ri ($Rr \in Tr$ & $Ri \in Tr$).

Phase 3: Training IKSVM:

Step 1a: The set Rr is modified by including the query image selected by user in it.

Step 1b: Calculate feature vector of Rr and Ri . Feature vectors of Rr forms the positive set and feature vectors of Ri forms the negative set of data points for training the classifier.

Step 1c: The dimensionality of these feature vectors are reduced by calculating the principle components and rejecting all those dimensions which reduces to zero due to sparseness of pixel concentration in the bins of histogram calculated.

Step 2: These sets are then given as input to IKSVM. It draws the hyperplane on the basis of them.

Step 3: The feature set of Tr is calculated and then fed to IKSVM for classification so that the hyperplane formed can separate data points in training set as positive or negative. Those data points which falls on the positive side of the hyperplane forms a set F_{Tr} and similarly the negative side forms another set F_{Ti} . Since an image is a data point in a feature space we use the term data point and feature vector interchangeably.

Phase 4: Selecting sample images for next feedback round:

Step 1: $k/2$ data points are selected from the set F_{Tr} , which are the most nearest to the hyperplane and similarly $k/2$ data points are selected from the set F_{Ti} . The images corresponding to these k data points are kept in set F .

Step 2: If this is not the final iteration then, Iterate from Step 2 of Phase 2 to Step 2 of Phase 4. Else iterate from Step 2 of Phase 2 to Step 3 of Phase 3.

7.2 Retrieval using histogram intersection based similarity measure

Step 1: Collect the set F_{Tr} from the last iteration of training by IKSVM_{Active}.

Step 2: Find from F_{Tr} , the reduced dimensional feature vector of the query image Now plot it in reduced dimensional space feature space.

Step 3: Plot all the data points in F_{Tr} in the same space (reduced dimensional).

Step 4: Now, say user wants g number of images from the database which are most relevant to the query image, then on the basis of histogram intersection based similarity measure calculate the distance of query data point with data points in the set F_{Tr} .

Step 5: Sort in descending order the distances as two of the most relevant images will have histogram distance close to 1 and the dissimilar ones will have distance close to 0 and then retrieve first g images from that sorted list.

The experimental results are discussed in the next section.

8. EXPERIMENTAL RESULTS

We have done our experiments on a 6 category image database, each category having 120 images. The categories include field, sea, sunset, white rose, red rose and sunflower.



FIGURE 1: Some Sample Images from the Database Used for Experimentation.

The categories have ambiguous RGB values. In sunset category the images have mainly red, yellow and orange colors. In sunflower and red rose category, the images have mainly yellow and red color values. Similarly in white rose category the images have mainly white and green colors. In sea and field category white and green are predominant. The database (<http://www.flickr.com/photos/48753989@N03/>) is created in this way by collecting images from Google search engine, mainly to confuse the classifier and check how it behaves in this environment. Two images from each category of the database created is shown in Fig 1.

The training set is computed in such a way that it contains 80 images from each category. The test set contains remaining 40 images from each category. One of the performance measures for image retrieval is precision which is defined below,

$$\text{Precision} = (\text{Number of relevant images retrieved}) / (\text{No of images retrieved})$$

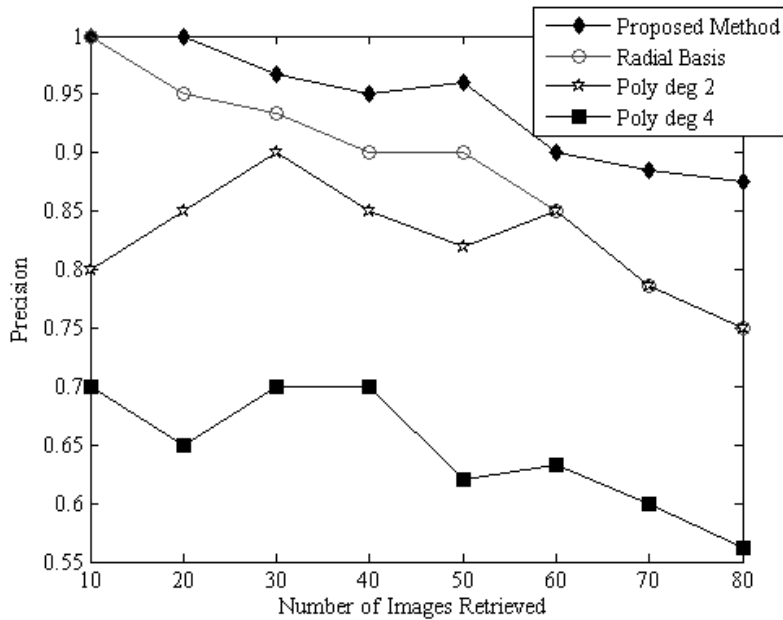


FIGURE 2: Comparative study of Existing Kernel Functions Used With SVM_{Active} and the Proposed Method After Three Feedback Rounds Before Dimensionality Reduction.

The graph of number of images retrieved vs. precision of the system in Stage 1 is shown in fig 2. It shows that the proposed method performs better than other 3 methods for instance when 20 images are retrieved the proposed method has 100% precision whereas SVM_{Active} used with Radial Basis kernel gives 95% precision, with polynomial degree 2 kernel gives 85% precision and with polynomial degree 4 kernel gives 65% precision.

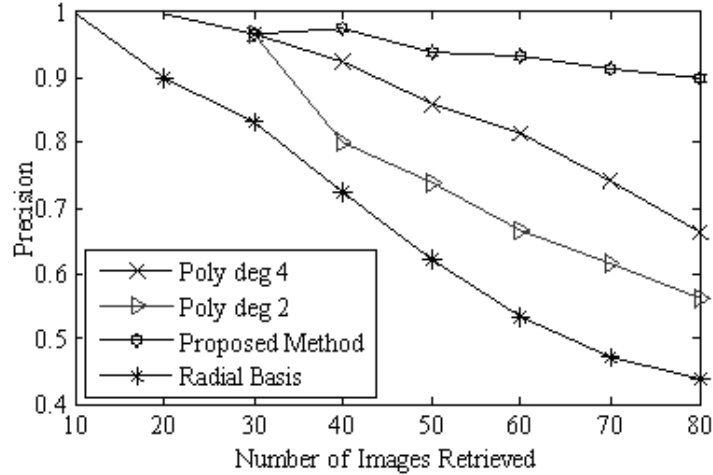


FIGURE 3: Comparative Study of Existing Kernel Functions Used With SVM_{Active} and the Proposed Method After Three Feedback Rounds After Dimensionality Reduction With PCA.

The graph of number of images retrieved vs. precision of the system in Stage 2 is shown in fig 4. The graph shows that the proposed method performs better in reduced dimensionality condition, than other 3 methods for instance when 40 images are retrieved the proposed method has 98% precision whereas SVM_{Active} used with polynomial (degree 4 kernel) gives 92% precision, with polynomial degree 2 kernel gives 80% precision and with polynomial degree 4 kernel gives 72% precision.

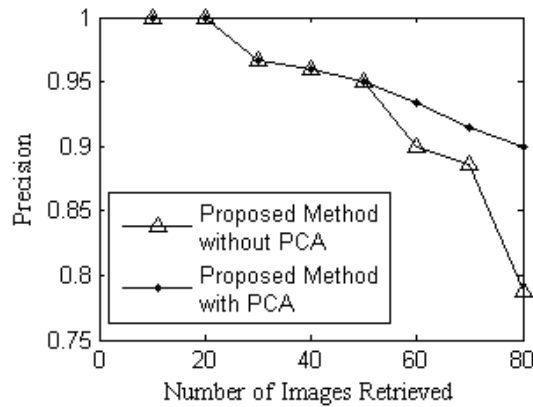


FIGURE 4: Proposed Method With and Without Dimensionality Reduction by PCA.



Query Image Q1



First 5 Results with Proposed Method Before Dimensionality Reduction



FIGURE 5: Comparative Study of Proposed Method Before and After Dimensionality Reduction by PCA.

The number of images retrieved vs. precision graph in fig 4 shows a comparison of the performance of the proposed method in Stage 1 and Stage 2.

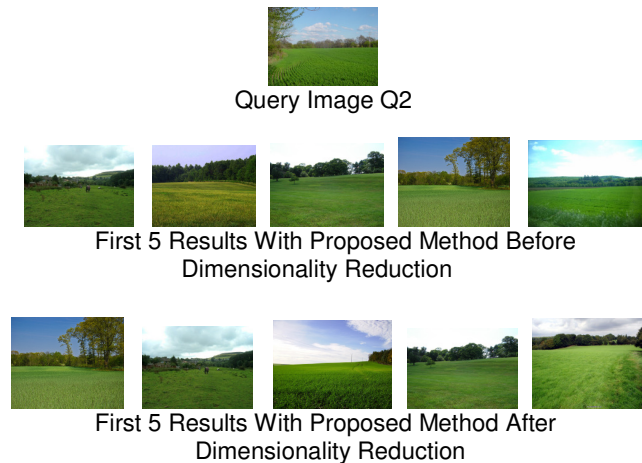


FIGURE 6: Another Comparative Study of Proposed Method Before and After Dimensionality Reduction by PCA.

The method without using PCA shows a sharp fall in performance whereas the one with using PCA gradually degrades its performance. Fig 5 and 6 also shows the comparison in the form of simulation results. In fig 5 the query image is a close up view of the white rose. Here dimensionality reduction phase returns those images which have more such type of images as the query image. Similarly in fig 6 the breadth of the field and sky in the query image is preserved more in the returned images after dimensionality reduction.

9. CONCLUSION AND DISCUSSIONS

The proposed method is efficient because unnecessary zero dimensions are deleted. Hence number of computations lessens for drawing the hyperplane using SVM as well as for nearest neighbor computation.

For large databases, the complexity can be reduced by sampling the database as we have done. Further sub sampling may further reduce the complexity. A future direction of research may be to use large databases and efficiently sub sampling them for using the proposed method and the results could be verified.

10. REFERENCES

1. S. Tong and E. Chang. "Support vector machine active learning for image retrieval". In Proceedings of ACM Multimedia. Ottawa, Canada, 2001
2. A. Tzotsos and D. Argialas. "A support vector machine approach for object based image analysis". Support Vector Machine Classification for Object-Based Image Analysis, Springer-Verlag, Chapter 7.2, 2008
3. T. Joachims. "Text categorization with support vector machines". In Proceedings of the European Conference on Machine Learning. Springer - Verlag, 1998
4. C. Burges. "A tutorial on support vector machines for pattern recognition". Data Mining and Knowledge Discovery, 2:121-167, 1998
5. S. Maji, A.C. Berg and J. Malik "Classification using intersection kernel support vector machine is efficient". Appeared in IEEE Computer Vision and Pattern Recognition Anchorage, 2008
6. M. J. Swain and D. H. Ballard. "Color indexing". IJCV, 7(1):11-32, 1991
7. K. Grauman and T. Darrell. "The pyramid match kernel: discriminative classification with sets of image features". ICCV, 2: 2005
8. L. Lazebnik, C. Schmid, and J. Ponce. "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories". In Proceedings of CVPR, 2006
9. I. Steinwart. "Sparseness of support vector machines-some asymptotically sharp bounds". In Proceedings of NIPS, 2003
10. Y. Rui, T.S. Huang, S.F. Chang. "Image Retrieval: Past, Present, and Future". Journal of Visual Communication and Image Representation, 10:1-23, 1999
11. L. I. Smith. "A tutorial on Principal Component Analysis".
12. J. Bi, Y. Chen and J.Z. Wang. "A sparse support vector machine approach to region-based image categorization". In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR '04), 2004
13. I.J. Cox, T.P. Minka, T.V. Papatomas, P.N. Yianilos. "The bayesian image retrieval system, pichunter: theory, implementation, and psychophysical experiments". IEEE Transaction on Image Processing, 9:20-37, 2000
14. K.-S. Goh, E.Y. Chang, and W.-C. Lai. "Multimodal concept-dependent active learning for image retrieval". In Proceedings of 12th Annual ACM International Conference of Multimedia, 2004
15. N.S. Chang and K.S. Fu. "A relational database system for images". Technical Report. TR-EE. 79-28, 1979
16. N.S. Chang and K.S. Fu "Query-by pictorial- example". IEEE Transaction On Software Engineering, 6(6):1980
17. S. K. Chang and A. Hsu. "Image information systems: where do we go from here?" IEEE Transaction on Knowledge and Data Engineering, 4(5):1992
18. J. Huang, R. Kumar, M. Mitra, W. J. Zhu and R. Zabih. "Image Indexing Using Color Correlograms". In Proceedings of CVPR, 1997
19. N. Sebe and M. S. Lew. "Robust color indexing". In Proceedings of the 7th ACM International conference on Multimedia, 1999
20. B.S. Manjunath, W.Y. Ma. "Texture features for browsing and retrieval of image data". IEEE Transaction on Pattern Analysis and Machine Intelligence, 18: 837-842,1996
21. J.Z. Wang, G. Wiederhold, O. Firschein, X.W. Sha. "Content-based image indexing and searching using daubechies' wavelets". International Journal of Digital Libraries, 1(4):311-328, 1998
22. C. Schmi, R. Mohr. "Local grayvalue invariants for image retrieval". IEEE Transaction on Pattern Analysis and Machine Intelligence. 19: 530-535, 1997
23. X. S. Zhou, Y. Rui and T. Huang. "Water-Filling: a novel way for image structural feature extraction". In Proceedings of IEEE International Conference on Image Processing, 2: 570-574, 1999
24. P. Hong, Q. Tian and T.S. Huang. "Incorporate support vector machines to content-based image retrieval with relevance feedback". In Proceedings of IEEE International Conference on Image Processing, Vancouver, Canada, 2000

25. S. D. MacArthur, C. E. Brodley and C. Shyu. "*Relevance feedback decision trees in content-based image retrieval*". In Proceedings of IEEE Workshop CBAIVL, South Carolina, 2000
26. B. Roberto and M. Ornella. "*Image retrieval by examples*". IEEE Transaction on Multimedia, 2: 2000
27. K. B. Duan and S. S. Keerthi. "*Which is the best multiclass svm method? an empirical study*". Springer-Verlag, 2005
28. M. Kirby and L. Sirovich. "*Application of the Karhunen-Loeve procedure for the characterization of human faces*". IEEE Transaction on Pattern Analysis and Machine Intelligence. 12:103–108, 1990
29. C. Faloutsos and K. Lin. "*Fastmap: A fast algorithm for indexing, data-mining and visualization of traditional and multimedia*". In Proceedings of SIGMOD, 1995
30. S. Chandrasekaran, B. S. Manjunath, Y. F. Wang, J. Winkeler and H. Zhang. "An eigenspace update algorithm for image analysis". CVGIP: Graph. Models Image Process. J., 1997

Improving the Accuracy of Object Based Supervised Image Classification using Cloud Basis Function Neural Network for High Resolution Satellite Images

Imdad Ali Rizvi

*Centre of Studies in Resources Engineering,
Indian Institute of Technology Bombay,
Mumbai 400076, India.*

imdadrizvi@iitb.ac.in

B.Krishna Mohan

*Centre of Studies in Resources Engineering,
Indian Institute of Technology Bombay,
Mumbai 400076, India.*

bkmohan@iitb.ac.in

Abstract

A lot of research has been undertaken and is being carried out for developing an accurate classifier for extraction of objects with varying success rates. Most of the commonly used advanced classifiers are based on neural network or support vector machines, which uses radial basis functions, for defining the boundaries of the classes. The drawback of such classifiers is that the boundaries of the classes as taken according to radial basis function which are spherical while the same is not true for majority of the real data. The boundaries of the classes vary in shape, thus leading to poor accuracy. This paper deals with use of new basis functions, called cloud basis function (CBF) neural network which uses a different feature weighting, derived to emphasize features relevant to class discrimination, for improving classification accuracy. Multi layer feed forward and radial basis function (RBF) neural network are also implemented for accuracy comparison sake. It is found that the CBF NN has demonstrated superior performance compared to other activation functions and it gives approximately 3% more accuracy.

Keywords: Accuracy assessment, Image Segmentation, Image classification, Object based image analysis, Radial basis functions neural network.

1. INTRODUCTION

Object based image classification methods are increasingly used for classification of land cover/use units from high resolution images, and often the final result is close to the way a human analyst would interpret the image. Object based image classification does not operate directly on single pixels, but image objects which refer to homogeneous, spatially contiguous regions. These are obtained by dividing an image, namely image segmentation, which is a challenging problem due to the fact that it is no longer meaningful to carry out this task on a pixel-by-pixel basis. The fine spatial resolution implies that each object is now an aggregation of a number of pixels in close spatial proximity, and accurate classification requires that this aspect be considered. To deal with the problem of complexity of high resolution images, the image is first segmented into homogeneous regions, and a set of features are computed for each region segment. These

segments are classified using one or more of the machine learning algorithms. In the present study, various activation functions for artificial neural network classification are considered. This method basically includes three steps. 1) Image segmentation to extract the regions from the pixel information based on homogeneity criteria. 2) Calculation of spectral parameters like mean vector, texture, NDVI and spatial/shape parameters like aspect ratio, convexity, solidity, roundness and orientation for each region. 3) Classification of image using the region feature vectors using suitable classifiers such as NN.

Supervised classification is one of the most commonly undertaken analyses of remotely sensed data. The output of a supervised classification is effectively a thematic map that provides a snapshot representation of the spatial distribution of a particular theme of interest such as land cover. The goal of a supervised image classification system is to group images into semantic categories giving thus the opportunity of fast and accurate image search. To achieve this goal, these applications should be able to group a wide variety of unlabelled images by using both the information provided by unlabelled query image as well as the learning databases containing different kind of images labelled by human observers. In practice, a supervised image classification solution requires three main steps: pre-processing, feature extraction and classification [1]. Based on this architecture, many image classification systems have been proposed, each one distinguished from others by the method used to compute the image signature and/or the decision method used in the classification step. Artificial Neural Network (ANN) and Support Vector Machine (SVM) are commonly used advanced methods for supervised classification of remotely sensed data [2]. The serious drawback of SVM is that the boundaries of the classes as taken according to radial basis function networks are spherical while the same is not true for majority of the real data. The boundaries of the classes vary in shape, thus leading to poor accuracy. This work is developed on the modified RBFs neural network based classifier for object based classification of high resolution satellite remotely sensed images. The new basis functions, called cloud basis functions use a different feature weighting, derived to emphasize features relevant to class discrimination as discussed in [3]. Further, these basis functions are designed to have multiple boundary segments, rather than a single boundary as for RBFs. This new enhancement to the basis function along with a suitable training algorithm allows the neural network to better learn the specific properties of the problem domain. The boundaries of classes considered are not spherical but a set of boundaries is considered for each class, which promises higher accuracy theoretically. This technique is emphasized specifically for multi-spectral satellite images. Thus it was aimed to propose a suitable classifier for the high resolution satellite remotely sensed images and to test the applicability of modified cloud basis functions for the field of remote sensing.

This paper is discussed under five different headings. In Section 2, proposed methodology for object based image segmentation and classification are elaborated. Some of the feature vectors are also discuss in the same section. In Section 3, comprises of experimental results and discussions. Section 4 summarizes the research findings and points out avenues for possible future works.

2. METHODOLOGY

The proliferation of high-spatial resolution multispectral imagery from satellite and aerial sensors (e.g. IKONOS from GeoEye, Inc., QuickBird from DigitalGlobe, Inc., ADS40 from Leica Geosystems, Inc.) has significantly changed the level of sophistication required in digital image processing [4]. In this paper we propose an approach for improving the accuracy of object based supervised image classification using Cloud Basis Functions Neural Network for high resolution remotely sensed multi-spectral satellite images, such as IKONOS or QuickBird.

Proposed Methodology for Object Based Image Segmentation

Over the last decade the analysis of Earth observation data has evolved from what were predominantly per-pixel multispectral-based approaches, to the development and application of

multiscale object-based methods. To empower users with these emerging object-based approaches, methods need to be intuitive, easy to use, require little user intervention, and provide results closely matching those generated by human interpreters. In an attempt to facilitate this, we present object-specific segmentation as an integrative object-based approach for automatically delineating image-objects (i.e., segments) from a high-spatial resolution remotely sensed image [5]. Object-based image analysis subdivides the image into meaningful homogeneous regions not only based on spectral properties but also on shape, texture, size, and other topological features, and organizes them hierarchically as image objects (also referred to as image segments) [6]. The segmentation procedure (extraction of the image objects) is controlled by the user-specified scale (size) or resolution of the expected objects. Object-based approaches have been successful for land-use and land-cover classification [7][8]. Classification of high-resolution satellite images using standard per-pixel approaches is difficult because of the high volume of data, as well as high spatial variability within the objects. One way to deal with this problem is to reduce the image complexity by dividing it into homogenous segments prior to classification. This has the added advantage that segments can not only be classified on basis of spectral information but on a host of other features such as neighborhood, size, texture and so forth. The proposed methodology for object based image segmentation is shown in Fig 1.

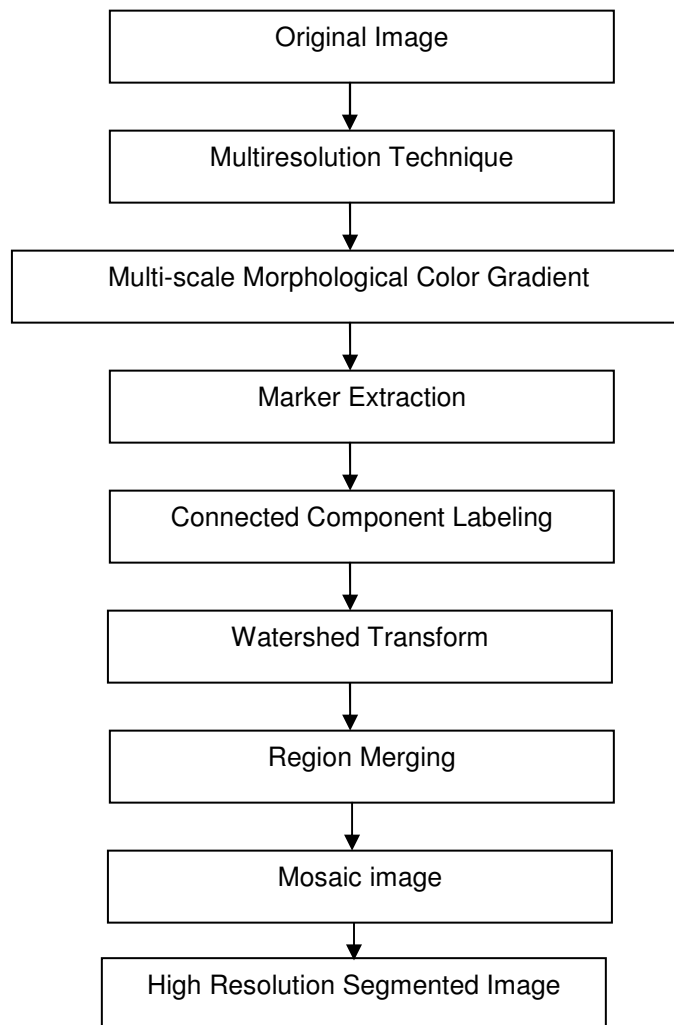


FIGURE 1: Proposed methodology for object based image segmentation

Segmentation of the images is carried out using the region based algorithms such as morphological marker based watershed transform by employing the advantages of multi-resolution framework and multi-scale gradient algorithms. The segmentation of the color images is obtained using watershed transform to get its homogenous regions. Classification technique is then applied into these homogenous regions taking the shape, texture and spectral properties of the regions. The proposed algorithm is given below

- Apply multi-resolution framework (here Daubech6 family of wavelet transform is used) to input image.
 - Use multi-scale gradient algorithms to calculate color gradient.
- The morphological gradient of each band of the image is calculated using equation (1)

$$G(f) = (f \oplus B) - (f \ominus B) \quad (1)$$

where $G(f)$ = Morphological color gradient,
 f = Given image
 B = Structuring element.

$$G(f) = \sqrt{(G_r(f))^2 + G_g(f)^2 + G_b(f)^2} \quad (2)$$

$G_r(f)$ = Gradient of the red band,
 $G_g(f)$ = Gradient of the green band and
 $G_b(f)$ = Gradient of the blue band.

The multi-scale morphological color gradient is dilated with a square structuring element of size 2x2.

- The markers can be extracted from white top-hat or black top-hat transform. But extracted markers from either white or black top-hat will miss some of the objects. So, to utilize the advantage of both top-hat, markers are extracted using morphological laplacian [9], which can be defined as:

$$L(f) = g^+(f) - g^-(f) \quad (3)$$

where $g^+(f)$ = White top hat transform and
 $g^-(f)$ = Black top hat transform

For utilizing the spectral property of the image, markers are extracted from morphological color laplacian of the image; and is calculated using equation (4)

$$L(f) = \sqrt{(L_r(f))^2 + L_g(f)^2 + L_b(f)^2} \quad (4)$$

where $L(f)$ = Morphological color gradient,
 $L_r(f)$ = Gradient of the red band,
 $L_g(f)$ = Gradient of the green band and
 $L_b(f)$ = Gradient of the blue band.

- Apply connected component labeling to connect various labels.
- Morphological marker based watershed transform algorithm is used for region segmentation [10].
- Region merging is done to avoid over-segmentation.
- Mosaic image is generated.
- Inverse wavelet transform is used to generate high resolution image.

The output of the watershed transform may result in over-segmentation. To merge the adjacent region or the homogenous regions; region merging criterion is implemented. Each segmented object or region is assigned the average grayscale of each band to generate the

mosaic color image. To get the final segmentation at high resolution image, low frequency coefficient of the wavelet is replaced with mosaic image, while detailed coefficients of the wavelet are modified so as to avoid noise introduced back into the finer image. Inverse wavelet transform is then applied on these modified images to get the high resolution segmented image.

2.2. Feature Vector Extraction

In general, physical features have certain associations with spectral features, hence they can be identified by using multi-spectral information from the remotely sensed images. Features of objects can be further divided into three categories

- Geometric
- Spectral or thematic
- Textural

A feature vector of all the regions present in the image is calculated. For this work totally 8 features were calculated. The first three values correspond to the values of region's average color in multi-spectral space. The next three features are related to the shape of the region such as solidity, aspect ratio and eccentricity. The next features correspond to the texture features of each region like contrast ASM etc.[11].

2.3. Object Based Image Classification

Many classifiers are available for classification of multi-spectral satellite images. These include discriminate analysis, maximum likelihood classification scheme, etc. A major disadvantage of these classifiers is that they are not distribution free. This has prompted significant increase in use of ANN for classification of remotely sensed images [12]. Several other reasons can be cited in favor of Neural Network (NN) based classifiers as listed below [13].

- Each of the (region) parameters will be in a different numerical range, some in [0,1], some in [0, 255], etc. Rescaling all parameters to a single range can affect the inter-class and intra-class separation.
- NN classifiers can detect and use to their advantage non-linearity in data patterns.
- Ancillary data can be included in NN classifiers.
- NN architectures are flexible which can be easily optimized for performance.
- NN can handle multiple subcategories per class.

Multi Layer Feed Forward (MLFF) and Radial Basis Function (RBF) NN classification techniques are widely used remote sensing applications. In this study we consider one more type of NN classifier called Cloud Basis Function (CBF) NN.

2.3.1. Multi Layer Feed Forward Neural Networks

Typically an MLFF NN consists of a set of sensory units (source nodes) that constitute the input layer, one or more hidden layers of computation nodes, and an output layer of computation nodes. The input signal propagates through the network in a forward direction on a layer-by-layer basis. Learning in MLFF NN consists of two passes through the different layers of the network: a forward pass and a backward pass. In the forward pass, an input pattern is applied to the sensory nodes of the network and its effect propagates through the network layer by layer. Forward pass is followed by a backward pass. During backward pass the error signal (difference between actual output of the network and the desired output) is propagated backward through the network, against the direction of synaptic connections. Hence it is named as back-propagation [14]. The synaptic weights are adjusted to make the actual response of the network move closer to the desired response. Following parameters are considered while implementing RBF ANN,

- Number of input nodes = 7
- Number of output nodes = 9
- Number of hidden layers = 1
- Number of nodes in each hidden layer = 8
- Learning rate = 0.79
- Momentum = 0.5
- Normalization factor for patterns = 255

2.3.2. Radial Basis Function Neural Networks

Design of RBF NN can be viewed as a curve fitting approximation in a high dimensional space. Learning is equivalent to finding a surface in a multidimensional space that provides a best fit to the training data. Correspondingly generalization is equivalent to the use of this multidimensional surface to interpolate the test data. The construction of a RBF NN, in its most basic form, involves three layers with entirely different roles. The input layer is made up of source nodes (sensory units) that connect the network to its environment. The second layer, the only hidden layer in the network, applies a nonlinear transformation from the input space to the hidden space; where hidden space is in general of high dimensionality. The output layer is linear, supplying the response of the network to the activation pattern applied to the input layer. A mathematical justification for the rationale of a nonlinear transformation followed by a linear transformation may be traced back to an early paper by [15]. According to this paper, a pattern-classification problem cast in a high dimensional space is more likely to be linearly separable than in a low-dimensional space. Following parameters are considered while implementing RBF ANN,

- Number of input nodes = 7
- Number of output nodes = 9
- Number of hidden layers = 1
- Number of nodes in each hidden layer = 8
- Learning rate = 0.85
- Momentum = 0.5
- Normalization factor for patterns = 255

2.3.3. Proposed Object Based Supervised Image Classification using Cloud Basis Function Neural Networks

Rather than treating image as set of pixels if we treat it as a set of objects more information can be extracted, as with pixels only intensity values can be used. And with the construction of regions, knowledge is given to the system to classify. This is similar to the way human brain analyzes an image by breaking it down into various objects and uses features such as shape, texture, color and context along with the its cognizance powers to interpret the image. Therefore, dividing the image into regions and then opt for classification is better than per pixel classification. Hence cloud basis function neural network is used which is essentially a form of neural network with modification in radial basis function neural network, the algorithm is as follows:

Creating the modified radial basis function neural network

- Define the input nodes, which take in as input the data from the images.
- Define the intermediate nodes for basis function mapping, which map the inputs to the basis space through the Gaussian functions.
- Define the output nodes, which form the classes in the image.

Programming the training algorithm for the neural network

- Apply k-means clustering for initial data to find the possible basis function centers, μ .
- Form the basis function mappings.
- Calculate the scale factors, for each of the basis function centers with respect to each of the other basis function centers.

$$\omega_{p,j} = \sqrt{\frac{1}{2} \sum_{i=1}^d (\mu_{ip} - \mu_{ij})^2} \tag{5}$$

where, $\omega_{p,j}$ = Scale factor of the boundary segment between pth mean and jth mean

$(\mu_{ip} - \mu_{ij})$ = Euclidian distance between pth mean and jth mean

d = Number of features in each object

And the default scale factor as the mean of all the scale factors as

$$\omega_{0,j} = \frac{1}{k} \sum_{p=1}^{k} \omega_{p,j} \tag{6}$$

where, $\omega_{0,j}$ = Default scale factor for j^{th} mean

- Compute the output matrix of the basis function mapping, ϕ , for the input samples.

$$\phi_j(x | \mu_j, \{\omega\}_j, U_j) = \exp \left(- \frac{\sum_{i=1}^d u_{i,j} (x_i - x_j)^2}{(Sel(\{\omega\}_j | x))^2} \right) \tag{7}$$

where, $\phi_j(x | \mu_j, \{\omega\}_j) =$ Basis function output for sample x belonging j^{th} to cluster

$Sel(\{\omega\}_j | x) =$ Boundary segment selected for sample x

$(x_i - u_{ij})^2 =$ Euclidian distance between sample x and j^{th} mean

- Compute the post basis function weight matrix, W .

$$W = (\phi)^{\neg} T \tag{8}$$

where $(\phi)^{\neg} =$ Pseudo inverse of the output of the basis function matrix
 $T =$ Target Vector

- Compute the output of the network for the input samples and the error in the output with respect to the target vector T as the Euclidean distance from the target vector.
- Update the scale factors and the basis function centers based on the error in the output of the network using the supervised iterative gradient descent algorithm.

$$\{\omega\}_j^m = \{\omega\}_j^{m-1} - \frac{\partial E}{\partial \{\omega\}_j^{m-1}} \tag{9}$$

- After iterative gradient descent is complete for the training iteration, the network output for all the training samples is calculated.

$$Network_Output = \Phi * W \tag{10}$$

where, $\Phi =$ Basis function output matrix and
 $W =$ Post basis function weight matrix

- According to the network output, classify the pixels and partition the training set into two sets of classified $\{X^C\}$ and misclassified samples $\{X^M\}$.
- If the number of misclassified samples is less than a set threshold, or if the number of misclassified samples doesn't change in successive cycles, stop training.
- For all the classes for which the number of misclassified samples is greater than the set threshold, add a basis function to improve the representation of the class.
- Repeat the training algorithm till a maximum number of epochs are completed or till the number of misclassified samples do not change with the increasing basis functions

Classifying the test images using the network

- Input the test images for classification
- Obtain the output matrix for the classification details of the image

- Calculate the classification accuracy of the network

Following parameters are considered while implementing CBF NN,

- Total number of training samples taken = 135
- Maximum number of training iterations = 10
- Maximum number of iterations for the iterative gradient descent for updating the scale factors calculated during each training iteration = 2
- Learning rate for the iterative gradient descent for updating the scale factors calculated during each training iteration = 0.2
- Maximum number of neighboring functions, for each basis function = 7 to 11
- Maximum number of misclassified samples, which when exceeded, a new basis function is to be added to the network = 10
-

3. RESULTS AND DISCUSSION

We have implemented object based Multi Layer Feed Forward, Radial Basis Functions and Cloud Basis Function Neural Network in order to compare the accuracies with different activation functions. All these algorithms are implemented using C/C++ on Windows platform with Pentium 4 processor machine. The methodology is tested on a QuickBird window (2000 x 2000 pixels) of an urban fringe area comprising a few buildings, a quarry site, ponds, road, vegetation and foot paths. This image was retrieved on August 2001. On ground, it covers the Powai Area of Mumbai City. The study area is located between latitude (19 07' 14.69"N - 19 06' 39.98" N) and longitude (72 53' 43.07"E - 72 54' 29.08" E) as shown in Fig.2.

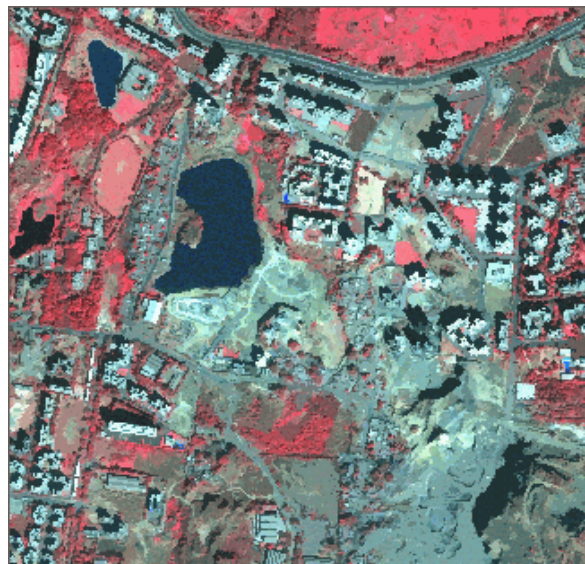


FIGURE 2: High resolution satellite image used as Study Area

The image was classified into 9 prominent classes covering a majority of the land cover features, Lake, Pool, Vegetation, Field, Road, Shadow, Bright Roof, and Dark Roof and Mountain. Accuracy and error statistics were computed for each activation function. Fig. 3 and Fig 4 depict output of object based classification using MLFF and RBF NN respectively which clearly indicates that object based classification is not a universal remedy, it is evident that regions are misclassified. For example, the roads and buildings or the grass and tree are spectrally similar and have a significant amount of spectral overlap. This is the primary reason for the large number of misclassifications between these classes. Similarly a part of the lake is being classified as a pool; an entire lake is classified as shadow, etc. This happens due to spectral

closeness of these regions. In order to reduce misclassification we need to take into consideration ancillary data (contextual data) available about the image.

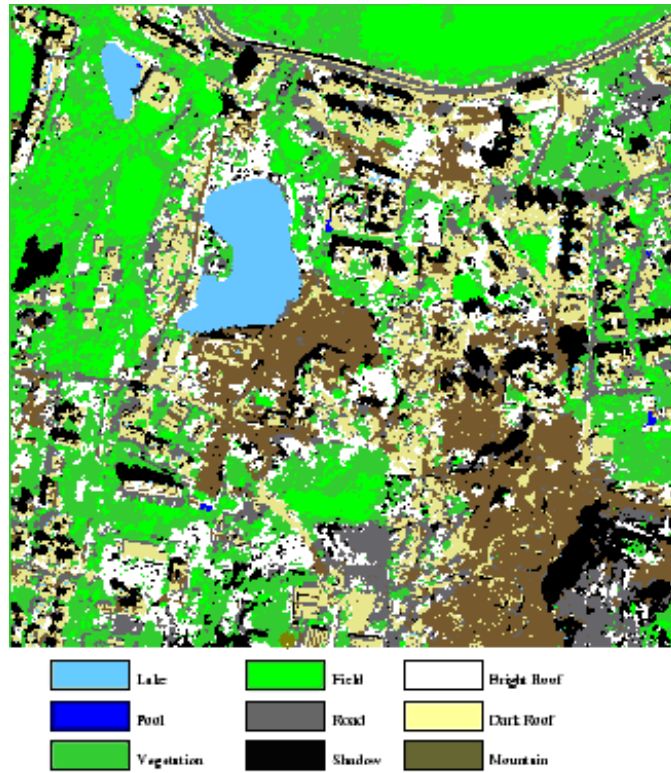


FIGURE 3: Classified image using MLFF

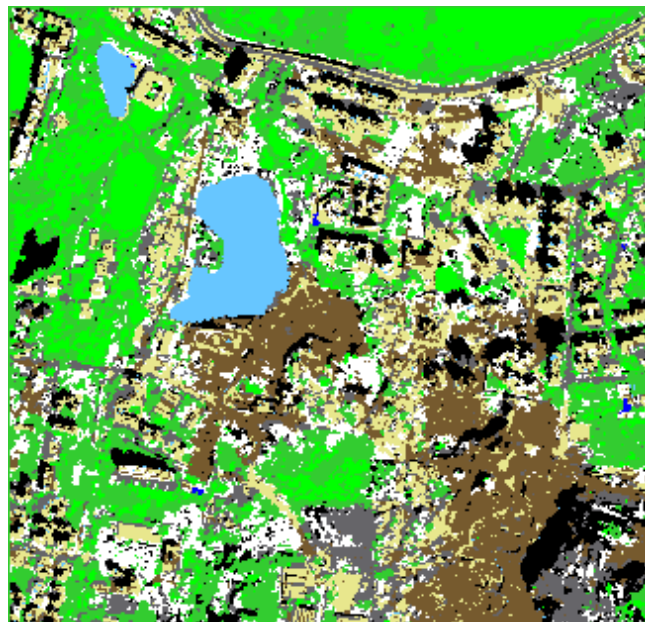


FIGURE 4: Classified image using RBF

For example, if a region is classified as a shadow then there has to be tall structure in the vicinity of the shadow. Hence some improvement can be observed in Fig 5 which is classified image by CBF. There are plenty of geometrical information such as object feature, shape feature, texture, and contextual relation feature and so on. In this paper we add other feature information into feature space, which is area, entropy, shape index and contextual relation feature. Accuracy assessment is a general term for comparing predicted (i.e., classification) results to geographical reference data that are assumed to be true. This comparison is typically achieved by a basic subjective assessment of the observed difference in accuracy but should be undertaken in a statistically rigorous fashion. A set of reference objects representing geographic points on the classified image is required for the accuracy assessment. Randomly selected reference objects lessen or eliminate the possibility of bias.

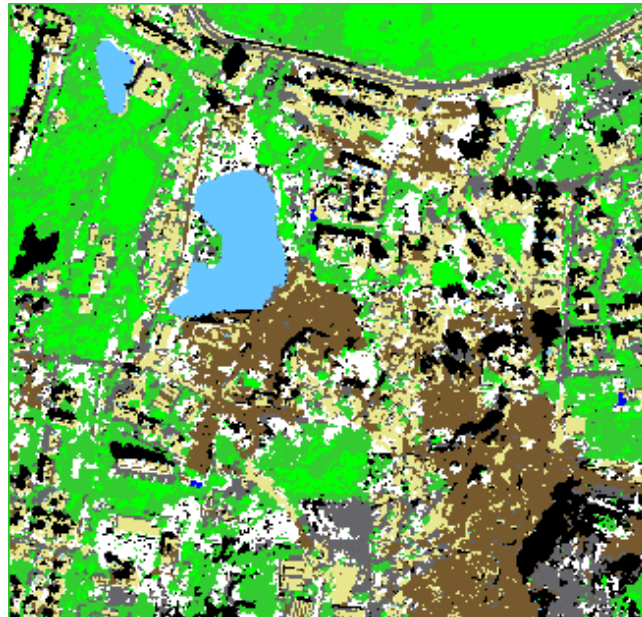


FIGURE 5: Classified image using CBF

A random stratified sampling method was used to prepare the ground reference data. This sampling method allocates the sample size for each land use based on its spatial extent. A summary of accuracy and error statistics of all mentioned kernels can be found in Table 1.

Classes	ANN (MLFF)		ANN (RBF)		ANN (CBF)	
	CA	PA	CA	PA	CA	PA
Lake	0.8434	0.9357	0.8974	0.9127	0.8105	0.8045
Pool	0.9820	0.9658	1.0000	1.0000	1.0000	0.8750
Vegetation	0.8456	0.8834	0.8473	0.8494	0.9000	0.8750
Field	0.9248	0.8452	0.9308	0.8242	0.8954	1.0000
Road	0.5414	0.6143	0.5064	0.6873	0.9000	1.0000
Shadow	0.9632	0.9658	1.0000	0.9711	0.7988	0.8873
Bright Roof	0.9873	0.9921	1.0000	1.0000	0.9059	0.7660
Dart Roof	0.5360	0.6898	0.5760	0.6388	0.8514	0.8000
Mountains	0.7255	0.6585	0.5395	0.6445	0.9808	0.8750
Accuracy	0.8272		0.8542		0.8962	
Kappa Coefficient	0.8319		0.8491		0.8735	

CA = Consumer's Accuracy

PA = Producer's Accuracy

Table 1. Accuracy and error statistics of object based supervised image classifier

The object based classification using CBF outruns the other kernel based NN classifiers in overall accuracy. The kappa coefficient which is 0.8319 and 0.8491 are low indicating the MLFF and RBF method are still an unsatisfactory one to classify remotely sensed images, where as for CBF it is recorded as 0.8735.

