

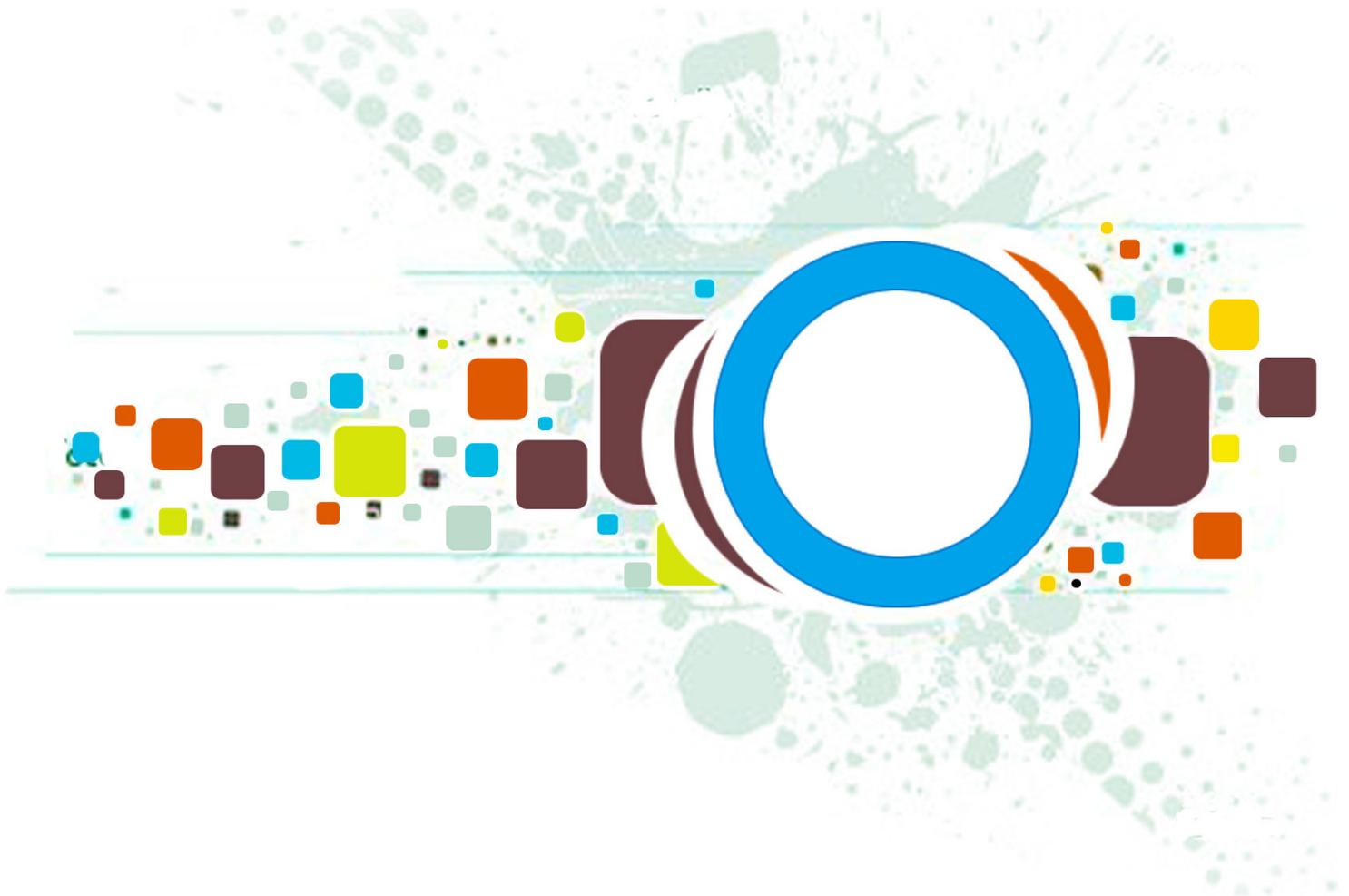
Volume 6 • Issue 6 • December 2012

Editor-in-Chief
Professor Hu, Yu-Chen

INTERNATIONAL JOURNAL OF
IMAGE PROCESSING (IJIP)

ISSN : 1985-2304

Publication Frequency: 6 Issues Per Year



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

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

VOLUME 6, ISSUE 6, 2012

**EDITED BY
DR. NABEEL TAHIR**

ISSN (Online): 1985-2304

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

IJIP Journal is a part of CSC Publishers

Computer Science Journals

<http://www.cscjournals.org>

INTERNATIONAL JOURNAL OF IMAGE PROCESSING (IJIP)

Book: Volume 6, Issue 6, December 2012

Publishing Date: 31-12- 2012

ISSN (Online): 1985-2304

This work is subjected to copyright. All rights are reserved whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication of parts thereof is permitted only under the provision of the copyright law 1965, in its current version, and permission of use must always be obtained from CSC Publishers.

IJIP Journal is a part of CSC Publishers

<http://www.cscjournals.org>

© IJIP Journal

Published in Malaysia

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

CSC Publishers, 2012

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 sixth issue of volume six of IJIP. The Journal is published bi-monthly, with papers being peer reviewed to high international standards. IJIP emphasizes on efficient and effective image technologies, and provides a central for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing. IJIP comprehensively cover the system, processing and application aspects of image processing. Some of the important topics are architecture of imaging and vision systems, chemical and spectral sensitization, coding and transmission, generation and display, image processing: coding analysis and recognition, photopolymers, visual inspection etc.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 7, 2013, IJIP will be appearing in more focused issues. Besides normal publications, IJIP intends to organize special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

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

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

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

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

Editorial Board Members

International Journal of Image Processing (IJIP)

EDITORIAL BOARD

EDITOR-in-CHIEF (EiC)

Professor Hu, Yu-Chen
Providence University (Taiwan)

ASSOCIATE EDITORS (AEiCs)

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

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

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

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

Professor Yen-Wei Chen
Ritsumeikan University
Japan

Associate Professor Tao Gao
Tianjin University
China

Dr Choi, Hyung Il
Soongsil University
South Korea

EDITORIAL BOARD MEMBERS (EBMs)

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 Chin-Feng Lee

Chaoyang University of Technology
Taiwan

Professor Santhosh.P.Mathew

Mahatma Gandhi University
India

Dr Hong (Vicky) Zhao

Univ. of Alberta
Canada

Professor Yongping Zhang

Ningbo University of Technology
China

Assistant Professor Humaira Nisar

University Tunku Abdul Rahman
Malaysia

Dr M.Munir Ahamed Rabbani

Qassim University
India

Dr Yanhui Guo

University of Michigan
United States of America

Associate Professor András Hajdu

University of Debrecen
Hungary

Assistant Professor Ahmed Ayoub

Shaqra University
Egypt

Dr Irwan Prasetya Gunawan

Bakrie University
Indonesia

Assistant Professor Concetto Spampinato

University of Catania
Italy

Associate Professor João M.F. Rodrigues

University of the Algarve
Portugal

Dr Anthony Amankwah
University of Witswatersrand
South Africa

Dr Chuan Qin
University of Shanghai for Science and Technology
China

AssociateProfessor Vania Vieira Estrela
Fluminense Federal University (Universidade Federal Fluminense-UFF)
Brazil