4. CONCLUSION

This paper attempts to study and compare the accuracy of object based image classifiers. The object based image analysis greatly reduced the salt-and-pepper classification effect in the classified image without adversely affecting the classified image accuracy. This greatly improves the visual effect of the classified image.

ANN has the advantages mainly of more tolerance to noise inputs and representation of boolean function apart from others, but too many attributes may result in over fitting. It was found that the neural network classifier trained using the standard back-propagation algorithm produced marginally better results compared to the other methods. The study shows that CBF NN improves the classification accuracy, though the CBF, being a relatively new technique in the remote sensing arena requires further study. A combined approach to classification using object based methods and contextual information available about the image, seems promising and needs further exploration.

5. REFERENCES

- [1] R.O. Duda, P.E. Hart P.E. and D.G. Stork, "*Pattern Classification*". 2nd edition, John Wiley & Sons (2001).
- [2] B. Tso and P.M. Mather, "*Classification methods for remotely sensed data*" London: Taylor and Francis (2001).
- [3] C.R. De Silva, S. Ranganath and L.C. De Silva, "*Cloud Basis Function Neural Network: A modified RBF network architecture for holistic for holistic Facial Expression Recognition*". Elsevier Pattern Recognition 41:1241-1253, 2008.
- [4] J. Im, J.R. Jensen and J.A.Tullis, "*Object-based change detection using correlation image analysis and image segmentation*". International Journal of Remote Sensing, 29(2), 399–423 (2008).
- [5] G.J Hay, G. Castilla, M. Wulder, and J.R Ruiz, "*An automated object-based approach for the multiscale image segmentation of forest scenes*" International Journal of Applied Earth Observation and Geoinformation 7, 339–359, 2005.
- [6] T. Blaschke, "*A framework for change detection based on image objects*", In Go'ttinger Geographische Abhandlungen, S. Erasmi, B. Cyffka and M. Kappas (Eds), 113, 1–9, Go'ttingen, 2005. Available online at http://www.definieris.com/pdf/publications/GGRS2004_Blaschke_G001.pdf (accessed 18 June 2009).
- [7] A.S. Laliberte, A. Rango, K.M. Havstad, J.F. Paris, R.F. Beck and Y.R. McNeely, "*Object-oriented image analysis for mapping shrub encroachment from 1937 to 2003 in southern New Mexico*". Remote Sensing of Environment, 93, 198–210, 2004.
- [8] R.C. Frohn, K.M. Hinkel, and W.R. Eisner, "*Satellite remote sensing classification of thaw lakes and drained thaw lake basins on the North Slope of Alaska*". Remote Sensing of Environment, 97, 116–126, 2005.
- [9] J. Eo and H. Kim, "*A Detail Extraction Technique For Image Coding Using morphological Laplacian Operator*". In Proceedings of the IEEE TENCON, pp.140-147, Korea,1997.
- [10] F. Meyer and S. Beucher, "*Morphological Segmentation*". Journal of Visual Communication and Image Representation, 11: 21–46, 1990.
- [11] L. Costa and R. Cesar, "*Shape Classification and Analysis*", 2nd edition CRC Press, 259- 310, (2009).
- [12] P. M. Mather, "*Computer Processing of Remotely Sensed Images*," John Wiley, NY, (2004).

- [13] J. Han, S. Lee, K. Chi, K. Ryu, "*Comparison of Neuro-Fuzzy, Neural Network, and Maximum Likelihood Classifiers for Land Cover Classification using IKONOS Multispectral Data*". In Proceedings of the IEEE International Geoscience and Remote Sensing Symposium, vol 6, pp. 3471-3473, Toronto, Canada, 2002.
- [14] S. Haykin, "*Neural Networks: A Comprehensive Foundation*" 2nd edition, Prentice Hall pp. 156- 256, (1999)
- [15] T. Cover, "*Geometrical and Statistical Properties of Systems of Linear Inequalities with Applications in Pattern Recognition*" IEEE Transactions on Electronic Computers, vol. EC-14(3),326-334, June 1965. Reprinted in Artificial Neural Networks: Concepts and Theory, IEEE Computer Society Press, Los Alamitos, Calif., 1992, eds. P. Mehra and B. Wah.

Interferogram Filtering Using Gaussians Scale Mixtures In Steerable Wavelet Domain

Gh.S.El-Taweel

Faculty of Computers and Informatics / C.S. Dep.
Suez Canal University
Ismailia, Egypt

g.sami@scuegypt.edu.eg

Ashraf. K. Helmy

National Authority for Remote Sensing
and Space Sciences,
Cairo, Egypt

akhelmy@narss.sci.eg

Abstract

An interferogram filtering is presented in this paper. The main concern of the proposed scheme is to lower the residues count mean while preserving the location and jump height of the lines of phase discontinuity.

The proposed method uses Steerable wavelet decomposition. At each scale, a noise covariance matrix is estimated for the neighborhood of each pixel using an interferogram power spectral density. The estimated covariance is then used to produce a maximum a-posteriori estimate of the noise-free value of the pixel. After the image has been "de-noised" at each scale-level, it is then reconstructed.

Based on Portilla et. al., neighborhoods of coefficients at adjacent positions and scales are modeled as the product of two independent random variables, the Bayesian least squares estimate of each coefficient reduces to a weighted average of the local linear estimates over all possible values of the hidden multiplier variable. The performance of this method substantially has the advantages of reducing number of residuals without affecting line of height discontinuity.

Keywords: Radar Imaging, Phase Unwrapping, Interferogram, Gaussian scale mixtures

1. INTRODUCTION

Synthetic aperture radar (SAR) enables imaging of the ground by processing microwave backscattering data collected along the flight path of an airborne or space borne platform. This results in high-resolution images of the local complex reflectivity of the ground. For the SAR images obtained from slightly different flight paths, the complex-valued pixels of one image are multiplied with the co-registered complex conjugate pixels of the other, the phase of the resulting product image constitutes a SAR interferogram. The significance of this phase image is that it contains information on terrain height. With proper processing, it yields a so-called digital elevation model (DEM) which represents the topography of the terrain. Other significant of using phase image is to measure land subsidence, snow motion, monitoring sand dunes movements, and many others.

The SAR interferogram presents two main features preventing its direct use. The first one is that the phase is wrapped within the interval $(-\pi, \pi)$, due to the periodic nature of the phase signal. In order to obtain the absolute phase, and therefore, to obtain the height information, it

is necessary to unwrap the phase signal [1]. From the other side, the interferometric phase signal is corrupted by noise. This noise, on a first stage, will affect the unwrapping process as it induces phase residues, which make the unwrapping process difficult. On a second stage, the own noise degrades the derived height information. In order to overcome these problems it is necessary to introduce a phase filtering process before phase unwrapping. This filtering process can be also thought as an estimation process.

The key computational problem in obtaining the DEM lies in the fact that the measured phase differences are undetermined within multiples of 2π . They are given as a wrapped phase field of principal values with range from $-\pi$ to π . It is necessary to perform a 2-D phase unwrapping operation, which removes phase jumps between neighboring pixels larger than π by adding or subtracting multiples of 2π so that the resulting distribution can be considered as samples of an underlying smooth function. This is a noise-sensitive problem and subject of recent research [2]. The 2-D phase unwrapping problem can be stated as follows:

$$\psi(n_1, n_2) = \text{Arg}\{e^{i[\phi(n_1 n_2 + \text{noise})]}\} \quad (1)$$

where ψ is the noisy wrapped interferogram, ϕ are samples of a smooth phase surface, and $|\text{Arg}(\cdot)| \leq \pi$ is determination of an estimate ϕ from ψ . A major challenge of 2-D phase unwrapping comes from the fact that the noise in (1) due to, temporal de-correlation introduces local inconsistencies of the data referred to as residues. Any phase unwrapping algorithm is sensitive to the presence of residues that make contour integrations path-dependent.

Many interferogram filtering are introduced. Lee et. al. demonstrated that the interferometric phase noise can be modeled as an additive noise [3]. Other algorithm is based on a new interferometric phase noise model in the complex plane [2]. Some estimate the InSAR phase within a local estimation window based on the InSAR sample statistics [4, 5]. In general it can be stated that filtering process is mandatory as it not only reduces the error probability in subsequent phase unwrapping but also increases processing speed considerably [3].

This paper is a modification version of Portilla et. al. where local residual removal solution based on Bayesian least square estimator is used. A model for neighborhoods of oriented pyramid coefficients based on Gaussian scale mixture is used and noise covariance matrix is estimated for the neighborhood of each pixel using an interferogram power spectral density via different sets of window shapes. We assume based on [6, 7] that the amplitude of coefficients of similar position, orientation and scaled are highly correlated. These higher order dependencies may be modeled by augmenting a factor for local dependences (Gaussian) with random variables that govern the parameters (variance). The remainder of the paper is organized as follows: section 2 focus on the interferometric residual model, section 3 describe a brief introduction to steerable pyramid wavelet transform, section 4 depict the interferogram probability model, section 5 emphasis the proposed filtering process, section 5 shows the experiment result, and section 6 represents the conclusion.

2. INTERFEROMETRIC RESIDUAL MODEL

The interferometric phase is due to the interaction between two SAR images. The statistical behavior of the interferometric phase depends on this interaction. The interferometric coherence is the amplitude of the correlation coefficient between the two complex SAR images:

$$\rho = \frac{E\{I_1 I_2^*\}}{\sqrt{\{ |I_1|^2 \} E\{ |I_2|^2 \}}} |\rho| e^{j\phi} \quad (2)$$

Where I_1 and I_2 represent the two SAR images, ρ is the coherence and ϕ_x is the corresponding interferometric phase. The interferometric phase has been completely

characterized in the real domain. For Gaussian scattering model and distributed scatterers, the interferometric phase follows the following distribution [1].

$$P_{\phi}(\phi) = \frac{\Gamma(n + \frac{1}{2})(1 - |\rho|^2)^n \beta}{2\sqrt{\pi}\Gamma(n)(1 - \beta^2)^{n+1/2}} + \frac{(1 - |\rho|^2)^n}{2\pi} {}_2F_1(n, 1; 1/2; \beta^2) \quad (3)$$

Where $\beta = |\rho| - \cos(\phi - \theta)$, ${}_2F_1$ represents the Gauss hyper-geometric functions and n is the number of looks. The probability density function (pdf) eq. (3) is symmetric (mod 2π) about its mode, which occurs at ϕ . Based on the distribution of the interferometric phase eq. (2), [3] demonstrated that the interferometric phase noise can be modeled as an additive Gaussian noise [8, 9]

$$\phi_z = \phi_x + v \quad (4)$$

Where ϕ_z is the measured interferometric phase, ϕ_x is the real interferometric phase and v represents the noise with the distribution listed in eq. (3). The measured phase ϕ_z can be encoded in the complex plane as a point in the unit circle:

$$e^{j\phi_z} = \cos(\phi_z) + j \sin(\phi_z) \quad (5)$$

Using the distribution (3) together with the properties of the trigonometric functions, the real and imaginary parts of (5) can be modeled as [9].

$$\cos(\phi_z) = N_c \cos(\phi_x) + v_c \quad (6)$$

$$\sin(\phi_z) = N_c \sin(\phi_x) + v_s \quad (7)$$

Based on (2), it can be demonstrated that the value of N_c , for $n = 1$, is:

$$N_c = \frac{\pi}{4} |\rho| {}_2F_1(1/2, 1/2; 2; |\rho|^2) \quad (8)$$

As equations (6), (7), and (8) show, N_c only depends on ρ . The terms v_c and v_s can be considered as noise terms, as their means are zero. These noise terms depend on the interferometric phase ϕ_x . This dependence does not affect the mean, which is zero, but only the standard deviation [9]. The dependence with the interferometric phase ϕ_x can be neglected, and then, the variance of the noise terms can be approximated by the function:

$$\sigma_{v_c}^2 = \sigma_{v_s}^2 = \frac{1}{2} (1 - |\rho|^{0.68}) \quad (9)$$

Eq's (6) and (7) can be seen respectively as a noise model for the real and imaginary parts of the interferometric phase in the complex plane. In each case, the signals to recover are $\cos(\phi_x)$ and $\sin(\phi_x)$. These signals are multiplied by N_c . As shown before, this parameter behaves in the same way as the coherence ρ does, so instead assuming it as a noise parameter, it can be considered as a useful parameter to recover. It was demonstrated that N_c can be employed to estimate the coherence ρ [9]. The main feature of this way to estimate the coherence is that the coherence information can be estimated with high spatial resolution.

3. STEERABLE PYRAMID

The Steerable Pyramid is a linear multi-scale, multi-orientation image decomposition that provides a useful front-end for image-processing and computer vision applications. It was developed in order to overcome the limitations of orthogonal separable wavelet decompositions that were becoming popular for image processing (specifically, those representations are heavily aliased, and do not represent oblique orientations well). Once the orthogonality constraint is dropped, it makes sense to completely reconsider the filter design

problem (as opposed to just re-using orthogonal wavelet filters in a redundant representation, as is done in cycle-spinning or undecimated wavelet transforms!). Detailed information may be found in the references listed below.

The basis functions of the steerable pyramid are directional derivative operators that come in different sizes and orientations. An example decomposition of an image of a white disk on a black background is shown to the right. This particular steerable pyramid contains 4 orientation sub-bands, at 2 scales. The number of orientations may be adjusted by changing the derivative order (for example, first derivatives yield two orientations). The smallest sub-band is the residual low-pass information. The residual high-pass sub-band is not shown.

Figure (1) shows the block diagram for the decomposition (both analysis and synthesis). Initially, the image is separated into low and high-pass sub-bands, using filters L_0 and H_0 . The low-pass sub-band is then divided into a set of oriented band-pass sub-bands and a low pass sub-band. This lower pass sub-band is sub-sampled by a factor of 2 in the X and Y directions. The recursive (pyramid) construction of a pyramid is achieved by inserting a copy of the shaded portion of the diagram at the location of the solid circle (i.e., the low-pass branch) [10].

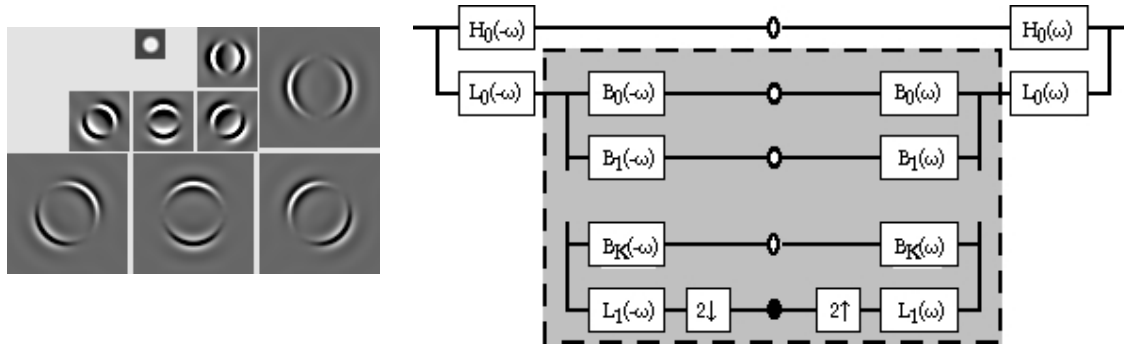


Figure: (1) Block diagram for Steerable Pyramid decomposition (analysis and synthesis).

4. INTERFEROGRAM PROBABILITY MODEL

Multiscale representations provide a useful front-end for representing the structures of 2-D signals. But they are critically sampled (the number of coefficients is equal to the number of image pixels), and this constraint leads to disturbing visual artifacts (i.e., “aliasing” or “ringing”). A widely followed solution to this problem is to use basis functions designed for orthogonal or biorthogonal systems, to reduce or eliminate the decimation of the subbands [11]. Once the constraint of critical sampling has been dropped, there is no need to limit oneself to these basis functions. Significant improvement comes from the use of representations with a higher degree of redundancy, as well as increased selectivity in orientation [12]. For the current paper, we have used a particular variant of an over complete tight frame representation known as a steerable pyramid [13]. The basis functions of this multiscale linear decomposition are spatially localized, oriented, and span roughly one octave in bandwidth. They are polar-separable in the Fourier domain, and are related by translation, dilation, and rotation.

4.1 Gaussian Scale Mixtures

Let an interferogram is decomposed into oriented subbands at multiple scales, and Let $x_{cs,o}(n,m)$ the coefficient corresponding to a linear basis function at scale (s) , orientation (o) , centered at spatial location $2sn, 2sm$. We denote as $x_{s,o}(n,m)$ a neighborhood of coefficients clustered around this reference coefficient . in our case the neighborhood include coefficients from three subbands at adjacent scale (i.e., corresponding to basis functions at nearby scales and orientations), as well as from the same subband. Thus taking advantage of the strong statistical coupling observed through scale in multiscale representations.

Lee et. al. proposed that the residuals are characterized with an additive model [3]. We assume the coefficients within each local neighborhood around a reference coefficient of a pyramid subband are characterized by a Gaussian scale mixture (GSM) model. Formally, a random vector is a Gaussian scale mixture if and only if it can be expressed as the product of a zero-mean Gaussian vector (u) and an independent positive scalar random variable (\sqrt{z})

$$X \equiv u * \sqrt{z} \quad (10)$$

the variable z is known as the multiplier. The vector x is thus an infinite mixture of Gaussian vectors, whose density is determined by the covariance matrix C_u of vector u and the mixing density $p_z(z)$, and (\equiv) mean equal in distribution.

4. 2 Modeling Wavelet Coefficients

In order to construct model for interferogram, one must specify both the neighborhood structure of the coefficients, and the distribution of the multipliers. Regardless to the neighborhood structure of the coefficients, the window selected should be aligned along fringes to include pixels having approximately the same elevation. For steep slope areas, the fringe rate is high. Square windows, will cover more than one fringe. This will destroy the continuity of fringes and make phase unwrapping difficult. Consequently, directional windows should be used, such as the ones shown in Fig. 2. This makes filtering more effective in preserving the fringe pattern [3].

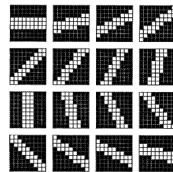


Figure 2: Sixteen directional masks for phase noise filtering. Only the white pixels are included in the computation.

For neighborhood structure GSM is used as a local description of the behavior of the cluster of coefficients centered at each coefficient in the pyramid. Since the neighborhoods overlap, each coefficient will be a member of many neighborhoods. The local model implicitly defines a Markov model, described by the conditional density of a coefficient in the cluster given its surrounding neighborhood, assuming conditional independence on the rest of the coefficients. In this paper, we simply use the estimation problem for the reference coefficient at the center of each neighborhood independently. To complete the interferogram model we need to specify the probability density, $p_z(z)$, and interferogram power spectral density (PSD).

4.3 Probability Density for Multiplier

Several authors have suggested the generalized Gaussian (stretched exponential) family of densities as an appropriate description of wavelet coefficient marginal densities [14, 15]: , where (α) the scaling variable controls the width of the distribution, and the exponent (p) controls the shape, and is typically estimated to lie in the range [0.5,0.8] for image subbands. Although these can be expressed as GSM's, the density of the associated multiplier has no closed form expression, and thus this solution is difficult to implement [16]. Others stated that the density of the log coefficient magnitude, $\log x$, may be expressed as a convolution of the densities of ($\log u$) and ($\log z$). Since the density of (u) is known, this means that estimation of the density of may be framed as a deconvolution problem. The resulting estimated density may be approximated by a Gaussian, corresponding to a lognormal prior for the z . This solution has two drawbacks. First, it is only valid to the case when all the neighbors have the same marginal statistics, which, in practice requires they all belong to the same subband. Second, it is estimated from the noise-free coefficients, and it is difficult to extend it for use in the noisy case.

In this paper we use what is a so-called non-informative prior [16-17], which has the advantage that it does not require the fitting of any parameters to the noisy observation. Such solutions have been used in establishing marginal priors for image denoising [18]. In the context of estimating the multiplier from coefficients, this takes the form:

$$p_z(z) \propto \sqrt{I(z)} \quad I(z) = E\left\{-\frac{\partial^2 \log p(x|z)}{\partial z^2}\right\} \quad (11)$$

Where $I(z)$ is the Fisher information matrix. Computing this for the GSM model is straightforward

$$\begin{aligned} -\frac{\partial^2 \log p(x|z)}{\partial z^2} &= \frac{\partial^2}{\partial z^2} \left[\frac{1}{2} \left(N \log(z) + \log C_u + \frac{x^T C_u^{-1} x}{z} \right) \right] \\ &= \frac{N}{2z^2} + \frac{x^T C_u^{-1} x}{2z^3} \end{aligned} \quad (12)$$

taking the square root of the expectation, and using the fact that $E\{x^T C_u^{-1} x\} = z$ we obtain Jeffrey's prior which corresponds to a constant prior on $\log(z)$.

$$p_z(z) \propto \frac{1}{z} \quad (13)$$

4.4 Interferogram PSD

The analysis presented in this section is based on the basic SAR interferometric system mode [19]. As shown in fig. 3, input to the system are the complex reflectivity functions $x_i(t)$, $i = 1; 2$. The impulse response functions are denoted with $h_{SAR_i}(t)$, $i = 1; 2$. Each of these point target responses describes the individual end-to-end SAR imaging system. Finally, the interferogram $\Phi_z(t; \tau)$ is formed from the focused SAR images $i_1(t; \tau)$ and $i_2(t; \tau)$. Here we introduce the variable $t[s]$ to denote a time dimension (range or azimuth), and the extension to the two-dimensional domain is straightforward.

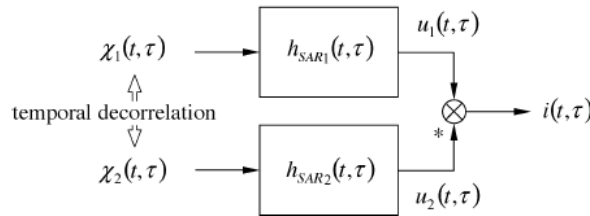


Figure 3: Interferometric system Model

The distributed scattered model constitutes the input to the interferometric system. The inputs $x_1(t)$ and $x_2(t)$ are correlated, unit variance complex, zero-mean Gaussian processes constructed from $a(t)$, $b(t)$, and $c(t)$ which are of uncorrelated Gaussian type

$$x_1(t) = \sqrt{1 - \gamma T} a(t) + \sqrt{\gamma T} c(t) \quad (14)$$

$$x_2(t) = (\sqrt{1 - \gamma T} b(t) + \sqrt{\gamma T} c(t)) \exp(-j2\pi\vartheta t)$$

a mutual correlation is realized by the factor γT (temporal coherence), and ϑ

$$S_{ii} = \frac{S_{u2u2}(f) \otimes S_{u1u1}(f)}{\int S_{u1u1}(f) df \int S_{u2u2}(f) df} |\gamma\vartheta|^2 \delta(f - \vartheta) \quad (15)$$

denotes a fringe frequency attached to the second interferometric channel for ease of formulation. All parameter cases discussed in the following section are based on

subsequently derived formulation of the interferogram PSD function. For transfer functions with finite bandwidth (which is true for any real SAR system), the interferometric PSD is easily calculated as the Fourier transform of the interferogram sample autocorrelation function (ACF) [20].

Equation (15) represents the normalized interferogram PSD so as to emphasize that the square of the coherence determines the line mass of its right-hand-side second term (\otimes denotes cross correlation).

$S_{I1 I1}(f)$ and $S_{I2 I2}(f)$ denote the PSDs of the SAR images, and $\delta(\cdot)$ denotes the Dirac Delta Distribution. As can be shown, the fringe frequency is found as frequency shift in the argument of the Delta function. From the definition of the interferogram ACF (Holzner 2002), it is clear that no phase bias can be measured neither from the interferogram ACF nor from the interferogram PSD. As described by [10], the individual terms of the PSD are defined as:

$$S(f) = P_S(f) + m \delta(f) \tag{16}$$

The first term of the PSD is referred to as pedestal $P_S(f)$ with an integral value of 1 (due to the normalization of the power spectrum), and m refers to the spectral line mass.

5. INTERFEROGRAM FILTERING

Recognizing the residuals sensitivity of 2-D phase unwrapping algorithms, it would be useful to reduce the residuals from the measured interferogram before processing them further. The main concern in the proposed scheme is not to affect the fringes (line of discontinuity) with respect to their location, sharpness, and jumps of height. Other requirements are low computational complexity, robustness, and ease of use.

Our filter procedures use the top-level structure as most previously published approaches and can be summarize as follows:

- 1) Decompose the interferogram into pyramid subbands at different scales and orientations;
- 2) Apply the filter at each subband, except for the low pass band.
- 3) Invert the pyramid transform, obtaining the denoised image.

The assumption that the residuals are exhibits a white Gaussian with known variance [3] is used in this paper. A vector y corresponding to a neighborhood of N observed coefficients of the pyramid representation can be expressed as

$$y = x + w = \sqrt{z}u + w \tag{17}$$

Based on GSM structure of N Coefficients, and the assumption that the residuals act as independent additive Gaussian. The parameters u , w , and z are independent. both u and w are zero-mean Gaussian vector [19], with associative covariance C_u and C_w . the density of the observed neighborhood vector conditioned on z is a zero-mean Gaussian, with covariance $C_{y|z} = zC_u + C_w$. The neighborhood noise covariance, C_w , is obtained by decomposing a function $\sigma \sqrt{N_y N_x} PSD_r(i)$ into pyramid subbands, where N_x, N_y are the interferogram dimensions, and $PSD_r(i)$ is the inverse Fourier transform of the square root of the interferogram power spectral density. This signal has the same power spectrum as the residual, but it is free from random fluctuations. Elements of C_w may then be computed directly as sample covariance (i.e., by averaging the products of pairs of coefficients over all the neighborhoods of the subband). Given C_w , the signal covariance C_u can be computed from the observation covariance matrix. C_y is computed from $C_{y|z}$ by taking expectations over z : taking an assumption that $E\{z\} = 1$.

$$C_y = E\{z\}C_u + C_w \tag{18}$$

5.1 Bayes Least Square Estimation

For each neighborhood, we wish to estimate x_c , the reference coefficient at the center of the neighborhood, from y , the set of observed coefficients. The Bayes least squares (BLS) estimate is just the conditional mean.

$$\begin{aligned}
 E\{x_c|y\} &= \int x_c p(x_c|y) dx_c & (19) \\
 &= \int \int_0^\infty x_c p(x_c, z|y) dz dx_c \\
 &= \int \int_0^\infty x_c p(x_c, y, z) p(z, y) dz dx_c \\
 &= \int p(z, y) E\{x_c|y, z\} dz
 \end{aligned}$$

Thus, the solution is the average of the Bayes least squares estimate of x when conditioned on z , weighted by the posterior density, $p(z|y)$. We now describe each of these individual components.

According to GSM model, the neighborhood coefficients of vector x is Gaussian when conditioned on z . Coupled with the assumption that the residuals are characterized as Gaussian noise. This means that the expected value inside the integral 8 is local linear (Wiener) estimate. Writing this for the full neighborhood vector

$$E\{x|y, z\} = zC_u(zC_u + C_w)^{-1}y \tag{20}$$

Portellia et. al, simplified this expression [10] result in:

$$E\{x_c|y, z\} = \sum_1^N \frac{zm_{cn}\lambda_n v_n}{z\lambda_n + 1} \tag{21}$$

where m_{ij} represents an element (i -th row, j -th column) of the matrix M , s is the symmetric square root of the positive definite matrix C_w and A is the eigenvalue expansion of the matrix $S^{-1}C_uS^{-T}$, λ_n are the diagonal elements of A , v_n is the elements of $v = M^{-1}y$, and c is the index of the reference coefficient within the neighborhood vector, which considered a first part of eq. (8). The other component of the solution given in (8) is the distribution of the multiplier, conditioned on the observed neighborhood values. We use Bayes' rule to compute this

$$p(z|y) = \frac{p(yz)p_z(z)}{\int_0^\infty p(y\alpha)p_z(\alpha)d\alpha} \tag{22}$$

we choose a non-informative Jeffrey's prior, corrected at the origin, for the function $p_z(z)$. The conditional density $p(y|z)$ is given in eq. (6), and its computation may be simplified using the relationship in eq. (10) and the definition of v

$$p(y|z) = \frac{\exp(-\frac{1}{2} \sum_{n=1}^N \frac{v_n^2}{z\lambda_n + 1})}{\sqrt{(2\pi)^N |C_w| \prod_{n=1}^N (z\lambda_n + 1)}} \tag{23}$$

5.2 Filtering Process

According to the proposed scheme and [16], Filtering interferogram process can be summarized as follows:

- 1) Decompose the interferogram into subbands.
- 2) For each subband (except the low-pass):
 - a) Compute neighborhood noise covariance from the image-domain noise covariance.
 - b) Estimate noisy neighborhood covariance.
 - c) Estimate C_u from C_w and C_y using eq. (18).
 - d) For each neighborhood:
 - i) For each value z in the integration range:
 - A) Compute $E\{x_c y, z\}$ using eq. (21).
 - B) Compute $p(y|z)$ using eq. (23).
 - ii) Compute $p(z|y)$ using eqs. (22) and (13).
 - iii) Compute $E\{x_c y\}$ numerically using eq. (19).
- 3) Reconstruct the denoised image from the processed subbands and the lowpass band.

6. EXPERIMENTAL RESULTS

The proposed scheme for interferogram filtering is applied to real interferometric SAR (InSAR) image. The size of the image in pixels is 512 (azimuth) \times 512 (range). The estimated surfaces are compared with the reference digital Elevation Model (DEM). Figure 4-a) shows the phase image of original interferogram while figure 4-b) shows the reference DEM. In the available literature on interferogram filtering, the method of choice is often judged by visually comparing original and processed interferogram. In addition to visual judgment, quantitative evaluations are used during this work includes:

- a) Peak Signal to Noise Ratio (PSNR) between filtered and original phase.
- b) Reduction of the residuals counts.
- c) Quantitative effect on phase unwrapping.

In order to evaluate the effect of complex interferogram filtering on 2-D phase unwrapping quantitatively, we consider determination of the phase field estimation by means of Goldstein algorithm [1]. The unwrapped phase values are compared with the reference data shown in figure 4(b). Three other filtering algorithms are used for evaluation, they are: pivoting median filter [1]. With 3X3 window size, 2) max-flat filter defined in [1] and modified Goldstein filter [21].

In general, it can be stated that the Peak Signal to Noise Ratio (PSNR) for proposed scheme reaches 45 db. Moreover the number of residuals after filtering process decreased by 78.6%. Table 1 summarizes the PSNR and residual counts for different techniques

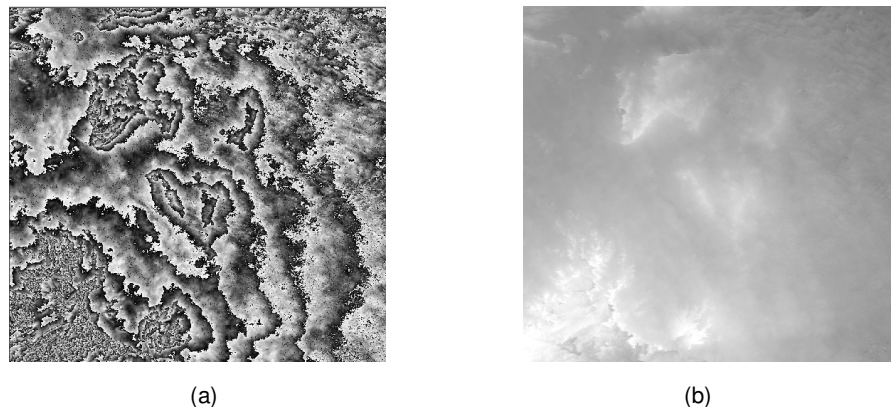


Figure 4: Demonstrate the Input Phase Images a) Original Unwrapped Phase. b) Reference Wrapped Phase.

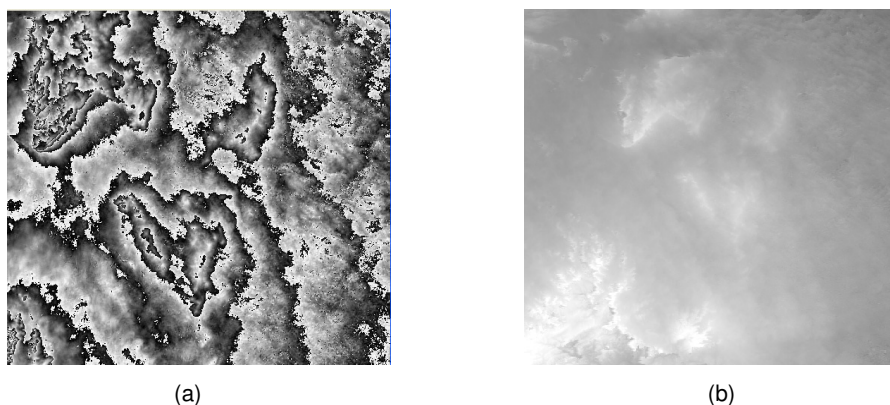


Figure 5: Demonstrate the output Images a) filtered Unwrapped Phase. b) Wrapped Phase.

Another evaluation parameter applied to the complex interferogram filtering is subsequent phase unwrapping including height inversion (DEM). This is done by taking the interferogram with and without preprocessing as input to Goldstein algorithm [1]. Table 2 lists the average errors between wrapped phase with and without preprocessing and the original height values. It is verified that the proposed scheme introduces the best average error, affecting directly the accuracy of terrain height.

	Original Interferogram	Proposed	Max-flat	Pivoting Median	Modified Goldstein
PSNR (dB)		45	11.6	15.7	38
Number of Residues	13785	2956	12306	1843	3456

Table 1: Performance Evaluation Measures

	DEM without filtering	DEM after proposed filtering	DEM after Max-flat	DEM after Pivoting Median	Modified Goldstein
Average error(Meter)	63.1	12.3	42.3	97.6	18

Table 2: Average Error of Height Inversion of Input Interferogram using Different Filters

Qualitative measures of the output phase data, as shown in figure (3), indicate the superiority of the proposed scheme. It demonstrates good adaptively of proposed scheme as it reduces the residues count while preserving edges and fine details. The pivoting median filter gives good results of residues reduction but introduces smoothness at level of phase details. Max-flat filter is better than pivoting median filter. The output wrapped phase data is shown in figures (6), while figures (7,8) shows the 3d and close look images for better evaluation process.

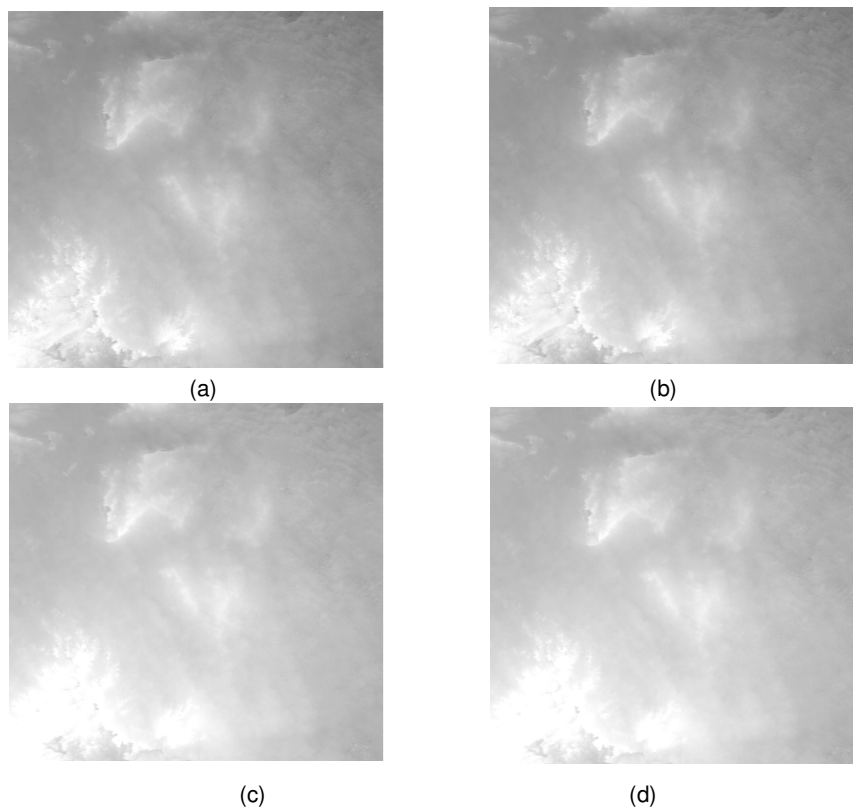


Figure 6: Output Unwrapped Phase Data after Different Filtering Process
a) proposed b) Max-Flat c) Pivoting Median d) Modified Goldstein

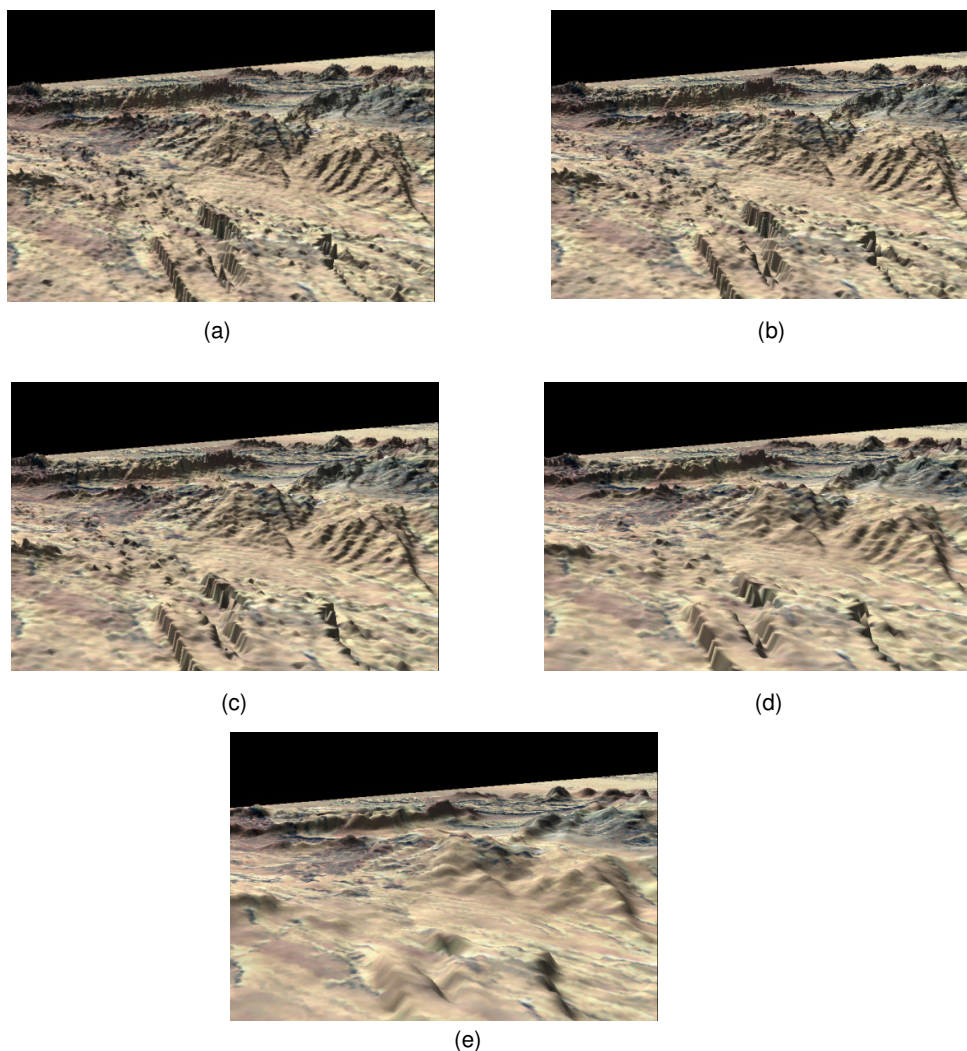


Figure 7: Perspective Phase Unwrapped Images a) Reference DEM b) proposed filter c) modified Goldstein d) Max-Flat e) Median filter

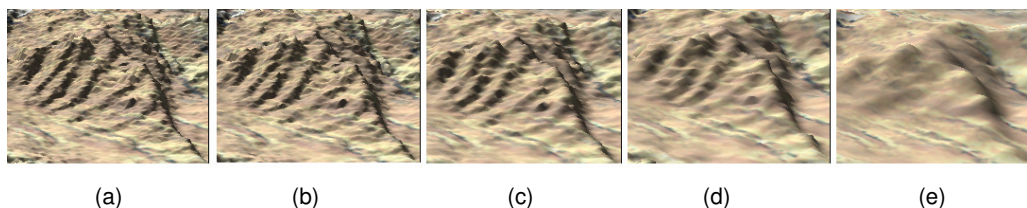


Figure 8: Closer Look of Perspective Phase Unwrapped Images a) Reference DEM b) proposed filter c) modified Goldstein d) Max-Flat e) Median filter

CONCLUSIONS

To obtain a more accurate unwrapped phase, a new scheme has been proposed based on a local Gaussian scale mixture model for reduction residual counts in interferogram. The approach uses Steerable wavelet decomposition. At each scale, a noise covariance matrix is estimated for the neighborhood of each pixel using interferogram power spectral density. The covariance is then used to produce a maximum a-posteriori estimate of the noise-free value of the pixel. After the image has been de-noised at each scale-level, it is then reconstructed.

Comparing with other interferogram filtering, the proposed scheme achieves 78.6% in reduction of residual count. On the other hand Max-Flat, Pivoting Median, and modified Goldstein has also the ability to reduce the residues count by 10.7%, 86.7 % and by 74.9 respectively.

Comparing the unwrapped phase with reference DEM, median filter achieves higher average error followed by Max-flat, modified goldstien and the proposed method comes with the lowest average error. The results of the unwrapped phase image show that the proposed scheme has the ability to reduce the residues count while preserving the phase discontinuity.

It has been proved that lowering residues count does not guarantee high unwrapping accuracy; since the median filter has the ability to reduce the residual count by 86.7% meanwhile introduces the worst DEM accuracy with 97.7 meter as an average error.