Dr Zayde Alcicek
firat university
Turkey

Dr Irwan Prasetya Gunawan
Bakrie University
Indonesia

TABLE OF CONTENTS

Volume 6, Issue 6, December 2012

Pages

- 397 - 402 Image Inpainting Using Cloning Algorithms
Shrilaxmi Deshpande & Shalini Bhatia
- 403 - 412 A Fuzzy Set Approach for Edge Detection
Pushpajit A.Khaire, Dr. Nileshsingh V. Thakur
- 413 - 421 Land Boundary Detection of an Island using improved Morphological Operation
Shreetam Behera, Mihir Narayan Mohant
- 422 - 440 Active Shape Model based On A Spatio-Temporal A Priori Knowledge: Applied To Left Ventricle Tracking In Scintigraphic Sequences
Said Eттаieb, Kamel Hamrouni, Su Ruan
- 441 - 466 Survey of The Problem of Object Detection In Real Images
Dilip K. Prasad
- 467 - 477 Hyperspectral Data Compression Using Spatial-Spectral Lossless Coding Technique
Ayman Mahmoud Ahmed, Salwa ElRamly, Mohamed El Sharkawy
- 478 - 487 Joint, Image-Adaptive Compression and Watermarking by GABased Wavelet Localization: Optimal Trade-Off between Transmission Time and Security
M. M. Ramya, R. Murugesan

Image Inpainting Using Cloning Algorithms

Ms. Shrilaxmi Deshpande

*Computer Engineering Department
Thadomal Shahani Engineering College
Mumbai-400050, India*

shrivd6686@gmail.com

Ms. Shalini Bhatia

*Computer Engineering Department
Thadomal Shahani Engineering College
Mumbai-400050, India*

shalini.tsec@gmail.com

Abstract

In image recovery image inpainting has become essential content and crucial topic in research of a new era. The objective is to restore the image with the surrounding information or modifying an image in a way that looks natural for the viewer. The process involves transporting and diffusing image information. In this paper to inpaint an image cloning concept has been used. Multiscale transformation method is used for cloning process of an image inpainting. Results are compared with conventional methods namely Taylor expansion method, poisson editing, Shepard's method. Experimental analysis verifies better results and shows that Shepard's method using multiscale transformation not only restores small scale damages but also large damaged area and useful in duplication of image information in an image.

Keywords: Image inpainting; Multiscale transformation; seamless cloning; poisson editing;

1. INTRODUCTION

Since from the period of Renaissance the technique of inpainting for missing area has been in practice. Image inpainting is nothing but restoring damaged region with information available in the surrounding area. It also includes modifying image with the help of inpainting algorithms. It is essential to fill the damaged region in such a way that it should not be identifiable by the viewer.

The artist starts canvassing from the external damaged region towards internal region to fill the damaged region, so that the image looks even. Based on this idea Bertalmio has proposed PDE-based algorithm for image inpainting [1] and [2], called as BSCB model. Chan and Shen presented a Total Variation model based on Rudin-Osher-Fatemi's Denoising bounded variation image model [3]. Atzori and de Natale introduced edge based algorithm for image inpainting [4]. A number of approaches have been introduced in later stages. But these can apply to repair small cracks and small scale damages of inpainted image. Another method has been proposed by using block of image to fill damaged region by Criminisi called block-based structure synthesis [5] and [6] and by Efros [7]. The main idea of this technique is to select appropriate size block of the damaged image in such a way that block should fill the damaged region. But this kind of algorithms need long time to repair damaged area.

In recent year mean-value-coordinate-theory [8] and adaptive grid techniques are used for seamless cloning which is used in image fusion and image reproduction. But these algorithms need lot of preprocessing and take much duration for seamless cloning. Image inpainting using Shepard's method using multiscale transformation method takes less time for cloning of a target region and as well as filling block within damaged region. The results are compared with conventional methods namely Taylor Series Method, Poisson editing, cloning using multiscale transformation. Experimental results show validity of image inpainting technique using Shepard's method.

2. IMAGE INPAINTING METHODS

Consider area R and area to be inpainted as Ω . Following methods are used for image inpainting.

2.1 Taylor Series Method

Taylor series expansion [9] for image inpainting is used to repair damaged image and to remove unwanted objects in an image. Taylor series expansion uses heat equation of the Partial Differential Equation (PDE). This fills the Ω by using the information present on the left and the right side of the damaged area depending on the shape, color and texture. Second order Taylor series is obtained by approximating u_{xx} on Taylor series expansion given as:

$$u_{xx}(x, y, t) \approx \frac{2}{h^2} [u(x+h, y, t) - u(x, y, t) - u_x(x, y, t)h] \tag{1}$$

This method recovers damaged area and removes unwanted objects of image but the process is slow and recovered area is not seamless.

2.2 Cloning Algorithms

In recovery from the damaged image cloning algorithms are used, in which the user specifies the co-ordinates for the area known as source domain. Damaged area is target domain. To interpolate the source domain with target domain seamlessly cloning algorithms are used. Steps of algorithms:

1. Identify the co-ordinate of target domain.
2. Specify the approximate co-ordinates of source domain which is to be used for recovery.
3. Mask source domain and select the region of interest.
4. Apply cloning algorithms namely poisson editing, Multiscale transformation method, Shepard's method to get the seamless inpainted image.

2.2.1 Poisson Editing

Poisson editing method [10] is a mathematical tool used for seamless editing and deriving cloning of selected region. Poisson editing includes Poisson equation:

$$\Delta f = \text{div } \nabla g \tag{2}$$

With Dirichlet boundary condition:

$$f|_{\partial P_A} = f^* \tag{3}$$

where g is target cloning domain and f is source cloning domain. Point p_A represents point on boundary ∂ for interpolation of source and target domain. This algorithm uses laplacian pyramid [11]. This incorporates cloning to remove and add objects seamlessly.

2.2.2 Multiscale transformation method

The method consists of multiscale scheme which resembles the Laplacian pyramid[12]. Repeatedly upsampling and downsampling are performed over image and convolved with and fixed width kernels, so as to operate on all scales of images [9].

Multiscale transformation is performed on both source region and masked target region. The forward transformation consists of convolving signal with filter h_1 and by factor of two it is subsampled. On subsampled data the process is repeated. At each level unsampled and unfiltered data is kept and compute:

$$a_0^l = a^l \tag{4}$$

$$a^{l+1} = \downarrow (h_1 * a^l) \tag{5}$$

Where l denotes level and a_0^l denotes unfiltered data. \downarrow represents subsampling operator. $a^0 = a$ initiates the transfer where a represent input signal.

The backward signal consists of upsampling signal by adding zero in between two samples and convolving with filter h_2 . Combine the upsampled data with stored data at each level after convolving it with another filter g as :

$$\hat{a}^l = h_2 * (\uparrow \hat{a}^{l+1}) + g * a_0^l \tag{6}$$

\uparrow denotes upsampling by zero. Choose h_1, h_2, g filter so as to accurately isolate and reconstruct lower frequency bands of original data. To keep number of operations $O(n)$ the filter must be small and finite.

2.2.3 Shepard’s method

By constructing smooth membrane it is possible to formulate seamless image cloning as boundary value problems can be effectively solved. By approximating Shepard’s scattered data interpolation method using a convolution pyramid is easy to construct suitable membrane faster. If region of interest is denoted by Ω , and $b(x)$ is the boundary value to interpolate these values inside Ω , Shepard’s method defines the interpolant r at x as a weighted average of known boundary values:

$$r(x) = \frac{\sum_k w_k(x) b(x_k)}{\sum_k w_k(x)} \tag{7}$$

Where x_k are the boundary points. The weight function of satisfactory membrane interpolation is obtained by:

$$w_k(x) = w(x_k, x) = 1/d(x_k, x)^2 \tag{8}$$

Defining \hat{f} as an extension of b to the entire domain, to rewrite Shepard’s method in terms of convolutions is given by:

$$\hat{f}(x_i) = \begin{cases} b(x_k), & \text{for } x_i = x_k \text{ on the boundary} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

If $\chi_{\hat{f}}$ is the characteristic function corresponding to \hat{f} , the ratio of convolutions is as follows:

$$r(x_i) = \frac{\sum_{j=0}^p w(x_i, x_j) \hat{f}(x_j)}{\sum_{j=0}^p w(x_i, x_j) \chi_{\hat{f}}(x_j)} \tag{10}$$

3. RESULTS AND COMPARISON

In order to verify the image inpainting algorithm based on cloning concept is used on more than ten images with different level of damages. Consider image1 of figure 1 is damaged image which shows the image of a wall but some text has been written on image. The algorithms discussed in this paper are applied on image1 to remove the text. Images of figure 2 (a), (b), (c), (d) show inpainted image after applying Taylor Series method, Poisson editing method, multiscale transformation method and Shepard’s method respectively. Inpainted images are compared with original images using mean square error (MSE) and peak signal noise ratio (PSNR). By comparing results of cloning algorithms and traditional method it shows that Shepard’s methods achieve high efficiency in terms of time and error factors. These methods used for recovering large scale damaged region and for duplicating objects of images unlike Taylor expansion method which is used only on small scale damages. Hence cloning algorithm used for image inpainting achieves good results and overcomes the issues of traditional method to repair damaged image.

Multiscale transformation, Shepard’s method, poisson editing algorithm give results more seamless than and Taylor series. Time complexity of MVSC and Taylor series are higher. Taylor Series algorithms are used only for object removal but algorithms of convolution pyramid can replicate the objects within the image as modification in the image. User interaction is more in Taylor expansion.

Figure 3 is considered as a sample image for object duplication. Images of figure 4 (a), (b), (c) are the inpainted images using poisson editing, multiscale transformation and Shepard's method respectively for manipulation in a sample image with duplicated small bird which is seamless.



Figure1: Damaged image Considered as image1.

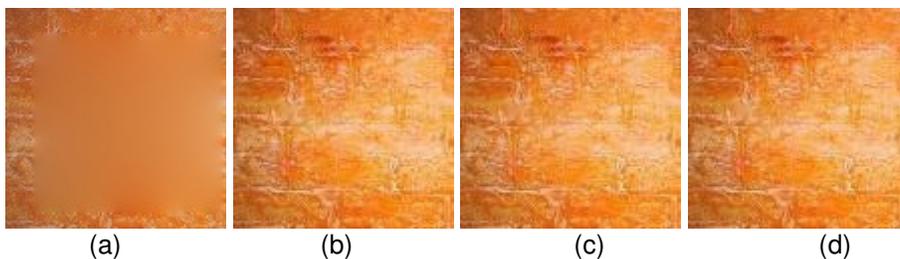


Figure2: Inpainted images with removed distorted area.

	Taylor series		Poisson Editing		Multiscale Transformation		Shepard's Method	
	PSNR	MSE	PSNR	MSE	PSNR	MSE	PSNR	MSE
Image 1	37.34	0.2678	46.62	0.2173	66.35	0.1507	66.4	0.1506
Image 2	24.18	0.4135	32.96	0.3282	37.29	0.2681	37.31	0.25
Image 3	65.68	0.1167	68.65	0.1056	93.46	0.1053	95.13	0.0954

Table1: MSE and PSNR values of inpainted images



Figure 3: Sample image

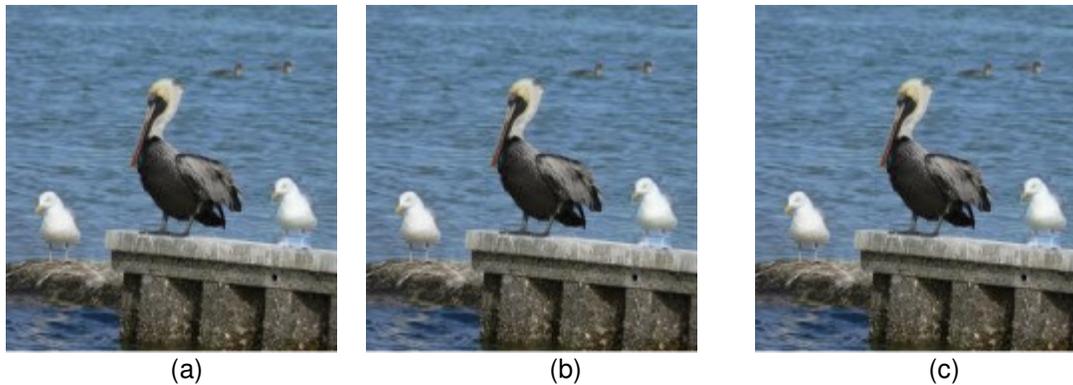


Figure 4: Inpainted images with duplicated object for image manipulation.

4. CONCLUSION

This paper discusses not only methods for removing occlusion as well as scratches from image but also replicate the objects. For cloning of target region which is to be filled in damaged region is cloned using Poisson editing, multiscale transformation and Shepard's method. Taylore series method uses PDE based algorithm and poisson editing algorithm uses Laplacian membrane and Dirichlet equation where as Multiscale Transformation and Shepard's method use optimized filters which are commonly used in computer graphics. Time complexity of Multiscale transformation, Shepard's method, poisson editing algorithm is less than Taylor series algorithm. The experimental result shows that the Shepard's method using multiscale transformation is more efficient than other methods and algorithm is fast, iterative, simple to implement and provides good results.

5. FURTHER WORK

Using above mentioned methods it is possible to duplicate objects in an image and recover damaged image. But it is not possible to deform the objects of an image. Future work includes embedding mesh deformation algorithm for an object, so that inpainted image will have deformed object in it.

6. REFERENCES

- [1] Bertalmio M.; Sapiro G.; V. Caselles et al; "Image Inpainting" In Kurt Akeley, editor, Siggraph 2000, Computer Graphics Proceedings. ACM Addison Wesley Longman, 2000. Page: 417-424.
- [2] M. Bertalmio, Processing of flat and non-flat image information on arbitrary manifolds using Partial Differential Equations. PhD thesis, 2001.
- [3] Chan F T, Shen J H. "Variational Image Inpaining" [EB/OL]. <http://www.math.ucla.edu/imagers/htmls/inp.html>, 2004-11-28.
- [4] L.Atzori and F. G. D Natale, "Error Concealment in video Transmission over Packet networks by a Sketch-based Approach," Image Communication, vol. 15 sep 1999. Page: 57-76.
- [5] A. Criminisi, P. Pérez, and K. Toyama, "Region filling and object removal by exemplar- based image inpainting," IEEE Transactions on Image Processing sep 2004, vol. 13. Page(s): 1200-1212.
- [6] C. P. Perez, Toyama K. "Region filling and object removal by Exemplar- based image inpainting[jj]." IEEE Trans. Image Processing, 2004, 13(9). Page: 1200-1212.