REFERENCES

- [1] Ghiglia, D.C. and Pritt, M. D. "Two-Dimensional Phase Unwrapping: Theory, Algorithms, and Software". New York: Wiley 1998.
- [2] López, C.M., Eric, P."Coherence estimation in synthetic aperture radar data based on speckle noise modeling.". *Applied Optics*, 46(4):544-558, 2007
- [3] Lee, J. S., Papathanassiou, K. P., Ainsworth, Grunes, T. L., M. R, Reigber, A. "A new technique for noise filtering of SAR interferometric phase images". *IEEE Transactions on Geoscience and Remote Sensing*, 36(5):1456–1465, 1998
- [4] Chang, Ge. L. and Rizos, C. "Mine subsidence monitoring using multi-source satellite SAR images". *Photogrammetric Engineering and Remote Sensing*, 73(3):259-266, 2007
- [5] Ferraiuolo, G. and Poggi, G. "A Bayesian filtering technique for SAR interferometric phase fields". *IEEE Transactions on Image Processing*, 13(10):1368 – 1378, 2004
- [6] Andrews, A. F. and Mallows, C. "Scale mixtures of normal distributions". *Journal of the Royal Statistical Society. Series B*, 36(1), 99-102, 1974
- [7] Guarnieri, A. M. "SAR interferometry and statistical topography". *IEEE Trans. Geoscience and Remote Sensing*, 40(12): 2567–2581, 2002.
- [8] Kingsbury, N. G. "Image processing with complex wavelets". *Philosophical Transactions: Mathematical, Physical and Engineering Sciences Royal Society London*, 357 (1760):2543-2560, 1999
- [9] Kingsbury, N. G. "Complex wavelets for shift invariant analysis and filtering of signals". *Journal of Applied and Computational Harmonic Analysis*, 10(3): 234-253, 2001
- [10] Portilla, J, Simoncelli, E P. "A Parametric Texture Model based on Joint Statistics of Complex Wavelet Coefficients". *Int. Journal of Computer Vision*, 40(1):49-71, 2000
- [11] Liu, J., Moulin, P. "Complexity-Regularized Image Denoising". *IEEE Transaction on Image Processing*,10(6):841-851, 2001
- [12] Starck, J .L., Candes, E. J., Donoho, D. L. "The curvelet transform for image denoising". *IEEE Transaction on Image Processing*, 11(6):670–684, 2002
- [13] Simoncelli, E. P., Freeman. T., Adelson, E. H., Heeger, D. J. "Shiftable multi-scale transforms". *IEEE Transaction on Information Theory*, 38(2):587–607, 1992

- [14] Simoncelli, E. P. and Adelson, E. H. "Noise removal via Bayesian wavelet coring". Proceedings of 3rd IEEE International Conference on Image Processing, 1996
- [15] Moulin, P., Liu, J."Analysis of multiresolution image denoising schemes using a generalized Gaussian and complexity priors". *IEEE Transaction on Information Theory*, 45(3): 909–919, 1999
- [16] Portilla, J., Strela, V., Martin J., Simoncelli, E. P. "Denoising Using Scale Mixtures of Gaussians in the Wavelet Domain". *IEEE transactions on image processing*,12(11):1338-1350, 2003
- [17] Box, G. and Tiao, C. "Bayesian Inference in Statistical Analysis". New York: Addison-Wesley, 1992.
- [18] Figueiredo , M., Nowak, R. "Wavelet-based image estimation: An empirical Bayes approach using Jeffrey's noninformative prior". *IEEE Transaction on Image Processing*, 10(9):1322–1331, 2001
- [19] Holzner,J. "Analysis and Statistical Characterization of Interferometric SAR Signals Based on the Power Spectral Density Function". *IEEE transactions on geoscience and remote sensing*, 42(5):1116-1121, 2004
- [20] Holzner, J., Bamler, R., "Burst-mode and ScanSAR interferometry". *IEEE Transaction on Geoscience and Remote Sensing*, 40(9):1917–1934, 2002
- [21] Li, Z.W., Ding, X.L. Huang, C. Zhu, J.J., Chen, Y.L. "Improved filtering parameter determination for the Goldstein radar interferogram filter". *International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS)*, 63(6):621-634, 2008

New Approach: Dominant and Additional Features Selection Based on Two Dimensional-Discrete Cosine Transform for Face Sketch Recognition

Arif Muntasa

*Informatics Department Trunojoyo University
Kamal, Bangkalan, East Java,
Indonesia*

arifmuntasa@trunojoyo.ac.id

Abstract

Modality reduction by using the Eigentransform method can not efficiently work, when number of training sets larger than image dimension. While modality reduction by using the first derivative negative followed by feature extraction using Two Dimensional Discrete Cosine Transform has limitation, which is feature extraction achieved of face sketch feature is included non-dominant features. We propose to select the image region that contains the dominant features. For each region that contains dominant features will be extracted one frequency by using Two Dimensional-Discrete Cosine Transform. To reduce modality between photographs as training set and face sketches as testing set, we propose to bring the training and testing set toward new dimension by using the first derivative followed by negative process. In order to improve final result on the new dimension, it is necessary to add the testing set pixels by using the difference of photograph average values as training sets and the corresponding face sketches average as testing sets. We employed 100 face sketches as testing and 100 photographs as training set. Experimental results show that maximum recognition is 93%.

Keywords: Face sketch, one frequency, new dimension, dominant and additional features selection.

1. INTRODUCTION

During the last several years, feature detection was conducted by many researches [1], [2], [3], [4]. Besides feature detection, extraction feature has been also conducted for face recognition, such as Principal Component Analysis [10], Linear Discriminant Analysis [11], [12], Radial Basis Function [6], Linear Locality Preserving Projection [7], Elastic Bunch Graph Matching [8], Kernel Principal Component Analysis [5], Kernel Linear Discriminant Analysis [9] and Independent Component Analysis [13]. However, face sketch recognition by using photograph as training set has not been conducted by researchers [1], [14], [15]. Besides high dimension, problem in face sketch recognition is difference modality, where face photograph images as training set and face sketch images as testing set have big difference modality. Popular method used to reduce the difference in modality is the Eigentransform [15], but this method can not efficiently work, when number of training sets larger than image dimension, because number of covariance value dimension larger than image dimension. To overcome this problem, we propose to improve modality reduction and feature extraction process. Our dimension reduction proposed method does not depend on number of training set. To improve modality reduction process, it is

necessary to bring photograph as training set and sketch as testing set toward new dimension by using the first derivative followed by negative process. Lastly, to enhance result of modality reduction on new dimension, the difference between training and testing set average value is added to the corresponding testing set. To achieve dominant feature on new dimension of the training and testing set, we propose to select regions that contain face dominant features and additional features to eliminate of feature's location errors.

2. FACE SKETCH RECOGNITION [15]

Face recognition has become interesting issue in biometrics research field. Unfortunately, there are many researchers who have not thought about face sketch interpretation [1] and recognition [15]. In fact, when there was a criminal case, while the camera is not installed around the scene, the witnesses and sketcher will be main key to solve this case. Therefore, it is necessary to conduct research about 'face sketch recognition'. In face sketch recognition, face photograph images as training set and face sketch images as testing set have different modality. Distance between two face photograph images of different person is smaller than distance between face photograph images and face sketch images of same person [1], [15]. Therefore, to recognize face sketch without reducing modality is not supported.

Xiaou Tang transformed face photograph images to face sketch images by using the Eigentransform [15]. It is method for reducing difference of modality based on the Principal Component Analysis [15]. Consider face photograph images $M \times N$, N represents number of dimension and M represents number of training set $[P_1, P_2, \dots, P_M]$ and P_i is the Eigenface of training set, covariant of training set can be computed by using the following equation

$$C = \sum_{i=1}^M (P_i - m_p)(P_i - m_p)^T = AA^T \tag{1}$$

m_p represents face photograph image average, A_p represents training set vector, which is composed in the matrix form

$$A_p = [P_1 - m_p, P_2 - m_p, \dots, P_m - m_p] \tag{2}$$

Based on the Singular Value Decomposition (SVD) method, E_p can be expressed by using the following equation

$$E_p = A_p V_p \lambda_p^{-1/2} \tag{3}$$

V and λ are the Eigenvector and Eigenvalue respectively. New image P can be reconstructed by using the following equation

$$P_r = E_p W_p + m_p \tag{4}$$

W_p represents weight vector that can be computed from image projection to *Eigen faces*

$$W_p = E_p^T (P - m_p) \tag{5}$$

Based on equation (4) and (5), reconstruction of training set can be re-represented by using the following the equation

$$\begin{aligned} P_r &= A_p V_p \lambda_p^{-1/2} w_p + m_p \\ &= A_p C + m_p \end{aligned} \tag{6}$$

However, the Eigentransform has a weakness, if number of image dimensions is smaller than number of training set, then this method can not efficiently work.

3 PROPOSED METHOD

Photograph and sketch image are two images with different modality. Two photograph images of different person have smaller distance than photograph and sketch of same person. It shows that, besides feature extraction process, modality is main key to recognize face sketch before feature extraction is conducted. To maximize feature extraction on face photograph and sketch image, it is necessary to remove non-face and non-dominant face feature regions. Examples of non-face image are the image background, neck and clothes. Non dominant face feature, included hair and ear. Dominant face features, included left eyebrow, right eyebrow, left eye, right eye, nose and lips. To reduce error in feature extraction, it is necessary to add additional features region closed to dominant face features region. We proposed framework as seen in Figure 1.

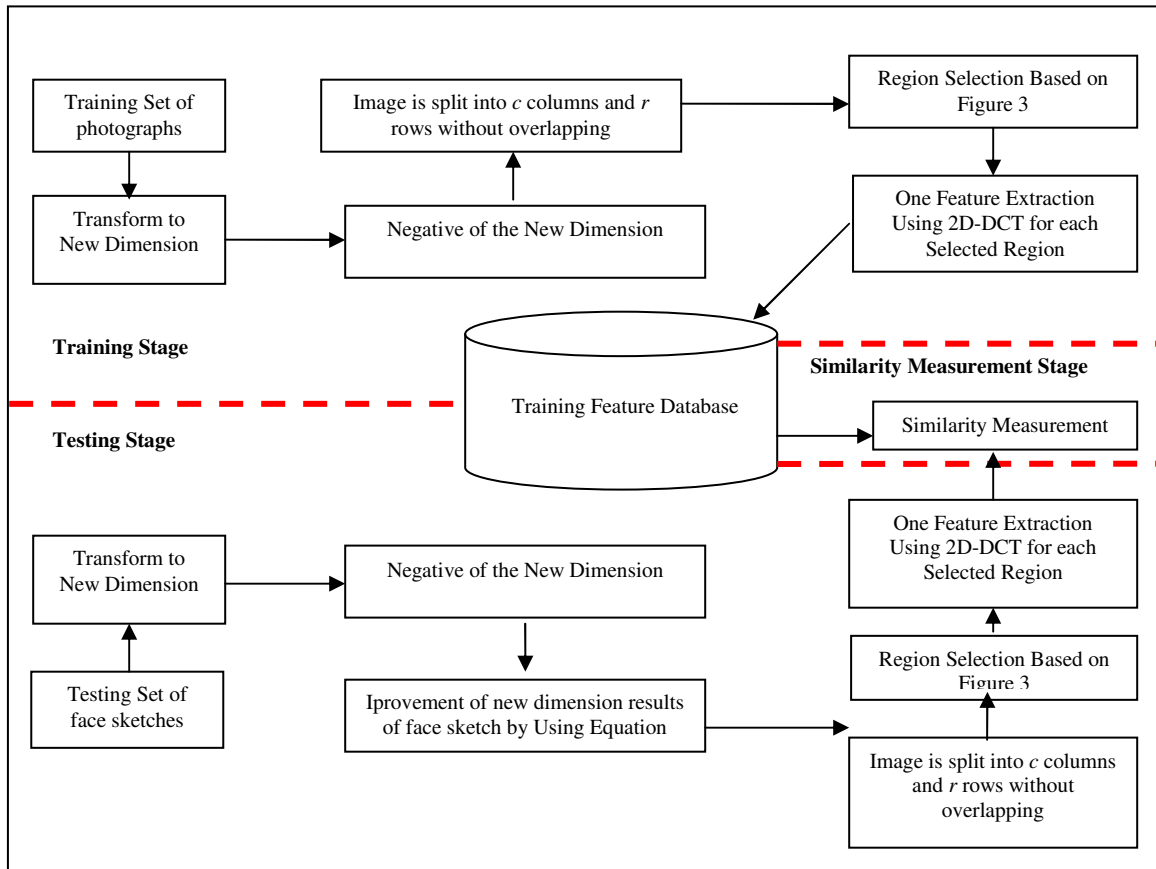


FIGURE 1 Framework of Proposed Method

3.1. Transformation toward New Dimension

Consider $f(x, y)$ and $g(x, y)$ are training and testing set respectively. To reduce the differences in the modality, for both training and testing set should be brought toward new dimension. First derivative is used to apply it. To improve the result of the first derivative, it is necessary to bring negative form. The first derivative of the training set can be expressed by using the following equation

$$f' = \begin{bmatrix} f'_x \\ f'_y \end{bmatrix} = \begin{bmatrix} \frac{\partial f(x, y)}{\partial x} \\ \frac{\partial f(x, y)}{\partial y} \end{bmatrix} \quad (7)$$

where

$$\frac{\partial f(x, y)}{\partial x} = \frac{f(x + \Delta x, y) - f(x, y)}{\Delta x} \quad (8)$$

$$\frac{\partial f(x, y)}{\partial y} = \frac{f(x, y + \Delta y) - f(x, y)}{\Delta y} \quad (9)$$

The Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of computations. Vertical and horizontal derivative can be approximated by using the following equation

$$f'_x = [f(x+1, y-1) + 2f(x+1, y) + f(x+1, y+1)] - [f(x-1, y-1) + 2f(x-1, y) + f(x-1, y+1)] \quad (10)$$

$$f'_y = [f(x-1, y+1) + 2f(x, y+1) + f(x+1, y+1)] - [f(x-1, y-1) + 2f(x, y-1) + f(x+1, y-1)] \quad (11)$$

Equation (10) and (11) can be written in the matrix form as follows

$$f'_x = \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (12)$$

$$f'_y = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (13)$$

The magnitude of these equations can be calculated by using the following equation

$$f'[f(x, y)] = \sqrt{(f'_x)^2 + (f'_y)^2} \quad (14)$$

To simplify the convolution result using the Sobel Operator on face photograph images as training set, it can be written in the following equation

$$f(x, y) \xrightarrow{\text{1st Derivative}} f'(x, y) \quad (15)$$

And negative of Equation (15) can be written in the following equation

$$\sim f'(x, y) = 255 - f'(x, y) \quad (16)$$

Similarly, it can be applied into face sketch as testing set. The result of the sobel operator on face sketch images as testing set can be expressed as follows

$$g(x, y) \xrightarrow{\text{1st Derivative}} g'(x, y) \quad (17)$$

And negative of Equation (17) can be written in the following equation

$$\sim g'(x, y) = 255 - g'(x, y) \quad (18)$$

To improve the result of new dimension on face sketch, it is necessary to add the testing set by using the difference of average photograph values as training sets and face sketches average on the corresponding testing sets. It can be modeled by using the following equation

$$\sim g'(x, y) \approx g'(x, y) + | \bar{f}_{training} - \bar{g}_{testing} | \quad (19)$$

Where $\bar{f}_{training}$ represents training sets average value and $\bar{g}_{training}$ is the corresponding testing set average value, therefore $\bar{f}_{training}$ is constant, while $\bar{g}_{training}$ based on testing set.

3.2. Localizing and Feature Selection

If photograph image is represented by using $f(x, y)$ as training set and face sketch is represented by using $g(x, y)$ as testing set, then images will be split into c columns and r rows. To achieve feature dominant, it is important to select regions that contain dominant features, such as left eyes, right eyes, left eyebrow, right eyebrow, nose and lips. In this research, image will be split into 7 columns and 18 rows, so there are 126 regions for each image as seen in Figure 1. Based on Figure 1, we have selected 13 regions that contain dominant and 12 additional features regions to reduce feature's location errors as seen in Table 1. In order to reduce feature's location error, we added 2 regions above right eyebrow location (region feature labels are 31 and 38), 2 regions above left eyebrow location (region feature labels are 33 and 40), 2 regions above nose feature (region feature labels are 46 and 53) and 6 regions below lips feature (region feature labels are 87, 88, 89, 94, 95 and 96). The result of feature selection for Figure 2 can be seen in Figure 3.

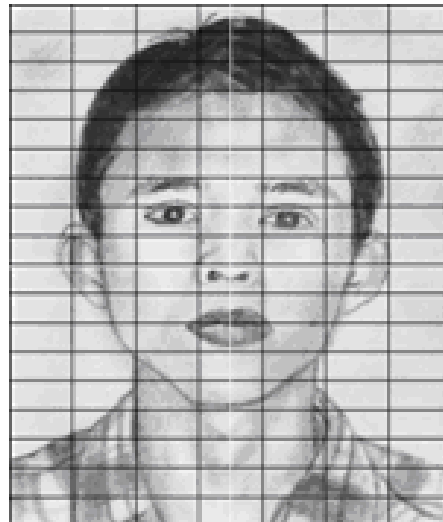


FIGURE 2 Image was split into 126 regions (7 columns and 18 rows)

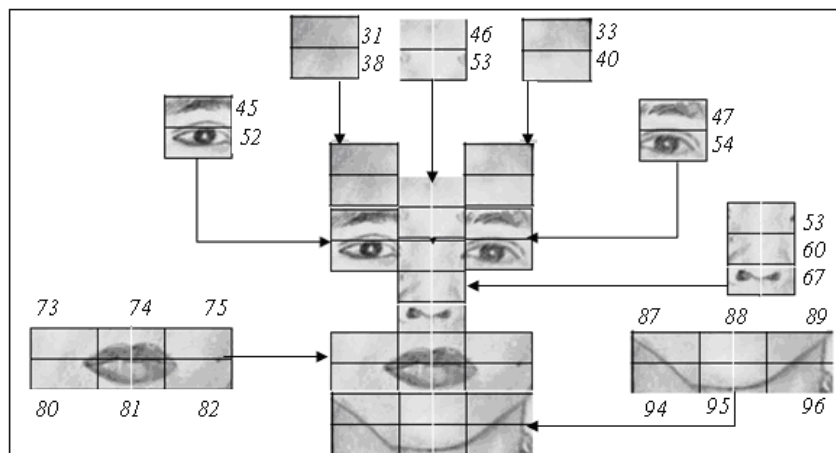


FIGURE 3 Selection of Feature's Regions to achieve Feature's Dominant and Their Additional to Reduce Feature's Error Location

While feature's region labels can be seen in Table 1

No	Features	Region Feature Labels	Additional Region Feature Labels
1	Right Eyebrow	45	31 and 38
2	Right Eye	52	-
3	Left Eyebrow	47	33 and 40
4	Left Eye	54	-
5	Nose	53 , 60 and 67	46 and 53
6	Lips	73, 74, 75, 80, 81 and 82	87, 88, 89, 94, 95 and 96.

TABLE 1 Label of Feature's Region and Their Additional

3.3. Feature extraction by Using Two Dimensional Discrete Transform (2D-DCT)

In order to achieve image features, we proposed to apply 2D-DCT for each region selected. To preserve region selected, for both feature and additional feature regions, we extracted features based on original feature location. Feature extraction resulted is based on location for each region. If number of regions selected are 25 region, then number of frequencies resulted are 25. 2D-DCT can be written in the following equation

$$C(u,v) = \frac{2}{\sqrt{M.N}} \alpha(u)\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} f(x,y) \cdot \cos\theta \cdot \cos\vartheta \tag{20}$$

Where

$$\theta = \left[\frac{\pi(2x+1)u}{2N} \right] \tag{21}$$

$$\vartheta = \left[\frac{\pi(2y+1)v}{2M} \right] \tag{22}$$

$$\alpha(u) = \begin{cases} \sqrt{\frac{1}{N}} & \text{Untuk } u = 0 \\ \sqrt{\frac{2}{N}} & \text{Untuk } u > 0 \end{cases} \tag{23}$$

$$\alpha(v) = \begin{cases} \sqrt{\frac{1}{N}} & \text{Untuk } v = 0 \\ \sqrt{\frac{2}{N}} & \text{Untuk } v > 0 \end{cases} \tag{24}$$

Consider an image $f(x, y)$, w and h are width and height of image respectively. If it is divided into c columns and r rows, then the area for each region (L_i) and i expresses region index, for each region can be written in the following equation

$$L_i = \frac{w.h}{c.r} \tag{25}$$

If number of selected regions and additional regions are R and A respectively, then number of processes required to extract an image with one frequency on each region using equation (20), (21), (22), (23) and (24) can be written by using the following equation

$$nL = \frac{w.h}{c.r} * (R + A) \tag{26}$$

If the features extraction used is selected regions only, then number of processes required to extract an image can be expressed by using the following equation

$$nL = \frac{w.h}{c.r} * R \quad (27)$$

Based on equation (26) and (27), the greater number of selected regions and additional regions, then the longer time the process required.

3.4. Similarity Measurement

To classify the result of feature extraction, four similarity measurement methods are used, which are the Euclidian Distance (d_1), Angular Separation (d_2) and the Canberra (d_3), as written in the following equation

$$d_1(f(x, y), g(x, y)) = \sqrt{\sum_{k=1}^{nF} (f(x, y)_k - g(x, y)_k)^2} \quad (28)$$

$$d_2(f(x, y), g(x, y)) = \sum_{k=1}^{nF} \frac{f(x, y)_k \cdot g(x, y)_k}{\|f(x, y)_k\| \cdot \|g(x, y)_k\|} \quad (29)$$

$$d_3(f(x, y), g(x, y)) = \sum_{k=1}^{nF} \frac{|f(x, y)_k - g(x, y)_k|}{|f(x, y)_k| + |g(x, y)_k|} \quad (30)$$

4 EXPERIMENTAL RESULTS AND DISCUSSION

In this research, we employed 100 face photograph images as training set (image samples can be seen in Figure 4) and 100 face sketch images as testing set (image samples can be seen in Figure 5). We perform two experiments. The first experiment was conducted by using 13 regions that contain feature dominant and secondly experiment was conducted by using 25 regions, which are 13 regions that contain feature dominant and 12 additional regions as seen in Table 1. Experimental results of our proposed method can be seen In Table 2.

On the first experiment, errors occurred are caused by deviation on feature selection of face sketch, for both the training and the testing set. Error of Features selection has caused the frequency difference between features of the corresponding training and the testing set. This difference will cause errors in classification. Based on Table 2 can be shown that the first experiment accuracies for the Euclidian Distance, the Angular Separation and the Canberra are 91%, 90% and 91%.

The weakness of the first experiments can be improved by additional features for each the corresponding features. On the second experiment, additional features of the area around of the corresponding feature causes number of generated frequencies more than the first experiment. However, the frequency difference between the corresponding training and testing set produces smaller frequencies than the first experiment. It can reduce the error rate that occurred in the first experiment. The Accuracies of the second experimental results are 91%, 92% and 93%. It shows that additional region selection can improve face sketch recognition of the first experimental.

Our proposed method has higher recognition than our previous research. On our previous research obtained maximum recognition accuracy reached only 89% [3].

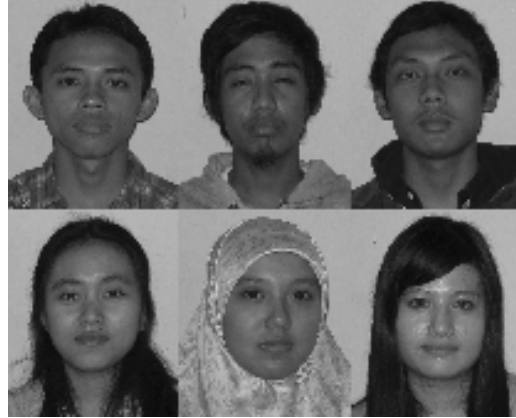


FIGURE 4 samples of 6 person face photograph images



FIGURE 5 samples of 6 person face sketch images

Experimental Results	The number of Regions Used	Face Sketch Recognition (%)		
		Euclidian Distance	Angular Separation	Canberra
1 st	13	91	90	91
2 nd	25	91	92	93
Average		91	91	92
Maximum		91	92	93

Table 2 Experimental Results of Our Proposed Method

5 CONCLUSION AND FUTURE WORK

In this paper, we can show that

1. Combining of feature extraction between dominant features and additional features can produce higher face sketch recognition than feature extraction using dominant features only.
2. Improvement of new dimension transformation results on face sketch by adding the difference of average photograph values face sketches average and dominant features selection followed by dominant features selection, additional features selection and

- feature extraction can obtain higher face sketch recognition than without dominant feature and additional features selection
3. Errors that occurred in face sketch recognition are caused error in dominant features and additional features selection.
 4. Generally, Errors occurred on the first and the second experiment are caused by errors in the dominant features selection of face sketch, it can produce high frequency difference value of corresponding data training and data testing. This difference will cause errors in classification.

For future work, we will detect feature location first to obtain more accurate position of features before feature extraction is conducted. To improve similarity measurement, the Support Vector Machine method will be used.

6 REFERENCES

- [1]. A. Muntasa, M. Hariadi., M. H. Purnomo. "*Maximum Feature Value Selection Of Nonlinear Function Based On Kernel Pca For Face Recognition*". In Proceeding of The 4th Conference On Information & Communication Technology and Systems, Surabaya, Indonesia, 2008
- [2]. A. Muntasa, M. Hariadi., M. H. Purnomo. "*A New Formulation of Face Sketch Multiple Features Detection Using Pyramid Parameter Model and Simultaneously Landmark Movement*". IJCSNS International Journal of Computer Science and Network Security, 9(9): 2009
- [3.] A. Muntasa. "*A Novel Approach for Face Sketch Recognition Based on the First Derivative Negative and 2D-DCT with Overlapping Model*". International Journal of Computer Science, (Accepted), 2010
- [4]. D. Cristinacce and T.F. Cootes. "*Facial Feature Detection using ADABOOST with Shape Constraints*". In Proceeding .BMVC, Vol.1, 2000
- [5]. Lu J., P. K.N. , V. A.N.. "*Face Recognition Using Kernel Direct Discriminant Analysis Algorithms*". IEEE Trans. Neural Networks, 14(1):117-126, 2003
- [6]. Su, H., Feng D., and Zhao R.-C. "*Face Recognition Using Multi-feature and Radial Basis Function Network*". In Proceeding of the Pan-Sydney Area Workshop on Visual Information Processing, Sydney, Australia, 2002
- [7]. X. He, S. Yan, Y. Hu, P. Niyogi, Hong-Jiang Zhang. "*Face Recognition Using Laplacianfaces*". IEEE Transactions On Pattern Analysis And Machine Intelligence, 27(3):328-340, 2005
- [8]. L. Wiskott, J.M. Fellous, N. Kruger, C. von der Malsburg. "*Face Recognition by Elastic Bunch Graph Matching*". IEEE Trans. on Pattern Analysis and Machine Intelligence, 19(7): 775- 779, 1997
- [9]. A. Muntasa, M. Hariadi., M. H. Purnomo. "*Maximum Feature Value Selection Of Nonlinear Function Based On Kernel Pca For Face Recognition*". In Proceeding of The 4th Conference On Information & Communication Technology and Systems, Surabaya, Indonesia, 2008
- [10] M. Turk, A. Pentland. "*Eigenfaces for recognition*". *Journal of Cognitive Science*, 71–86, 1991
- [11] J.H.P.N. Belhumeur, D. Kriegman. "*Eigenfaces vs. fisherfaces: Recognition using class specific linear projection*" IEEE Trans. on PAMI, 19(7):711–720, 1997
- [12]. Yambor, W.S. "*Analysis of PCA-Based and Fisher Discriminant-Based Image Recognition Algorithms*". Tesis of Master, Colorado State University, 2000
- [13]. Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. "*Face recognition by independent component analysis*". IEEE Trans. on Neural Networks, 13(6): 1450-1464, 2002
- [14]. R.G. Uhl and N.d.V. Lobo. "A Framework for Recognizing a Facial Image from A Police Sketch". In *Proceedings of CVPR*, 1996
- [15]. X. Tang, X. Wang. "*Face Sketch Recognition*". IEEE Transactions on Circuits and Systems for Video Technology, 14(1):50-57, 2000

Performance Comparison of Face Recognition Using DCT Against Face Recognition Using Vector Quantization Algorithms LBG, KPE, KMCG, KFCG

Dr. H. B. Kekre

*Senior Professor, MPSTME, SVKM's
NMIMS University Mumbai,
400-056, India.*

hbkekre@yahoo.com

Dr. Tanuja K. Sarode

*Assistant Professor, TSEC, Bandra (W),
Mumbai, 400-050, India.*

tanuja_0123@yahoo.com

Prachi J. Natu

*Assistant Professor, GVAIET, Shelu,
Karjat, 410201, India*

prachi.natu@yahoo.com

Shachi J. Natu

*Lecturer, TSEC, Bandra (W),
Mumbai, 400-050, India.*

shachi_natu@yahoo.com

Abstract

In this paper, a novel face recognition system using Vector quantization (VQ) technique is proposed. Four different VQ algorithms namely LBG, KPE, KMCG and KFCG are used to generate codebooks of desired size. Euclidean distance is used as similarity measure to compare the feature vector of test image with that of trainee images. Proposed algorithms are tested on two different databases. One is Georgia Tech Face Database which contains color JPEG images, all are of different size. Another database used for experimental purpose is Indian Face Database. It contains color bitmap images. Using above VQ techniques, codebooks of different size are generated and recognition rate is calculated for each codebook size. This recognition rate is compared with the one obtained by applying DCT on image and LBG-VQ algorithm which is used as benchmark in vector quantization. Results show that KFCG outperforms DCT and other three VQ techniques and gives better recognition rate up to 85.4% for Georgia Tech Face Database and 90.66% for Indian Face Database. As no Euclidean distance computations are involved in KMCG and KFCG, they require less time to generate the codebook as compared to LBG and KPE

Keywords: Face Recognition, DCT, VQ, KFCG, LBG

1. INTRODUCTION

Information is an important asset and can be in any form. Widespread use of an internet is making access and transmission of the information much easier. Particularly, the ability to access the information in an efficient manner has become a crucial issue. Hence some security measures must be provided so that only authorized users can access the intended information. Fingerprint recognition, speech recognition, iris recognition etc. are different recognition techniques that can be used for this purpose. But each of them has its own drawbacks. Special equipments to capture the input for recognition system and willingness of an individual to undergo testing procedure can hinder the implementation of these security measures. Face recognition is one of the important techniques and has become an interesting research topic due to its enormously commercial and law enforcement applications [1]. Face recognition is challenging because it is a real world problem. The human face is a complex, natural object that tends not to have easily (automatically) identified edges and features. Because of this, it is difficult to develop a mathematical model of the face that can be used as prior knowledge when analyzing a particular image. A human face is complex object with features that can vary over the time due to aging. Other factors that influence face recognition include shape, pose, poor lighting, occlusion, sunglasses, long hair, or other objects partially covering the subject's face and low resolution images [1].

Face recognition as a security measure has an advantage that it does not require any extensive equipment to capture the image of an individual and also explicit co-operation from an individual is not expected [10, 34]. Face recognition can also be used in human- computer interaction. A face recognition system is a process of automatically identifying or verifying a person from a digital image. In face recognition applications, original input image is of high dimension. Feature extraction is the most important step in face recognition, which reduces high dimensional image data to low dimensional feature vectors. A face recognition system can operate in following two modes:

Verification: A one to one comparison of a captured biometric with a stored template to verify that the individual is who he claims to be.

Identification: A one to many comparisons of the captured biometric against a biometric database in attempt to identify an unknown individual. The identification only succeeds in identifying the individual if the comparison of the biometric sample to a template in the database falls within a previously set threshold [15].

In this paper, the goal is to present effective face recognition system by using vector quantization technique. Vector Quantization (VQ) is simpler method for data compression and has been successfully used in various applications involving VQ based encoding and VQ based recognition. VQ consists of three steps Codebook design, Encoding and Decoding.

The rest of the paper is organized as follows: in section 2 we present related work carried out in the field of face recognition. In section 3 we discuss vector quantization. In section 4 our proposed approach is presented. Section 5 elaborates the experiment conducted. Results are tabulated in section 6. Conclusion has been outlined in section 7.

2. RELATED WORK

In face recognition system feature extraction plays an important role to create more robust systems [1].

Various techniques are proposed till now to extract feature vector from the image. PCA [15], Wavelet analysis, LDA [16], EBGM [15] are to name the few. Principle Component Analysis (PCA) approach is used to reduce the dimension of data by means of data compression basics and reveals the most effective low dimensional structure of facial patterns [17]. This reduction in dimension removes information that is not useful and precisely decomposes the face structure into orthogonal components known as eigenfaces. Each face image is represented as a feature

vector which is stored in a 1D array. A distance between the respective feature vectors of images is compared to find the match [18]. In LDA samples of unknown classes are classified based on training samples of known classes. Whereas nonlinear characteristics of face are considered in EBGm. DCT [26] has been used as a feature extraction step in various studies on face recognition. Other transforms like Walsh-Hadamard transform [30, 31, 32], Wavelet transform also have been proposed [9, 24, 25, 28, 35].

For Recognizing objects from large image databases, histogram based methods were proposed in last decade. Initially, this idea was based on color histograms that were launched by Swain and Ballard [19]. Following this idea numerous developments were made by different people, exploiting this idea, such as texture histograms for 2D object recognition suggested by Gimelfarb and Jain [20], shape-index histograms for range image recognition proposed by Dorai and Jain [21] and relational histograms used by Huet and Hancock [22] for line-pattern recognition. Similarly, one dimensional (1D) and two dimensional (2D) histograms are also proposed with diverse variations like 1D shape index histogram, 2D maximum and minimum curvature histogram, 2D mean and Gaussian curvature histogram and 2D shape index and curvedness histogram in [23]. Based on vector quantization algorithm, VQ histogram method is proposed by kotni [1].

Closest distance between histograms of different face images can be used for recognition purposes. Different distance measures may affect the recognition rate [22]. Euclidean distance can be used as it produces stable and satisfactory results [23].

2.1 Feature Extraction Using 2D-DCT on the Image

DCT technique that we have applied is as follows:

- 1) Read the color image from the database.
- 2) Convert the image into grayscale and resize it to size 128x128.
- 3) Apply 2D-DCT on this preprocessed image to obtain feature vector of an image.
- 4) Repeat steps 1 through 3 for each database image.
- 5) Read query image.
- 6) Repeat steps 2 and 3 for each query image to obtain its feature vector.
- 7) Calculate Euclidean distance of query image with each image in the database using equation (1).

$$ED = \sqrt{\sum_{i=1}^n (V_{pi} - V_{qi})^2} \dots\dots\dots (1)$$

where V_{pi} and V_{qi} be the feature vectors of image P and Q respectively. Minimum Euclidean distance gives the closest matching image from the database. Accuracy of this algorithm is discussed in result section.

3. Vector Quantization [2, 7, 8]

VQ can be defined as a mapping function that maps k dimensional vector space to a finite set $CB = \{C_1, C_2, C_3, \dots, C_N\}$. The set CB is called codebook consisting of N number of codevectors and each codevector $C_i = \{c_{i1}, c_{i2}, c_{i3}, \dots, c_{ik}\}$ is of dimension k. The key to VQ is the good codebook. There are various codebook generation algorithms available in literature [7].

Vector quantization is composed of three operations: Codebook design, Encoding and Decoding. The input to the encoder is a vector and output is the index of the codevector that shows closest matching with that of input vector. In this case the closest match is found by evaluating the Euclidean distance between the input vector and each codevector in the codebook. Once the closest codevector is found, the index of that codevector is sent through a channel. When the decoder receives the index of the codevector, it replaces the index with the associated codevector [3]. Various techniques to generate the codebook are available. Codebook can be generated in spatial domain by clustering algorithms or using transform domain techniques [11].

The method most commonly used to generate codebook is the Linde-Buzo-Gray (LBG) algorithm [2, 7, 8].

3.1 Linde-Buzo-Gray (LBG) Algorithm [2,3,4]

Initially entire training vector of an image acts as one cluster. Centroid of this cluster is calculated. This is the first codevector (CV). Constant error is added to CV which gives two new vectors V1 and V2. Euclidian distance of all training vectors with V1 and V2 is calculated. Training vectors having minimum Euclidian distance with V1 than V2 are grouped into one cluster and remaining are grouped into another cluster. Now centroids of these two clusters are calculated which are nothing but the two new codevectors. These vectors overwrite the value CV in the codebook. Thus two new clusters are formed and codebook size has become two. In second iteration above procedure is repeated for two new clusters. It generates four new clusters. Then this procedure is repeated for each new cluster till codebook of desired size is generated [2].

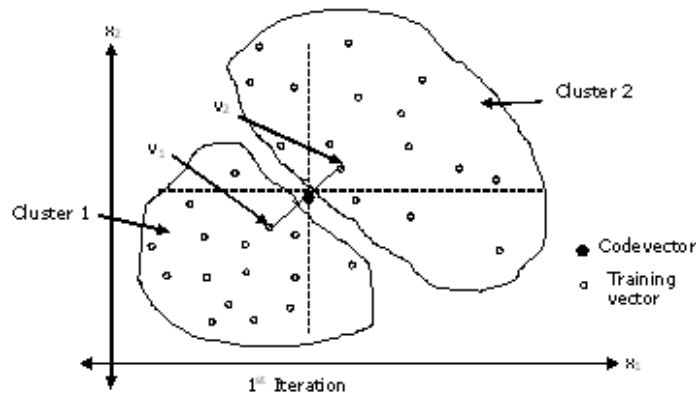


FIGURE 1: LBG for two dimensional case

3.2 Kekre's Proportionate Error Algorithm (KPE) [6, 33]

This algorithm is same as LBG, but the difference is, instead of adding constant error to codevector, we add proportionate error. Number of Euclidean distance computations required in LBG and KPE are $2N$ and numbers of comparisons are also $2N$. Here N is the total number of training vectors in each iteration to generate clusters.

3.3 Kekre's Median Codebook Generation Algorithm (KMCG) [6, 29]

In this algorithm the training vector matrix T is sorted with respect to first column of the training vector. The median of training vector is selected and put into the codebook. Now codebook size is one. Matrix T is divided into two equal parts. Each part is sorted with respect to second column of the matrix. Median of each sorted part is chosen and written into the codebook. Codebook size is now increased to two. Each half obtained in previous step is again divided into two half parts. Thus four parts are obtained. Each of these parts is sorted with respect to next column number i.e. with respect to third column. Median of each cluster is chosen and written into the codebook. This gives codebook of size four. This process is repeated till the codebook of desired size is generated. As Euclidean distance computation is not required while generating the codebook, this algorithm generates codebook in a faster way.

3.4 Kekre's Fast Codebook Generation Algorithm (KFCG) [5, 27, 30]

KFCG does not use Euclidean distance to generate the codebook. Let T is the initial training vector. T contains m training vectors each of size k . $T = \{X_1, X_2, \dots, X_m\}$ and $X_1 = \{x_{11}, x_{12}, \dots, x_{1k}\}$.

Initial codevector C_1 is computed as centroid of training vector T . In first iteration, first element of each training vector of T is compared with the first element of codevector. i.e. if $x_{i1} < C_{11}$ then put X_i in cluster 1 else in cluster 2.

In second iteration, split cluster1 into two by comparing x_{i2} with second element of the centroid of cluster 1 and cluster 2 into two by comparing x_{i2} with second element of the centroid of cluster 2. Repeat this procedure till codebook of desired size is generated.

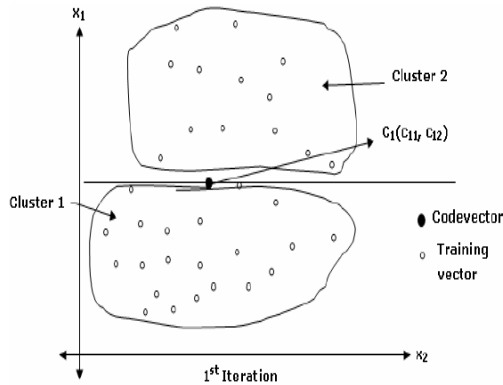


FIGURE 2: KFCG for two dimensional case with first iteration

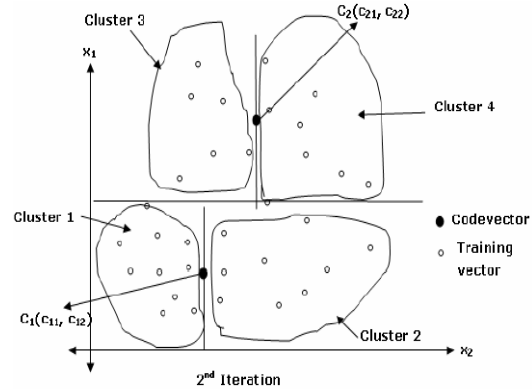


FIGURE 3: KFCG for two dimensional case with second iteration

3.5 Proposed Method

Figure 4 shows steps involved in proposed method.

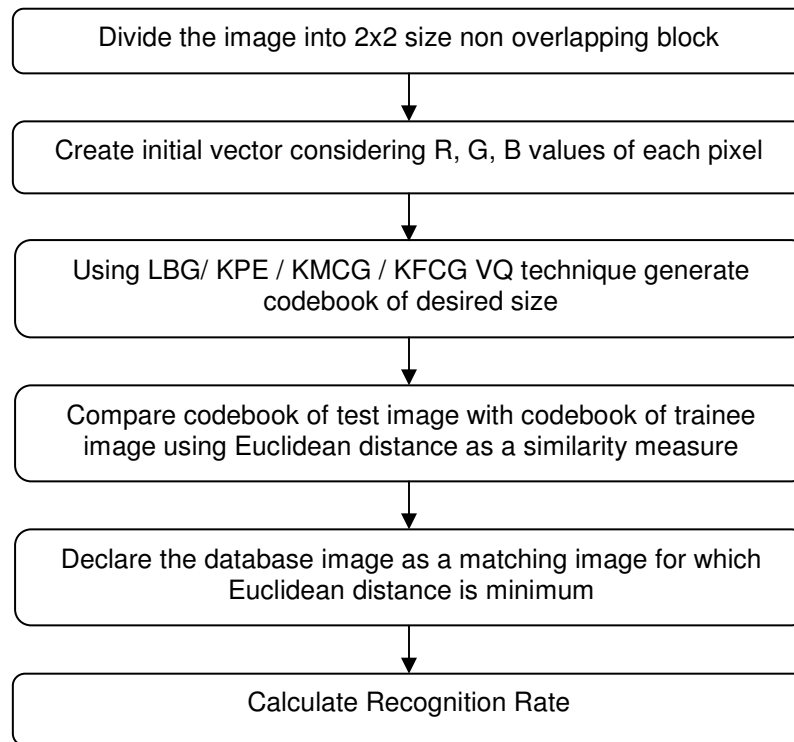


FIGURE 4: Flowchart for face recognition system using proposed VQ techniques

3.6 Query execution

Here the codebook of query image is generated by applying proposed VQ technique. This codebook is compared with other codebooks in codebook database using direct euclidean distance as similarity measure which is given in equation (1). Codebooks of size 4, 8, 16, 32, 64, 128, 256, 512 are generated using LBG, KPE, KMCG and KFCG algorithms. Recognition rate in each case is calculated and compared.

Recognition rate is calculated as:

$$(\%) \text{ Recognition Rate} = \frac{\text{Number of images correctly recognized} \times 100}{\text{Total number of images tested}} \dots\dots\dots(2)$$

4. Experiments and Results

In implementation of these algorithms, we have used two different databases, Georgia Tech Face Database and Indian Face Database. No preprocessing is done on the images in the database. These algorithms were implemented in MATLAB 7.0 on Intel Core 2 Duo Processor (2.0 GHz), 3GB RAM on Windows XP.