- [7] A. A. Efros and W. T. Freeman, "Image quilting for texture synthesis and transfer," in SIGGRAPH 2001, Computer Graphics Proceedings (E. Fiume, ed.), ACM Press / ACM SIGGRAPH, 2001. Page(s): 341-346.
- [8] Z Hua; Y Li; Jinjiang L; "Image inpainting algorithm based on contour features and improved MVSC" In Computer Design and Applications (ICCDAS), 2010 International Conference Page: V1-212-V1-216.
- [9] S. Xiaoli; X. Chen; "Image denoising and image inpainting model based on Taylor Expansion" Computational Intelligence and Security, 2009. International Conference, Page(s): 666-670.
- [10] P. Perez, M. Gangnet, A. Blake; "poisson Image Editing" ACM Trans Graph 22, (3). Page:313-318.
- [11] Burt, P. J., and Adelson, E. H. 1983. "The Laplacian Pyramid as a image code." IEEE Trans. Comm 31, (4). Page(s): 532-540.
- [12] Z. Farbman, R. Fattal, D. Lischinski." Convolution pyramids" ACM Trans Graph 2011 30,(6). Page(s):771-779

A Fuzzy Set Approach for Edge Detection

Pushpajit A. Khaire

*Department of Computer Technology,
Karmavir Dadasaheb Kannamwar College of Engineering,
Nagpur-440009, India*

pushpjitkhaire@gmail.com

Dr. Nileshsingh V. Thakur

*Department of Computer Science & Engineering,
Prof Ram Meghe College of Engineering and Management,
Badnera-Amravati-444701, India*

thakurnisvis@rediffmail.com

Abstract

Image segmentation is one of the most studied problems in image analysis, computer vision, pattern recognition etc. Edge detection is a discontinuity based approach used for image segmentation. An edge detection using fuzzy set is proposed here, where an image is considered as a fuzzy set and pixels are taken as elements of fuzzy set. The proposed approach converts the color image to a partially segmented image; finally an edge detector is convolved over the partially segmented image to obtain an edged image. The approach is implemented using MATLAB 7.11. (R2010b). In this paper, an attempt is made to evaluate edge detection using ground truth for quantitative and qualitative comparison. 30 BSD (Berkeley Segmentation Database) images and respective ground truths are used for experimentation. Performance parameters used are PSNR (dB) and Performance ratio (PR) of true to false edges. Experimental results shows that the proposed approach gives higher PSNR and PR values when compared with Canny's edge detection algorithm under almost all scenarios. The proposed approach reduces false edge detection and identification of double edges are minimum.

Keywords: Edge Detection, Fuzzy Set, BSD (Berkeley Segmentation Database), Ground Truth, PSNR.

1. INTRODUCTION

Image Segmentation is an important and difficult task in low level image processing, image analysis etc. Edge detection is one of the important techniques used for image segmentation. Earlier the segmentation algorithms were divided into two groups. 1) Discontinuity based approach (Edge detection) and 2) Similarity based approach (Thresholding, Region Growing). Each of these methods has their own advantages and disadvantages. At earlier stages of research on image segmentation, edge detection (Like Prewitt, Sobel) was gaining more attention compared to region growing. Image Segmentation process simplifies, further analysis of images by reducing the amount of data to be processed significantly, at the same time useful structural information of object boundaries are preserved. There are numerous applications of image segmentation like Remote Sensing, Analysis of Medical Images, Industrial Machine Vision for Product Assembly and Inspection, Automated Target Detection and Tracking, Fingerprint Recognition, Face Recognition, Astronomical Study etc. As a result it remains an active area of research.

1.1 Edge Detection

An edge is a sudden change in the pixel intensity of the image. It contains the critical characteristics and important features of an image. An edge is a boundary between the object and its background, also the process of detecting boundaries between object and background in image is known as edge detection. It facilitates, further processing of image like feature selection etc. These all put together edge detection as one of the most important task in computer vision

and image processing. In recent years, researchers have applied various soft computing techniques for edge detection to improve segmentation results for various images and to enhance edge detection technique. Canny [1] proposed a method which is able to detect both strong and weak edges and look more promising to detect edges under noisy conditions. In [2] comparative analysis of various edge detection techniques is given. It is shown that Canny, LOG, Sobel, Prewitt, Roberts's exhibit better performance, respectively.

1.2. Characteristics of Edge Detector.

1. To identify less number of false edges and detection of real edges should be maximum.
2. The marked pixels should be closer to the true edge.
3. Error of detecting more than one response to single edge (double edges) should be less.
4. To design one edge detector that performs well in several contexts (Satellite images, face recognition, medical images, natural images etc.)

This paper is organized as follows: Section (2) emphasizes on work done on edge detection and image segmentation using soft computing approaches with images and parameters used for evaluation. Proposed approach is presented in Section (3). Experimental setup and results are shown in Section (4) and conclusion and future scope are discussed in Section (5).

2. RELATED WORK

Several approaches have been proposed for edge detection, a few of them are discussed here. Konishi and *et al.* [12] formulate edge detection as a statistical inference. They used pre-segmented images to learn the probability distributions of filter responses conditioned on whether they are evaluated on or off an edge. Ground truths of images are considered and performance is measured on Receiver Operator Characteristic (ROC) curves basis. The main disadvantage of this method is, it uses pre-segmented images for learning on one dataset of images and then it is applied on other dataset. J Patel and *et al.* [7] proposed an algorithm based on fuzzy systems and fuzzy rules, where Sobel and Laplacian values are computed and applied to fuzzy system. The proposed approach reduces false edge detection and detection of multiple responses to a single edge is less when compared to Sobel and Laplacian methods. Ground truth evaluation was not discussed here. An algorithm to detect continuous and smooth edges using particle swarm optimization was proposed by Mahdi Setayesh and *et al.* [14]. The results showed that the algorithm performs better and less sensitive to impulsive noise than Canny. The algorithm takes much longer time to execute when compared to Canny method. An approach for edge detection using independent component analysis is proposed by Mendhurwar and *et al.* [15], the proposed approach works well under noisy conditions when compared with Canny's method. The performance is compared on PSNR and no ground truth evaluations of images are considered. The method is robust to noise and detect better edges under noisy conditions. Abdallah A. Alshennawy and Ayman A. Aly [8] proposed a fuzzy logic technique for edge detection without determining the threshold value. The algorithm works well and gives line smoothness and straight for the straight lines, corners get sharper and less detection of double edges when compared to Sobel method, Ground truth evaluation was absent. Many of these approaches discussed here evaluate edge detection without using ground truth of images, results in perplexity for quantitative and qualitative performance evaluation of approaches.

3. PROPOSED APPROACH

In this paper, an approach for edge detection using fuzzy set theory is proposed. In Psychological terms, when humans view a color object, we tend to describe it by its hue, saturation and intensity (H, S, I). Keeping in mind these terms, first RGB color image is converted into HSI image. We, as humans perceive image primarily due to dominant wavelength of light reflected by an object i.e. Hue and amount of light reflected by that object i.e. Intensity. Using this fact, saturation component is removed from HSI image and hue and intensity components are added to form a new hue and intensity (HI) image. The pixel values in the range [0 to 1] are mapped to [0 to 255] to make computations easier to understand. The obtained (hue and intensity) HI image looks like a gray image with pixel values from 0 to 255.