4.1 Georgia Tech Face Database

It contains total 750 images, 15 images each for 50 individuals. The images are in JPEG format and all are of different sizes. The images in the database show variations in illumination conditions, facial expression, and appearance. The images show frontal and/or tilted faces with different facial expressions, lighting conditions and scale. Out of 15 images per person, 8 images were taken as trainee images and remaining 7 images were taken as test images. Overall 350 images are test images and 400 images are trainee images. For second case, numbers of trainee images were increased to 10 images per person so that 250 images were used as test images and same methodology was applied. Sample images from Georgia Tech Face Database are shown below in figure 5.



FIGURE 5: Sample images from Georgia Tech Face Database

Steps mentioned in Section 2.1 are applied on the images in case 1 and case 2 to obtain the results of 2D-DCT. These results are mentioned in result section in 5.3. Steps mentioned in Section 3.5 are applied for each image in the database. For every image, codebooks of size 4x12, 8x12, 16x12, 32x12, 64x12, 128x12, 256x12, and 512x12 are generated using LBG, KPE, KMCG and KFCG algorithms. These codebooks are feature vectors of the corresponding images and are stored in the database. Feature vector of test image is compared with feature vectors of all trainee images using Euclidean distance as a measure of similarity to obtain the closest recognized image. After giving query image as input, the closest match is sent as output. Recognition rate is calculated for different codebook size using all four VQ algorithms.

4.2 Indian Face Database

Local database is generated to test efficiency of algorithms. This database has been prepared using Indian faces. The faces have been selected from long video clips where the object is asked to move the face with different angle and expressions. The shooting has been carried out at different locations with wide difference in illumination. From these video clips various frames have been selected for each object by taking care that variation of expression and angle are included. This database contains images of 50 persons and is stored in bitmap format. There are 10 color images of each person. Database contains images with rotation of face at different angles, different facial expressions and with considerable difference in illumination. Images are taken at different locations. Out of 10 images per person, 3 images were used in testing set and 7 in training set. In second case 2 test images were considered and 8 trainee images were taken. Sample faces from Indian Face Database are shown low. For this database also, codebooks of size 4x12, 8x12, ..., 512x12 are generated using LBG, KPE, KMCG and KFCG algorithms and recognition rate obtained for these algorithms are compared. Sample images from Indian Face Database are shown below in figure 6. Steps mentioned in Section 2.1 and Section 3.5 are followed for Indian Face Database also.



FIGURE 6: Sample images from Indian Face Database

4.3 Results obtained by applying 2D-DCT on the image:

Technique Used	(%) Recognition Rate			
	Georgia Tech Face Database		Indian Face Database	
2D-DCT	400 Trainee Images	500 Trainee Images	350 Trainee Images	400 Trainee Images
		70%	70%	88.66%

Table 1: Comparison of (%) Recognition Rate for different number of trainee images, by applying DCT on the image.

Table 1 shows percentage accuracy obtained by applying 2D- DCT on the gray scale resized image. Images from two different databases are used. It is observed that, for Georgia tech Face Database 70% accuracy is obtained when test sets of 350 and 250 images are used. For Indian Face Database, accuracy of 88.66% is obtained with 150 test images and 90% accuracy is obtained with 100 test images.

4.4 Results obtained by applying VQ Algorithms:

4.4.1 Results obtained for Georgia Tech Face Database:

VQ Technique	Codebook Size							
	4x12	8x12	16x12	32x12	64x12	128x12	256x12	512x12
LBG	73.42	79.14	79.14	77.14	76.85	72.28	70.57	64.57
KPE	73.42	79.42	78.57	77.14	76.85	74.85	72.85	64
KMCG	43.71	48.85	52.57	56.28	66.29	64.86	76.57	78.57
KFCG	78.28	78.57	74.57	75.71	76	76.86	76.86	76.28

Table 2: Comparison of (%) Recognition Rate for different codebook sizes by applying different VQ Techniques on Georgia Tech Face Database with 400 trainee images

Table 2 shows comparison of LBG, KPE, KMCG and KFCG algorithms. For each of these algorithms codebooks of size 4, 8,16,32,64,128,256 and 512 are generated. From Table 2 it is clear that, for different VQ techniques, codebook size that gives maximum recognition rate is different. Optimal codebook size for LBG is 16x12 whereas for KPE and KFCG it is 8x12. Due to overclustering, recognition rate decreases when codebook size is increased in LBG, KPE, and KFCG. For KMCG, recognition rate goes on increasing as codebook size is increased.

VQ Technique	Codebook Size							
	4x12	8x12	16x12	32x12	64x12	128x12	256x12	512x12
LBG	78.4	84	83.2	83.2	83.2	79.2	76.4	69.6
KPE	79.2	84	84.6	84	84	80.8	76.8	70
KMCG	51.6	48	66.8	69.2	75.6	80.4	84	84.4
KFCG	82.8	84	82.4	83.2	82.8	84	85.4	82.8

Table 3: Comparison of (%) Recognition Rate for different codebook sizes by applying different VQ Techniques on Georgia Tech Face Database with 500 trainee images.

By changing the number of trainee images for the same database, recognition rate is calculated. Here 5 test images and 10 trainee images per individual are used. Table 3 shows that if number of trainee images in the database is increased, the recognition rate also changes.

4.4.2 Results obtained for Indian Face Database:

VQ Technique	Codebook Size							
	4x12	8x12	16x12	32x12	64x12	128x12	256x12	512x12
LBG	66.66	78.66	84	86	86	86.66	82	58.66
KPE	66	76.66	77.33	80	82	76	67.33	53.33
KMCG	43.33	53.33	68.66	76	82.66	83.33	86	87.33
KFCG	79.33	80.66	84.66	86	86.66	90.66	84	77.33

Table 4: Comparison of (%) Recognition Rate for different codebook sizes by applying different VQ Techniques on Indian Face Database with 350 trainee images.

Table 4 shows the comparison of recognition rates for Indian Face Database. Here out of 10 images per person, 7 images are used as trainee images and 3 images are used for testing. It shows that KFCG outperforms all three VQ techniques in terms of recognition rate, giving 90.66% accuracy.

VQ Technique	Codebook Size							
	4x12	8x12	16x12	32x12	64x12	128x12	256x12	512x12
LBG	68	81	86	88	87	87	80	61
KPE	65	77	78	71	62	54	36	12
KMCG	45	55	64	77	80	82	86	86
KFCG	80	82	88	89	89	90	84	76

Table 5: Comparison of (%) Recognition Rate for different codebook sizes by applying different VQ Techniques on Indian Face Database with 400 trainee images

Table 5 shows comparison of recognition rate when 8 images per person in Indian Face Database are considered as trainee images. In this case also KFCG shows highest recognition rate than other VQ algorithms.

4.5 Complexity Analysis

Let M be the total number of training vectors,

k be the vector dimension,

N be the codebook size,

1 CPU unit is required for addition of 8 bit numbers

1 CPU unit for comparison

For multiplication/division of two 8 bit number 8 CPU units are required

To compute one squared Euclidean distance (ED) of k dimensional vector k multiplications and 2k-1 additions are required and hence 8k + 2k -1 CPU units are needed.

To compute Centroid for M training vectors of k dimension, M-1 additions and k divisions are required and therefore M-1 + 8k CPU units are needed.

Let $P = \log_2 N$

Complexity Parameters	LBG	KPE	KFCG	KMCG
Total Comparisons	2MP	2MP	MP	$M \cdot \sum_{i=0}^{P-1} \log_2 (M/2^i)$
Total No. of ED	2MP	2MP	0	0

Table 6: Comparison of LBG, KPE, KFCG and KMCG algorithm with respect to total number of Comparisons and total number of ED computations required.

Table 6 shows comparison of all four VQ algorithms in terms of two complexity parameters: total number of comparisons required and total number of Euclidean Distance computations required. It is observed that, numbers of comparisons required in KFCG are exactly half of the number of comparisons required in LBG and KPE and number of Euclidean Distance computations in KMCG and KFCG are zero. Hence generation of codebooks using KFCG and KMCG is faster.

Parameter	Algorithm applied on Georgia Tech Face Database with 400 Trainee Images				
	DCT	LBG	KPE	KMCG	KFCG
Codebook size for maximum recognition rate	-	16x12	8x12	512x12	8x12
Recognition Rate (%)	70	79.14	79.42	78.57	78.57

Table 7: Comparison of Recognition Rate by applying DCT and different VQ techniques on Georgia tech Face Database with 400 trainee images.

Table 7 shows the comparison of highest recognition rates obtained using VQ algorithms with the recognition rate obtained by applying 2D DCT technique on Georgia Tech Face Database with 400 trainee images. It is observed that, maximum rate of 79.42% is obtained for codebook size 8x12 using KPE algorithm. For same codebook size KFCG gives 78.57% accuracy. DCT gives 70% rate for same set of database.

Parameter	Algorithm applied on Georgia Tech Face Database with 500 Trainee Images				
	DCT	LBG	KPE	KMCG	KFCG
Codebook size for maximum recognition rate	-	8x12	16x12	512x12	256x12
Recognition Rate (%)	70	84	84.6	84.4	85.4

Table 8: Comparison of Recognition Rate by applying DCT and different VQ techniques on Georgia tech Face Database with 500 trainee images.

Similarly, from table 8 it is observed that, KFCG gives highest recognition rate of 85.4% for 256x12 size codebook with 500 trainee image set of Georgia Tech Face Database. DCT gives 70% for same case and peak value obtained by LBG, KPE and KMCG is 84%, 84.6% and 84.4% respectively.

Parameter	Algorithm applied on Indian Face Database with 350 Trainee Images				
	DCT	LBG	KPE	KMCG	KFCG
Codebook size for maximum recognition rate	-	128x12	64x12	512x12	128x12
Recognition Rate (%)	88.66	86.66	82	87.33	90.66

Table 9: Comparison of Recognition Rate by applying DCT and different VQ techniques on Indian Face Database with 350 trainee images.

Table 9 shows comparison of accuracy for Indian Face Database with 350 trainee images and 150 test images. DCT gives 88.66% accuracy. KFCG gives highest accuracy of 90.66% with 128x12 codebook size. KMCG gives maximum rate of 87.33%.

Parameter	Algorithm applied on Indian Face Database with 400 Trainee Images				
	DCT	LBG	KPE	KMCG	KFCG
Codebook size for maximum recognition rate	-	32x12	16x12	256x12	128x12
Recognition Rate (%)	90	88	78	86	90

Table 10: Comparison of Recognition Rate by applying DCT and different VQ techniques on Indian Face Database with 400 trainee images.

Result for second set of Indian Face Database is presented in table 10. Here KFCG gives maximum 90% recognition rate for 128x12 size codebook.

According to table 11, maximum accuracy obtained by KMCG is for codebook size 512x12. Using KFCG algorithm, maximum recognition rate is obtained at codebook size 256x12. Number of CPU units required by KMCG and KFCG are 368639 and 112723 respectively. Here also, KFCG requires least number of CPU units to perform all calculations.

Parameter	Algorithm applied on Georgia Tech Face Database with 500 Trainee Images with image size 128x128				
	DCT	LBG	KPE	KMCG	KFCG
Codebook size giving the maximum recognition rate	0	8x12	16x12	512x12	256x12
Number of multiplications required	4210688	29599	393420	6144	3084
Number of additions required	4194303	614579	803187	12287	55283
Number of comparisons required	0	24576	32768	307200	32768
Number of CPU units required	37879807	875947	3983315	368639	112723

Table 11: Comparison of number of calculations and CPU units required for applying DCT and different VQ techniques on Georgia Tech Face Database with 500 trainee images.

Though the size of codebook in case of LBG and KPE is smaller for maximum Recognition Rate than the codebook size in case of KFCG, the numbers of CPU units required by these two algorithms are higher than the CPU units required by KFCG.

5. CONCLUSION

In this paper, we have discussed face recognition using DCT and vector quantization. LBG, KPE, KMCG and KFCG-VQ algorithms are used to generate codebook of an image. No preprocessing is done on the images and their performance is compared with that of DCT technique in terms of recognition rate and number of CPU units required. DCT gives maximum accuracy of 70% for Georgia Tech face database and 90% for Indian face database. For Georgia Tech face database KFCG gives maximum accuracy of 85.4%. LBG gives 83.2%, KMCG gives 84.4% and KPE gives 84.6% accuracy. For Indian face database KFCG gives the best performance with the accuracy of 90.66%. LBG gives 88%, KMCG gives 87.33% and KPE gives 82% accuracy. Thus KFCG outperforms DCT algorithm and other VQ techniques namely LBG, KPE and KMCG. Numbers of CPU units required in KFCG are almost 336 times less than that of required in DCT and KFCG gives highest recognition rate. Also the CPU units required by KFCG are 7.77 times less than CPU units needed by LBG, 35 times less than that of needed by KPE and 3.27 times less than the CPU units required by KMCG. This is because the numbers of comparisons are reduced to half in KFCG as compared to LBG and KPE and no Euclidean distance computations are required in KMCG and KFCG which makes their execution much faster.

6. REFERENCES

- [1] K. kotani, C. Qiu and T. Ohmi. "Face Recognition Using Vector Quantization Histogram Method". International Conference on Image Processing, 2002
- [2] H.B.Kekre, K. Shah, T. K. Sarode, S. Thepade. "Performance Comparison of Vector Quantization Technique-KFCG with LBG, Existing Orthogonal Transforms and PCA For Face Recognition". International Journal of Information Retrieval, II(1):64-71, 2009
- [3] Y. Linde, A. Buzo, R. M. Gray. "An algorithm for vector quantizer design". IEEE Transaction on Communication, COM-28(1):84-95, 1980
- [4] A. Gersho, R.M. Gray. "Vector Quantization and Signal Compression", Kluwer Academic Publishers, Boston, 1991.
- [5] H.B. Kekre, T. Sarode." An Efficient Fast Algorithm to Generate Codebook for Vector Quantization". First International conference on Emerging Trends In Engineering and Technology (ICETET), 2008

- [6] H.B. Kekre, T. Sarode. "Fast Codebook Search Algorithm For Vector quantization Using Sorting technique". ACM International Conference on Advances in Computing, Communication And Control (ICAC3) 2009
- [7] Y. Linde, A. Buzo, R. M. Gray. "An algorithm for vector quantizer design". *IEEE Trans. Commun.*, COM-28(1):84-95, 1980
- [8] C. chan, L. Po. "A Complexity reduction Technique For Image Vector Quantization". *IEEE Transactions on Image Processing*, 1(3):312-321, 1992
- [9] H. K. Ekenel, R. Stiefelhagen. "Analysis of Local Appearance Based Face recognition: Effects of Feature Selection and Feature Normalization". International Conference on Computer vision and Pattern Recognition Workshop, 2006
- [10] S. Lin. "An Introduction to Face Recognition Technology". Informing Science Special Issue on Multimedia Informing Technologies- Part 2, 3(1): 2000
- [11] J. Choi, Y. Chung, K. Kim, J. Yoo. "Face Recognition using Energy Probability in DCT Domain". IEEE International Conference on Multimedia and Expo, 2006
- [12] H. B. Kekre, Ms. Tanuja K. Sarode, S. D. Thepade. "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation". *ICGST-International Journal on Graphics, Vision and Image Processing (GVIP)*, 9(5):1-8, 2009. Available at [http:// www.icgst.com/gvip/Volume9/Issue5/ P1150921752.html](http://www.icgst.com/gvip/Volume9/Issue5/P1150921752.html).
- [13] "Georgia Tech Face Database". Available at: <http://www.face-rec.org/databases>.
- [14] "Face Recognition". Available at: <http://www.biometrics.gov/Documents/FaceRec.pdf>
- [15] P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman. "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711-720, 1997
- [16] M. Begum, N. Nahar, K. Fatimah, M. K.Hasan and M. A. Rahaman. "An Efficient Algorithm for Codebook Design in Transform Vector Quantization". WSCG, 2003
- [17] F. Basit, M. Y. Javed and U. Qayyum. "Face Recognition Using Processed Histogram and Phase-only Correlation (POC)". International conference on Emerging Technologies (ICET), 2007
- [18] M. J. Swain and D. H. Ballard, "Indexing via color histogram", In Proceedings of third international conference on Computer Vision (ICCV), Osaka, Japan, 1990.
- [19] G. L. Gimel'farb, A. K. Jain. "On retrieving textured images from an image database". *Pattern Recognition*, 29(9):1461-1483, 1996
- [20] C. Dorai, A. K. Jain, Cosmos. "A representation scheme for free form surfaces". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 19(10):1115-1130, 1997
- [21] B. Huet, E. R. Hancock. "Line pattern retrieval using relational histograms". *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 21(12):1363-1370, 1999
- [22] Y. Li, E. R. Hancock. "Face Recognition using Shading-based Curvature Attributes". *IEEE Proceedings of the 17th International Conference on Pattern Recognition (ICPR'04)* 1051-4651/04

[23] H.B.Kekre, S. D. Thepade. "Improving the Performance of Image Retrieval using Partial Coefficients of Transformed Image". International Journal of Information Retrieval (IJIR), Serials Publications, 2(1):72-79, 2009

[24] H.B.Kekre, T. Sarode, S. D. Thepade. "DCT Applied to Row Mean and Column Vectors in Fingerprint Identification". In Proceedings of International Conference on Computer Networks and Security (ICCNS), VIT, Pune, 2008

[25] H.B.Kekre, S. D. Thepade, A. Athawale, A. Shah, P. Verlekar, S. Shirke. "Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors". International Journal of Computer Science and Network Security (IJCSNS), 10(1): 2010, (ISSN: 1738-7906) Available at www.IJCSNS.org.

[26] H.B.Kekre, S. D. Thepade, A. Athawale, A. Shah, P. Verlekar, S. Shirke. "Performance Evaluation of Image Retrieval using Energy Compaction and Image Tiling over DCT Row Mean and DCT Column Mean". Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 2010

[27] H.B.Kekre, T. K. Sarode, S. D. Thepade, V. Suryavanshi. "Improved Texture Feature Based Image Retrieval using Kekre's Fast Codebook Generation Algorithm". Springer-International Conference on Contours of Computing Technology (Thinkquest-2010), Babasaheb Gawde Institute of Technology, Mumbai, 2010

[28] H. B. Kekre, T. K. Sarode, S. D. Thepade. "Image Retrieval by Kekre's Transform Applied on Each Row of Walsh Transformed VQ Codebook". (Invited), ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), Thakur College of Engg. And Tech., Mumbai, 2010

[29] H. B. Kekre, T. Sarode, S. D. Thepade. "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook". International Journal on Imaging (IJI), 2(A09):55-65, 2009. Available at www.ceser.res.in/iji.html (ISSN: 0974-0627).

[30] H. B. Kekre, Ms. T. K. Sarode, S. D. Thepade. "Image Retrieval using Color-Texture Features from DCT on VQ Codevectors obtained by Kekre's Fast Codebook Generation". ICGST-International Journal on Graphics, Vision and Image Processing (GVIP), 9(5):1-8, 2009. Available at: <http://www.icgst.com/gvip/Volume9/Issue5/P1150921752.html>.

[31] H. B. Kekre, S. Thepade, A. Maloo. "Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform". International Journal of Engineering Science and Technology, 2(4): 362-371, 2010

[32] H. B. Kekre, S. Thepade, A. Maloo. "Performance Comparison of Image Retrieval Using Fractional Coefficients of Transformed Image Using DCT, Walsh, Haar and Kekre's Transform". CSC-International Journal of Image processing (IJIP), 4(2)142-155, 2010

[33] H. B. Kekre, T. Sarode. "Two Level Vector Quantization Method for Codebook Generation using Kekre's Proportionate Error Algorithm". CSC-International Journal of Image Processing, 4(1):1-10, 2010

[34] H. B. Kekre, S. Thepade, A. Maloo. "Eigenvectors of Covariance Matrix using Row Mean and Column Mean Sequences for Face Recognition". CSC-International Journal of Biometrics and Bioinformatics (IJBB), 4(2):42-50, 2010

[35] H. B. Kekre, T. Sarode, S.i Natu, P. Natu. "Performance Comparison Of 2-D DCT On Full/Block Spectrogram And 1-D DCT On Row Mean Of Spectrogram For Speaker Identification". CSC International Journal of Biometrics and Bioinformatics (IJBB), 4(3).

A Novel Method for De-warping in Persian Document Images Captured by Cameras

Hadi Dehbovid

Electrical Engineering Department, Faculty of Engineering, Islamic Azad University, Science and Researches Branch, Tehran, Iran.

hadi.dehbovid@gmail.com

Farbod Razzazi

Electrical Engineering Department, Faculty of Engineering, Islamic Azad University, Science and Researches Branch, Tehran, Iran

farbod_razzazi@yahoo.com

Shahpour Alirezaee

Electrical Engineering Department, Faculty of Engineering, Zanzan University, Zanzan, Iran Zanzan, Iran

sh_alirezaee@yahoo.com

Abstract

In this Paper, We proposed a novel algorithm for de-warping of Persian document images captured by the cameras. The aim of de-warping is to remove page distortions and to straighten document images captured by the cameras, so that the documents are readable to the OCR system. Recently, the industrial implementation of the images captured by digital cameras has significantly expanded. Most of the studies carries out so far in this regard have focused on the documents written in Latin and few researches have been conducted regarding Persian documents. The original idea of the proposed algorithm is based on the segmentation of the components of texts. In this algorithm, an effective technique is offered for detection of the upper and lower baselines, which is used in estimation of the slope of the words. Moreover, vertical shift of the warped words is done through fitting a quadratic curve fitted to the centers of the words in a line in relation to the horizontal line. The suggested algorithm is examined by qualitative and quantitative measures and the results of its implementation on various documents indicate a 92% accuracy of the proposed technique in correction of the location and angle of the words.

Keywords: Geometric Distortion, OCR, camera based OCR, Image Archives

1. INTRODUCTION

Recently, the significant role of digital cameras in preparation and analysis of document images is expanding. The reason behind this development is the fact that compared to scanners; such cameras are lighter, more applicable and less expensive. Cameras can be employed in cases where scanners are not of any use. Imaging of thick books and hand-written historical books that have to not be touched are examples of such cases. Such advantages have resulted in

applicability of digital cameras in a wide range of applications such as printed documents digital library and natural scenes containing textual contents.

The analysis of the document images captured by these cameras can be divided into different categories based on the type of the document, the technology used in the process of imaging and the desired application.

Generally, images can be captured from any context containing textual contents (such as a paper document, a subtitled video frame, a car plate number, or an open book). In fact, the differences in method of the imaging, will determine the complexities of the textual contents extraction.

The conditions of capturing images with cameras are more complex and different, in comparison to scanners. Therefore, techniques that are more effective and robust have to be applied for the analysis of camera-based document images. Novel algorithmic technologies used for document analysis, generate desirable results with pure, flat text documents. By pure texts, we mean those documents that do not include images or formulas and, and those in which a single language is used. Flat documents refer to documents that are free from geometric distortions. In such techniques, the assumption is a high-quality simple-structured document image (i.e. a straightened black colored text on white background).

Unfortunately, there are no such assumptions for camera-based systems, resulting in new challenges for their applications. Examples of such challenges include low resolution, variable brightness in the page, perspective distortion, non-planer surfaces, complex backgrounds, zooming and focusing on a specific part, movements of the objects while a photo is captured, as well as warping of the content. The significant challenge that we are going to address here is geometric warping of the document images. Warping of the document images is one of the main complexities of the pre-processing stage, and different techniques have been devised so far in order to remove this problem. In fact, the actual aim of de-warping of the pages is to align the document image captured by camera before the OCR stage, so that the document is readable by an OCR system. The most important types of such distortions are those resulting from perspective and paper curvatures.

In the recent years, various techniques are proposed to recover the distortion of the warped document images [1]. These techniques are categorized into hardware versus software-based page reconstruction sets. An example of the first type is the reconstruction of a 3D shape by means of hardware such as stereo cameras [2,3], photometric devices [4], and laser-based scanners [5]. Page reconstruction using a single camera in an uncontrolled environment [6, 7, and 8] can be mentioned as an instance of the second group. In the present article, our assumption is an image taken by a digital camera, without any hardware post-processing. Therefore, the focus of our study is on the second type of algorithms.

Among the second type of techniques mentioned above, three main techniques of restoration of the warped images are mostly focused on. These are representation of a continuous skeletal image for document images de-warping (SKEL)[9], segmentation-based recovery of the warped document images (SEG) [10], and a coordinates-translation model for guiding the documents in one direction and recovery of the distorted document (CTM) [11].

The central idea of the skeletal technique is based on the extraction of the surface skeleton of the document in which connected branches estimated by means of Bezier curves, determine the internal space of the text lines. CTM is a model based on the translation of the cylindrical coordinates and through biasing the document, restores the warped book. Besides, the basis of SEG technique lies upon the segmentation of the words and correction of the skew of these words, using the baselines estimated for them, and finally the vertical translation of these warped words. The three models are used in de-warping Latin document images; however, none of the previously proposed methods can remove the distortion of such documents completely. The SKEL method is able to remove the distortions resulting from page curvatures, but fails to do so

for perspective distortions, nor can it restore the distortions occurred to formulas and some components of the documents belonging to the adjacent page. CTM can recover formulas to some extent and also works for the components of the document which are apparent from the neighboring pages, however, it fails to remove the perspective and page curvature distortions. SEG model also is applicable for the parts appearing from the adjacent pages as well as distortions of formulas and page curvatures, yet unable to work for perspective distortions. Among these three models, the SEG method is of more applicability with regard to de-warping documents[13]. Therefore, in the present article, this model is considered as the base model and its improved version is proposed as a technique for restoration of the distortions occurred to documents written in Persian or Arabic. It has to be pointed that since the SEG model is based on the features of Latin orthography, it is not directly applicable to Persian texts. The experimental results presented in this study also confirm such a fact.

The remaining of present paper is structured as follows: in section 2, the related literature is reviewed and. In the 3rd section, the proposed algorithm is represented. Section 4 includes the experimental results and discussions on the recovered images quality and finally the article will be summarized in section 5.

2. SEGMENTATION-BASED MODEL FOR DE-WARPING

2.1 SEG Model

As was mentioned before, SEG model is considered as the base model of this study. Such a technique improves the quality of the images captured by digital cameras through the following four stages:

In the first stage, noisy areas and the black margins as well as the noisy components of the text belonging to adjacent pages are detected and reduced. The original idea of black margins discrimination is based on vertical as well as horizontal projections. First, the image is smoothed, and then the initial and final offsets from the corners and textual areas are calculated with regard to vertical and horizontal projections.

Finally, the black margins are removed through the analysis of the connected segments of the image. In the next step, the words are reconstructed through smoothing.

During the second stage, words and baselines of the text are detected through implementing the segmentation model for the documents for which noise reduction has been carried out. The adjacent words are linked to each other consecutively in order to define the text lines.

This is performed by the sequential extraction of the words on the right and left sides of the first detected word after a stage by stage top-down scan of the document. For each determined word, the upper and lower baselines are calculated [12].

In the third stage, the initial binary warped image is restored through de-warping and translation of the words with regard to the upper and lower baselines. For each word, the skew is calculated using the slopes of the corresponding upper and lower baselines. Then, all the detected words are de-warped and transformed, so that a primary estimation of the recovered binary image is at hand. Ultimately, it is in the fourth stage that the smoothed image is reconstructed.

2.2 Limitations of the original model with regard to Persian documents

SEG technique have a number of limitations when used for Latin texts and even more problems emerge when this model is used for documents written in Persian. Regarding English texts, there is no appropriate filter for accurate detection of noises and removing them. Besides, in such texts, the estimation of the upper and lower baselines is not always error free.

In case the original model is implemented with Persian texts, none of its stages works appropriately and the only output will be some noisy pixels. Thus, the original model is merely considered as a basis for offering a novel algorithm. The original technique bears many problems if used for restoration of the Persian document images. It faces lots of complexities in case of detection and elimination of noisy areas. Moreover, due to the differences in the writing structure of English and Persian, segmentation of the words in Persian texts also will be problematic, since unlike English texts, in Persian documents the height and slopes of the letters of the words do not follow a specific order. Therefore, again the original SEG model will face difficulties in determining the upper and lower baselines. In other words, in Persian, the slope detection and transformation of the words in a geometrically warped text have to be done in a structured manner relying on the whole page. Whereas, in English, these are conducted for each word separately, independent from other words on the page. Another critical limitation of the original model is that the translation of words in each line is based on the highest point in the written form of the first word of the same line. This is less problematic for English texts, since their writing structure is more organized; however, with regard to Persian texts, difficulties are abundant due to the fact that in Persian the height of the text line does not correspond to the height of the words. In figure 1, implementation of the base model on a Persian imaged document is discussed. As we can observe, the original method has no application for Persian texts and the distortions resulting from warping of the image cannot be restored. This limitation made us to propose a new technique in order to remove the above mentioned problems. In our suggested model, robust solutions are presented for the stated limitations.

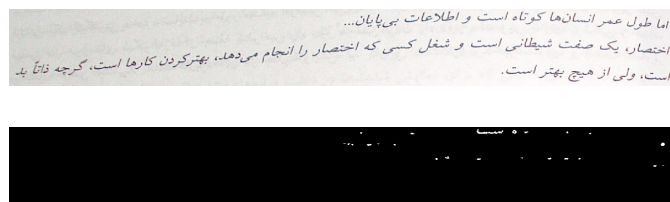


FIGURE 1: Implementing the original model on a Persian sample, from top to down, a. the original document image, b. the output image

3. THE PROPOSED ALGORITHM

The flow chart of the proposed algorithm is demonstrated in figure 2.

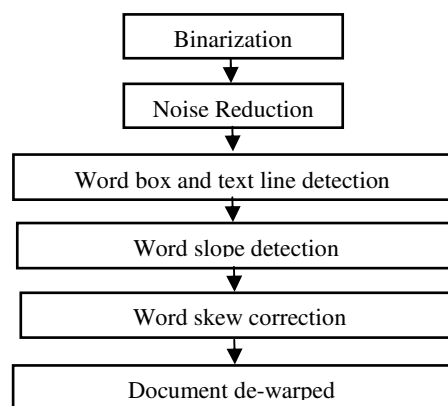


FIGURE 2: Flowchart of the proposed algorithm

As the chart indicates, in this algorithm, first, the image is binarized (black as background and white as foreground) by means of the threshold extracted from the Otsu algorithm, so that the text components of the image are indicated. Then, the bounding boxes of the words are extracted. In

the next step, based on the height of the bounding boxes of all words in the text, a histogram is calculated. The maximum value of the histogram corresponds to the average character height M . Those histograms whose heights are more than $a.M$ or less than M/b are removed. In addition, those with a width less than M/c are also removed, so that in the next stage of the algorithm, we do not face a difficulty. In fact, at this stage, small letters and segments are eliminated. Empirically, a , b , and c are selected in a way that undesired text components and noises are removed and only the main words remain.

Further in the procedure, two low-pass filters are used in order to horizontally smooth the words within the bounding boxes. The thresholds of these filters are determined empirically. If the thresholds are considered lower than the determined desirable values, the noisy areas and the redundant letters within the bounding boxes that have a small width and length, will not be removed. In addition, if the determined thresholds are increased, then some of the words that are, in fact, the main components of the texts, will be missed. After the horizontal smoothing, once more, the noisy areas are removed. Then, the remained words are detected. In this stage, through a top-down scanning, the word in the highest point of the page is detected and is labeled as K , to which, this to the first text line L is assigned. Next, the words are scanned from right to left and the first word, the word that neighbors at a small distance in the left side of word K , is detected. Furthermore, it is assumed that these two words which are closest to each other horizontally, should have something in common vertically too, therefore the word which is detected as the closest one is not selected from the following line. This closest word is labeled as K_f . Continuing our scanning, we got to the last word on the left. The same procedure is followed for all the words on the right side of the first word and label all words found in a left to right scanning order and assign them to the first text line. Furthermore, a flag is enabled for these components to indicate that they will not participate to further calculations and labeling. Subsequently, all of the above procedures were repeated for the remained words till we got to the last word of the image. In this way, all of the words of the image were detected and for each word, the corresponding lines were also determined, which, in turn, are used for detection and correction of the skew of the words, as is explained bellow.

The second stage of the proposed algorithm, includes the determination of the upper and lower baselines for each word which delimit the main body of the words (see Fig. 3), by which, we are able to find the slope of each word. For each word, the upper and lower baseline is defined via a simple linear equation as is shown in equation (1), (2).

$$y = a_{ij}x + b_{ij} \quad (1)$$

$$y = a'_{ij}x + b'_{ij} \quad (2)$$

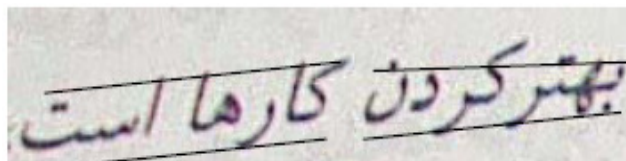


FIGURE 3: Example of upper and lower baseline estimation.

In these equations, the $a_{ij}, b_{ij}, a'_{ij}, b'_{ij}$ coefficients are obtained based on the characteristics of the word. Due to the atypical written forms of the Persian words, and because of the fact that the height of the words does not follow a specific order, it is not possible to estimate the upper and lower baselines for each word correctly, thus it is necessary to consider a number of successive words as one phrase and find the upper and lower baselines for that phrase. This is done through transforming the bounding box of a single word to a bigger one, which encloses a sequence of words. Experimentally, we came up with the conclusion that if we consider each word together with its two preceding words and three following words as a phrase, and regard them in a common bounding box, the baselines will be estimated with better accuracy.

In the next stage, the skew of the words was estimated and also rotated. Here, the slope of each word is derived from the corresponding baseline slopes. Upper and lower baseline slopes of word K_{ij} are denoted as:

$$\theta_{ij}^u = \arctan(a_{ij}^u), \theta_{ij}^l = \arctan(a_{ij}^l) \tag{3}$$

Using the estimated slope for each word, the angle of that word in relation to the horizontal line was also determined through equation 4. This angle, of course, derived from the comparison we drew between the estimated angles of the upper and lower baselines. This procedure was carried out through equation 4, for the first word of each line, and equation 5, for the other words.

$$\theta_{io} = \begin{cases} \theta_{io}^u, & \text{if } |\theta_{io}^u| < |\theta_{io}^l| \\ \theta_{io}^l, & \text{otherwise} \end{cases} \tag{4}$$

$$\theta_{ij} = \begin{cases} \theta_{ij}^u, & \text{if } |\theta_{ij}^u - \theta_{ij-1}^u| < |\theta_{ij}^l - \theta_{ij-1}^l| \\ \theta_{ij}^l, & \text{otherwise} \end{cases} \tag{5}$$

Then we re-formed the bounding box of each word and in each bounding box, we transformed those pixels that were translated in relation to the horizontal line, using the following formula (6).

$$\begin{cases} x_{\min} = x \\ y_R = \max(1, \text{round}((x - x_{\min}) * \sin(-\theta_{ij}) + y * \cos(\theta_{ij}))) \end{cases} \tag{6}$$

Estimation and correction of the slope of each word, was carried out based on the slope of previous word in the same line, and when the words in a line finished, this procedure also is ended, and for the new line, the correction of the skews was performed based on the words of the same line and their estimated slopes.

The final stage of the proposed algorithm, consists of transformation of the warped words which was corrected in terms of their slopes. This stage is of paramount importance and, in fact, the main part of the proposed algorithm is related to the transformation of the words with a corrected angle, and sequencing the words in a line, so that all the words in a line are in the same direction and on a horizontal line.

In this part, the criterion is not the maximum value of the bounding box of each word. In fact, here, first the geometric center of all bounding boxes of the words are found, then a polynomial curve of degree two or three is fitted to these centers, using an MMSE scale. In the present study, experiments indicated that the most desirable results will be gained if a quadratic curve is used. As figure 4 indicates, after fitting the curve, the difference between the centers of the bounding boxes with the curve, symbolized as d (see Fig. 4), should be estimated and recovery should be carried out with regard to the horizontal line, through the following formula(7).

$$\begin{cases} y_{Rs} = \max(1, (\text{round} (y + d))) \\ x_{Rs} = x \end{cases} \tag{7}$$

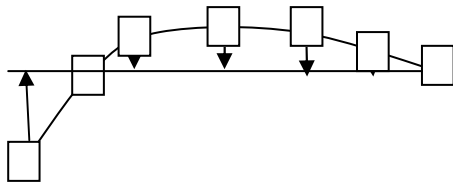


FIGURE 4: Words alignment model.

4. EXPERIMENTAL RESULTS

In this section, the implementation of the proposed algorithm on the Persian document images and the resulting findings are examined. Then a comparison is done between the original algorithm and the one proposed in this paper with regard to their application for the same samples. The samples are captured by a Sony digital camera (S950), with a resolution of 10 megapixels. Totally, a sum of 8 samples were provided and the suggested algorithm is implemented on all of them, the result of which are presented bellow.

Figure 5(a) & 6(a) & 7(a) show three Persian warped samples.



FIGURE 5: from top to down – a. the original image, b. elimination of the noisy areas, determination of the boxes and the text lines, c. estimation and correction of the skews of the boxes, d. transformation of the boxes and the output image of the algorithm.

In figure 5, part a, the original document image containing geometric distortion is presented, in b, the noisy areas and the small words are removed, in fact, the points and symbols such as slash, comma, etc. are removed, however, as it can be observed, in the end of second line, a main word is also removed. In part c, for each phrase within a bounding box, its slope is determined and based on the formulas presented in the previous part, the slopes are recovered. In d, the words, for which skew correction was conducted, were transformed. As presented by the figure, this stage was successful for all the three lines, but the dot of the last word of the second line is not places where it should.

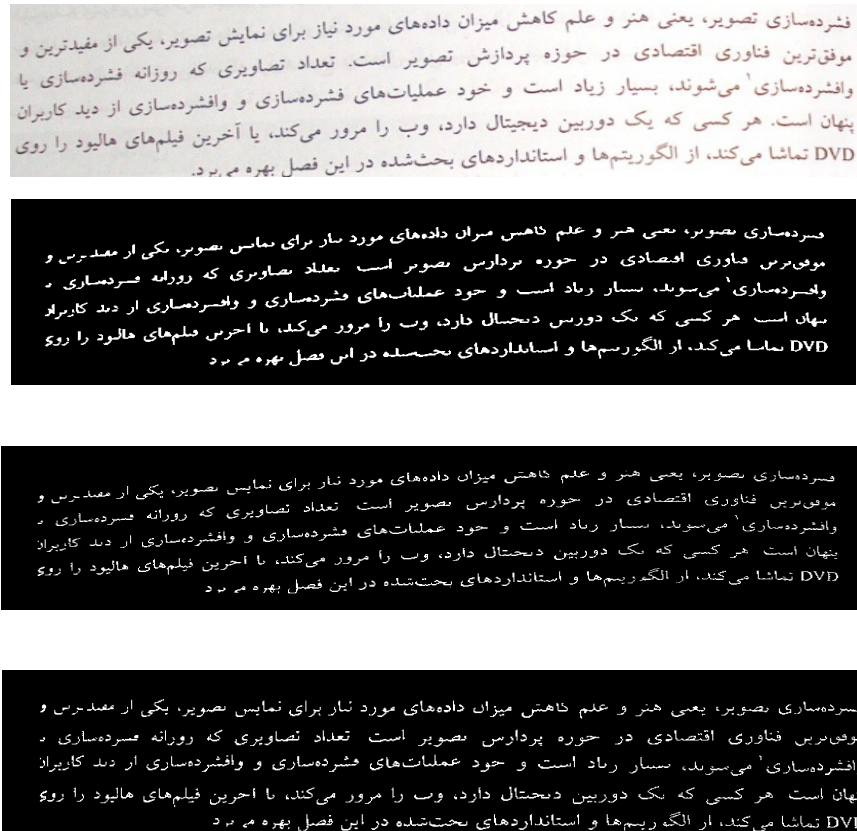


FIGURE 6: from top to down- a, the original image, b. elimination of the noisy areas- determination of the boxes and the text lines, c. estimation and reformation of the skews of the boxes, d. transformation of the boxes and the final output of the algorithm

In figure 6, part b, which is related to elimination of the noisy areas, a limited number of words in the end of the second line as well as those at the end of the last line are missed. In c, the slopes of the phrases are recovered accurately and in d, the transformation process was also successfully conducted.

شکل ۱۴-۸ فرآیند رمزگذاری دوبعدی را برای پیمایش یک خط نشان می‌دهد. توجه کنید که مراحل اولیه این روش، به یافتن چندین عنصر متغیر^۳ مهم هدایت شده‌اند: a_0 ، a_1 ، a_2 و b_1 و b_2 . عنصر متغیر، توسط استاندارد، به عنوان پیکسلی تعریف می‌شود که مقدارش متفاوت از پیکسل قبلی در همان خط است. مهمترین عنصر متغیر a_0 (عنصر مرجع) است، که یا برابر با مکان عنصر متغیر سفید فرضی در سمت چپ اولین پیکسل هر خط رمزگذاری می‌شود یا از حالت رمزگذاری قبلی تعیین می‌شود. حالت‌های رمزگذاری در ادامه بحث می‌شود. پس از تعیین مکان a_0 ، a_1 به عنوان مکان عنصر متغیر بعدی در سمت راست a_0 روی خط رمزگذاری فعلی، a_2 به عنوان عنصر متغیر بعدی در سمت راست a_1 روی خط رمزگذاری فعلی، b_1 به عنوان عنصر متغیر با مقدار مخالف (با a_0) و در سمت راست a_0 در خط مرجع (با خط قبلی)، و b_2 به عنوان عنصر متغیر بعدی در سمت راست b_1 روی خط مرجع مشخص می‌شود. اگر هر کدام از این عناصر متغیر تشخیص داده نشوند، مکان آن، نسبت به یک فرضی در سمت راست آخرین پیکسل، روی خط مناسبی تعیین می‌شود. شکل ۱۵-۸ دو نمونه از روابط کلی بین عناصر متغیر مختلف را نشان می‌دهد.

شکل ۱۴ ۸ فرآیند رمزگذاری دوبعدی را برای پیمایش یک خط نشان می‌دهد. توجه کنید که مراحل اولیه این روش، به یافتن چندین عنصر متغیر^۳ مهم هدایت شده‌اند: a_0 ، a_1 ، a_2 و b_1 و b_2 . عنصر متغیر، توسط استاندارد، به عنوان پیکسلی تعریف می‌شود که مقدارش متفاوت از پیکسل قبلی در همان خط است. مهمترین عنصر متغیر a_0 (عنصر مرجع) است، که یا برابر با مکان عنصر متغیر سفید فرضی در سمت چپ اولین پیکسل هر خط رمزگذاری می‌شود یا از حالت رمزگذاری قبلی تعیین می‌شود. حالت‌های رمزگذاری در ادامه بحث می‌شود. پس از تعیین مکان a_0 ، a_1 به عنوان مکان عنصر متغیر بعدی در سمت راست a_0 روی خط رمزگذاری فعلی، a_2 به عنوان عنصر متغیر بعدی در سمت راست a_1 روی خط رمزگذاری فعلی، b_1 به عنوان عنصر متغیر با مقدار مخالف (با a_0) و در سمت راست a_0 در خط مرجع (با خط قبلی)، و b_2 به عنوان عنصر متغیر بعدی در سمت راست b_1 روی خط مرجع مشخص می‌شود. اگر هر کدام از این عناصر متغیر تشخیص داده نشوند، مکان آن، نسبت به یک فرضی در سمت راست آخرین پیکسل، روی خط مناسبی تعیین می‌شود. شکل ۱۵ ۸ دو نمونه از روابط کلی بین عناصر متغیر مختلف را نشان می‌دهد.

FIGURE 7: from top to down- a. the original image, b. final output of the algorithm

Sample	Figure 5	Figure 6	Figure 7	Average of the total samples
Accuracy of the original algorithm	5%	3%	5%	4%
Accuracy of the proposed algorithm	95%	90%	92%	92%

TABLE 1: A comparison between the proposed and the original approaches.

Table 1 indicates the results of applying the suggested algorithm on 8 samples. Greater efficiency of the proposed method is also presented numerically. The results reveal the high applicability of the proposed algorithm presented in this paper. References should be indicated in the text by consecutive numbers in square brackets, as [1], [2] etc.

5. CONCLUSION & FUTURE WORK

The suggested algorithm is of much efficiency with regard to Persian texts. This algorithm can be implemented on the Persian document images captured by cameras, which have dramatic geometric distortions, as well as photometric distortions. The results of our studies support the effectiveness of this algorithm. This algorithm is generalizable to texts containing formulas and figures as well, about which, researches are being conducted. Besides, reduction of photometric distortions is another area, which can play a complementary role to the practically applications of this algorithm.

6. REFERENCES

- [1] J. Liang, D. Doermann, H. Li. *"Camera-based analysis of text and documents: a survey"*. Int. Jour. Of Document Analysis and Recognition, 7(2-3): 84–104, 2005
- [2] A. Ulges, C. Lampert, and T. M. Breuel. *"Document capture using stereo vision"*. In Proceedings of the ACM Symposium on Document Engineering, Milwaukee, Wisconsin, USA, 2004
- [3] A. Yamashita, A. Kawarago, T. Kaneko and K.T.Miura. *"Shape reconstruction and image restoration for non-flat surfaces of documents with a stereo vision system"*. In Proceedings of 17th International Conference on Pattern Recognition (ICPR) Cambridge UK, 2004
- [4] M.S. Brown and W.B. Seales. *"Document restoration using 3d shape: A general deskewing algorithm for arbitrarily warped documents"*. In International Conference on Computer Vision (ICCV), Vancouver, B.C., Canada, 2001
- [5] M. Pilu. *"Deskewing perspectively distorted documents: An approach based on perceptual organization"*. In HP Technical Reports, 2001
- [6] L. Zhang and C.L. Tan. *"Warped image restoration with applications to digital libraries"*. In Proc. Eighth Int. Conf. on Document Analysis and Recognition, Washington, DC, USA, 2005

- [7] A. Ulges, C.H. Lampert and T.M. Breuel. "*Document image dewarping using robust estimation of curled text lines*". In Proc. Eighth Int. Conf. on Document Analysis and Recognition, Washington, DC, USA, 2005
- [8] J. Liang, D.F. DeMenthon, and D. Doermann. "*Flattening curved documents in images*". In Proc. Computer Vision and Pattern Recognition, San Diego, 2005
- [9] A. Masalovitch and L. Mestetskiy. "*Usage of continuous skeletal image representation for document images de-warping*". In 2nd Int. Workshop on Camera- Based Document Analysis and Recognition, Curitiba, Brazil, 2007
- [10] B.Gatos, I. Pratikakis, and K. Ntirogiannis. "*Segmentation based recovery of arbitrarily warped document images*". In Proc. Int. Conf. on Document Analysis and Recognition, Curitiba, Brazil, 2007
- [11] B. Fu, M.Wu, R. Li, W. Li, and Z. Xu. "*A model-based book de-warping method using text line detection*". In 2nd Int. Workshop on Camera-Based Document Analysis and Recognition, Curitiba, Brazil, 2007
- [12] U.V. Marti, H. Bunke. "*Using a statistical language model to improve the performance of an HMMbased cursive handwriting recognition system*". Int. Jour. of Pattern Recognition and Artificial Intelligence, 15(1): 65–90, 2001
- [13] F. Shafait and T. M. Breuel. "*Document Image Dewarping Contest*". In proc CBDR, 2007

A Novel and Robust Wavelet based Super Resolution Reconstruction of Low Resolution Images using Efficient Denoising and Adaptive Interpolation

Liyakathunisa

*Ph.D Research Scholar
Dept of Computer Science & Engg
S. J. College of Engineering
Mysore, India.*

liyakath@indiatimes.com

C.N .Ravi Kumar

*Professor & Head of Department
Dept of Computer Science & Engg
S. J. College of Engineering
Mysore, India.*

kumarcnr@yahoo.com

Abstract

High Resolution images can be reconstructed from several blurred, noisy and aliased low resolution images using a computational process know as super resolution reconstruction. Super resolution reconstruction is the process of combining several low resolution images into a single higher resolution image. In this paper we concentrate on a special case of super resolution problem where the wrap is composed of pure translation and rotation, the blur is space invariant and the noise is additive white Gaussian noise. Super resolution reconstruction consists of registration, restoration and interpolation phases. Once the Low resolution image are registered with respect to a reference frame then wavelet based restoration is performed to remove the blur and noise from the images, finally the images are interpolated using adaptive interpolation. We are proposing an efficient wavelet based denoising with adaptive interpolation for super resolution reconstruction. Under this frame work, the low resolution images are decomposed into many levels to obtain different frequency bands. Then our proposed novel soft thresholding technique is used to remove the noisy coefficients, by fixing optimum threshold value. In order to obtain an image of higher resolution we have proposed an adaptive interpolation technique. Our proposed wavelet based denoising with adaptive interpolation for super resolution reconstruction preserves the edges as well as smoothens the image without introducing artifacts. Experimental results show that the proposed approach has succeeded in obtaining a high-resolution image with a high PSNR, ISNR ratio and a good visual quality.

Keywords: Adaptive Interpolation, DWT, Denoising, Super Resolution, Thresholding.

1. INTRODUCTION

Super Resolution Reconstruction (SRR) is a process of producing a high spatial resolution image from one or more Low Resolution (LR) observation. It includes an alias free up sampling of the image thereby increasing the maximum spatial frequency and removing the degradations that arises during the image capture, Viz Blur and noise. An image is often corrupted by noise during acquisition and transmission. For instance in acquiring images with a CCD camera, light levels and sensor temperature are the major factors affecting the amount of noise in the resulting image. Images are also corrupted during transmission due to interference in the channel [3][6]. Viewing low resolution images limits many tasks, for example, one might have difficulty identifying, resolving or reading notices in front of a building, features or objects carried by someone in an image. Image quality is also limited by a sensors native resolution or other factors caused by under sampling images. Many low cost visible and thermal sensors spatially or electronically under sample an image. Under sampling results in aliased imagery in which the high frequency components are folded into low frequency components in the image, high frequency components is corrupted in these images[4].

Many approaches have been developed to increase image resolution, one of which uses 'micro dither' a special sensor or scanner captures multiple images, where each image is captured by displacing the sensor a known fraction of a pixel. The individual frames are combined into a higher resolution image [4]. An alternative approach employs super resolution image reconstruction, this software technology produces high resolution images using low cost, low resolution video equipment on the basis of a very small number of digital input frames, and resolution can be doubled or even quadrupled. Image quality can thus be improved using sensors without the need for hardware upgrades [4]. Super resolution reconstruction can increase image resolution without changing the design of optics and the detectors, by using a sequence of low resolution images.

Super resolution imaging has proved to be useful in many practical cases where multiple frames of the same scene are obtained. Some of the applications of super resolution imaging are for example in Astronomical imaging where it is possible to obtain different looks of the same scene, in satellite imagery when there is need to obtain higher resolution pictures than what is available, super-resolution imaging is a good choice. Medical imaging is a very important application area for image super-resolution, in medical imaging it is easier to obtain many different "looks" of the same scene with sub-pixel shifts in a controlled manner. CCD size limitations and shot noise prevents obtaining very high resolution images directly [3][5]. Super-resolution reconstruction can be used to get desired high resolution image.

The flow of the topics is as follows, In section II Mathematical Formulation for Super Resolution Model is created, section III presents Wavelet Transforms and Wavelet based Decomposition, section IV presents the proposed super resolution reconstruction of low resolution images, section V provides the proposed Super Resolution Reconstruction Algorithm, in VI simulation results and discussion of super resolution reconstruction is presented, and finally section VII consists of conclusion.

2. MATHEMATICAL FORMULATION FOR THE SUPER RESOLUTION MODEL

In this section we give the mathematical model for super resolution image reconstruction from a set of Low Resolution (LR) images. LR means that the pixel density with an image is less, offering fewer details. The CCD discretizes the images and produces a digitized noisy image. Imaging systems do not sample the scene according to Nyquist criterion. As a result of this the high frequency contents of images are destroyed and appear as LR.

In this section we give the mathematical model for super resolution image reconstruction from a set of Low Resolution (LR) images.

Let us consider the low resolution sensor plane by M_1 by M_2 . The low resolution intensity values are denoted as $\{y(i, j)\}$ where $i = 0 \dots M_1 - 1$ and $j = 0 \dots M_2 - 1$; if the down sampling parameters are

q_1 and q_2 in horizontal and vertical directions, then the high resolution image will be of size $q_1 M_1 \times q_2 M_2$. We assume that $q_1 = q_2 = q$ and therefore the desired high resolution image Z will have intensity values $\{z(k,l)\}$ where $k=0..qM_1-1$ and $l=0..qM_2-1$.

Given $\{z(k,l)\}$ the process of obtaining down sampled LR aliased image $\{y(i, j)\}$ is

$$y(i, j) = \frac{1}{q^2} \sum_{k=qi}^{(q+1)i-1} \sum_{l=qj}^{(q+1)j-1} z(k, l) \tag{1}$$

i.e. the low resolution intensity is the average of high resolution intensities over a neighborhood of q^2 pixels. We formally state the problem by casting it in a Low Resolution restoration frame work. There are P observed images $\{Y_m\}_{m=1}^P$ each of size $M_1 \times M_2$ which are decimated, blurred and noisy versions of a single high resolution image Z of size $N_1 \times N_2$ where $N_1 = qM_1$ and $N_2 = qM_2$.

After incorporating the blur matrix, and noise vector, the image formation model is written as

$$Y_m = H_m D Z + \eta_m \quad \text{Where } m=1..P \tag{2}$$

Here D is the decimation matrix of size $M_1 M_2 \times q^2 M_1 M_2$, H is PSF of size $M_1 M_2 \times M_1 M_2$, η_m is $M_1 M_2 \times 1$ noise vector and P is the number of low resolution observations Stacking P vector equations from different low resolution images into a single matrix vector

$$\begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ y_p \end{bmatrix} = \begin{bmatrix} D H_1 \\ \cdot \\ \cdot \\ D H_p \end{bmatrix} Z + \begin{bmatrix} \eta_1 \\ \cdot \\ \cdot \\ \eta_p \end{bmatrix} \tag{3}$$

The matrix D represents filtering and down sampling process of dimensions $q^2 M_1 M_2 \times 1$ where q is the resolution enhancement factor in both directions. Under separability assumptions, the matrix D which transforms the $qM_1 \times qM_2$ high resolution image to $N_1 \times N_2$ low resolution images where $N_1 = qM_1$, $N_2 = qM_2$ is given by

$$D = D_1 \otimes D_1 \tag{4}$$

Where \otimes represents the kronecker product, and the matrix D_1 represents the one dimensional low pass filtering and down sampling. When $q=2$ the matrix D_1 will be given by

$$D_1 = \frac{1}{2} \begin{bmatrix} 11 & 00 & 00 & \dots & 00 \\ 00 & 11 & 00 & \dots & 00 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 00 & 00 & 11 & & \end{bmatrix} \tag{5}$$

and

$$D = \frac{1}{2^2} \begin{bmatrix} 11 & 00 & 00 & \dots & 00 \\ 00 & 11 & 00 & \dots & 00 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 00 & 00 & & & 11 \end{bmatrix} \tag{6}$$

The square matrix H of dimensions $P_{N_1} \times P_{N_2}$ represents intra channel and inter channel blur operators. i.e. 2D convolution of channel with shift-invariant blurs. The blur matrix is of the form

$$H_I = \begin{bmatrix} H_{(0)} & H_{(1)} & \dots & H_{(M-1)} \\ H_{(M-1)} & H_{(0)} & \dots & H_{(M-2)} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ H_{(1)} & H_{(2)} & \dots & H_{(0)} \end{bmatrix} \tag{7}$$

and it is circulant at the block level. In general each $H_{(i)}$ is an arbitrary $P_{M_1} \times P_{M_2}$, but if shift invariant circular convolution is assumed H(i) becomes

$$H_{(i)} = \begin{bmatrix} H_{(i,0)} & H_{(i,1)} & \dots & H_{(i,M-1)} \\ H_{(i,M-1)} & H_{(i,0)} & \dots & H_{(i,M-2)} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ H_{(i,1)} & H_{(i,2)} & \dots & H_{(i,0)} \end{bmatrix} \tag{8}$$

which is also circulant at the block level $H_{(i,j)}$. Each $P \times P$ sub matrix (sub blocks) has the form.

$$H_{(i,j)} = \begin{bmatrix} H_{11(i,j)} & H_{12(i,j)} & \dots & H_{ip(i,j)} \\ H_{21(i,j)} & H_{22(i,j)} & \dots & H_{2p(i,j)} \\ \vdots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots \\ H_{p1(i,j)} & H_{p2(i,j)} & \dots & H_{pp(i,j)} \end{bmatrix} \tag{9}$$

Where $H_{ii(m)}$ is intra channel blurring operator, $H_{ij(m)}^{i \neq j}$ is an inter channel blur i.e. $P \times P$ non circulant blocks are arranged in a circulant fashion, it's called Block Semi-Block Circulant (BSBC); which can be easily solved using blind deconvolution or regularization.

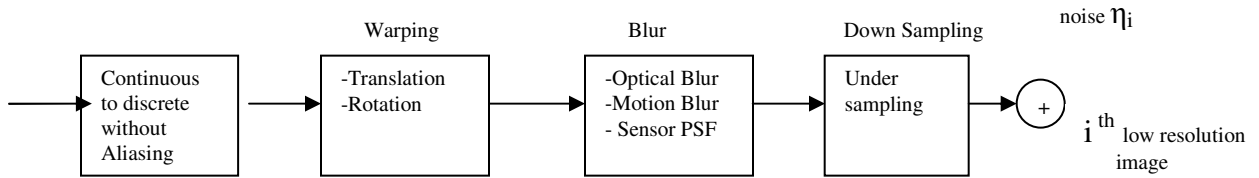


FIGURE1: Low Resolution Observation Model

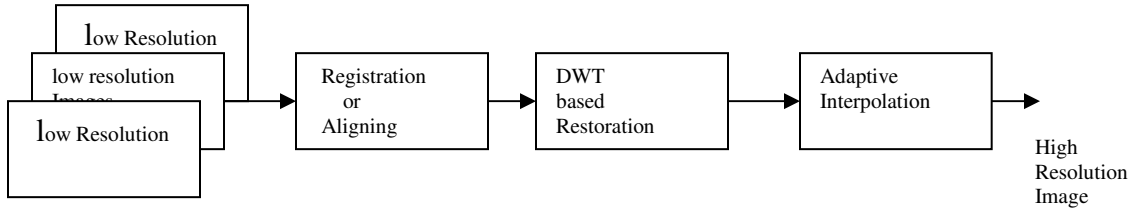


FIGURE 2: Super Resolution Reconstruction Model

3. Wavelet Transforms and Wavelet based Decomposition

Wavelet transforms decomposes a signal into a set of basis functions. Unlike Fourier transforms whose basis functions are sinusoids, wavelet transforms are based on small waves called wavelets of varying frequency and limited duration[6][7].

The Discrete wavelet transform of function $f(x,y)$ of size $M \times N$ is

$$W_{\phi}(j_0, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \phi_{j_0, m, n}(x, y) \quad (10)$$

$$j \geq j_0$$

$$W_{\psi}^i(j, m, n) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x, y) \psi_{j, m, n}^i(x, y) \quad i = \{H, V, D\} \quad (11)$$

j_0 is an arbitrary starting scale and $W_{\phi}(j_0, m, n)$ coefficients define an approximation of $f(x, y)$ at scale j_0 . The $W_{\psi}^i(j, m, n)$ coefficients add horizontal, vertical and diagonal details for scales $j \geq j_0$. Given the W_{ϕ} , W_{ψ}^i of Eq.10 and Eq.11, $f(x,y)$ is obtained via the inverse DWT.

$$f(x, y) = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} W_{\phi}(j_0, m, n) \phi_{j_0, m, n}(x, y) + \frac{1}{\sqrt{MN}} \sum_{i=H, V, D} \sum_{j=0}^{\infty} W_{\psi}^i(j, m, n) \psi_{j, m, n}^i(x, y) \quad (12)$$

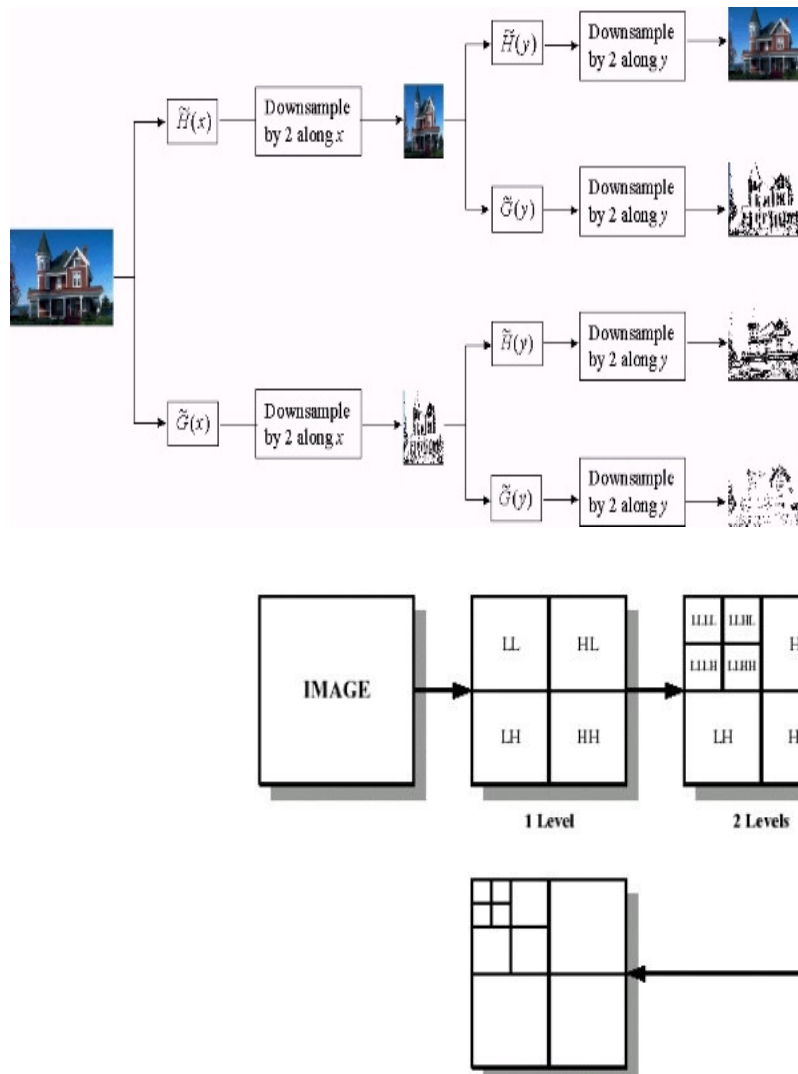


FIGURE 3: Discrete Wavelet Transformed down sampled image

The discrete wavelet transform uses low-pass and high-pass filters, $H(x)$ and $G(x)$, to expand a digital signal. They are referred to as analysis filters. These filters correspond to $\varphi(t)$ and $\psi(t)$ of Eq. 10. and Eq.11. The dilation performed for each scale is now achieved by a decimator. The coefficients c_k and d_k are produced by convolving the digital signal, with each filter, and then decimating the output. The c_k coefficients are produced by the low-pass filter, $H(x)$, and called coarse or approximate coefficients. The d_k coefficients are produced by the high-pass filter and called detail coefficients. Coarse coefficients provide information about low frequencies, and detail coefficients provide information about high frequencies. Coarse and detail coefficients are produced at multiple scales by iterating the process on the coarse coefficients of each scale. The entire process is computed using a tree-structured filter bank as shown in Fig.3.

The choice of filter depends on the type of applications we are using biorthogonal filter based wavelet transform. In biorthogonal case, there are two scaling functions which may generate different multiresolution analysis and accordingly two different wavelet functions.

4. Proposed Super Resolution Reconstruction of Low Resolution Images

The Low Resolution observation model of Eq. 2 is considered. The undersampled low resolution images are captured by natural jitter or some kind of controlled motion of the camera. Our proposed super resolution algorithms work in three phases: registration, restoration and interpolation.

4.1 Image Registration

For super resolution reconstruction, image registration is performed first in order to align the LR images as accurately as possible. It is a process of overlaying two or more images of the same scene taken at different times, from different view points and or by different sensors. Typically one image called the base image is considered the reference to which the other images called input images are compared. The objective is to bring the input image into alignment with the base image by applying a spatial transformation to the input image. Spatial transformation maps locations in one image into a new location in another image. Image registration is an inverse problem as it tries to estimate from sampled images Y_m , the transformation that occurred between the views Z_m considering the observation model of Eq 2. It is also dependent on the properties of the camera used for image acquisition like sampling rate (or resolution) of sensor, the imperfection of the lens that adds blur, and the noise of the device. As the resolution decreases, the local two dimensional structure of an image degrades and an exact registration of two low resolution images becomes increasingly difficult. Super resolution reconstruction requires a registration of high quality.

Different methods exist for estimating the motion or sub pixel shift between the two images[8]. The registration technique considered in our research is based on Fast Fourier Transform proposed by Fourier Mellin and DeCastro [9].

The transformation considered in our research is rotation, translation and shift estimation. Let us consider the translation estimation, the Fourier transform of the function is denoted by $F\{f(x, y)\}$ or $\hat{f}(w_x, w_y)$. The shift property of the Fourier transform is given by

$$F\{f(x+\Delta x, y+\Delta y)\} = \hat{f}(w_x, w_y) e^{i(w_x \Delta x + w_y \Delta y)} \quad (13)$$

Eq. 13 is the basis of the Fourier based translation estimation algorithms. Let $I_1(x, y)$ be the reference image and $I_2(x, y)$ is the translated version of the base image, i.e.

$$I_1(x, y) = I_2(x + \Delta x, y + \Delta y) \quad (14)$$

By applying the Fourier transform on both the sides of Eq. (14). We get

$$\hat{I}_1(w_x, w_y) = \hat{I}_2(w_x, w_y) e^{i(w_x \Delta x + w_y \Delta y)} \quad (15)$$

or equivalently,

$$\frac{\hat{I}_1(w_x, w_y)}{\hat{I}_2(w_x, w_y)} = e^{i(w_x \Delta x + w_y \Delta y)} \quad (16)$$

$$\text{corr}(x, y) \cong F^{-1} \left(\frac{\hat{I}_1(w_x, w_y)}{\hat{I}_2(w_x, w_y)} \right) = \delta(x + \Delta x, y + \Delta y) \quad (17)$$

For discrete images we replace the FT in the computation above with FFT, and $\delta(x + \Delta x, y + \Delta y)$ is replaced by a function that has dominant maximum at $(\Delta x, \Delta y)$ as

$$(\Delta x, \Delta y) = \arg \max \{ \text{corr}(x, y) \} \quad (18)$$

Calculate the cross power spectrum by taking the complex conjugate of the second result. Multiplying the FT together element wise, and normalizing this product element wise.

$$corr(w_x, w_y) \cong \frac{\hat{I}_1(w_x, w_y)}{\hat{I}_2(w_x, w_y)} \bullet \left| \frac{\hat{I}_1(w_x, w_y)}{\hat{I}_2(w_x, w_y)} \right| \tag{19}$$

$$corr(w_x, w_y) = R \cong \frac{\hat{I}_1(w_x, w_y) \hat{I}_2^*(w_x, w_y)}{\left| \hat{I}_2(w_x, w_y) \right| \left| \hat{I}_1(w_x, w_y) \right|} = e^{i(w_x \Delta x + w_y \Delta y)} \tag{20}$$

where * denotes the complex conjugate. Obtain the normalized cross correlation by applying the inverse FT. i.e. $r = F^{-1}\{R\}$, determine the location of the peak in r. This location of the peak is exactly the displacement needed to register the images

$$(\Delta x, \Delta y) = \arg \max \{r\} \tag{21}$$

The angle of rotation is estimated by converting the Cartesian coordinates to log polar form. We observe that the sum of a cosine wave and a sine wave of the same frequency are equal to phase shifted cosine wave of the same frequency. That is if a function is written in Cartesian form as

$$v(t) = A \cos(t) + B \sin(t) \tag{22}$$

Then it may also be written in polar form as

$$v(t) = c \cos(t - \phi) \tag{23}$$

We may write the Eq (23) in polar form as

$$Y = y(x) = \frac{a_0}{2} + \sum_{k=1}^N m_k \cos(2\pi f_k x - \phi_k) \tag{24}$$

Where

$$m_k = \sqrt{a_k^2 + b_k^2} \dots (\text{magnitude}) \tag{25}$$

$$\phi_k = \tan^{-1} \left(\frac{b_k}{a_k} \right) \dots (\text{Phase})$$

The Shift is estimated by finding cross power spectrum and computing Eq. (20). We obtain the normalized cross correlation by applying the inverse FT. i.e. $r = F^{-1}\{R_2\}$, determines the location of the peak in r. This location of the peak is exactly the shift $I(x_0, y_0)$ needed to register the images. Once we estimated the angle of rotation and translation and shift, a new image is constructed by reversing the angle of rotation, translation and shift.



FIGURE 4: Registered Image

In the second phase the low resolution registered images are fused using fusion rule, then restoration is performed using our proposed novel denoising technique. In the final phase proposed adaptive interpolation is performed to obtain an image with double, quadrupled the resolution of the original.

4.2 Wavelet based Restoration

Images are obtained in areas ranging from everyday photography to astronomy, remote sensing, medical imaging and microscopy. In each case there is an underlying object or scene we wish to observe, the original or the true image is the ideal representation of the observed scene. Yet the observation process is never perfect, there is uncertainty in the measurement occurring as blur, noise and other degradations in the recorded images. Image restoration aims to recover an estimate of the original image from the degraded observations. Classical image restoration seeks an estimate of the true image assuming the blur is known, whereas blind image restoration tackles the much more difficult but realistic problem where the degradations are unknown.

The low resolution observation model of Eq.2 is considered. We formally state by casting the problem in multi channel restoration format, the blur is considered as between channels and within channel of the low resolution images. In order to remove the blur and noise from the LR images, we have proposed an efficient wavelet based denoising using thresholding, our proposed approach performs much better when compared to other approaches.

4.2.1. Proposed Efficient wavelet based Denoising for SRR using Thresholding

Image denoising techniques are necessary to remove random additive noises while retaining as much as possible the important image features. The main objective of these types of random noise removal is to suppress the noise while preserving the original image details [6]. Statistical filter like average filter, wiener filter can be used for removing such noises but wavelet based denoising techniques proved better results than these filters. The wavelet transforms compresses the essential information in an image into a relatively few, large coefficients which represents image details at different resolution scales. In recent years there has been a fair amount of research on wavelet thresholding and threshold selection for image denoising [10] [11] [12].

Let Z be an $M \times M$ image from Eq. (2), during transmission the image Z is corrupted by zero mean white Gaussian noise η with standard deviation σ . At the receiver end the noisy observation Y of Eq. (2) is obtained. The goal is to obtain the image Z from noisy observation Y such that the MSE is minimum. Wavelet Transforms decomposes the image into different frequency subbands. Small coefficients in the subbands are dominated by noise while coefficients with large absolute value carry more image information than noise. Replacing noisy coefficients by zeros and an inverse wavelet transform may lead to reconstruction that has lesser noise. Normally hard thresholding and soft thresholding techniques are used for denoising.

Hard Thresholding

$$D(X, T) = \begin{cases} X & \text{if } |X| > T \\ 0 & \text{if } |X| < T \end{cases} \quad (26)$$

Soft Thresholding

$$D(X, T) = \text{Sign}(X) * \max(0, |X| - T) \quad (27)$$

Where X is the input subband, D is the denoised band after thresholding and T is the threshold level. The denoising algorithms which are based on thresholding suggests, that each coefficient of every detail subband is compared to threshold level and is either retained or killed if its magnitude is greater or less respectively.

In wavelet decomposition of an image we obtain one approximate (LL) and three details (LH, HL and HH) subbands. The approximate coefficients are not submitted in this process. Since on one

hand they carry the most important information about the image, on the other hand the noise mostly affects the high frequency subbands. Hence the HH subband contains mainly noise. For estimating the noise level we use the Median Absolute Deviation (MAD) as proposed by Donoho [13].

$$\sigma = \frac{\text{Median} |Y_{ij}|}{0.6745}, Y_{ij} \in LH, HL, HH \tag{28}$$

The factor 0.6745 in the denominator rescales the numerator, so that σ is also a suitable estimator for standard deviation for Gaussian white noise.

Selecting an optimum threshold value (T) for soft thresholding is not an easy task. An optimum threshold value should be selected based on the subband characteristics. In wavelet subbands as the level increases, the coefficients of the subband become smoother. For example when an image is decomposed into 2 level DWT using Daubechies 4 tap or bio-orthogonal wavelet transform, we get 8 subbands Fig.3, the HH subband of first level contains large amount of noise, hence the noise level is estimated for the HH subband of level1 using equation (28).Once the noise level is estimated, we select the threshold value T. The Threshold value T is

$$T = \sigma - (|HM - GM|) \tag{29}$$

Here σ is the noise variance of the corrupted image. As given in [6], the Harmonic mean and geometric mean are best suited for the removal of Gaussian noise, hence we use the absolute difference of both the Harmonic Mean (HM) and Geometric Mean (GM) or either of the means can also be considered for denoising the image corrupted by Gaussian noise. The harmonic mean filter is better at removing Gaussian type noise and preserving edge features than the arithmetic mean filter. Hence we have considered harmonic mean than arithmetic mean. The process is repeated for LH and HL bands and threshold is selected for all the three bands once threshold is estimated, soft thresholding of Eq.27, is performed to denoise the image.

$$HM = \frac{M^2}{\sum_{i=1}^M \sum_{j=1}^M \frac{1}{g(i,j)}} \tag{30}$$

$$GM = \left[\prod_{i=1}^M \prod_{j=1}^M g(i,j) \right]^{\frac{1}{M^2}} \tag{31}$$

4.3. Proposed Adaptive Interpolation for Super Resolution reconstruction

The algorithm works in four phases: In the first phase the wavelet based fused image is expanded. Suppose the size of the fused image is n x m. The image will be expanded to size N (2n-1) x (2m-1). In the Fig.5 solid circles show original pixels and hallow circles show undefined pixels. In the remaining three phases these undefined pixels will be filled

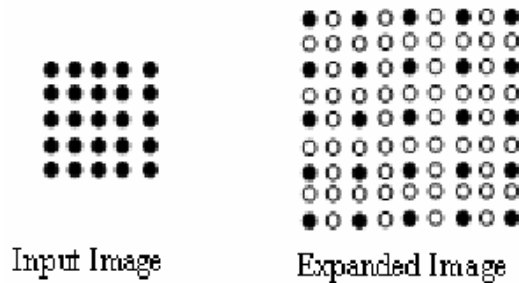


FIGURE 5 : High Resolution Grid

The second phase of the algorithm is most important one. In this phase the interpolator assigns value to the undefined pixels

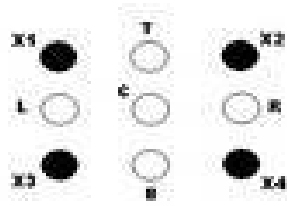


FIGURE 6: HR unit cell with undefined pixels Top, Center, Bottom, Left, Right denoted by T,C,B,L,R.

The undefined pixels are filled by following mutual exclusive condition.

Uniformity: select the range (X_1, X_2, X_3, X_4) and a Threshold T.

if range $(X_1, X_2, X_3, X_4) < T$ Then

$$C = \frac{(X_1 + X_2 + X_3 + X_4)}{4} \tag{32}$$

if there is edge in NW-SE Then

$$C = (X_1 + X_2) / 2$$

if there is edge in NS Then

$$T = (X_1 + X_2) / 2 \text{ and } B = (X_3 + X_4) / 2$$

if there is edge in EW Then

$$L = (X_1 + X_3) / 2 \text{ and } R = (X_2 + X_4) / 2 \tag{33}$$

In this phase, approximately 85% of the undefined pixels of HR image are filled. In the third phase the algorithm scans the magnified image line by line, looking for those pixels which are left undefined in the previous phase.

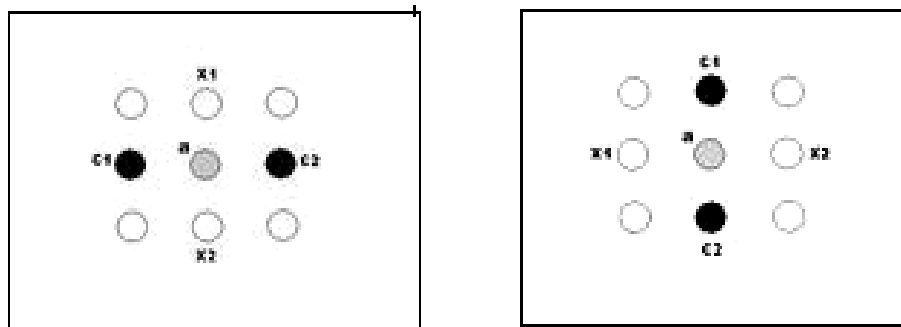


FIGURE 7 : Layout referred in the phase 3

In the third phase ,the algorithm checks for the layout as shown in Figure(7).

$$\begin{aligned} & \text{if there is edge } c_1 c_2 \text{ then } a = (c_1 + c_2) / 2 \\ & \text{else if there is edge } X_1 X_2 \text{ then } a = (X_1 + X_2) / 2 \end{aligned} \quad (34)$$

In the fourth phase all the undefined pixels will be filled. If there are any undefined pixels left then the median of the neighboring pixels is calculated and assigned. We call our interpolation method adaptive as the interpolator selects and assigns the values for the undefined pixels based on mutual exclusive condition.

5. Proposed Algorithm for Super Resolution Reconstruction of Low Resolution Images.

Our proposed novel super resolution reconstruction consists of following consecutive steps:

Step 1: Three input low resolution blurred, noisy, under sampled, rotated, shifted images are considered.

$$\text{i.e. } I_1(i, j), I_2(i, j), I_3(i, j) \text{ where } i=1 \dots N, j=1 \dots N \quad (35)$$

Step 2: The images are first preprocessed, i.e. registered using FFT based algorithm, as explained in section(4.1).

Step 3: The registered low resolution images are decomposed using DWT to a specified number of levels. At each level we will have one approximation i.e. LL sub band and 3 detail sub bands, i.e. LH, HL, HH coefficients.

Step 4: The decomposed images are fused using the fusion rule i.e. Maximum Frequency Fusion: "fusion by averaging for each band of decomposition and for each channel the wavelets coefficients of the three images is averaged". That is maximum frequencies of approximate and detail coefficients are fused separately

$$A_j^o I = \max(A_j^o I_1 + A_j^o I_2 + A_j^o I_3) / 3 \quad D_j^d I = \max(D_j^d I_1 + D_j^d I_2 + D_j^d I_3) / 3 \quad (36)$$

Step 5: The fused image contains LL, LH, HL and HH subbands.

- a) Obtain the noise variance (σ) using Eq.(28) for LH, HL and HH subbands of level one.
- b) Compute Eq. (29) and select the threshold (T) for LH, HL and HH subbands of level 1.
- c) Denoise all the detail subband coefficients of level one (except LL) using soft thresholding given in Eq. (27) by substituting the threshold value obtained in step (5b).

Step 6: Most of the additive noise will be eliminated during the fusion process by denoising using our proposed soft thresholding approach as explained in section (4.2.1), where as the image is deblurred using Iterative Blind Deconvolution Algorithm (IBD)[14].

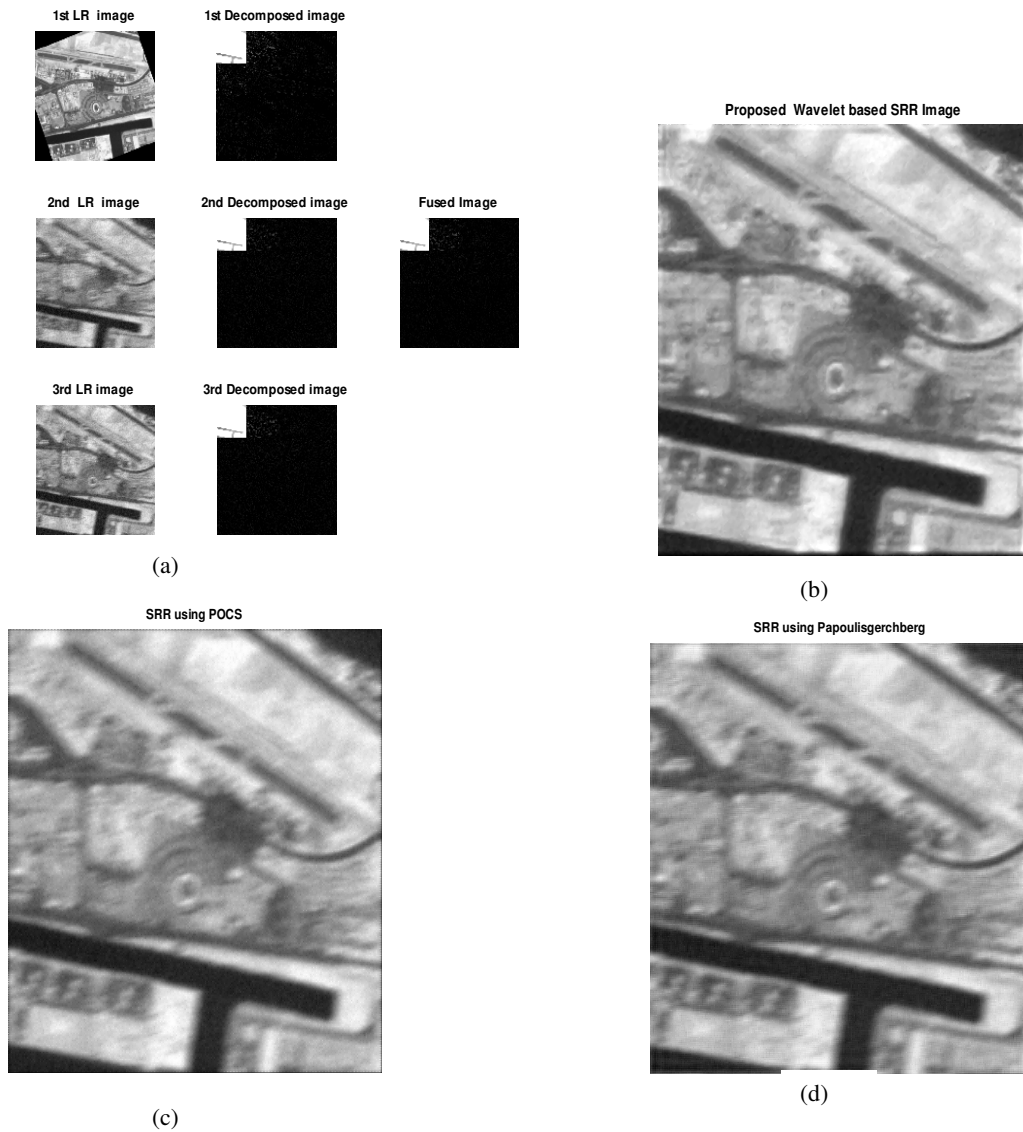
Step 8: Inverse DWT is applied to obtain a high resolution restored image.

Step 9: Finally in order to obtain a super resolved image, an image with double the resolution as that of the original image, our proposed adaptive interpolation as explained in section (4.3) is applied.