3.1 Fuzzy Membership of Pixels

This HI image is considered as a fuzzy set and the pixels are taken as elements of a fuzzy set. Fuzzy membership of pixel elements is defined based on their constant gray (HI) value. Maximum number of pixels having a constant gray value has the highest degree of membership i.e. 1. Similarly, second maximum set of pixels having constant gray value (pixel value) has the next membership i.e. less than 1. Each pixel in an image holds their membership value depending upon number of pixels having same pixel (gray) value. Now a pixel in this Fuzzy image (Set) has three features:

1. Spatial co-ordinates i.e. (x, y) co-ordinates.
2. Pixel Value (gray value).
3. Fuzzy membership (membership value).

The fuzzy Set F of image is defined as follows:

$F = \{(x, \mu F(x), x \in X\}$, where $\mu F(x)$ denotes the membership value of (pixel) element x in (Image) Fuzzy Set F.

The next step is to employ fuzzy rule on all set of pixels, which results in a partially segmented image. Let A be the set of pixels in fuzzy set F with constant gray value g_1 and membership value m_1 . Similarly, let B be the set of pixels in the same fuzzy set F with constant gray value g_2 and membership value m_2 . Let C (g_3, m_3) be the union of the two sets A and B holds true if it satisfies following conditions:

- 1) If difference between membership values of A and B is less than or equal to 0.2 ($|m_1 - m_2| \leq 0.2$).
- 2) Difference between gray values of A and B is less than or equal to 32 ($|g_1 - g_2| \leq 32$).

If the pixel sets satisfies above two conditions then a new set C(m_3, g_3) is created using set A and B i.e. $C = (A \cup B)$, where $m_3 = \max(m_1, m_2)$ and $g_3 =$ respective gray value of $\max(m_1, m_2)$. The two pixel sets A and B are replaced by pixel set C in image. This procedure is repeated for all set of pixels, results in partially segmented image. Histograms of HI image and partially segmented image are shown in Figure (1) and Figure (2) respectively.

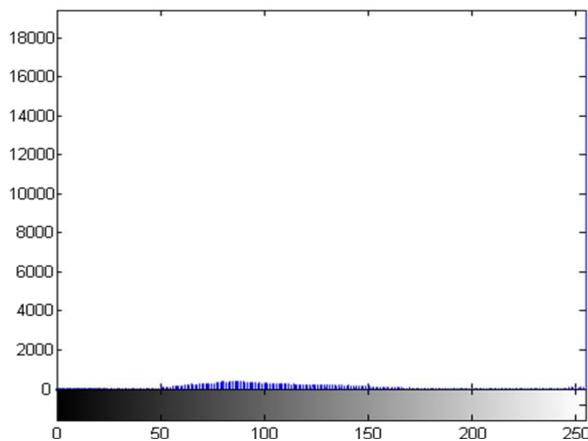


FIGURE 1: Histogram of HI image

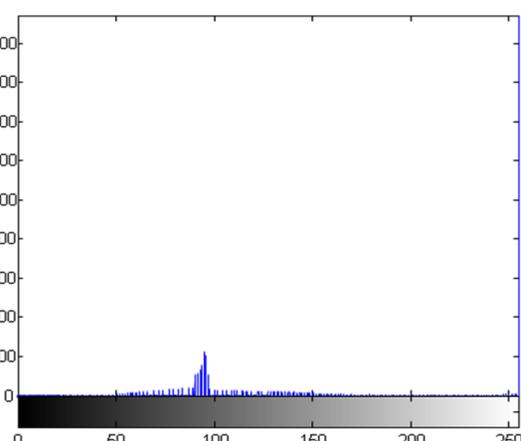


FIGURE 2: Histogram of Partially Segmented Image.

3.2 Edge Detection of Obtained Fuzzy Image

A 3x1 gradient operator in horizontal and vertical direction is shown in Figure (a). These masks are convolved over partially segmented image obtained in step 3.1. Gx, Gy are used to detect edges in horizontal and vertical directions respectively.



FIGURE (a): 3x1 Edge operator

The resultant magnitude of edge pixels are calculated using equation (3.1)

$$G = \sqrt{(Gx)^2 + (Gy)^2} \tag{3.1}$$

These 3x1 masks requires less computations to detect edges compared to other 3x3 masks used (Like Prewitt, Sobel). It also reduces blurring effect while detecting edges. Generally, real image comprises of both strong and weak edges. Here, two thresholds are set for edges, higher threshold and lower threshold. Edges above higher threshold are strong edges and edges above lower threshold are weak edges. Higher threshold value used is 0.3 for strong edges and for weak edges lower threshold is 0.4 × high threshold. Figure (3) shows the original, partially segmented, ground truth and obtained edged image in (a) (b) (c) and (d) respectively.

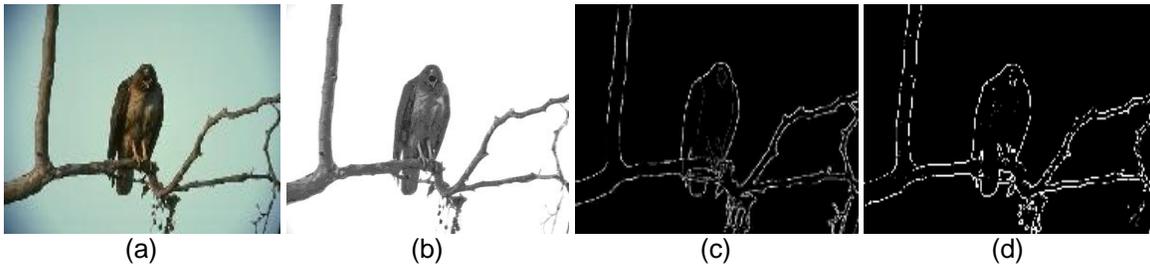


FIGURE 3: (a) BSD image, (b) Partially Segmented, (c) Ground Truth, (d) Proposed Approach

4. EXPERIMENTAL SETUP AND RESULTS

The approach is simulated using MATLAB 7.11 (R2010b). BSD (Berkeley Segmentation Dataset) images [5] and respective ground truths are used for experimentation. Performance parameters used are PSNR and PR (Ratio of true to False Edges). Results shows that the proposed approach detect real edges as shown in ground truth and gives higher PR. Performance Ratio (PR) is the ratio of true to false edges. It is calculated as given in equation (4.1).

$$PR = \frac{\text{True Edges (Edge pixels identified as Edges)}}{\text{False Edges (Non edge pixels identified as edges) + (Edge pixels identified as Non-Edge pixels)}} \times 100 \tag{4.1}$$

The performance and comparative results are shown in Table 4.1. The proposed approach is compared with Canny's algorithm using Ground Truth of respective images. Results show that the proposed approach gives higher PSNR and PR than Canny's approach. After number of

experiments it is found that the default sigma value available in Matlab 7.11 i.e. 1 and threshold=0.3 for Canny approach offer better result than other sigma and threshold values. Here threshold value =0.3 and default sigma=1 for Canny is used for comparison. In proposed approach higher threshold value used is 0.3 for strong edges and for weak edges lower threshold is $0.4 \times$ high threshold. As thickness of edge determines whether an edge is strong or weak edge, to distinguish between strong and weak edges thinning operation is not performed on the resultant edged image. Resultant edged image and respective ground truth of images are shown in Figure (4) through Figure (6).