6. Simulation Results and Discussions

In computer simulations, three LR (256x 256) images are considered. We have tested using both motion blur with an angle of (10, 20 and 30) and Gaussian blur of 3x3, 5x5 and 7x7 is considered. Gaussian white noise with SNR (5, 10,15 and 20 dB) is added to the blurred low resolution images. A Super Resolution image of size 512 x512 and 1024x1024 is reconstructed from three noisy, blurred, under sampled, mis-registered images

Case 1: Abudabi SAR grayscale Low resolution Images with a Super resolution factor of 2 (enlarged from 256x256 to 512 by 512)



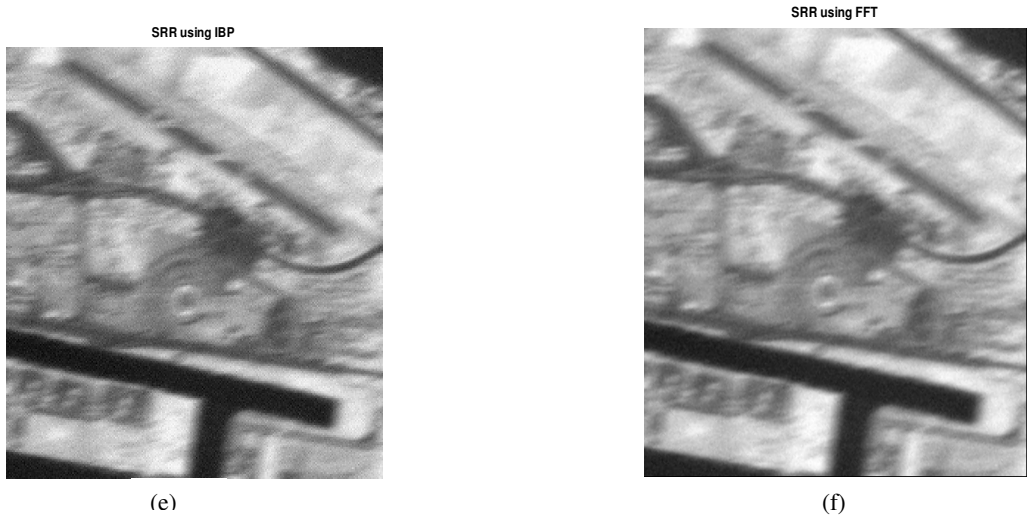
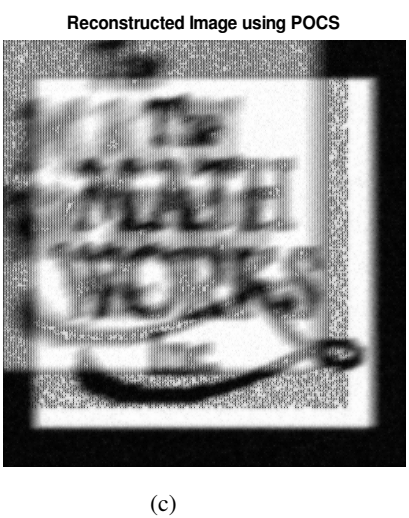
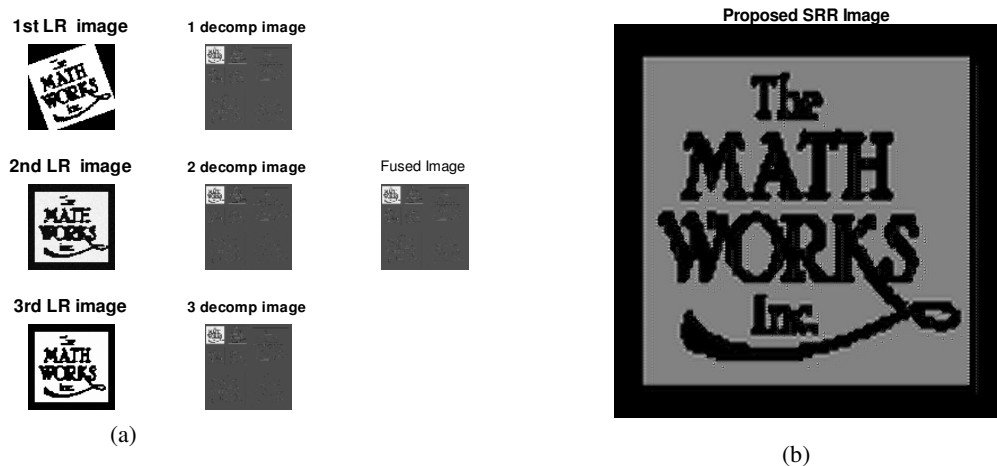


FIGURE 8: (a) Low Resolution Images (b) Proposed wavelet SRR (c) Projection on to Convex Sets(POCS) SRR (d) Papoulis Gerchberg SRR (e) Iterative Back Propagation (IBP) SRR (f) FFT based SRR
 Case 2: Logo grayscale Low resolution Images with a Super resolution factor of 4 (enlarged from 128x128 to 512 by 512)



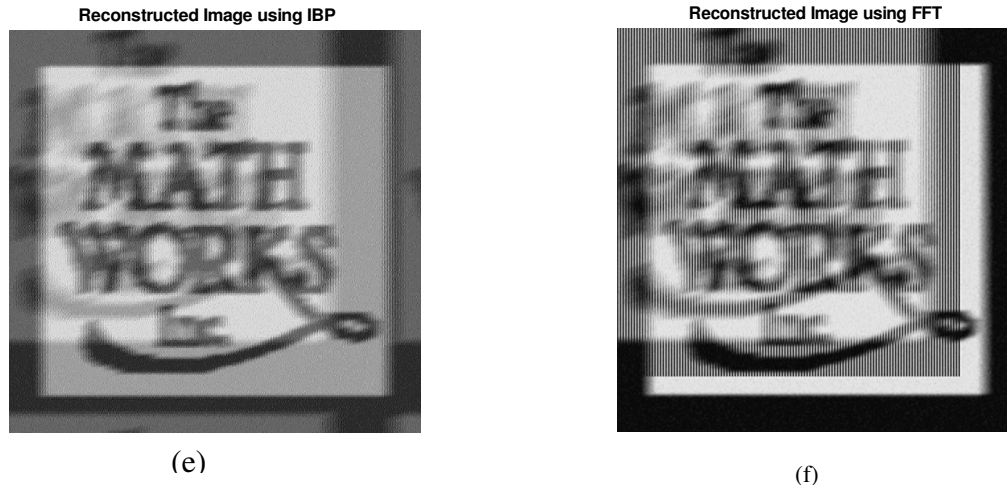


FIGURE 9: (a) Low Resolution Images (b) Proposed wavelet SRR (c) Projection on to Convex Sets (POCS) SRR (d) Papoulis Gerchberg SRR (e) Iterative Back Propagation (IBP) SRR (f) FFT based SRR

Simulations are carried out to verify noise and blur removing capability of the proposed super resolution reconstruction and the results are compared with several existing state of art techniques. A quantitative comparison is performed between proposed wavelet based super reconstruction and several existing techniques in terms of Improvement in Signal to Noise Ratio (ISNR), Peak Signal to Noise ratio (PSNR), Mean Square Error (MSE), Super Resolution Factor (SRF) and Mean Structural Similarity (MSSIM) Index. Our method produced results are superior to other methods in both visual image quality and quantitative measure. Simulations were made on several gray scale and color images corrupted with Gaussian noise and motion blur.

The performance of the algorithm for various images at different blur and noise levels is studied and the results for two cases are shown in Fig.8. and Fig 9. The quantities for comparison are defined as follows and Table I and II display the quantitative measures.

1) Improvement in Signal-to-Noise Ratio (ISNR)

For the purpose of objectively testing the performance of the restored image, Improvement in signal to noise ratio (ISNR) is used as the criteria which is defined by

$$ISNR = 10 \log_{10} \frac{\sum_{i,j} [f(i, j) - y(i, j)]^2}{\sum_{i,j} [f(i, j) - g(i, j)]^2} \tag{37}$$

Where j and i are the total number of pixels in the horizontal and vertical dimensions of the image; f(i, j), y(i, j) and g(i, j) are the original, degraded and the restored image.

2) The MSE and PSNR of the reconstructed image is

$$MSE = \frac{\sum [f(i, j) - F(I, J)]^2}{N^2} \tag{38}$$

Where f(i, j) is the source image F(I, J) is the reconstructed image, which contains N x N pixels

$$PSNR = 20 \log_{10} \left(\frac{255}{RMSE} \right) \quad (39)$$

3) Super Resolution Factor

$$SRF = \frac{\sum_{i=1}^M \sum_{j=1}^N (F(i, j) - f(i, j))^2}{\sum_{i=1}^M \sum_{j=1}^N (y(i, j) - f(i, j))^2} \quad (40)$$

4) MSSIM

The structural similarity (SSIM) index is defined in [29] by equations

$$SSIM(f, F) = \frac{(2\mu_f \mu_F + C_1)(2\sigma_f + C_2)}{(\mu_f^2 + \mu_F^2 + C_1)(\sigma_f^2 + \sigma_F^2 + C_2)} \quad (41)$$

$$MSSIM(f, F) = \frac{1}{G} \sum_{p=1}^G SSIM(f, F) \quad (42)$$

The Structural SIMilarity index between the original image and reconstructed image is given by SSIM, where μ_f and μ_F are mean intensities of original and reconstructed images, σ_f and σ_F are standard deviations of original and reconstructed images, f and F are image contents of p th local window and G is the number of local windows in the image.

The simulation results show that our approach provides a visually appealing output. The proposed wavelet based super resolution reconstruction with efficient denoising can reconstruct a super resolution image from a series of blurred, noisy, aliased and downsampled low resolution images, which is demonstrated by two cases as show in Fig. 8 and Fig.9.

In case 1, Fig 8, three LR SAR images of Abu Dhabi stadium are considered which are corrupted by Gaussian noise of standard deviation ($\sigma = 10, 15$ & 20) and motion blur of angle ($10, 20$ & 30). The results of the proposed SRR are compared with Projection on to Convex Sets (POCS)[21], Papoulis Gerchberg algorithm [23] [24], Iterative Back Propagation (IBP) [22] and FFT based SRR [26] [27].

Table I and II gives the quantitative measures of abudhabi.gif, logo.tif, lean.jpg and baboob.png images. Fig10, a, b, c & d shows the comparison graph of various quantitative measures for the LR test images. When compared with standard state of art techniques, it is clear that the proposed wavelet based SRR with efficient denoising and adaptive interpolation can eliminate the noise and blur while preserving the edges and fine details, and can reconstruct a super resolution image with a super resolution factor of 2, 4 times than the original image.

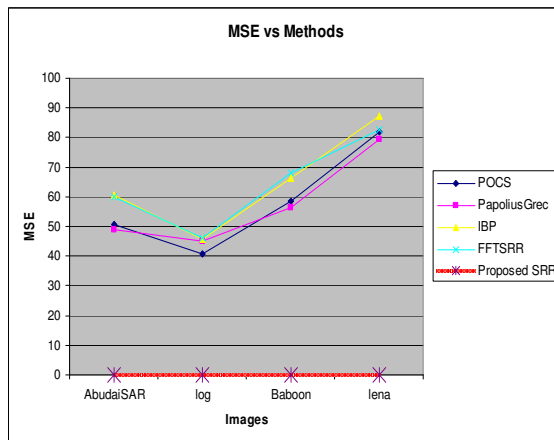
The results show that fig 8b and 9b has more sharp edges and less noise when compared to state of art techniques.

Table I: MSE and PSNR comparison of our proposed approach with other approaches

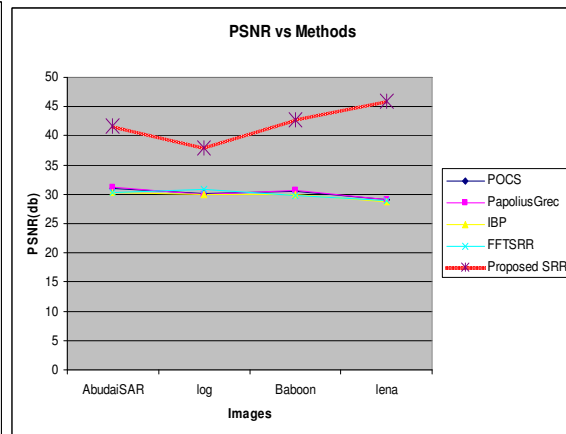
Source Image	POCS		Paploius Greberg		IBP		FFT SRR		Proposed Wavelet SRR	
	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR
AbudaiSAR	50.62	31.08	48.9	31.23	60.51	30.31	59.83	30.36	0.002	41.74
log	40.63	30.06	45.1	30.04	45.78	29.95	46.08	30.86	0.035	37.87
Baboon	58.64	30.44	56.28	30.62	66.46	29.9	68.04	29.8	0.086	42.8
lena	82.015	28.99	79.52	29.12	87.21	28.72	82.48	28.96	0.061	45.95

Table II: ISNR and MSSIM comparison of our proposed approach with other approaches

Source Image	POCS		Paploius Greberg		IBP		FFT SRR		Proposed Wavelet SRR	
	ISNR	MSSIM	ISNR	MSSIM	ISNR	MSSIM	ISNR	MSSIM	ISNR	MSSIM
AbudaiSAR	3.392	0.6783	3.96	0.625	3.35	0.761	3.39	0.68	6.49	0.833
log	3.58	0.723	4.61	0.76	3.69	0.79	3.94	0.73	5.05	0.871
Baboon	2.8	0.671	3.09	0.663	2.13	0.711	2.51	0.702	6.87	0.778
lena	3.58	0.683	3.43	0.617	3.04	0.786	4.08	0.711	6.01	0.858



(a)



(b)

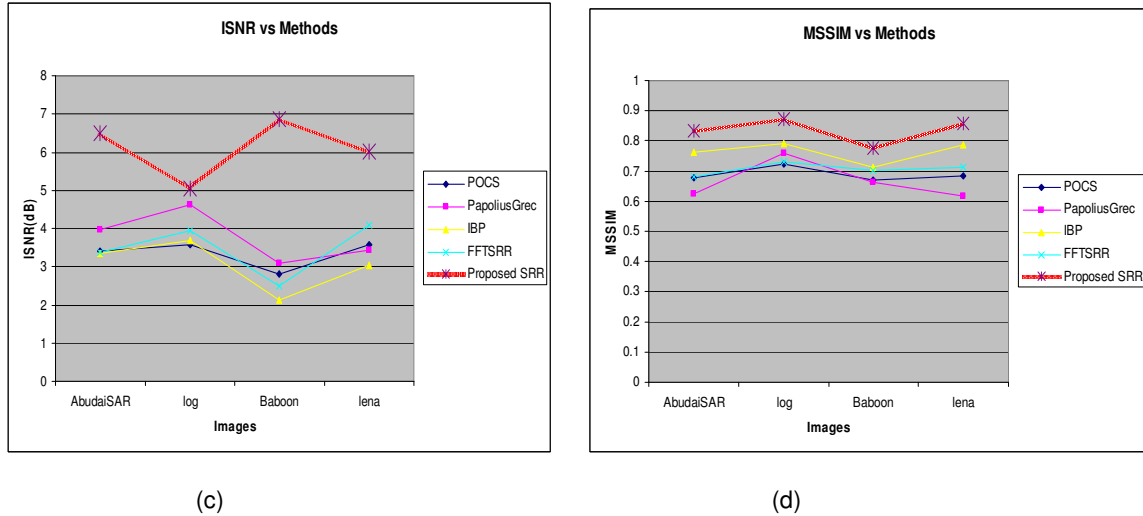


Figure 10: (a) Comparison graph of MSE at different Blur and noise densities of LR images (b) Comparison graph of PSNR at different Blur and noise densities of LR images. (c) Comparison graph of ISNR at different Blur and noise densities of LR images (d) Comparison graph of MSSIM at different Blur and noise densities of LR images

7. CONCLUSION

In this paper we have proposed a novel and robust wavelet based super resolution reconstruction of low resolution images using efficient denoising and adaptive interpolation. Wavelet transforms are suitable for achieving a high super resolution reconstruction image. The proposed wavelet based denoising is based on threshold estimation and analysis of statistical parameters like harmonic mean, geometric mean and standard deviation of the detail sub band coefficients. Harmonic mean filter performs better when compared to arithmetic mean filters and preserve the edges which are most significant for super resolution reconstruction. Experiments are conducted on different natural images corrupted by various noise levels to access the performance of the proposed thresholding method in comparison with other methods. The advantage of using wavelets and adaptive interpolation for super resolution reconstruction is blocking artifacts are reduced and we are able to obtain very high PSNR and ISNR value when compared to the state of art super resolution reconstruction. With our proposed Adaptive interpolation technique we are able to obtain a very high super resolution image with super resolution factor of 2, 4 times than that of the original image. Experimental results show that our proposed method performs quite well in terms of robustness and efficiency.

REFERENCES

- [1] Liyakathunisa and C.N.Ravi Kumar "Advances in Super Resolution Reconstruction of Low Resolution Images" International Journal of Computational Intelligence Research ISSN 0973-1873 Volume 6, Number 2 (2010), pp. 215–236.
- [2] Liyakathunisa, C.N.Ravi Kumar and V.K. Ananthashayana , "Super Resolution Reconstruction of Low Resolution Images using Wavelet Lifting schemes" in Proc ICCEE'09, " 2nd International Conference on Electrical & Computer Engineering", Dec 28-30TH 2009, Dubai, Indexed in IEEE Xplore.
- [3] S. Chaudhuri, Ed., Super-Resolution Imaging. Norwell, A: Kulwer, 2001.
- [4] S.Susan Young, Ronal G.Diggers, Eddie L.Jacobs, "Signal Processing and performance Analysis for imaging Systems", ARTEC HOUSE, INC 2008
- [5] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical review," IEEE Signal Processing Mag., vol. 20, pp. 21–36, May 2003.
- [6] Gonzalez Woods, "Digital Image Processing", 2nd Edition.
- [7] A. Jensen, A.Ja Cour-Hardo, "Ripples in Mathematics" Springer publications.
- [8] Barbara Zitova, J.Flusser, "Image Registration: Survey", Image and vision computing, 21, Elsevier publications, 2003.

- [9] E.D. Castro, C. Morandi, Registration of translated and rotated images using finite Fourier transform, IEEE Transactions on Pattern Analysis and Machine Intelligence (1987) 700–703.
- [10] S. Grace Chang, Bin Yu and M. Vattereli. (2000). Adaptive Wavelet Thresholding for Image denoising and compression IEEE Transaction, Image Processing, vol. 9, pp. 1532-15460.
- [11] S.K.Mohiden, Perumal, Satik, "Image Denoising using DWT", IJCSNS, Vol 8, No 1, 2008.
- [12] D. Gnanadurai, V.Sadsivam, "An Efficient Adaptive Thresholding Technique for Wavelet based Image Denoising" IJSP, Vol 2, spring 2006.
- [13] D.L. Donoho and I.M John Stone (1995), Adapting to unknown smoothness via wavelet shrinkage, Journal of American Association, Vol 90, no, 432, pp1200-1224 .
- [14] Grayer and J.C.Danity, "Iterative Blind Deconvolution", July 1988, vol13, No 7, optics letters
- [15] G. Nason, "Choice of the threshold parameter in wavelet function estimation," in Wavelets in Statistics, A. Antoniadis and G. Oppenheim, Eds. Berlin, Germany: Springer-Verlag, 1995.
- [16] Le Moigne J. and Crop R. F., 1996, "The use of wavelets for remote sensing image registration and fusion". Technical Report TR-96-171, NASA Goddard Space Flight Center.
- [17] I. M. Johnstone and B.W. Silverman, "Wavelet threshold estimators for data with correlated noise," J. R. Statist. Soc., vol. 59, 1997.
- [18] D. L. Donoho, "De-noising by soft-thresholding," IEEE Trans. Inform. Theory, vol. 41, pp. 613–627, May 1995
- [19] H. H. Wang, "A new multi wavelet-based approach to image fusion", Journal of Mathematical Imaging and Vision, vol.21, pp.177-192, Sep 2004
- [20] S. Borman and R.L. Stevenson, "Super-resolution from Image sequences -A Review", in Proc. 1998 Midwest Symp. Circuits and Systems, 1999, pp.374-378.
- [21] A.M. Tekalp, M.K. Ozkan, and M.I. Sezan, "High-resolution image reconstruction from lower-resolution image sequences and space varying image restoration," in Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing (ICASSP), San Francisco, CA., vol. 3, Mar. 1992, pp. 169-172
- [22] M. Irani and S. Peleg, "Improving resolution by image registration," CVGIP: Graphical Models and Image Proc., vol. 53, pp. 231-239, May 1991.
- [23] Papoulis, A. (1975) A new algorithm in spectral analysis and band-limited extrapolation. IEEE Transactions on Circuits and Systems, CAS-22, 735–742.
- [24] Gerchberg, R. (1974) Super-resolution through error energy reduction. Optical Acta, 21, 709–720.
- [25] Priyam chatterjee, Sujata Mukherjee, Subasis Chaudhuri and Guna Setharaman," Application of Papoulis – Gerchberg Method in Image Super Resolution and inpainting," The Computer Journal Vol 00 No 0, 2007.
- [26] P. Vandewalle, S. Susstrunk, and M. Vetterli, Lcav," A frequency domain approach to registration of aliased images with application to Super resolution", EURASIP Journal on applied signal processing pp 1-14, 2006.
- [27] Deepesh Jain, " Super Resolution Reconstruction using Papoulis-Gerchberg Algorithm" EE392J – Digital Video Processing, Stanford University, Stanford, CA
- [28] F. Sroubek, G.Cristobal and J.Flusser, "Simultaneous Super Resolution and Blind deconvolution ", Journal of Physics: Conference series 124(2008) 01204
- [29] Freeman, W. T., Jones, T. R., and Pasztor, E. C. Example-based super-resolution, IEEE Computer Graphics and Applications, 22, 56–65, 2000.
- [30] S.Baker and T.Kanade, "Limits on Super Resolution and How to Break Them ", Proc. IEEE conf. Computer Vision and Pattern Recognition, 2000.
- [31] S. Lertrattanapanich and N. K. Bose, "High Resolution Image Formation from Low Resolution Frames Using Delaunay Triangulation", IEEE Trans on image processing vol. 11, no. 12, December 2002.
- [32] R.Y. Tsai and T.S. Huang, "Multiple frame image restoration and registration," in Advances in Computer Vision and Image Processing. Greenwich, CT: AI Press Inc., 1984, pp. 317-339.
- [33] Wang, Bovik, Sheikh, et al. "Image Quality Assessment: From Error Visibility to Structural Similarity," IEEE Transactions of Image Processing, vol. 13, pp. 1-12, April 2004.
- [34] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," IEEE Trans. Pattern Anal. Machine Intell, vol. 11, pp. 674–693, July 1989.
- [35] X. Li and M. T. Orchard, "New Edge-Directed Interpolation," IEEE Trans. on Image Processing, vol. 10, no. 10, October 2001.
- [36] J.W. Hwang and H.S. Lee, "Adaptive Image Interpolation Based on Local Gradient Features," IEEE Signal Processing Letters, vol.11, no.3, March 2004.
- [37] S. Carrato and L. Tenze, "A High Quality 2x Image Interpolator," IEEE Signal Processing Letter, vol. 7, no. 6, June 2000.
- [38] T. Acharya, P.S. Tsai, "Image up-sampling using Discrete Wavelet Transform," in Proceedings of the 7th International Conference on Computer Vision, Pattern Recognition and Image Processing (CVPRIP 2006), in conjunction with 9th Joint Conference on Information Sciences (JCIS 2006), October 8-11, 2006, Kaohsiung, Taiwan, ROC, pp. 1078-1081.

- [39]Tight Frame: An Efficient way for High Resolution Image Reconstruction by Raymond H.Chan. ELSEVIER Publications Feb 2004.
- [40]On Ambiguities in Super Resolution Modeling by Zhao Hong Wang and Feihu Qi-IEEE Signal Processing Letters Vol11, Nov 8, Aug 2008.
- [41] A High efficiency Super Resolution Reconstruction Algorithm from Image/Video Sequences by Chaunghai Xia.
- [42]Performance Evaluation of Super Resolution Reconstruction Methods on real World Data by A.W.M.Van Eekern, K.Schuttle, O.R.Oudegeest and L.J.Van Vliet –EURASIP Journal on Advances in Signal Processing Vol2007
- [43] High Resolution Images from low resolution compressed video by C.Andrew Segall, Rafael Molina and Aggelos K.Katsaggelo –IEEE Signal Processing Magazine May 2003

Separation of mixed Document Images in Farsi Scanned Documents Using Blind Source Separation

Hossein Ghanbarloo

*Department of Electrical Engineering,
Faculty of Engineering, Islamic Azad
University, Science and Research
Branch, Tehran, Iran*

hosseinghanbarloo@gmail.com

Farbod Razzazi

*Department of Electrical Engineering,
Faculty of Engineering, Islamic Azad
University, Science and Research
Branch, Tehran, Iran*

Razzazi@srbiau.ac.ir

Shahpoor Alirezaee

*Department of Electrical Engineering,
Faculty of Engineering, Zanjan
University, Zanjan, Iran*

alirezaee@znu.ac.ir

Abstract

In the field of mixed scanned documents separation, various studies have been carried out to reduce one (or more) unwanted artifacts from the document. Most of the approaches are based on comparison of the front and back sides of the documents. In some cases, it has been proposed to analyze the colored images; however, because of the calculation complexity of the approaches, they are not well applicable in practical applications. Furthermore, none of them are tested on Farsi/Arabic documents. In this paper, an applicable approach to large size images is presented which is based on image block segmentation (mosaicing). The advantages of this approach are less memory usage, combining of the simultaneous and ordinal blind source separation methods in order to increase the algorithm efficiency, reducing calculation complexity of the algorithm into about twenty percents of the basic algorithm, and high stability in noisy images. In noiseless conditions, the average signal to noise ratio of the output images is reached up to 30.26 db. All of these cases have been tested on Farsi official documents. By applying the proposed ideas, considerable accuracy is achieved in the results, at minimum time. In addition, various parameters of the proposed algorithm (e.g. the size of each block, appropriate initial point, and the number of iterations) are optimized.

Keywords: Blind source Separation, Independent component Analysis, show-through, feed-through, background removing, scanned documents processing.

1. INTRODUCTION

In many cases, there is some additional information on scanned or photographed documents in addition to the main document image. Some of this extra information may or may not visually detectable. Depending on the user's goal, it may be interesting to highlight or remove them.

Generally, the noise affected on the documents, could be appeared as two following cases. One of them is pervading the color or no-color-equilibrium in the document, and another is the effects of the existed information behind of the document or another document on the under process document. In the scanning procedure, due to blazing radiation of light, the image of the back side of the document may be mixed with the front side in the resulted image. In this effect, if the back side ink affects the front side image of the document, the effect is named "bleed-through", and if the document is two-sided or there are consecutive thin documents in the imaging procedure, the effect is called as "show-through" effect. Sometimes, it may be required to separate the various layers of the scanned image in order to focus on each of them. Background removing is a remarkable example in document analysis [1, 2]. Undesired effects on background of a document consists of the various elements such as optical blurring and noise caused by scanning, dots, under writings and over writings [3, 4, 5]. These cases are strictly important in restoring the document images and retrieving data from the ancient documents.

The ideas which are proposed in this paper can be used in decoding of the security documents, in addition to the OCR application, electronic archives and libraries, handwritings and subscript detection, and generally, improvement of image quality.

Various techniques are available to increase the quality of the documents. The earliest methods for the document quality enhancement are binarization and image retrieving approaches which produce acceptable results in the case that one source is accompanied with noise [6]. Some researchers have tried to reduce the show-through effect by using blind source separation (BSS) approach. Although this approach requires registering the documents of both sides of the paper; however, the detection of this one by one correspondence is a hard and time-consuming task [7, 8]. A competing approach is employing Markov model for bleed-through removing, which has caused more readability in the resulted document [9, 10, and 11]. In addition, Markov model has been useful in enhancing the reconstructed image resulted BSS procedure [12].

In [13], a comparison has been performed between independent component analysis (ICA) and diffusion methods, and the advantages of the diffusion method are illustrated. In addition, in [14], various approaches among ICA has been compared, in order to evaluate the quality enhancement of the document based on providing a colored scan of the document, then analyzing it to RGB components, and finally testing on ancient documents containing watermark effect and hidden text. In the mentioned study, by analyzing the RGB components, the need to scan and register the both sides of the document has been removed.

In [15], this procedure has been used by focusing on show-through, ink-bleed, and palimpsests affected documents. In this way, sources have been separated by analyzing the colored image into the three RGB components. In the next step, ICA has been used to solve the equations set containing three equations (colored components) and three unknown elements (background, body text, foreground). In addition, instead of using only RGB components as the three sensors, invisible bands such as infra-red or ultra-violet may be used to gain more independent components [15].

Heretofore, image blocking (mosaicing) idea has been employed in few algorithms. However, the use of this idea has not been dynamic. In fact, they have selected the initial parameters of each block from a constant default values set. Therefore, they calculate each block notwithstanding neighboring blocks again and again. Hence, all respective activities, in this field, are time-consuming. Here, an important point is their less stability against noise. Furthermore none of them has been tested on Farsi documents.

The main proposed idea in this paper is based on mosaicing of the input colored document image, processing of each block separately using BSS algorithm by considering RGB components as the three sensed signals, and finally applying the results of each block as the initial values of the parameters of the next block. The method is inspired from ordinal blind source separation algorithm, while all of these are obtained at the least time and calculation complexity. The algorithm has applied to the two main categories. One of them is the real official documents. In this category, the goal of this paper is separation of the background effect from the body text, and separation of the information on front and backside of the documents from the body text, at least time and higher quality. Another category is the synthesized documents. In this category, the goal of this paper is full separation of the sources which has been synthesized and noise has been added to their combinations. These goals have been satisfied at the least time and high signal to noise ratio comparing with the basic algorithm. Of course, it should be considered that one of the important parts of the text extraction is the binarization procedure which is a complement process in order to gain the final version of the reconstructed document sources.

The structure of the paper is as follows. In the second section, the proposed method for separation of the mixed documents by conjointly use of simultaneous and ordinal BSS is explained and mathematically formulated. In section three, experimental results of applying the proposed method on Farsi official and synthesized documents is presented; and finally, the paper is concluded in the forth section.

2. PROPOSED MODEL FOR SEPARATION OF THE MIXED DOCUMENTS BASED ON BLIND SOURCE SEPARATION

Generally, there are two main methods for blind source separation named simultaneous and ordinal BSS. In simultaneous method, all sources would be separated from each other at the same time, while in ordinal method; the extraction of the sources is performed one by one in a greedy manner and the extraction of each source, improves the system for extracting the other sources.

In this paper, a combination of two mentioned methods has been proposed which on one hand, optimizes the calculation complexity and the separation accuracy of the mixed images, and on the other hand, reduces the defects of the each method. This approach uses simultaneous BSS method to process the each block of the mixed image, while optimally utilizes the knowledge gained in the previous block for the current block, due to data correlation of the various sections of the image. Equation (1) shows the formulation of the problem.

$$\mathbf{x}(k) = \mathbf{H} \cdot \mathbf{s}(k) + \mathbf{v}(k) \quad (1)$$

In this equation, “x” is the vector of observed signals, “H” is the mixing matrix, “s” is the sources which should be reconstructed, and “v” is the noise vector.

In the noiseless system, equation (1) could be written as equation (2).

$$\mathbf{X} = \mathbf{H} \cdot \mathbf{S} \quad (2)$$

$$\mathbf{X} = [\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(N)]^T \in \mathbb{R}^{m \times N} \quad (3)$$

$$\mathbf{S} = [\mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(N)]^T \in \mathbb{R}^{n \times N} \quad (4)$$

In order to blind source separation two cases could be defined. One of them is based on exactly extraction of the source signals with high computational complexity, and another is based on iterative operations with less computational complexity. In the first case, by assuming statistical independency of the original sources and linear independency of the columns of matrix H, generally, the ICA technique, could estimate all sources and their quantity, or recognize mixing

matrix (H) or the pseudo inverse matrix ($w = H^\dagger$). So, by multiplying the pseudo inverse matrix at the sensed signals, the source signals could be estimated as follows.

$$\hat{s}(k) = w \cdot x(k) \tag{5}$$

Here, $\hat{s}(k)$ is the vector of the reconstructed signals. The exact response of the reconstructed sources will illustrate later. If the number of the independent components is great and its distribution is sparse, it is better to use the following form.

$$X^T = S^T \cdot H^T \tag{6}$$

In this case, the rows of the S and the columns of the H should as independent as possible, and both of them should approach to similar statistical attributes (sparse presentation or laplacian distribution). In this way, the vectors of the matrix H^T can be considered as the independent components, the matrix S^T matrix as the mixing matrix and vice versa.

Now, instead of estimating the sources directly, the information should be gathered about the mixing system at first. Since w is the pseudo-inverse of H, the rows of H are corresponded with the columns of w. In most of cases, the sources have been mixed and filtered, linearly and simultaneously. In fact, the mutual vectors h_j of the mixing matrix H or pseudo inverse of it (w_j), should be recognized.

Basically, in blind source separation, it is desired to minimize the average square error between output signals and original sources. In other words, the following cost function should be minimized.

$$J_p(s) = \|x - H \cdot s\|_p = \|e(s)\|_p \quad , \quad (p \geq 1) \tag{7}$$

In this equation, p is the geometric norm and the error vector could be defined as the following equation.

$$e_i(s) = x_i - h_i^T \cdot s = x_i - \sum_{j=1}^p h_{ij} \cdot s_j \tag{8}$$

In order to minimize the cost function, by use of least linear squares approach, and considering norm 2, the equation (7) could be transformed as below.

$$J(s) = \frac{1}{2} \|X - H \cdot s\|_2^2 = \frac{1}{2} (X - H \cdot s)^T (X - H \cdot s) = \frac{1}{2} e^T \cdot e = \frac{1}{2} \sum_{i=1}^m e_i^2 \tag{9}$$

If the gradient of the cost function be equaled to zero, the exact response for reconstructed sources could be achieved as following.

$$\nabla J(s) = H^T (X - H \cdot s) = 0 \tag{10}$$

$$s_* = H^T (H \cdot H^T)^{-1} X = H^\dagger \cdot X \quad , \quad J(s_*) = 0 \tag{11}$$

In this case of BSS, because of unknown quantity of the pseudo inverse matrix, difficulty of finding it, inversion of the large matrixes and as the result more computational complexity, reconstructing the sources is very difficult. So, using an iterative equation could be applicable. By using gradient based approaches, the problem could be reformulated as below [14].

$$\frac{ds}{dt} = -\mu \cdot \nabla J(s) = \mu \cdot H^T (X - Hs) = \mu \cdot H^T \cdot e \tag{12}$$

In this equation, $\mu = [\mu_{ij}]$ is a positive definite $n \times n$ matrix which usually is diagonal. In order to the source estimation, this differential equation could easily and directly be converted to an iterative equation, as following.

$$\hat{s}(k+1) = \hat{s}(k) + \eta_s \cdot H^T [X - H \cdot \hat{s}(k)] \quad (13)$$

Generally, since the input signals applied to ICA should be independent, regardless of the inherent independency of the observed signals, the operation called whitening should affect on them. So, in order to provide independent signals, the correlation of the observed signals should be removed. The problem could be modeled as equation (14), which therein " $y(k)$ " is the whitened vector, and W is a $n \times m$ whitening matrix [14, 15]. Here, in order to prevent be mistaken for the pseudo inverse matrix, the whitening matrix has been indicated with upper case (W).

$$y(k) = W \cdot x(k) \quad (14)$$

After whitening, usually ICA or BSS approaches converge better [14]. This occurs because of describing the secondary separation system (non-mixing system) by the orthogonal matrix for real signals, and an identity matrix for weights of mixed signals.

In the whitening process, the selection of matrix W follows the following procedure. The covariance matrix of the whitened signals ($E\{y(k)y(k)^T\}$) should become the identity matrix (I_n). In this manner, the whitened vectors $y(k)$ will be mutually independent [14].

$$R_{yy} = E\{yy^T\} = E\{Wxx^TW^T\} = WR_{xx}W^T = I_n \quad (15)$$

In order to achieve an iterative equation, Similar to equation 13, based on gradient approaches, whitening operation could be illustrated as following [16].

$$W(k+1) = W(k) + \eta_w [I - y(k) \cdot y^T(k)]W(k) \quad (16)$$

In this equation, $W(k)$ is the quantity of matrix W at k^{th} iteration, $y(k)$ is the whitened vectors at k^{th} iteration, and η_w is the convergence rate of the equation. This equation will converge, if the general matrix as G (in equation 17) could be founded to satisfy either equations (18) or (19).

$$G = W \cdot H \quad (17)$$

$$G \cdot G^T = G^T \cdot G = I_n \quad (18)$$

$$G^{-1} = G^T \quad (19)$$

In this situation, matrix G will be orthogonal. Furthermore, by multiplying equation (16) at the mixing matrix (H) from right side, equation (20) would be obtained [16]:

$$\frac{d}{dt} W(k+1)H = G(k+1) = G(k) + \eta_g [I - G(k) \langle \hat{s}(k) \cdot \hat{s}^T(k) \rangle G^T(k)]G(k) \quad (20)$$

In this equation, $G(k)$ is the quantity of matrix G at k^{th} iteration, $\hat{s}(k)$ is the reconstructed sources at k^{th} iteration, and η_g is the convergence rate of this equation.

Without loss of generality, by assuming that self correlation matrix of the estimated sources (equation 21) is an identity matrix, it is clear that a learning algorithm which uses above equation is stable and converges, if the matrix $G(k)$ is orthogonal ($G^{-1} = G^T$).

$$R_{\hat{s}\hat{s}} = \langle \hat{s}(k) \hat{s}^T(k) \rangle \quad (21)$$

Finally, by considering the equations 14, 17 and applying them at equation 13, the basic equation for estimation of sources could be achieved as the following equation.

$$\hat{s}(k+1) = \hat{s}(k) + \eta_s \cdot G^T(k+1)[y(k) - (G(k+1) \cdot \hat{s}(k))] \tag{22}$$

In this equation, η_s as the learning rate should be a positive constant to guarantee the stability of the algorithm. In addition, the learning rate, η_s should satisfy $0 < \eta_s < 1/\lambda_{max}$ constraint, where λ_{max} is the maximum eigenvalue of $H^T H$. In summary, if η_s is near the upper boundary, the algorithm converges rapidly. In order to effectively apply the algorithm, the largest eigenvalue of $H^T H$ should be estimated to specify the upper boundary.

Fig. 1 shows the overall block diagram which is used for reconstruction of the sources. The goal of this paper is to utilize the BSS approach into blind separation of the image sources resulted from several records on the document such as show-through and background effects in official documents and full separation of the sources in synthesized images, with minimum computational complexity. The first idea is based on partitioning of image or “mosaicing”. In most of the cases, because of much length of the signals, a large amount of the memory is needed to apply BSS algorithm. Therefore, it requires the systems with large amount of the memory. On the other hand, in order to keep the correlation of the neighboring pixels, it is better to divide the image into the smaller blocks and process each block separately.

Of course, the choice of the block size is very important. The block size should be as small as possible to keep the correlation between neighboring pixels, and on the other hand, the desired block size should contain textual data. In other words, it should not be so small which its textual data could not be read. For final decision, the output signal to noise ratio and computational complexity of the algorithm should be evaluated.

Fig. 2 shows the block diagram of the proposed algorithm in this paper. It should be mentioned that the algorithms and equations which are expressed in this paper, are quite different from the other available algorithms in blind source separation for document separation [12].

As it is shown in the block diagram, in the first step, the colored scanned image of the document is decomposed into three red, green, and blue components [15]. These components are processed as three independent observed sensed signals (x_1, x_2, x_3).

After mosaicing the images, the corresponding mosaics of each sensor have to be applied to the separation algorithm. Before employing the main separation algorithm, each of the applied blocks should be changed to a one dimensional signal [13]. Then, in order to reduce the correlation between the elements of the obtained three one dimensional signals, whitening operation is unavoidable.

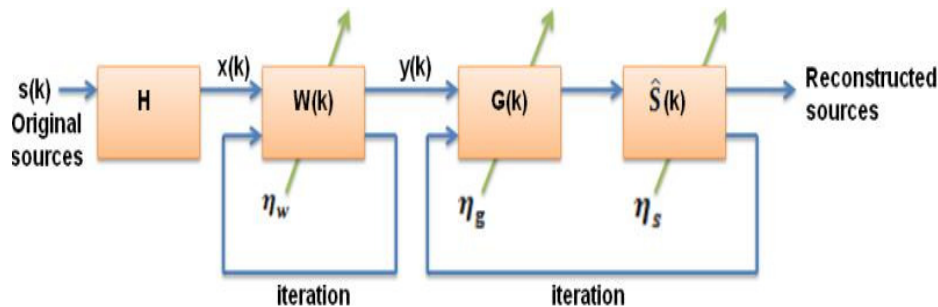


FIGURE 1: BSS Basic Algorithm Block Diagram.

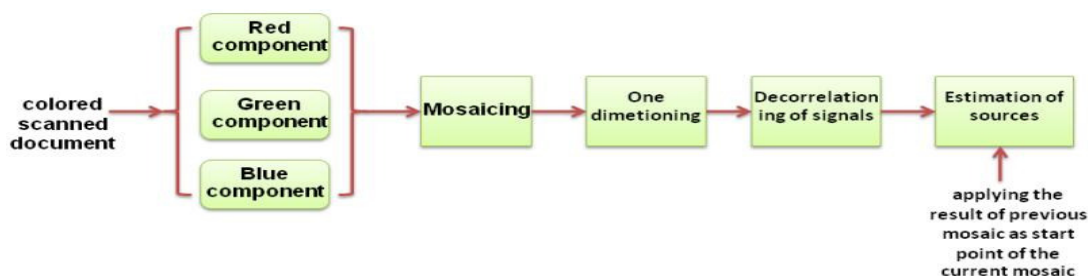


FIGURE 2: Block Diagram of Proposed BSS Approach and Optimization Ideas.

To select the initial values of the whitening and mixing matrixes, they should be assumed in the manner that in each row and column, only one element is dominant. This causes the algorithm converge faster. Then, the source estimation procedure is performed by using the reconstructed sources of each block as the initial values of the next block's estimation (equation 22). This procedure obtains a good initial point for the main algorithm, in order to have less computational complexity.

According to tests performed in the next section, most of the calculations occur in the estimator process, although after applying the proposed ideas, in addition to keep high visual quality and quantitative accuracy, the run time of the algorithm is considerably reduced. In this study, all of the proposed ideas, are examined on the synthesized and Farsi official documents, and as the good results, the stability of the proposed algorithm in noisy environments is demonstrated.

3. EXPERIMENTS AND RESULTS

3.1 Experiments Setup

In this section, the results of applying the proposed ideas are presented. All of the templates which are used in this section are selected from Farsi official documents. In this paper, the quantitative evaluation is performed on the synthesized images. The performance criterion is the ratio of the average SNR of the output images to the average SNR of the input images. It is possible to post-process the output image such as binarization and filtering operations. The dimensions of the synthesized images which are used in all of these experiments are 384*384 pixels and the mixing matrix is as following.

$$H = \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.3 & 0.3 & 0.4 \end{bmatrix}$$

In the synthetic experiments, three textual sources have been mixed by above mentioned mixing matrix and three gained mixtures have been applied to the algorithm as the three sensors. The original Images are prepared by scanning Farsi official documents in 300 dpi resolution and 24 bits color depth, while their dimensions is 512*896 pixels. After exploring and selecting the optimum size of the mosaics in the next section, for other experiments, the block size is considered as 64*64. The experiments are conducted on a Personal Computer 2.0 GHz Dual Core Due Intel processor. The simulation environment is Matlab R2008.

3.2 Optimizing the size of mosaics

In order to select the optimum size of the mosaics, various experiments based on the different mosaic sizes such as 4*4, 8*8, 16*16, 32*32, 64*64, 128*128, has been performed. It has been observed that for 64*64 block size, the reconstructed signals would be more acceptable than the other sizes. In this experiment, the base of the block size selecting, is the high separation accuracy with the minimum computational complexity. As it is observable at TABLE 1, for 4*4 block size, the operation time of algorithm is 1109 seconds, and average SNR of the output

images, is 24.82 db, while for 64*64 block size, these values are obtained as 206 seconds and 30.26 db.

The maximum SNR is obtained for 16*16 block sizes by amount of 30.58 db. As it will be presented in the next section, the operation time of the separation algorithm for 16*16 blocks mosaicing is 299 seconds while this time for 64*64 block size is 206 seconds. The average SNR of these two situations differs only about 0.25 decibel. Considering all aspects, 64*64 block size is selected as the best choice. Figure 3 shows the visual comparison between the results of considering 4*4, 16*16, and 64*64 block sizes. As it is observable, the results of 4*4 block size have low quality because of dot effects on them. The results of the 16*16 and 64*64 block sizes do not differ visually, because their SNR is very close together.