BSD Image No.	Proposed (T=0.3)		Canny (T=0.3, $\sigma=1$)	
	PSNR(dB)	PR	PSNR(dB)	PR
135069	23.9931	28.0419	23.9735	14.1513
176039	23.487	10.113	23.4721	6.9261
15088	23.3953	17.7663	23.3726	8.8939
12074	23.246	10.8485	23.2335	7.2393
210088	23.0247	13.2036	23.0153	8.7167
28075	22.981	8.1661	22.9767	8.2033
108073	21.5224	10.3232	21.5102	5.8069
3096	21.4337	9.5958	21.4189	5.5835
134052	21.2345	11.0943	21.2264	6.4736
189080	20.9949	12.4702	20.9795	8.993
189011	20.6976	10.5861	20.6882	6.3055
253036	20.4997	21.8664	20.4877	12.6724
8068	20.1987	4.951	20.1962	4.9881
310007	20.1268	15.2524	20.122	11.3261
3063	20.0554	4.3045	20.0496	4.0618
118035	19.9305	17.5689	19.9168	10.3822
41004	19.8741	15.4356	19.8647	9.2601
23025	19.7884	11.2603	19.7855	10.5085
176035	19.758	11.4857	19.7472	6.8046
113044	19.442	16.7319	19.4237	11.3814
197017	19.2163	14.7429	19.2133	11.1349
181018	19.121	10.4562	19.1163	7.4872
35070	18.9703	23.8817	18.9636	16.6493
163014	18.7961	16.2093	18.7877	10.9464
101087	18.423	11.1953	18.417	9.962
157055	18.174	11.6058	18.1672	9.8996
242078	18.1425	16.5446	18.1324	11.006
42049	17.9908	27.2644	17.9737	14.8333
245051	17.3586	26.2115	17.3408	13.9813
35010	17.2309	21.85	17.214	12.5393

TABLE 4.1: Comparison of Approaches.

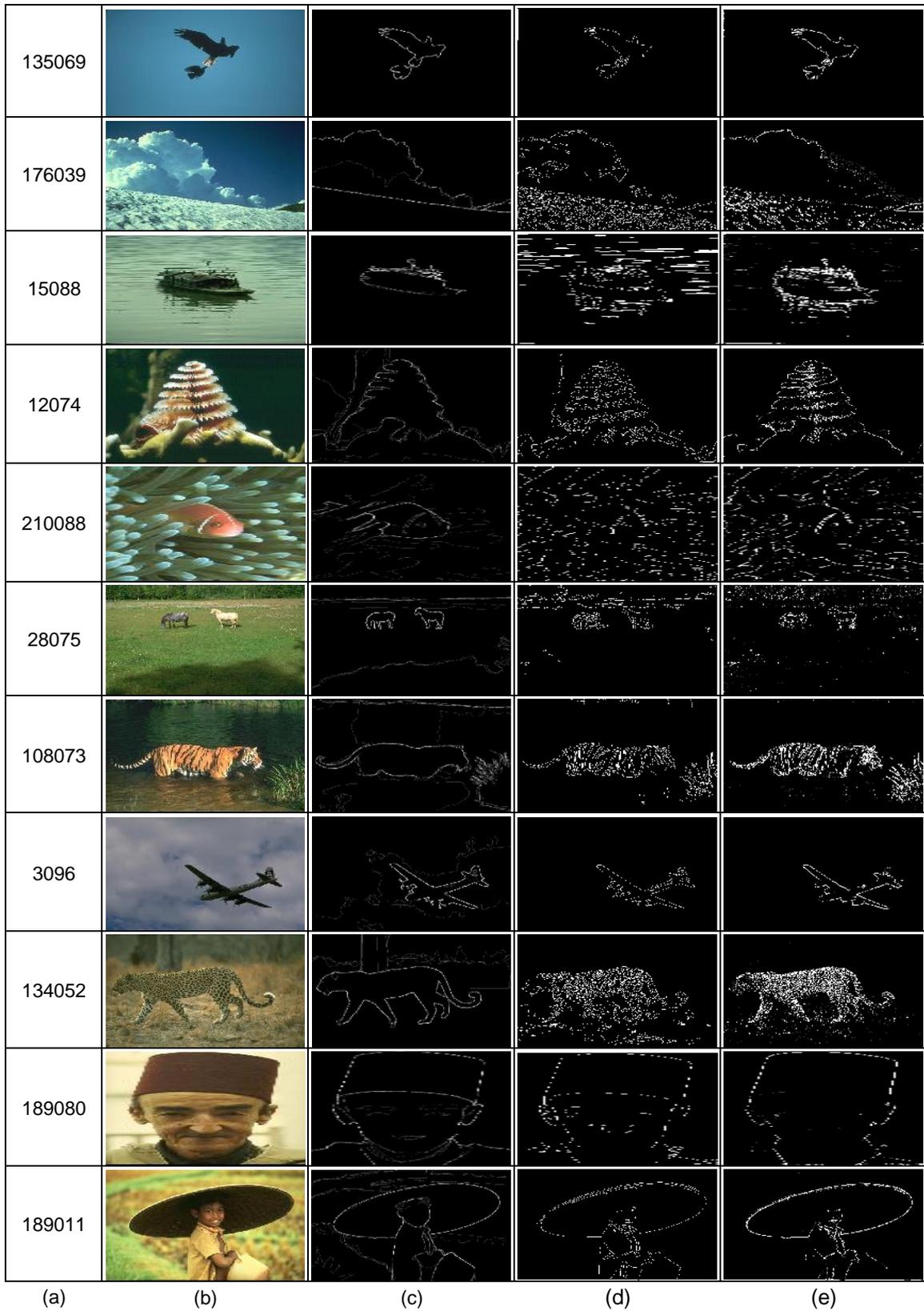


FIGURE 4: Column (a) Image No., Column (b) BSD image, Column (c) Ground Truth, Column (d) Canny's approach, Column (e) Proposed approach