SNR Block size	SNR of first output	SNR of second out put	SNR of third out put	Average SNR of outputs
4*4	27.82	22.99	23.67	24.82
8*8	31.66	29.18	30.12	30.32
16*16	32.46	28.96	30.32	30.58
32*32	31.85	28.68	29.68	30.07
64*64	31.06	30.45	29.28	30.26
128*128	31.57	27.62	27.87	29.02

TABLE 1: Variations of Average SNR of Reconstructed Sources in Terms of Different Block Sizes in Noiseless System.



FIGURE 3: a, b, c) original sources. d, e, f) three synthetic mixtures in noiseless system. g, h, i) reconstructed sources using BSS by 4*4 mosaic size. j, k, l) reconstructed sources using BSS by 16*16 mosaic size. m, n, o) reconstructed sources using BSS by 64*64 mosaic size.

3.3 Computational Complexity of the proposed algorithm

Figure 4, depicts the operation time of the each block versus the various block sizes. As it can be observed, by increasing in the mosaic sizes, the processing time of the each mosaic increases, while as it is observable in Figure 5, the processing time of the entire image decreases. In Figure 5, the maximum point of curve is assigned to the 4*4 block size which the related operation time is obtained as 1109 seconds. The operation time of 64*64 block size, has been calculated 206 seconds that is the minimum point of the curve. Considering these curves and visual results, presented in section 3.2, the best block size as the trade off point of visual quality, computational complexity, and SNR aspects is 64*64 block size.

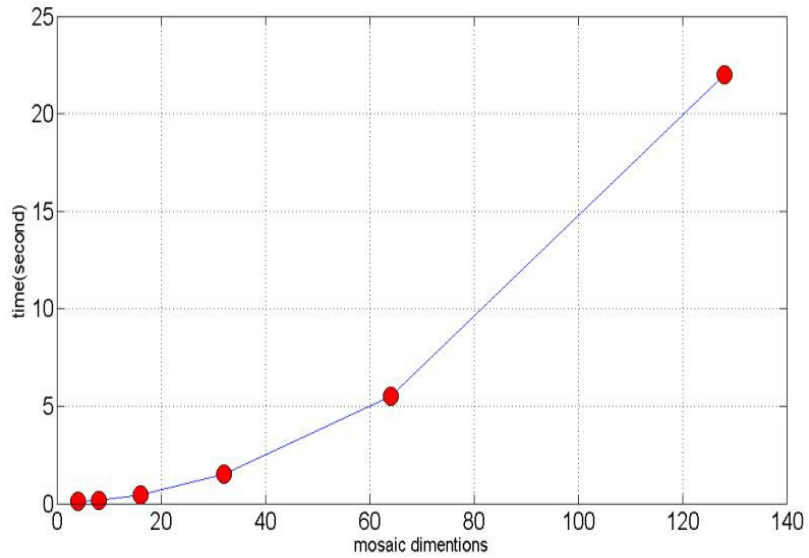


Figure 4: Computational Complexity of the each block versus the various block sizes

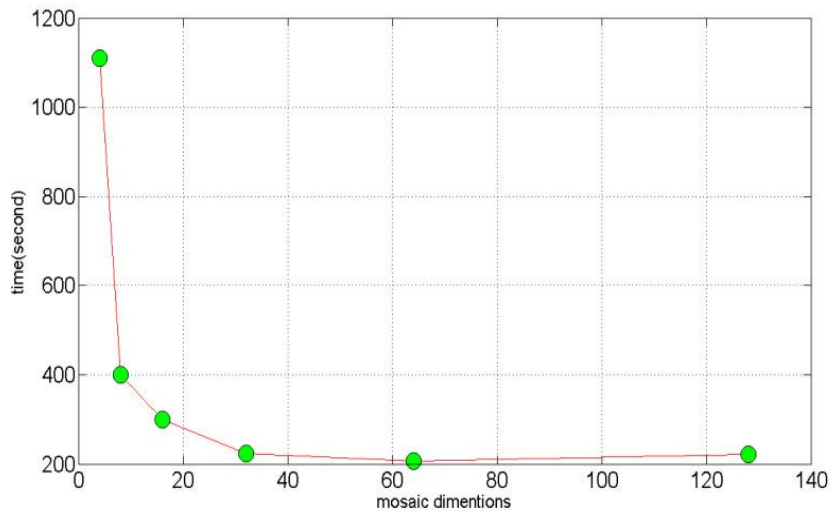


Figure 5: Total Computational Complexity of the typical image versus the various block sizes

3.4 Effect of the Number of the Iterations on the Computational Complexity

In BSS approach, computational complexity increases linearly with the number of iterations (in w, g, s). Thus, by optimizing the number of iterations, the computational complexity and operation time could be reduced. The convergence characteristic of the proposed algorithm is very good. In fact, after specific iterations, the results will not have considerable improvement. Before applying the results of previous block to the under process block, to achieve an acceptable output results, these iterations were as follows: 4000 iteration for whitening algorithm, and 200 iterations for G matrix and source estimator algorithms (Figure 6).

For a typical image, the operation time of the basic BSS is calculated as 750 seconds. After applying the proposed idea, the operation time reduced to 154 seconds, and the required number of iterations for convergence, except first block, decreased to 5 iterations for G matrix, and 100 iterations for whitening and source estimator algorithms (Figure 7). The operation time has decreased to 20 percent of the basic algorithm.

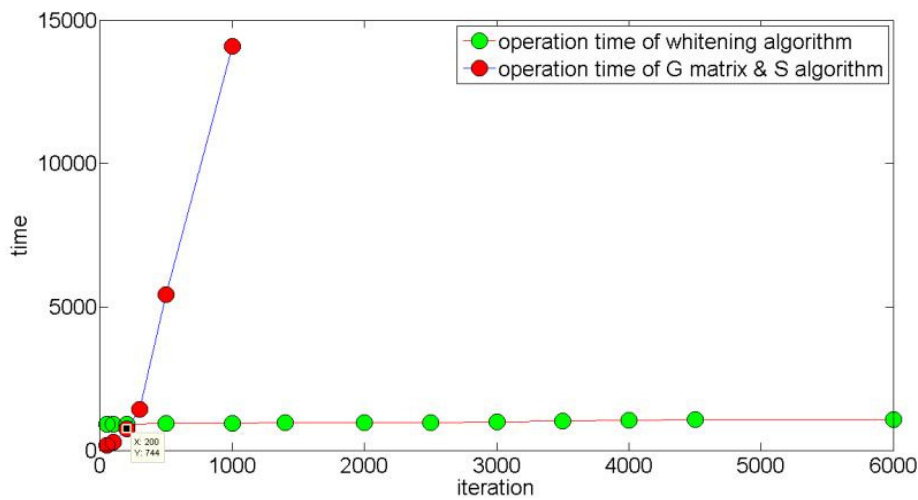


Figure 6: Effect of the Number of the Iterations on Computational Complexity in Basic BSS Algorithm

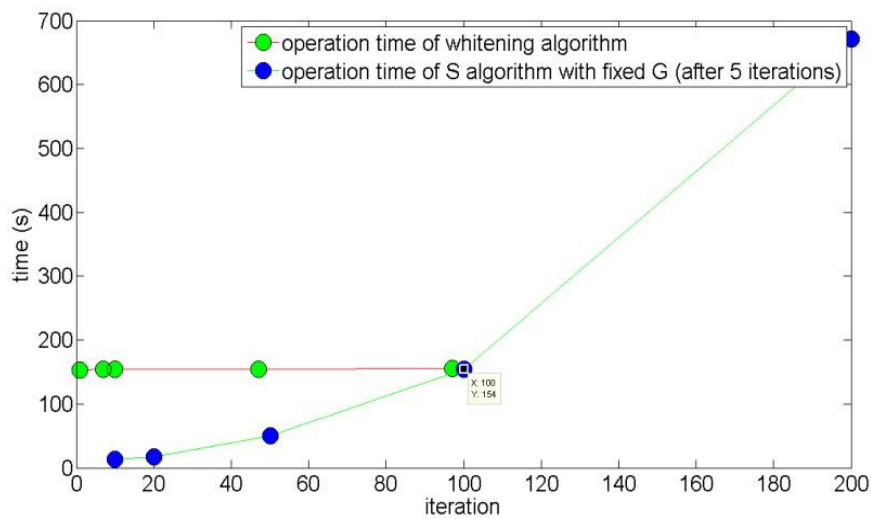


Figure 7: Effect of the Number of the Iterations on Computational Complexity after Applying Optimization Ideas

3.5 Separation of the Background Affected Documents

To investigate the ability of the proposed algorithm in the separation of the background effect of the image, the method is applied to the document that the watermarked logo of the owner corporation was on the background. The results of this experiment are shown in Figure 8. The operation time of the basic BSS, for this document is 825 seconds, with 4000 iterations for whitening and 200 iterations for G matrix and estimator algorithms, while after applying the optimization ideas, the operation time reduced to 136 seconds, with 10 iterations for whitening, 5 iterations for G matrix and 100 iterations for estimator algorithm (except first block). As it is observable in the output images of the algorithm (b, c, d), the algorithm has focused on one component at each output. After BSS operation, the main text could be extracted using binarization (e) and the background effect could be extracted using complementation and binarization of the first output (f). In this manner, the background effect and the main text are fully separated from each other, after post-processing.

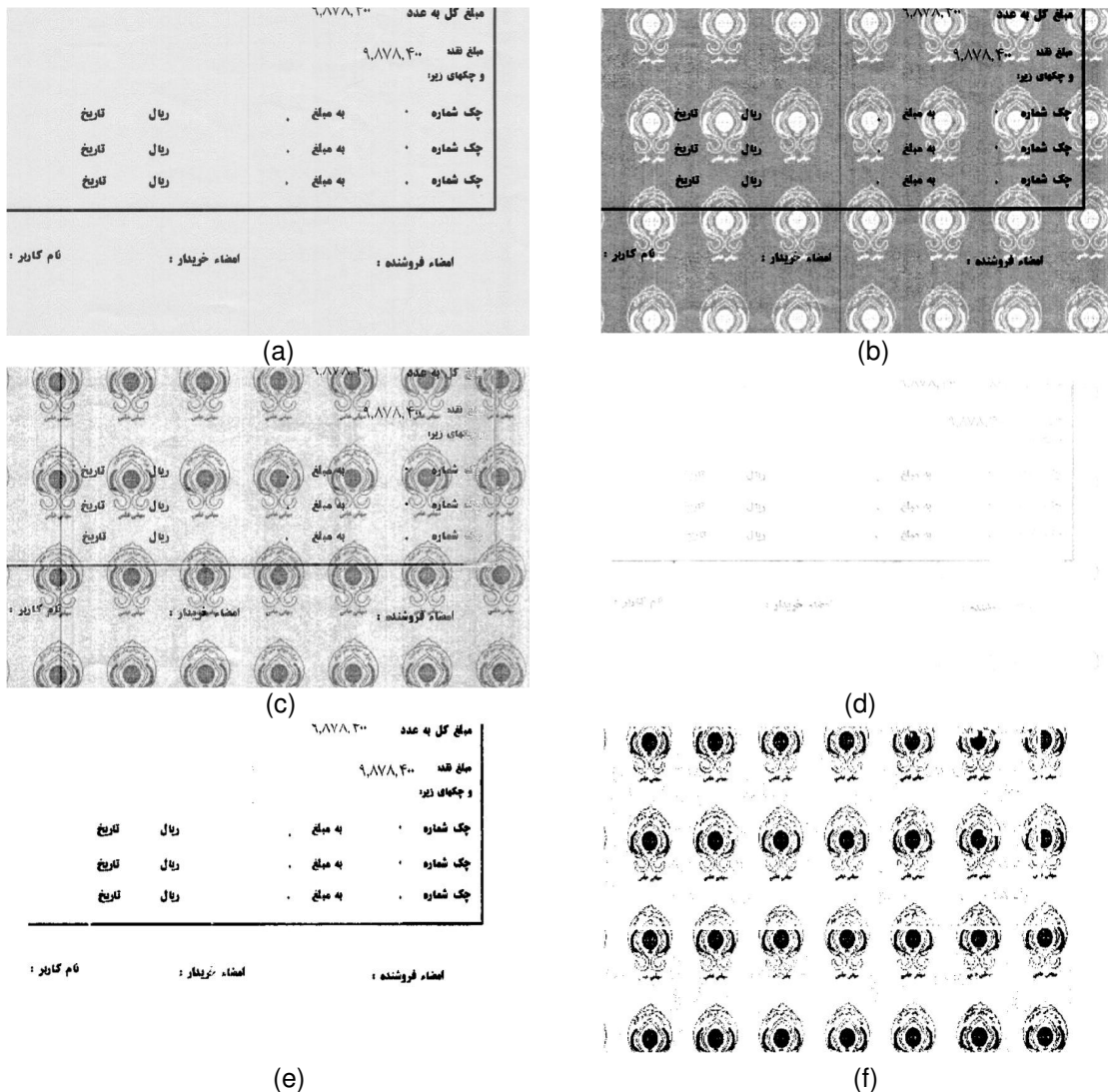


Figure 8: a) Colored scan of a typical official document. b, c, d) results of the extracting RGB components of the original image as three sensed signals. e, f) separated main text and background effect after post-processing operations.

3.6 Robustness of the Proposed Algorithm in Separating the Show-Through Effect

The proposed algorithm was applied to the colored image of the scanned official document with the show through effect. The original image has been demonstrated in Figure 9-a. The front of the document contains signature and main text, while there are some written information and stamp effect in behind. It is cleared that in this test, there are four sources, while only three RGB sensors would be available for applying to the algorithm. As it is observable in Figure 9-g, one of the sources might not been extracted singular. For this document, the operation time of the basic BSS, is 602 seconds, with 3000 iterations for whitening and 200 iterations for G matrix and estimator algorithms, while after applying the optimization ideas, the operation time reduced to 136 seconds, with 20 iterations for whitening, 5 iterations for G matrix and 100 iterations for estimator algorithm (except first block). Figures 9-b,c,d, are the outputs of the algorithm which in each of them, it has been focused on one (or more) component(s). After carrying out the algorithm, the stamp effect has been extracted using binarization, complementation, and median filter (Figure 9-e), the signature effect has been extracted using binarization and median filter (Figure 9-f), and finally, the written information behind of the document, has been extracted using binarization (Figure 9-g). All of these have been occurred at the least computational complexity.

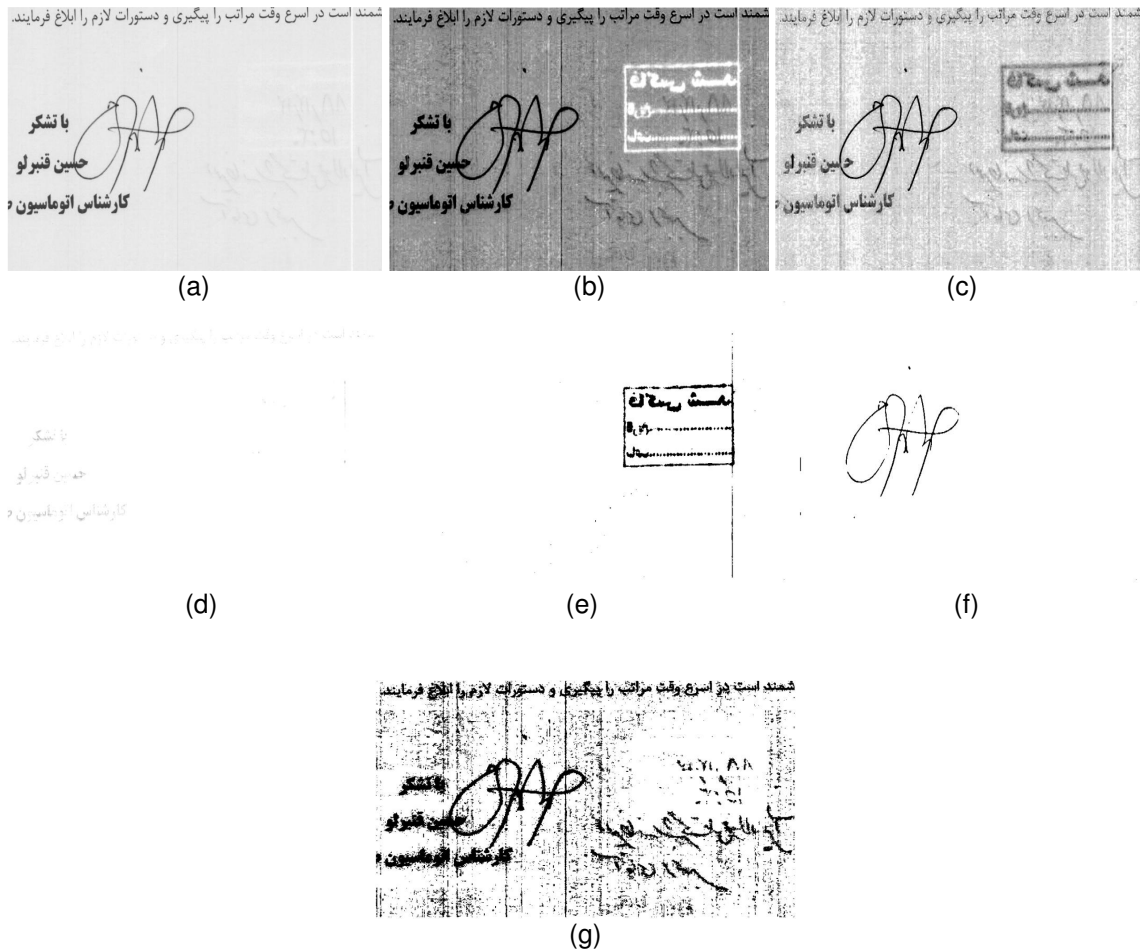


Figure 9: a) The Colored Scan of an Official Document containing show-through effect. b, c, d) The Three Results of Proposed BSS Algorithm. e, f, g) Extraction the Desired Data by Post Processing Operations Such as Binarization and Filtering.

3.7 Robustness of the Proposed Algorithm in Presence of the Additive Noise

This section has focused on the stability of the algorithm in noisy conditions. It should be considered that the noise has been added to the three synthetic mixtures, that is very dangerous kinds of the additive noises. In Figure 10, it is observable that in noiseless system, the SNR of the reconstructed sources has been increased up to 30.26 db that in comparison with the basic BSS, the amount of the improvement is about 2 decibel. In the proposed algorithm, the output signals with the SNR up to 15 decibel have been achieved versus the input signals with the zero decibel SNR. By increasing in the SNR of the input images, the saturation effect will occur. Table II, presents the SNR of the input and output signals, in terms of the different salt and pepper noise energies. While 50% salt and pepper added to the mixtures, the SNR of the output signals has been obtained up to 4 decibel. In Figure 11, a visual example has been given in the presence of 10% salt and pepper noise. It is observable that after binarization of the output signals of the algorithm, the original sources are fully separated.

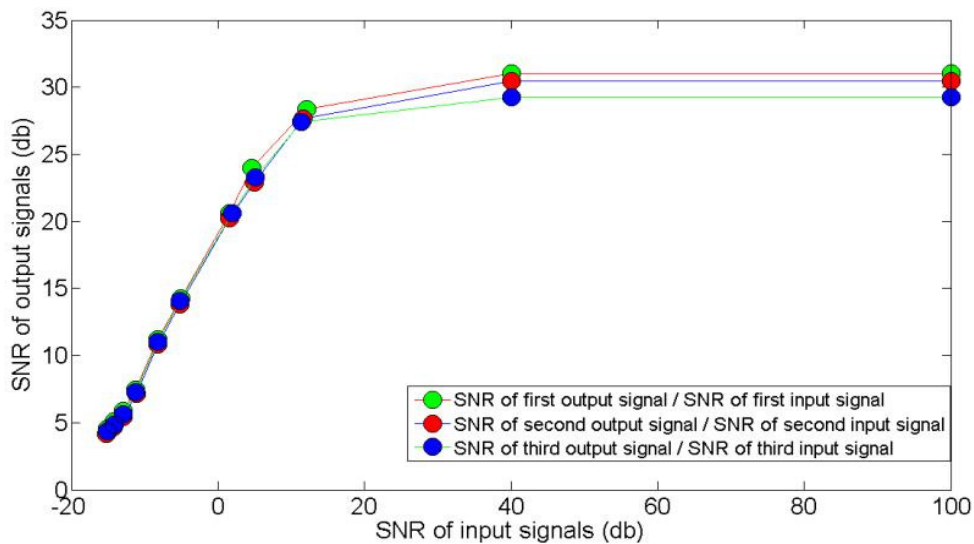


Figure 10: SNR of the Reconstructed Sources versus SNR of the Original Sources with Different Values of the Salt and Pepper Noise applied to the mixture sensed images

Noise (%)	SNR IN_1	SNR IN_2	SNR IN_3	SNR IN_Ave	SNR OUT_1	SNR OUT_2	SNR of OUT_3	SNR of OUT_Ave
0	+∞	+∞	+∞	+∞	31.06	30.45	29.28	30.26
5	-5.2	-5.15	-5.15	-5.16	14.19	13.87	13.99	14.01
10	-8.22	-8.19	-8.16	-8.19	11.22	10.85	11.05	11.04
20	-11.21	-11.09	-11.17	-11.15	7.44	7.17	7.29	7.3
30	-12.92	-12.9	-12.9	-12.9	5.88	5.45	5.65	5.66
35	-14.17	-14.2	-14.19	-14.18	5.1	4.66	4.88	4.88
50	-15.14	-15.17	-15.13	-15.14	4.55	4.13	4.34	4.34

TABLE II: SNR comparison of the input and output images



Figure 10: a, b, c) original sources. d, e, f) 3 mixtures added 10% salt & pepper noise. g, h, i) reconstructed sources using proposed algorithm. J, k, l) enhancement of the results using binarization method.

4. CONCLUSION & FUTURE WORK

In this paper, an optimization procedure on classic approaches for blind source separation is proposed that could be extended into many similar procedures in the field of the image processing. In this approach, by segmenting the colored image to the equal-sized mosaics and optimizing the algorithm for each mosaic, a considerable reduction in the computational complexity and run time of the algorithm has been achieved. In this approach, the results of the each block are used as the initial point of the separation algorithm in the next block. Therefore, both of the computational performance and accuracy of the algorithm are well increased. This algorithm could successfully separate the background effect from the body text and show-through effect. Of course, in most of the cases, it may be unavoidable to use the post processing operations for gaining desired results. In this approach, no registration operation is required. In addition, the high stability against noise is another considerable property of the proposed algorithm. The study is now continued on proposing better approaches to guarantee the stability of algorithm against other kinds of the image noises. In addition, by deriving a two dimensional BSS algorithm, to consider the spatial correlations of the pixels, more than available one dimensional algorithms, the results may become better.

5. REFERENCES

- [1] N. K. Kishore and P. P. Rege. “*Adaptive enhancement of historical document images*”. In Proceedings of IEEE International Symposium on Signal Processing and Information Technology. Giza, Dec 2007
- [2] B. Nija, N. G. Preethi, and S. S. Shylaja. “*Degraded document image enhancement using Hybrid Thresholding and Mathematical Morphology*”. In Proceedings of sixth Indian Conference on Computer Vision, Graphics & Image Processing. Bhubaneswar, Dec 2008
- [3] J. Kennard and W. A. Barrett. “*Separating lines of text in free-form handwritten historical documents*”. In Proceedings of Second International Conference on Document Image Analysis for Libraries (DIAL'06). Lyon, France, April 2006
- [4] F. Drira. “*Towards restoring historic documents degraded over time*”. In Proceedings of the Second International Conference on Document Image Analysis for Libraries (DIAL'06), Pages 350-357. Lyon, France. April 2006
- [5] R. Farrahi moghaddam and M. Cheriet. “*Restoration of single sided low quality document images*”. In Proceedings of Pattern Recognition, Volume 42, Issue 12, Pages 3355-3364. Dec 2009
- [6] M. Feng, Y. Tan. “*Adaptive binarization method for document image analysis*”. In Proceedings of the IEEE International Conference on Multimedia and Expo, ICME. Taipei, Taiwan. June 2004
- [7] F. Merrikh-Bayat, M. Babaie-Zadeh, and C. Jutten. “*A nonlinear blind source separation solution for removing the show-through effect in the scanned documents*”. In Proceedings of the 16th European Signal Processing Conference (EUSIPCO). Lausanne, Suisse. Sep 2008
- [8] J. Wang, M. S. Brown, and C. Lim Tan. “*Accurate alignment of double-sided manuscripts for bleed-through removal*”. In Proceedings of the 8th IAPR International Workshop on Document Analysis Systems (DAS), Pages 69-75. Nara, Japan. Sep 2008
- [9] Y. Huang, M. S. Brown, and D. Xu. “*A framework for reducing ink-bleed in old document*”. In Proceedings of the IEEE Computer Society Conference Vision and Pattern Recognition (CVPR 2008). Anchorage, Alaska, USA. June 2008
- [10] J. Banerjee. “*Document enhancement using text specific prior*”. Thesis of Master of Science Degree. Hyderabad, India. Dec 2008
- [11] C. Wolf. “*Document ink bleed-through removal with two hidden Markov random fields and a single observation field*”. In Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence Volume 32, no 3, Pages 431-447. Jan 2010
- [12] A. Tonazzini, I. Gerace, and F. Cricco. “*Joint blind separation and restoration of mixed degraded images for document analysis*”. In Proceedings of the International Conference on Image Processing (ICIP'04). Singapore. Oct 2004
- [13] M. Cheriet and R. Farrahi Moghaddam. “*Processing of low quality document images: issues and directions*”. In Proceedings of the 16th European Signal Processing Conference (EUSIPCO). Lausanne, Switzerland. August 2008
- [14] A. Tonazzini, E. Salerno, M. Mochi, and L. Bedini. “*Blind Source Separation Techniques for Detecting Hidden Texts and Textures in document Images*”. In Proceedings of the International Conference on Image Analysis and Recognition (ICIAR), Part II. Porto, Portugal. Sep-Oct 2004
- [15] A. Tonazzini, L. Bedini, and E. Salerno. “*Independent component analysis for document restoration*”. International Journal on Document Analysis and Recognition, Volume 7, Number 1, Pages 17-27. March 2004
- [16] A. Cichocki, and S. Amari. “*Adaptive blind signal and image processing*”. 2002.

Anisotropic Diffusion for Medical Image Enhancement

Nezamoddin N. Kachouie

nezam.nk@gmail.com

*Department of Systems Design Engineering,
University of Waterloo, Waterloo, ON, Canada
Present Affiliation: Harvard-MIT Health
Sciences and Technology Harvard
Medical School, Cambridge, MA, USA*

Abstract

Advances in digital imaging techniques have made possible the acquisition of large volumes of Trans-rectal Ultrasound (TRUS) prostate images so that there is considerable demand for automated segmentation of these images. Prostate cancer diagnosis and treatment rely on segmentation of TRUS prostate images. This is a challenging and difficult task due to weak prostate boundaries, speckle noise, and narrow range of gray levels which leads most image segmentation methods to perform poorly. Although the enhancement of ultrasound images is difficult, prostate segmentation can be potentially improved by enhancement of the contrast of TRUS images. Anisotropic diffusion has been used for image analysis based on selective smoothness or enhancement of local features such as region boundaries. In its conventional form, anisotropic diffusion tends to encourage within-region smoothness and avoid diffusion across different regions. In this paper we extend the anisotropic diffusion to multiple directions such that segmentation methods can effectively be applied based on rich extracted features. A preliminary segmentation method based on extended diffusion is proposed. Finally an adaptive anisotropic diffusion is introduced based on image statistics.

Keywords: TRUS Imaging, Deformable Models, Level Sets, Anisotropic Diffusion, Segmentation.

1. INTRODUCTION

As the most diagnosed cancer, prostate cancer is the second leading cause of the cancer death in North America [1]. Hence diagnosis of this cancer in its early stages is crucial. Prostate TRUS images, in comparison with the other modalities such as CT and MRI, are captured more easily, in real-time, and with lower cost, so they are widely used for the diagnosis of prostate cancer, cancer treatment planning, needle biopsy, and brachytherapy. The size and the shape of the prostate must be determined by prostate segmentation to diagnose the cancer stage. Although in the traditional approach, an expert infers this information manually from the TRUS images, such a manual method is tedious, expensive, time consuming, and subjective. Given the increasing amount of TRUS data being collected, automated methods of TRUS prostate segmentation are in high demand and different segmentation methods have been proposed [2, 3, 4, 5, 6, 7, 8, 9]. These methods include boundary segmentation, deformable models, and region segmentation approaches.

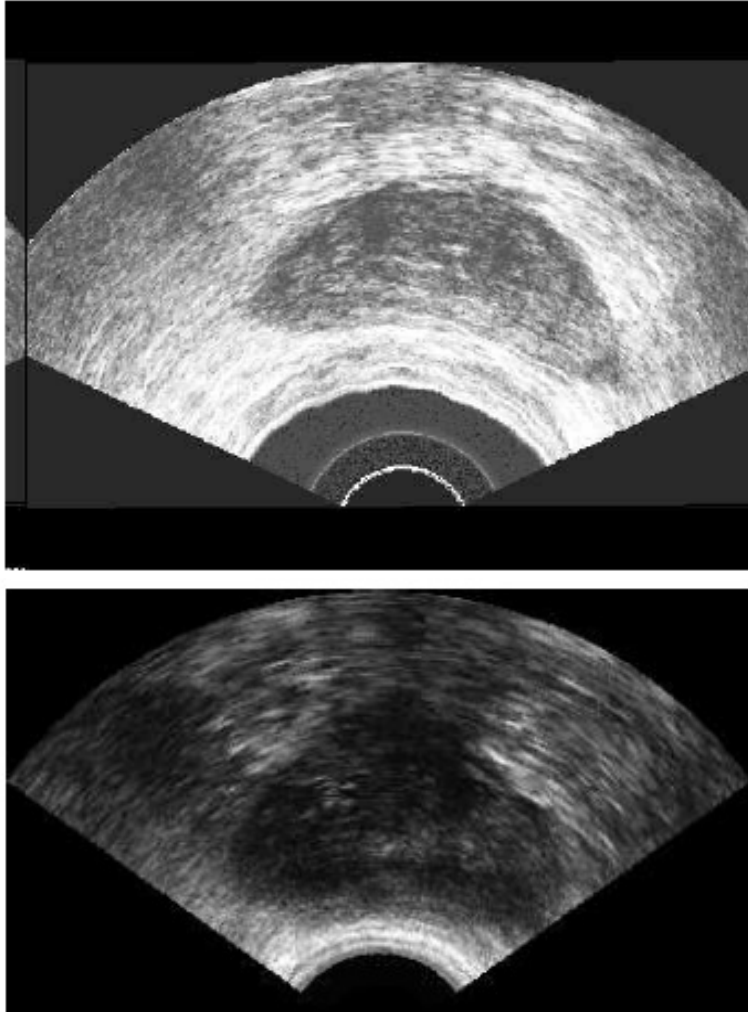


FIGURE 1: Two TRUS prostate images.

The prostate region in TRUS prostate images usually maintains a very weak contrast against the background. Because of the speckle noise, short range of gray levels, very weak prostate region texture, and shadow regions the conventional image processing and analysis techniques are not capable to effectively capture, discriminate, and segment the prostate region based on its intensity, texture, and gradient. There have been some attempts by employing the Gabor filter bank for prostate texture segmentation [10, 5], however prostate has a very weak texture and has not yet been investigated seriously.

Anisotropic diffusion [11] was introduced by Prona and Malik to perform edge preserving and within-region smoothing based on the differential structure of the image [12, 13]. Anisotropic diffusion considers no prior information about the regions and boundaries, and does selective diffusion based on local computation of a conduction term. Moreover, anisotropic diffusion computes group diffusion as a single diffusion value for each spatial location by summation of diffusions in four directions in each time step. There are some weak radial and angular features in ultrasound images which potentially can be used to reveal weak structures and textures. Our goal in this paper is extending anisotropic diffusion to multiple directions which are computed independently for contrast enhancement of TRUS images. Therefore, eight anisotropic diffusion values will be computed independently for each spatial location. This can provide a rich feature space with potential use in image analysis and segmentation.

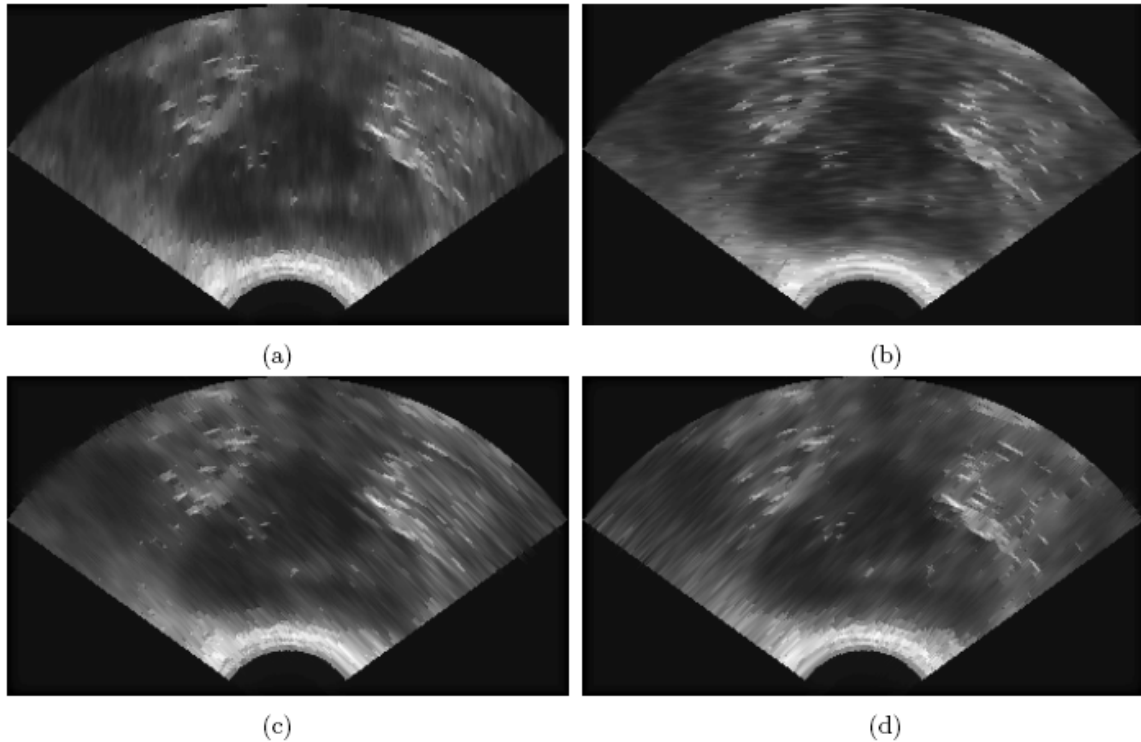


FIGURE 2: Application of extended anisotropic diffusion in eight directions. (a) North-South. (b) East-West. (c) NE-SW. (d) NW-SE.

Moreover, considering a semi-supervised segmentation such as deformable models to initialize a seed, we introduce an adaptive anisotropic diffusion in which the estimated statistics extracted from the region of interest can be used to adaptively switch between different conduction functions leading to better within-region smoothness while preserving the region boundaries.

2. The Proposed Method

Strong region boundaries are desired for image segmentation, however image denoising tends to smooth sharp boundaries of the image and reduces the image contrast. To overcome this drawback of image denoising methods, anisotropic diffusion method, as an alternative to linear-filtering was introduced by Perona and Malik [11]. Anisotropic diffusion considers a conduction term that is locally computed and depends on the differential structure of the image. Anisotropic diffusion filter was used by Gerig et al. [12] to enhance MR images. To perform edge preserving and within region smoothing of MR images, Sapiro and Tannenbaum [13] used a similar approach.

Anisotropic Diffusion

Perona and Malik [11] presented the anisotropic diffusion filter as a diffusion process that encourages intra-region smoothness while inhibits inter-region smoothness. Mathematically, the process is defined as follows:

$$\frac{\partial}{\partial t} I(\bar{x}, t) = \nabla \cdot (c(\bar{x}, t) \nabla I(\bar{x}, t)) \quad (1)$$

where $I(x,0)$ is the initial unprocessed image, x is the image coordinate and t is the iteration step. $c(x,t)$ is the diffusion function and is a monotonically decreasing function of the image gradient magnitude. To encourage smoothing within a region and discourage it across different regions, the conduction coefficient c must be set to one inside the region (smooth conduction) and set to zero otherwise. For edge estimation to locate the region boundaries, the gradient of intensity image is first obtained:

$$\hat{E}(\bar{x}, t) = \nabla(I(\bar{x}, t)) \quad (2)$$

The conduction coefficient of diffusion is then computed locally as a gradient magnitude of local image intensities:

$$c(\bar{x}, t) = f(\|\nabla(\bar{x}, t)\|) \quad (3)$$

By the proper selection of function f , not only region boundaries can be preserved but also edges maybe sharpened. Any monotonically decreasing continuous function of $\|\nabla I\|$ could be selected as a diffusion function. Two functions for local computation of the conduction to satisfy selective edge smoothness and enhancement were suggested by Perona and Malik [11]. The first one

$$c(\bar{x}, t) = \exp\left(-\left(\frac{\|\nabla(\bar{x}, t)\|}{\kappa}\right)^2\right) \quad (4)$$

encourages high contrast edges over low contrast ones while the second function

$$c(\bar{x}, t) = \frac{1}{1 + \left(\left(\frac{\|\nabla(\bar{x}, t)\|}{\kappa}\right)^2\right)} \quad (5)$$

encourages wide regions over smaller regions where k is the diffusion coefficient. The differential relation in (1) can be discretized and be numerically implemented as

$$I_{i,j}^{t+1} = I_{i,j}^t + \eta \left(N_c \cdot \nabla_N I + S_c \cdot \nabla_S I + W_c \cdot \nabla_W I + E_c \cdot \nabla_E I \right) \quad (6)$$

where N_c , S_c , W_c , and E_c are conduction in north, south, west, and east directions respectively and \square is the step size.

Extended Diffusion

We extended anisotropic diffusion to multiple directions to be used to reveal weak radial and angular features in ultrasound images. This rich feature set can be used for contrast enhancement, image analysis, and segmentation. We apply anisotropic diffusion in eight directions, generating four diffused images computed independently regarding four directional pairs for each spatial location. Thus the anisotropic diffusion is computed for North-South and East-West directions separately and is extended by introducing two new directions as North-East

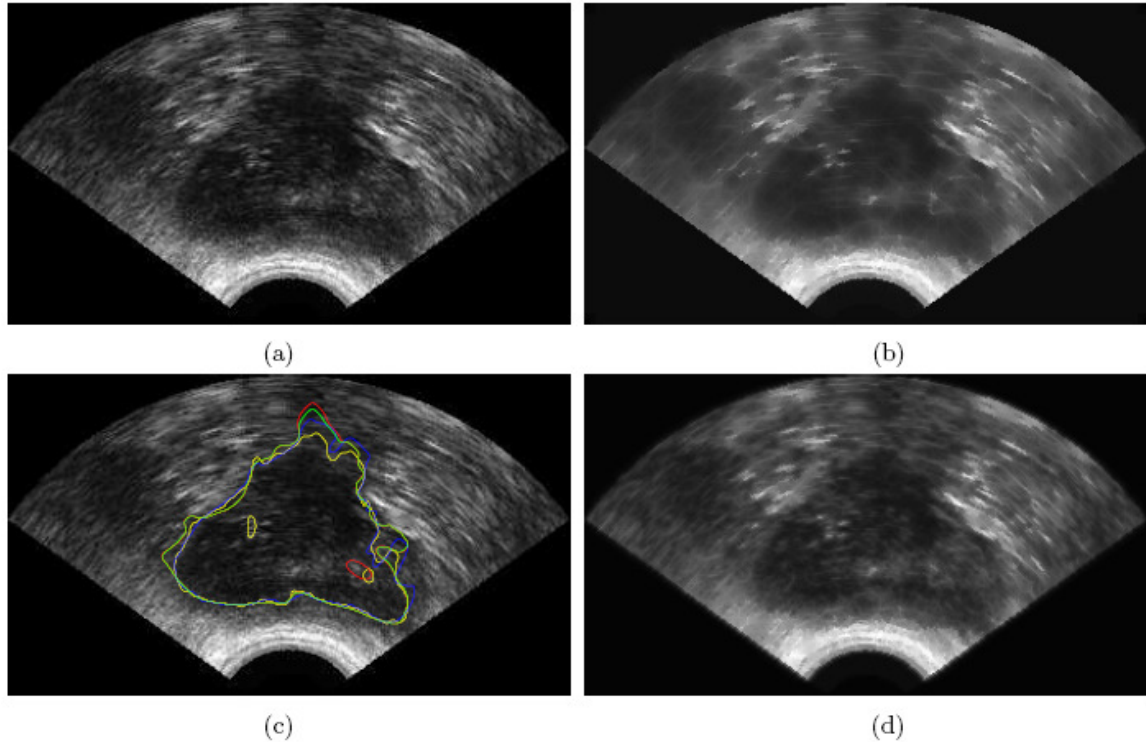


Figure 3: (a) Unprocessed TRUS image. (b) Extracted maximum diffusion over all diffused images. (c) Deformations of different contours over specific diffusion directions. (d) Application of adaptive anisotropic diffusion and extraction of maximum diffusion over all diffused images.

to South-West, i.e., 45 degree, and North-West to South-East, i.e., 135 degree. Recalling (1), considering the gradient in a specific single direction, diffusion can be simplified as

$$\nabla \cdot (c(\bar{x}, t) \nabla I_R(\bar{x}, t)) = \frac{\partial}{\partial R} (c(\bar{x}, t) I_R) \quad (7)$$

where R is any of four direction pairs, i.e., horizontal, vertical, 45 degree, and 135 degree. By definition, $c(x, t) I_R$ is called flux in R direction, $\Phi(I_R)$. One dimensional diffusion can be decomposed to the four independent fluxes as:

$$\frac{\partial}{\partial R} (\Phi(I_R)) = \Phi'(I_R) \cdot I_{RR} \quad (8)$$

where $c(x, t)$ is replaced with $f(\|\nabla(\bar{x}, t)\|)$.