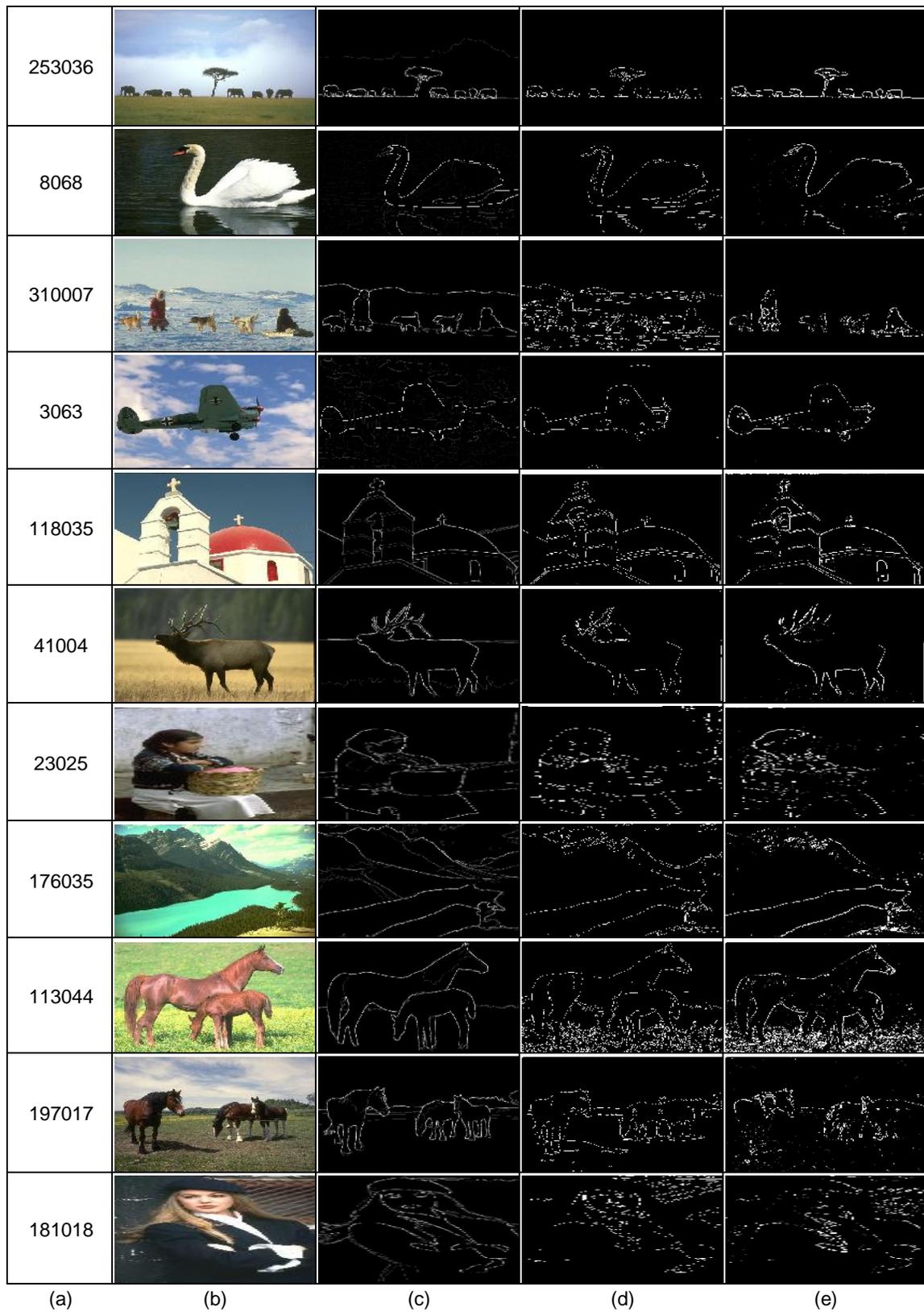


FIGURE 5: Column (a) Image No., Column (b) BSD image, Column (c) Ground Truth, Column (d) Canny's approach, Column (e) Proposed approach

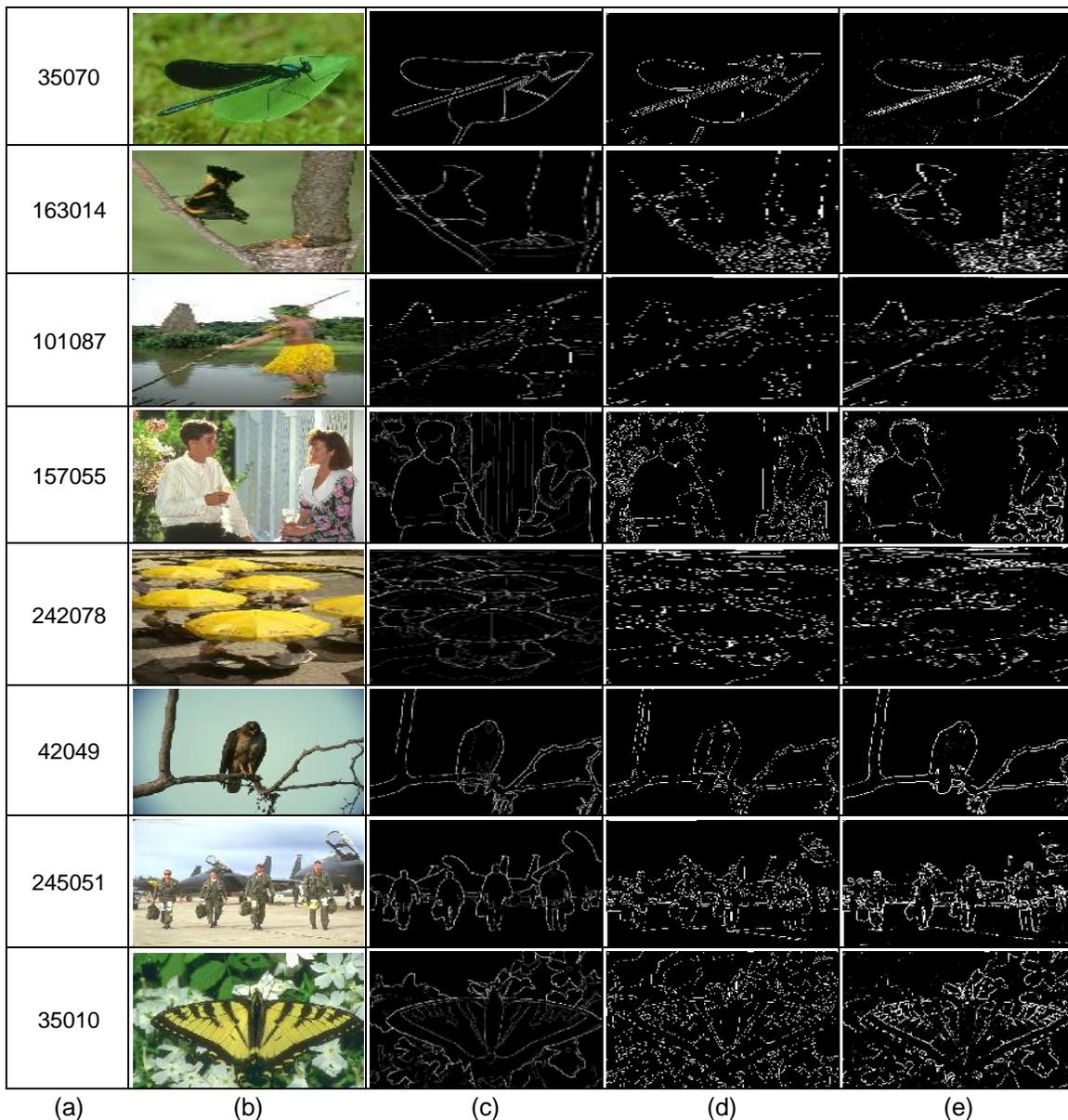


FIGURE 6: Column (a) Image No., Column (b) BSD image, Column (c) Ground Truth, Column (d) Canny's approach, Column (e) Proposed approach

5. CONCLUSION AND FUTURE SCOPE

Edge detection is one of the important techniques used for image segmentation. Image segmentation remains a puzzled problem even after four decades of research. In this paper, a soft computing approach based on Fuzzy Set is proposed for edge detection, where an image is considered as a Fuzzy Set and pixels are taken as elements of Fuzzy Set. The fuzzy approach converts the color image to a partially segmented image, finally an edge detector is convolved over the partially segmented image to obtain edged image. As, proposed edge operator does not perform blurring on image, double edges are less identified. Generally real images comprises of both strong and weak edges. The proposed approach gives both strong and weak edges having different edge strength using higher and lower thresholds.