Adaptive Anisotropic Diffusion

Most often a semi-supervised method is used for region segmentation for example by initializing a seed or initiate a contour. We propose an adaptive anisotropic diffusion such that the estimated statistics which are extracted from the region of interest can be used to adaptively select the conduction function which potentially may lead to a better within-region smoothness while could preserve the region boundaries. Therefore, to encourage the smoothness and discourage the diffusion based on image statistics and image features more effectively, an adaptive conduction is introduced.

$$c(\bar{x}, t) = \exp\left(-\left(\frac{\|\nabla(\bar{x}, t)\|}{\kappa}\right)^2\right) \quad \forall \bar{x} \in G \quad (9)$$

otherwise,

$$c(\bar{x}, t) = C \quad \forall \bar{x} \notin G \quad (10)$$

where C is a constant conduction factor and G is some definition of the region, for example

$$\bar{x} \in G \text{ if } I_{\bar{x}}^t \in [\mu - \sigma, \mu + \sigma]$$

for some region mean and standard deviation. Diffusion is encouraged within the region based on constant conduction factor. However, outside the region, diffusion will follow the formal conduction function (4).

Level Sets

A level set contour will be initialized and deformed toward prostate boundary. External force that controls the deformations of the zero level set toward the prostate boundary is extracted by applying the anisotropic diffusion in four directional diffusion pairs. To generate the external force for each spatial location of the TRUS image, the maximum diffusion among directional diffusions is selected.

Starting contour which is the zero level set of a 3D volume will be initialized by user interactions. It could also be initialized as an elliptical level set by selecting the ellipse centre. The statistics of the region of the interest are estimated over the initialized contour to be used for adaptive anisotropic diffusion. Having the zero level set initialized as an elliptical contour, velocity function F is designed based on the extracted features from diffused images so that the interface evolves toward the prostate boundary and converges in its vicinity.

$$F = -\exp\{\epsilon \kappa - \alpha \times D_v\} \quad (11)$$

where κ and \square are curvature and curvature coefficient respectively, D_v is velocity derived by external energy which is extracted from diffused images, α is a velocity constant coefficient and

$$\kappa = \nabla \cdot \frac{\nabla \Phi}{|\nabla \Phi|} = \left\{ \frac{\Phi_{xx}\Phi_y^2 - 2\Phi_x\Phi_y\Phi_{xy} + \Phi_{yy}\Phi_x^2}{(\Phi_x^2 + \Phi_y^2)^{\frac{3}{2}}} \right\} \quad (12)$$

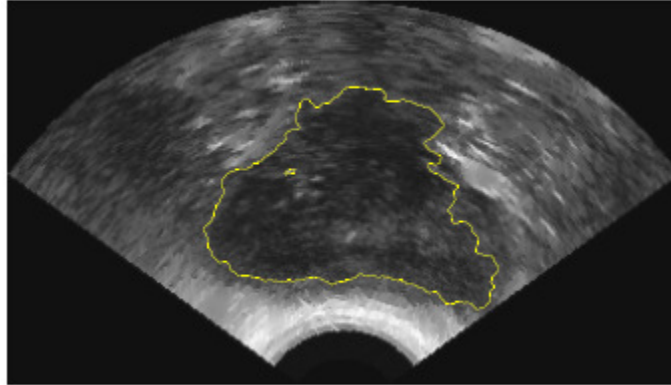


FIGURE 4: The deformations of the level set where the contour converges and stops in the vicinity of the prostate boundary.

3. Results

As depicted in Fig. 1, TRUS prostate images have very weak contrast, are corrupted with the speckle noise, have short range of gray levels, maintain very weak prostate region texture, and are defected by shadow regions. Although the segmentation of Fig. 1(Top) seems to be less challenging, Fig. 1(Bottom) is very difficult to be segmented. The ordinary image analysis methods perform very poorly for segmentation of prostate region in poor contrast TRUS images.

The proposed extended anisotropic diffusion and its adaptive version are applied for image enhancement and feature extraction. Application of extended anisotropic diffusion in eight directions is depicted in Fig. 2. Fig. 3(b) shows the maximum diffusion obtained over independently computed directional diffusions. As we can observe, independent directional diffused images provide rich features to be used for segmentation and analysis. Deformations of different contours over specific diffusion directions are superimposed on the unprocessed image in Fig. 3(c).

As it is depicted in Fig. 3, not only the contrast is improved, but also some features of prostate region are recovered. Moreover, the directional diffusion features can be effectively used to segment the prostate region. For example, these features can be used in conjunction with structure tensor and vector flow to design the diffusion based velocity functions. The application of adaptive conduction to TRUS image is depicted in Fig. 3(d). The adaptive conduction is computed based on prostate region statistics where mean and standard deviation are computed over the level set contour. A sample segmentation of the TRUS image superimposed on the original image is depicted in Fig. 4.

4. Conclusions and Discussions

Prostate texture analysis is a very difficult and challenging task due to the poor contrast, weak texture structure, the speckle noise, and shadow regions. Hence the standard and conventional image processing methods are not capable of segmentation of the prostate region. In this paper we introduced an adaptive anisotropic diffusion method to improve the selective within-region image smoothness while preserving the region boundaries. Moreover, the anisotropic diffusion is extended to eight directions, i.e., four directional pairs such that diffusion in each direction is computed independently yielding a rich feature set for ultrasound image enhancement and analysis. The preliminary results are quite promising, leading to our future work to extend the method to be employed for multiple region segmentation.

5. REFERENCES

1. Cancer facts and figures, <http://www.cancer.org>, American Cancer society.
2. C.K. Kwoh, M. Teo, W. Ng, S. Tan, and M. Jones, "Outlining the prostate boundary using the harmonics method", *Med. Biol. Eng. Computing*, 36:768-771, 1998.
3. R.G. Aarnink, R.J.B. Giesen, A. L. Huynen, J. J. de la Rosette, F.M. Debruyne, and H. Wijkstra "A practical clinical method for contour determination in ultrasound prostate images", *Ultrasound Med. Biol.*, 20:705-717, 1994.
4. R.G. Aarnink, S.D. Pathak, J. J. de la Rosette, F.M. Debruyne, Y. Kim, and H. Wijkstra, "Edge detection in ultrasound prostate images using integrated edge map", *Ultrasound Med. Biol.*, 36:635-642, 1998.
5. Y. Zhan and D. Shen, "Deformable segmentation of 3-d ultrasound prostate images using statistical texture matching method", *IEEE Transactions on Medical Imaging*, 25(3):256-272, 2006.
6. D. Freedman, R.J. Radke, T. Zhang, Y. Jeong, D.M. Lovelock, and G.T.Y. Chen, "Model-based segmentation of medical imagery by matching distributions", *IEEE Transactions on Medical Imaging*, 24:281-292, 2005.
7. A. Ghanei, H. Soltanian-Zadeh, A. Ratkewicz, and F. Yin, "A three dimensional deformable Model for segmentation of human prostate from ultrasound images", *Med. Phys*, 28:2147-2153, 2001.
8. C. Knoll, M. Alcaniz, V. Grau, C. Monserrat, and M. Juan, "Outlining of the prostate using snakes with shape restrictions based on the wavelet transform", *Pattern Recognition*, 32:1767-1781, 1999.
9. S. D. Pathak, V. Chalana, D. haynor, and Y. kim, "Edge guided boundary delineation in prostate ultrasound images", *IEEE Transactions on Medical Imaging*, 19:1211-1219, 2000.
10. S. S. Mohamed, T. K. Abdel-galil, M. M. Salma, A. Fenster, D. B. Downey, and K. Rizkalla, "Prostate cancer diagnosis based on gabor filter texture segmentation of ultrasound image", *IEEE CCECE*, 1485-1488, 2003.
11. P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(7):629-639, 1990.
12. G. Gerig, O. Kbler, R. Kikinis, , and F. Jolesz, "Nonlinear anisotropic filtering of MRI data", *IEEE Transactions on Medical Imaging*, 11(2):221-232, 1992.
13. G. Sapiro and A. Tannenbaum, "Edge preserving geometric smoothing of MRI data", Technical report, University of Minnesota, Dept. of Electrical Engineering, 1994.

Algorithm to Generate Wavelet Transform from an Orthogonal Transform

Dr. H. B. Kekre

Senior Professor, SVKM's NMIMS,
Mumbai-56, India

hbkekre@yahoo.com

Archana Athawale

Ph. D. Research Scholar,
MPSTME, SVKM's NMIMS, Mumbai-56.
Assistant Professor,
Thadomal Shahani Engineering College,
Mumbai-50, India

athawalearchana@gmail.com

Dipali Sadavarti

Lecturer, Computer Engineering Dept.,
Fr C .R.C. E, Mumbai-50, India

dipasadavarti@rediffmail.com

Abstract

This paper proposes algorithm to generate discrete wavelet transform from any orthogonal transform. The wavelet analysis procedure is to adopt a wavelet prototype function, called an *analyzing wave* or *mother wave*. Other wavelets are produced by translation and contraction of the mother wave. By contraction and translation infinite set of functions can be generated. This set of functions must be orthogonal and this condition qualifies a transform to be a wavelet transform. Thus there are only few functions which satisfy this condition of orthogonality. To simplify this situation, this paper proposes a generalized algorithm to generate discrete wavelet transform from any orthogonal transform. For an $N \times N$ orthogonal transform matrix T , element of each row of T is repeated N times to generate N Mother waves. Thus rows of original transform matrix become wavelets. As an example we have illustrated the procedure of generating Walsh Wavelet from Walsh transform. Since data compression is one of the best applications of wavelets, we have implemented image compression using Walsh as well as Walsh Wavelet. Our experimental results show that performance of image compression technique using Walsh Wavelet is much better than that of standard Walsh transform. More over image reconstructed from Walsh transform has some blocking artifact, which is not present in the image reconstructed from Walsh Wavelet. Similarly image compression using DCT and DCT Wavelet has been implemented. Again the results of DCT Wavelet have been proved to perform better than normal DCT.

Keywords: Wavelet Transform, Walsh Wavelet, DCT Wavelet, Image compression.

1. INTRODUCTION

In the past few years, the study of wavelets and the exploration of the principles governing their behavior have brought about sweeping changes in the disciplines of pure and applied mathematics and sciences. One of the most significant developments is the realization that, in addition to the canonical tool of representing a function by its Fourier series, there is a different representation more adapted to certain problems in data compression, noise removal, pattern classification and fast scientific computations [1, 2]. Wavelets are mathematical functions that cut up data into different frequency components, and then study each component with a resolution matched to its scale. They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes [1-4]. Fourier analysis is a global scheme. While we can analyze the frequencies that make up a signal, the local properties of the signal cannot be easily detected from the Fourier coefficients. To overcome this difficulty STFT (Short Time Fourier Transform) [5] was introduced. However it gives only local properties at the cost of global properties. Wavelets overcome this shortcoming of Fourier analysis [6, 7] as well as STFT. They provide a way to do a time-frequency analysis simultaneously. Another advantage of wavelet is that they process data at different *scales* or *resolutions*. If we look at a signal with a large "window," we would notice gross features. Similarly, if we look at a signal with a small "window," we would notice local features. The result in wavelet analysis is to see both the forest *and* the trees [1, 3]. The wavelet analysis procedure is to adopt a wavelet prototype function, called an *analyzing wave* or *mother wave*. Other wavelets are produced by translation and contraction of the mother wave. By contraction and translation infinite set of functions can be generated. This set of functions must be orthogonal and this condition qualifies a transform to be a wavelet transform. Thus a transform is qualified to be a wavelet transform if and only if it satisfies the condition of orthogonality. Thus there are only few functions which satisfy these conditions. Principal advantage of wavelet is that they provide time-frequency localization. Temporal analysis is performed with a contracted, high-frequency version of the prototype wavelet, while frequency analysis is performed with a dilated, low-frequency version of the same wavelet. The exact shape of the mother wave strongly affects the accuracy and compression properties of the approximation. Because the original signal or function can be represented in terms of a wavelet expansion (using coefficients in a linear combination of the wavelet functions), data operations can be performed using just the corresponding wavelet coefficients. And if you further choose the best wavelets adapted to your data, or truncate the coefficients below a threshold, your data is sparsely represented. This sparse coding makes wavelets an excellent tool in the field of data compression.

This paper presents a generalized algorithm to develop wavelet from any orthogonal transform. Purpose of this paper is to generate orthogonal discrete wavelet transform from existing orthogonal transform. In [8] Kekre's wavelet transform is generated from Kekre's transform [9]. We propose that the same algorithm can be used to generate wavelets from any orthogonal transform. As an example we have illustrated how Walsh Wavelet can be generated from standard Walsh transform. A necessary condition to generate a discrete wavelet transform from any transform is that the transform from which the wavelet is build must be orthogonal. For example Walsh, DCT, DST, Slant etc. are orthogonal transforms so we can generate Walsh Wavelet, DCT Wavelet, DST Wavelet, Slant Wavelet respectively. The paper is organized as follows. Section 2 discusses a generalized algorithm to generate Wavelet from an orthogonal transform; Section 3 illustrates generation of Walsh Wavelet from Walsh transform. Section 4 discusses properties of Walsh Wavelet. In Section 5 we have explored image compression technique using Walsh and Walsh Wavelet. In Section 6, experimental results of image compression using Walsh and Walsh Wavelet as well as DCT and DCT Wavelet are discussed. Finally in Section 7 the paper is concluded.

2. GENERATING WAVELET FROM ANY ORTHOGONAL TRANSFORM

From any $N \times N$ orthogonal transform T , we can generate wavelet transform matrix of size $N^2 \times N^2$. For example, from orthogonal transform T of size 5×5 , we can generate corresponding Wavelet transform matrix of size 25×25 . In general $M \times M$ Wavelet transform matrix can be generated from

$N \times N$ orthogonal basic transform T such that $M = N^2$. Consider any orthogonal transform T of size $N \times N$ shown in FIGURE 1.

T_{11}	T_{12}	T_{13}	...	$T_{1(N-1)}$	T_{1N}
T_{21}	T_{22}	T_{23}	...	$T_{2(N-1)}$	T_{2N}
T_{31}	T_{32}	T_{33}	...	$T_{3(N-1)}$	T_{3N}
.
.
T_{N1}	T_{N2}	T_{N3}	...	$T_{N(N-1)}$	T_{NN}

FIGURE 1: $N \times N$ orthogonal transform

1st Column of
T repeated N times

2nd Column of
T repeated N times

Nth Column of
T repeated N times

T_{11}	T_{11}	...	T_{11}	T_{12}	T_{12}	...	T_{12}	...	T_{1N}	T_{1N}	...	T_{1N}
T_{21}	T_{21}	...	T_{21}	T_{22}	T_{22}	...	T_{22}	...	T_{2N}	T_{2N}	...	T_{2N}
T_{31}	T_{31}	...	T_{31}	T_{32}	T_{32}	...	T_{32}	...	T_{3N}	T_{3N}	...	T_{3N}
.
.
T_{N1}	T_{N1}	...	T_{N1}	T_{N2}	T_{N2}	...	T_{N2}	...	T_{NN}	T_{NN}		T_{NN}
T_{21}	T_{22}	...	T_{2N}	0	0	...	0	...	0	0	0	0
0	0	...	0	T_{21}	T_{22}	...	T_{2N}	...	0	0	0	0
	.								.			
	.								.			
	.								.			
0	0	...	0	0	0	...	0	...	T_{21}	T_{22}	...	T_{2N}
T_{31}	T_{32}	...	T_{33}	0	0	...	0	...	0	0	...	0
0	0	...	0	T_{31}	T_{32}	...	T_{33}	...	0	0	...	0
	.								.			
	.								.			
0	0	...	0	0	0	...	0	...	T_{31}	T_{32}	...	T_{33}
		
		
		
T_{N1}	T_{N2}	...	T_{NN}	0	0	...	0		0	0	...	0
0	0	...	0	T_{N1}	T_{N2}	...	T_{NN}		0	0	...	0
0	0	...	0	0	0	...	0		T_{N1}	T_{N2}	...	T_{NN}

FIGURE 2: $M \times M$ Wavelet transform matrix ($M=N^2$) generated from orthogonal transform T of size $N \times N$.

FIGURE 2 shows $M \times M$ Wavelet transform matrix ($M = N^2$) generated from $N \times N$ orthogonal transform T . First N number of rows of Wavelet transform matrix is generated by repeating every column of transform T , N times. To generate next $(N+1)$ to $2N$ rows, second row of transform T is translated. To generate next $(2N+1)$ to $3N$ rows, third row of transform T is translated. Like wise to generate last $((N-1)N + 1)$ to N^2 rows, N^{th} row of transform T is used. Note that by repeating every column of the basic transform N times we are generating Mother wave. Other wavelets are generated by contraction and translation of the mother wave.

3. GENERATING WALSH WAVELET FROM WALSH TRANSFORM

In [8] Kekre's Wavelet transform is generated from Kekre's Transform [9]. From $N \times N$ Kekre's transform matrix we can generate $2N \times 2N$, $3N \times 3N$, ..., $N^2 \times N^2$ Kekre's Wavelet transform. Similar technique can be applied to any orthogonal transform of size $N \times N$ to generate wavelet of size $N^2 \times N^2$. Since Walsh is an orthogonal transform, we can generate Wavelet from Walsh transform. Consider 4×4 Walsh transform matrix shown in FIGURE 3. Procedure of generating 16×16 Walsh Wavelet from 4×4 Walsh transform is illustrated in FIGURE 4.

1	1	1	1
1	1	-1	-1
1	-1	-1	1
1	-1	1	-1

FIGURE 3: 4×4 Walsh Transform matrix W_4

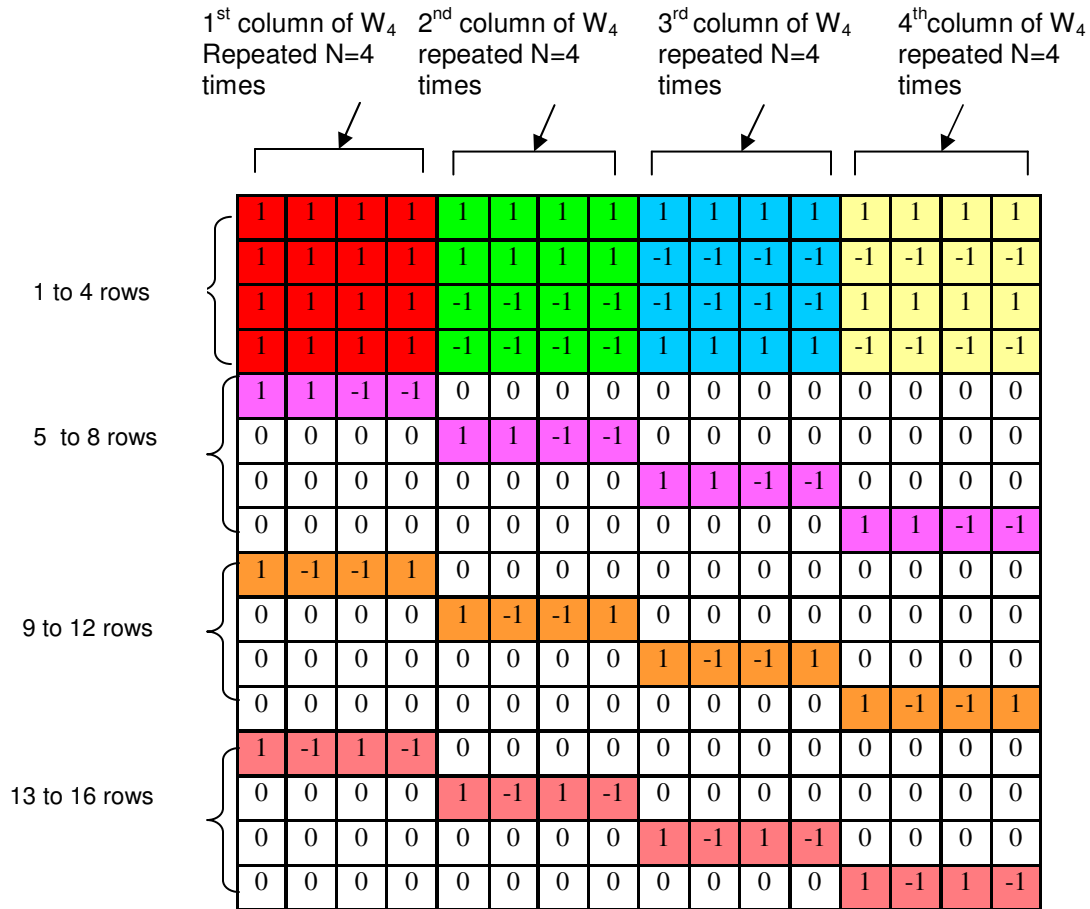


FIGURE 4: 16x16 Walsh Wavelet generated from 4x4 Walsh transform

4. PROPERTIES OF WALSH WAVELET

4.1 Orthogonal

The transform matrix M is said to be orthogonal if the following condition is satisfied.

$$[M][M]^T = [D]$$

Where D is the diagonal matrix. $N^2 \times N^2$ Wavelet generated from any orthogonal transform of size $N \times N$ satisfies this property and hence it is orthogonal.

4.2 Non Involutorial

An involutory function is a function that is its own inverse [7]. So involutorial transform is a transform which is inverse transform of itself. Walsh Wavelet is non involutorial transform.

4.3 Transform on Vector

The Walsh Wavelet transform (WLT) on a column vector f is given by

$$F = [WLT] f$$

And inverse is given by

$$f = [WLT]^T [D]^{-1} F$$

4.4 Transform on 2D Image

Walsh Wavelet transform on 2D image [f] is given by

$$[F] = [WLT] [f] [WLT]^T$$

Obtaining Inverse:

Calculate Diagonal matrix D (FIGURE 5) as,

$$[D] = [WLT] [WLT]^T$$

D1	0	0	0	0	0
0	D2	0	0	0	0
0	0	D3	0	0	0
0	0	0	...	0	0
0	0	0	0	...	0
0	0	0	0	0	DN

FIGURE 5: Diagonal matrix D

Inverse is calculated as

$$[f] = [WLT]^T [F_{ij} / D_{ij}] [WLT]$$

Where $D_{ij} = D_i * D_j$; $1 \leq i \leq N$ and $1 \leq j \leq N$

5. IMAGE COMPRESSION USING WALSH AND WALSH WAVELET

Image compression is the process of converting image files into smaller files for efficiency of storage and transmission [10-13]. Thus a compression is the process of representing data in compact form. In this section, we discuss the algorithm to compress an image using Walsh transform and Walsh Wavelet.

5.1 Image Compression using Walsh:

Step 1: Walsh transform is first applied on a full image. We get the transformed image.

Step 2: The transformed image is then divided into 64 equal non-overlapping blocks.

Step 3: Energy of each block is computed as summation of square of the coefficients within that block.

Step 4: Sort all 64 blocks in ascending order according to their energies. Thus the first block in the sorted list is the lowest energy block and the last block in the sorted list is the highest energy block.

Step 5: Input number of blocks to be compressed say M.

Step 6: Compress first M blocks (of the transformed image) from the sorted list. Compressing an image is nothing but making coefficients of the selected block zero.

Step 7: Apply Inverse Walsh transform to reconstruct the image.

5.2 Image Compression using Walsh Wavelet

Step 1: Walsh Wavelet Transform (WLT) is first applied on full image f of size $N \times N$. The resultant image F is,

$$F = [WLT] [f] [WLT]^T$$

Step 2: The diagonal matrix D is computed as $D = [WLT] [WLT]^T$

Step 3 : Compute $G = [F_{ij} / D_{ij}]$ i.e. every element of transformed image F is divided by corresponding D_{ij} value.

$$\text{Where } D_{ij} = D_i * D_j \quad ; 1 \leq i \leq N \text{ and } 1 \leq j \leq N$$

Step 4: Calculate energy of each element of G as,

$$E_{ij} = [F_{ij} / D_{ij}]^2 * D_{ij}$$

Step 5: Divide the whole G image into 64 equal non-overlapping blocks. Compute energy of each block as summation of energy of each element within the block.

Step 6: Calculate percentage energy of each block.

Step 7: Sort the blocks in ascending order of their percent energy.

Step 8: Input number of blocks to be compressed say M .

Step 9: Compress first M number of blocks from the sorted list (Make all coefficients of the block zero)

Step 10: Apply inverse Walsh Wavelet transform to reconstruct the image.

$$\text{Reconstructed}_f = [WLT]^T [G] [WLT]$$

6. RESULTS AND DISCUSSION

Algorithms discussed in previous section were implemented on a bitmap gray image of size 256×256 . Image was compressed using both Walsh as well as Walsh Wavelet transform as explained in section 5. 256×256 Walsh Wavelet was generated from 16×16 Walsh transform and applied on to the image. TABLE1 compares the results of Data compression using Walsh and Walsh Wavelet.

Image	Compression Ratio	No of Blocks compressed	Walsh		Walsh Wavelet		MSE Reduction factor	Increased PSNR
			PSNR	MSE	PSNR	MSE		
Horse	0.625	40	27.89	105.61	35.37	18.86	5.599	7.48
	0.75	48	27.46	116.61	32.7	34.9	3.32	5.24
	0.82	53	27.14	125.48	29.12	79.62	1.57	1.98
Temple	0.625	40	24.72	218.85	32.69	34.93	6.26	7.97
	0.75	48	24.46	232.41	29.09	80.16	2.89	4.63
	0.82	53	24.21	246.26	27.07	127.42	1.93	2.86
Rose	0.625	40	33.83	26.89	43.08	3.19	8.42	9.25
	0.75	48	33.45	29.35	38.47	9.23	3.17	5.02
	0.82	53	33.14	31.54	35.42	18.66	1.69	2.28
Puppy	0.625	40	30.04	64.34	34.2	24.68	2.60	4.16
	0.75	48	29.56	71.83	32.33	38	1.89	2.77
	0.82	53	29.19	78.17	30.97	51.94	1.50	1.78

TABLE 1: Results of image compression using Walsh and Walsh Wavelet.

Our experimental results clearly show that performance of Walsh Wavelet is far better than Walsh transform for compression. Compression using Walsh Wavelet achieves better PSNR and MSE values for all type of images. On an average MSE is reduced by factor of 5 to 8, 2 to 3 and 1.5 to 2.0 for compression ratio 0.625, 0.75 and 0.82 respectively. PSNR is increased by 7 to 9, 4 to 5 and 2 to 3 dB for compression ratio 0.625, 0.75 and 0.82 respectively.

FIGURE 6 shows the original and reconstructed image using Walsh and Walsh Wavelet.



FIGURE 6 (a): Original Image

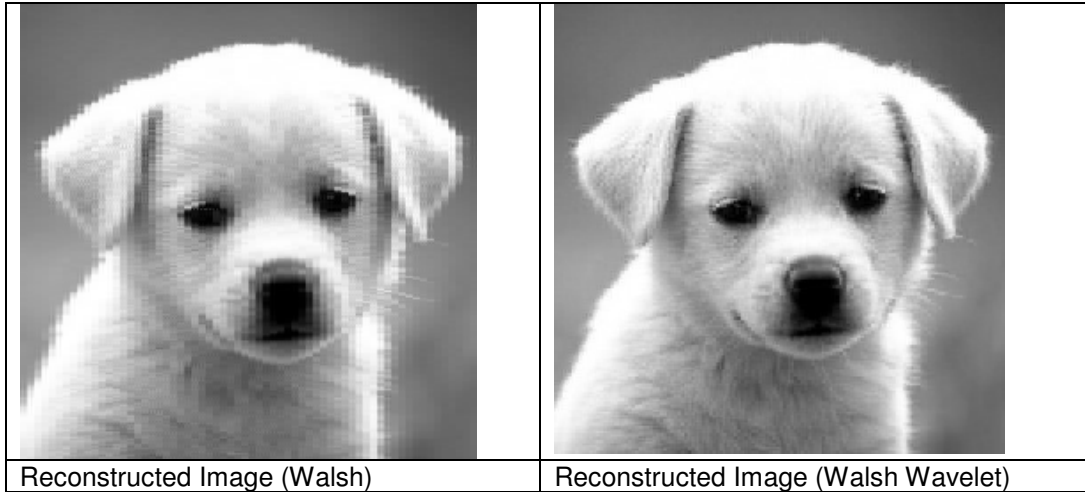


FIGURE 6 (b): Reconstructed Image (No. Of Blocks Compressed = 40)

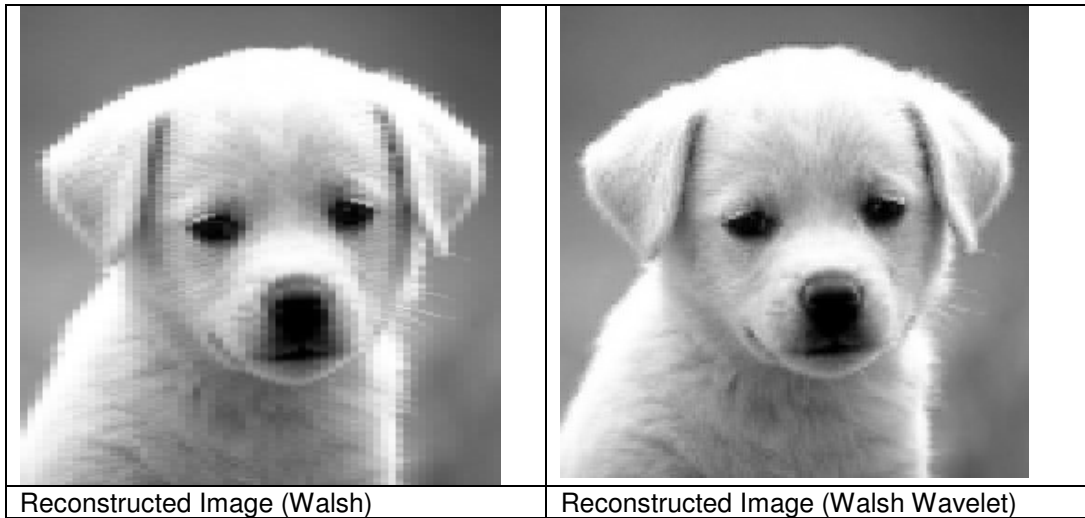


FIGURE 6 (c): Reconstructed Image (No. Of Blocks Compressed = 48)

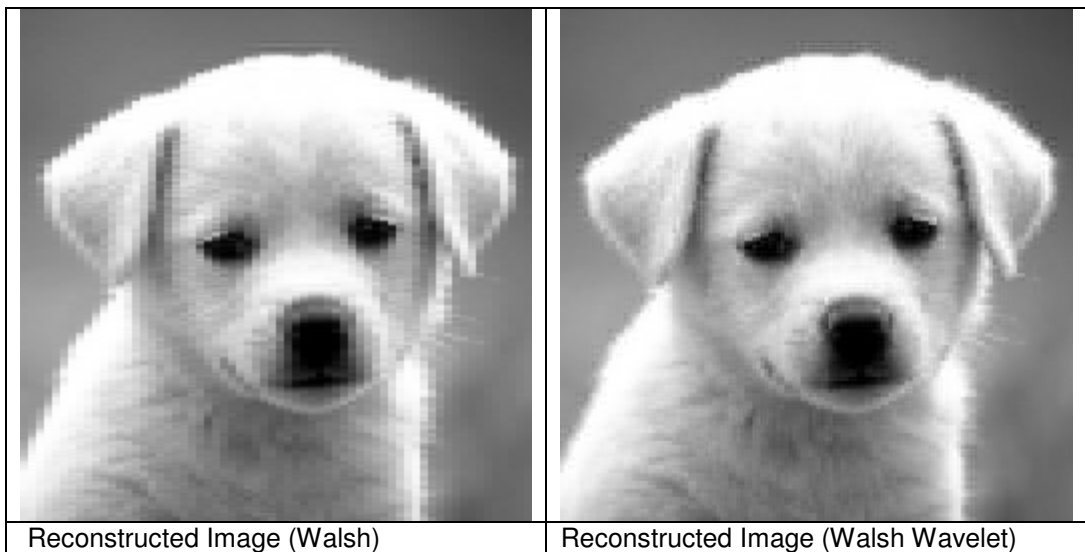


FIGURE 6 (d): Reconstructed Image (No. Of Blocks Compressed = 53)

Our results show that reconstructed image using Walsh has blocking artifact, where as image reconstructed from Walsh Wavelet is free from this type of blocking artifact.

Similarly we have implemented image compression using DCT and DCT Wavelet. Again DCT Wavelet of size 256x256 is generated from 16x16 DCT. TABLE 2 summarizes the results of image compression technique using DCT and DCT Wavelet.

Image	Compression Ratio	No of Blocks compressed	DCT		DCT Wavelet		MSE Reduction factor	Increased PSNR
			PSNR	MSE	PSNR	MSE		
Horse	0.625	40	31.21	49.2	40.66	5.58	8.81	9.45
	0.75	48	31.19	49.33	36.68	13.95	3.53	5.49
	0.82	53	31.07	50.78	33.9	26.44	1.92	2.83
Temple	0.625	40	27.74	109.32	38.02	10.24	10.67	10.28
	0.75	48	27.73	109.58	33.67	28.52	3.84	5.94
	0.82	53	27.65	111.56	30.66	55.85	1.99	3.01
Rose	0.625	40	38.83	8.51	52.23	0.38	22.39	13.4
	0.75	48	38.83	8.51	47.57	1.13	7.53	8.74
	0.82	53	38.75	8.66	42.24	3.87	2.23	3.49
Puppy	0.625	40	33.09	31.85	35.49	18.36	1.73	2.4
	0.75	48	32.79	34.18	33.86	26.69	1.28	1.07
	0.82	53	32.37	37.64	32.69	34.96	1.07	30.32

TABLE 2: Results of image compression using DCT and DCT Wavelet

Again the image compression using DCT Wavelet is proved to be superior to image compression using DCT. FIGURE 7 (a) compares the MSE values obtained from Walsh, Walsh Wavelet, DCT and DCT Wavelet for four different gray images for 62.5% compression that is 40 out of 64 blocks are compressed.

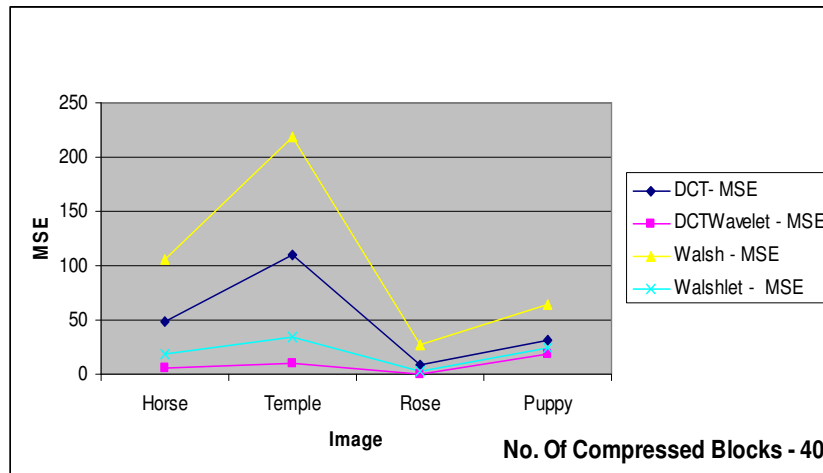


FIGURE 7 (a): Comparison of Data compression technique using DCT, DCT Wavelet, Walsh and Walsh Wavelet with respect to MSE (For Compression ratio 0.625)

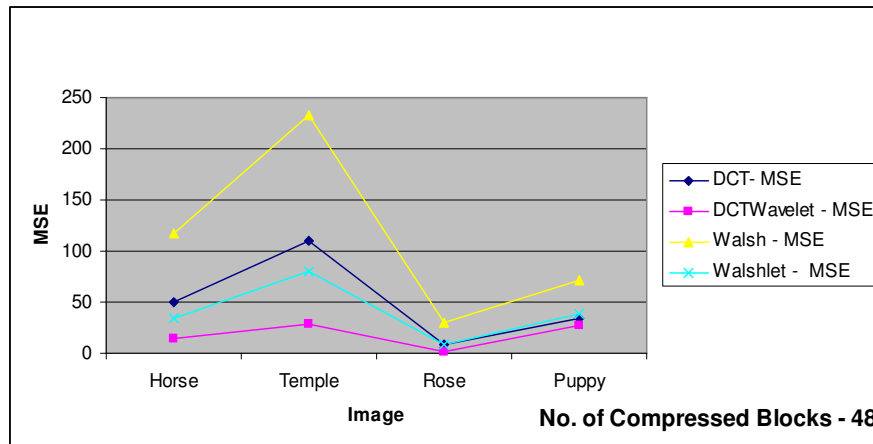


FIGURE 7 (b): Comparison of Data compression technique using DCT, DCT Wavelet, Walsh and Walsh Wavelet with respect to MSE (For Compression ratio 0.75)

From FIGURE 7 it is clear that MSE value between original image and reconstructed image of data compression technique using DCT Wavelet and Walsh Wavelet is much lower than that of using DCT and Walsh.

7. CONCLUSION

The wavelet analysis procedure is to adopt a wavelet prototype function, called an *analyzing wave* or *mother wave*. Other wavelets are produced by translation and contraction of the mother wave. A transform thus formed qualifies to be a wavelet transform if and only if it satisfies the condition of orthogonality. Thus there are only few functions which satisfy these conditions. To simplify this situation, this paper proposes a generalized algorithm to generate discrete wavelet transform from any orthogonal transform. For an $N \times N$ orthogonal transform matrix T , element of each row of T is repeated N times to generate N Mother waves. Thus rows of original transform matrix become wavelets. As an example we have illustrated how Walsh Wavelet is generated from Walsh transform. Image compression using Walsh Wavelet has been implemented and the results are compared with image compression using Walsh transform. Compression using Walsh Wavelet achieves much better PSNR and MSE values for all type of images. On an average MSE is reduced by factor of 5 to 8, 2 to 3 and 1.5 to 2.0 for compression ratio 0.625, 0.75 and 0.82 respectively. PSNR is increased by 7 to 9, 4 to 5 and 2 to 3 dB for compression ratio 0.625, 0.75 and 0.82 respectively. Results of compression have been observed for DCT and DCT Wavelet as well. Again the image compression using DCT Wavelet is proved to be superior to image compression using DCT. Moreover the reconstructed image from Walsh Wavelet is free from blocking artifact which appears in reconstructed image from Walsh Transform.

8. REFERENCES

- [1]. K. P. Soman and K.I. Ramachandran. "*Insight into WAVELETS From Theory to Practice*", Printice -Hall India, pp 3-7, 2005.
- [2]. Raghuvveer M. Rao and Ajit S. Bopardika. "*Wavelet Transforms – Introduction to Theory and Applications*", Addison Wesley Longman, pp 1-20. 1998.
- [3]. C.S. Burrus, R.A. Gopinath, and H. Guo. "*Introduction to Wavelets and Wavelet Transform*" Prentice-hall International, Inc., New Jersey, 1998.
- [4]. Amara Graps, "An Introduction to Wavelets", IEEE Computational Science and Engineering, Summer 1995, vol. 2, num. 2, published by the IEEE Computer Society,

10662 Los Vaqueros Circle, Los Alamitos, CA 90720, USA.

- [5]. Julius O. Smith III and Xavier Serra^P, "An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation", Proceedings of the International Computer Music Conference (ICMC-87, Tokyo), Computer Music Association, 1987.
- [6]. S. Mallat, "A Theory of Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674-693, 1989.
- [7]. Strang G. "Wavelet Transforms Versus Fourier Transforms." *Bull. Amer. Math. Soc.* 28, 288-305, 1993.
- [8]. Dr. H. B. Kekre, Archana Athawale and Dipali Sadavarti", "Algorithm to generate Kekre's Wavelet Transform from Kekre's Transform", *International Journal of Engineering Science and Technology, IJEST*, Vol. 2(5), 2010, 756-767
- [9]. Dr. H. B. Kekre and Sudeep Thepade", "Image Retrieval Using Non-Involutorial Orthogonal Kekre's Transform", *International J. of Multi discipline. Research & Advances in Engineering, IJMRAE*, Vol. 1, No. 1, November 2009, pp 189-203.
- [10]. DeVore, R.; Jawerth, B.; and Lucier, B. "Images Compression through Wavelet Transform Coding." *IEEE Trans. Information Th.* 38, 719-746, 1992.
- [11]. D. S. Taubman and M. W. Marcellin , "*JPEG2000: Image Compression, Fundamentals, Standards, and Practice* , Kluwer Academic Publishers, 2002.
- [12]. D. Salomon , "*Data Compression: The Complete Reference*", Springer Verlag, 2000.
- [13]. Amir Said and William A. Pearlman ", "An Image Multiresolution Representation for Lossless and Lossy Image Compression," *IEEE Transactions on Image Processing*, vol. 5, pp. 1303-1310, Sept. 1996.

CALL FOR PAPERS

Journal: International Journal of Image Processing (IJIP)

Volume: 4 **Issue:** 5

ISSN: 1985-2304

URL: <http://www.cscjournals.org/csc/description.php?JCode=IJIP>

About IJIP

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.

IJIP 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
- 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

- 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
- Storage and transmission
- 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
- Video coding algorithms and technologies for ATM/p

Important Dates

Volume: 4

Issue: 5

Paper Submission: September 30 2010

Author Notification: November 01, 2010

Issue Publication: November / December

CALL FOR EDITORS/REVIEWERS

CSC Journals is in process of appointing Editorial Board Members for ***International Journal of Image Processing (IJIP)***. CSC Journals would like to invite interested candidates to join **IJIP** network of professionals/researchers for the positions of Editor-in-Chief, Associate Editor-in-Chief, Editorial Board Members and Reviewers.

The invitation encourages interested professionals to contribute into CSC research network by joining as a part of editorial board members and reviewers for scientific peer-reviewed journals. All journals use an online, electronic submission process. The Editor is responsible for the timely and substantive output of the journal, including the solicitation of manuscripts, supervision of the peer review process and the final selection of articles for publication. Responsibilities also include implementing the journal's editorial policies, maintaining high professional standards for published content, ensuring the integrity of the journal, guiding manuscripts through the review process, overseeing revisions, and planning special issues along with the editorial team.

A complete list of journals can be found at <http://www.cscjournals.org/csc/byjournal.php>. Interested candidates may apply for the following positions through <http://www.cscjournals.org/csc/login.php>.

Please remember that it is through the effort of volunteers such as yourself that CSC Journals continues to grow and flourish. Your help with reviewing the issues written by prospective authors would be very much appreciated.

Feel free to contact us at coordinator@cscjournals.org if you have any queries.

Contact Information

Computer Science Journals Sdn Bhd

M-3-19, Plaza Damas Sri Hartamas
50480, Kuala Lumpur MALAYSIA

Phone: +603 6207 1607
 +603 2782 6991
Fax: +603 6207 1697

BRANCH OFFICE 1

Suite 5.04 Level 5, 365 Little Collins Street,
MELBOURNE 3000, Victoria, AUSTRALIA


Fax: +613 8677 1132

BRANCH OFFICE 2

Office no. 8, Saad Arcad, DHA Main Bulevard
Lahore, PAKISTAN

EMAIL SUPPORT

Head CSC Press: coordinator@cscjournals.org
CSC Press: cscpress@cscjournals.org
Info: info@cscjournals.org



COMPUTER SCIENCE JOURNALS SDN BHD
M-3-19, PLAZA DAMAS
SRI HARTAMAS
50480, KUALA LUMPUR
MALAYSIA