As mentioned in [4] decades of research on edge detection has produced N edge detectors without a solid basis to evaluate the performance. Many researchers compare edge detection algorithms without using ground truth of images, results in perplexity to evaluate and compare these algorithms. In this paper, an attempt is made to evaluate edge detection using ground truth for quantitative and qualitative comparison. Experimentation is carried out using BSD (Berkeley Segmentation Database) images [5] and respective Ground Truths. The performance evaluation parameters used are PSNR and PR (Ratio of True to false Edges). Experimental Results shows that the proposed approach gives higher PSNR and PR values compared to Canny's approach. It reduces false edge detection and identification of double edges are minimum, Also the marked pixel is closer to the true edge. Here memberships of pixels are calculated based on their constant gray (HI) value. In future, using spatial co-ordinates, different combinations color components of different color models, fuzzy membership of pixels can be calculated.

6. REFERENCES

1. J. Canny, "A Computational Approach to Edge Detection", IEEE Transactions on Pattern Analysis and Machine Intelligence. 8 (6), pp- 679-687, 1986.
2. Raman Maini and Himanshu Aggarwal, "Study and Comparison of Various Image Edge Detection Techniques", International Journal of Image Processing (IJIP), Volume (3), 2010, pp-1-12.
3. Adam Hoover and *et al.*, "An Experimental Comparison of Range Image Segmentation Algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.18 no.7, July 1996.pp- 673-689.
4. M. Heath, S. Sarkar, T. Sanocki, and K.W. Bowyer, "Comparison of Edge Detectors: A Methodology and Initial Study", Computer Vision and Image Understanding, vol. 69, no. 1, Jan. 1998, pp- 38-54.
5. D. Martin, C. Fowlkes, D. Tal and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics", in: Proceedings of the IEEE International Conference on Computer Vision (ICCV '01), 2001, pp- 416–425.
6. K. Bowyer, C. Kranenburg, and S. Dougherty, "Edge Detector Evaluation Using Empirical ROC Curves," Computer Vision and Image Understanding, vol. 84, no. 1, Oct. 2001, pp- 77-103
7. J. Patel, J. Patwardhan, K Sankhe and R Kumbhare, "Fuzzy Inference based Edge Detection System using Sobel and Laplacian of Gaussian Operators", ICWET'11, ACM 978-1-4503-0449-8, February 25–26, 2011, pp- 694-697.
8. Abdallah A. Alshennawy, and Ayman A. Aly, "Edge Detection in Digital Images Using Fuzzy Logic Technique", World Academy of Science, Engineering and Technology, 2009, pp-178-186.
9. Song Wang, Feng Ge, and Tiecheng Liu, "Evaluating Edge Detection through Boundary Detection", EURASIP Journal on Applied Signal Processing, Article ID 76278, June 2006 Pages 1–15.
10. Aborisade, D.O, "Fuzzy Logic Based Digital Image Edge Detection", Global Journal of Computer Science and Technology, Volume 10, November 2010, pp- 78-83.
11. Raman Maini, J.S.Sohal, "Performance Evaluation of Prewitt Edge Detector for Noisy Images", GVIP Journal, Volume 6, December, 2006.

12. S. Konishi, A. Yuille, J. Coughlan and S.C. Zhu, "Statistical Edge Detection: Learning and Evaluating Edge Cues", IEEE Transactions on Pattern Analysis and Machine Intelligence, Jan, 2003.
13. Bhojar and Kakde, "Color Image Segmentation using Fast Fuzzy C-Means Algorithm", Electronic letters on (CVIA) 2010, pp-18-31.
14. Mahdi Setayesh, Mengjie Zhang, and Mark Johnston, "Detection of Continuous, Smooth and Thin Edges in Noisy Images Using Constrained Particle Swarm Optimization", ACM 978-1-4503-0557-0, GECCO'11, Dublin, Ireland, July 12–16, 2011, pp- 45-52.
15. Kaustubha Mendhurwar, Shivaji Patil, Harsh Sundani, Priyanka Aggarwal, and Vijay Devabhaktuni, "Edge-Detection in Noisy Images Using Independent Component Analysis", ISRN Signal Processing, 9 pages, February 2011.
16. N. Pal and S. Pal, "A Review on Image Segmentation Techniques", Pattern Recognition, Vol. 26, no. 9, 1993, pp- 1,277-1, 294.
17. Evans, A. N. and Lin, X. U., "A Morphological Gradient Approach to Color Edge Detection", IEEE Transactions on Image Processing, 15 (6), pp. 1454-1463, 2006.
18. Chandra Sekhar Panda and Srikanta Patnaik, "Filtering Corrupted Image and Edge Detection in Restored Grayscale Image Using Derivative Filters", International Journal of Image Processing (IJIP), Volume 3, 2010, pp- 105-119.
19. O. J. Tobias and R. Seara, "Image segmentation by histogram thresholding using fuzzy sets", IEEE Transaction on Image Processing., Vol. 11, 2002, pp-1457-1465.
20. R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Towards objective evaluation of image segmentation algorithms", IEEE Transactions on Pattern Analysis and Machine Intelligence 29 (6), (2007), pp-929–944.
21. R. C. Gonzalez and R. E. Woods. "Digital Image Processing", Addison Wesley, 2nd edition, 1992.
22. George J.Klir and Bo Yuan, "Fuzzy sets and Fuzzy logic: Theory and Applications", Prentice Hall, 1995.
23. Francisco J. Estrada and Allan D. Jepson, "Benchmarking Image Segmentation Algorithms", International Journal of Computer Vision. Vol. 85, no. 2, Nov 2009, pp. 167-181.
24. Wenshuo Gao and *et al.*, "An Improved Sobel Edge Detection", 978-1-4244-5540-9 IEEE, ICICT 2010.

Land Boundary Detection of an Island using improved Morphological Operation

Shreetam Behera

Final year M. Tech Student

Department of Electronics & Instrumentation Engineering

Siksha O' Anusandhan University

Bhubaneswar, Odisha, 751030, India

shreetam@gmail.com

Mihir Narayan Mohanty

Associate Professor

Department of Electronics & Communication Engineering

Siksha O' Anusandhan University

Bhubaneswar, Odisha, 751030, India

mihir.n.mohanty@gmail.com

Abstract

Image analysis is one of the important tasks to obtain the information about earth surface. To detect and mark a particular land area, it is required to have the image from remote place. To recognize the same, the accurate boundary of that area has to be detected. In this paper, the example of remote sensing image has been considered. The accurate detection of the boundary is a complex task. A novel method has been proposed in this paper to detect the boundary of such land. Mathematical morphology is a simple and efficient method for this type of task. The morphological analysis is performed using structure elements (SE). By using mathematical morphology the images can be enhanced and then the boundary can be detected easily. Simultaneously the noise is removed by using the proposed model. The results exhibit the performance of the proposed method.

Keywords: *Remote Sensing images ; Edge detection; Gray- scale Morphological analysis, Structuring Element (SE).*

1. INTRODUCTION

Remote sensing includes the choice of sensors, the reception, and processing of signal data. It is important aspect to study about the image boundary for further processing like detection, classification, and segmentation etc. of the interested images. Mathematical Morphology is the process to analyze image and that derives from set theory. It is based on shape of the image. Morphological analysis is used with the structure element (SE) that has certain structures and features for measuring and processing image. Small-scale structure element is sensitive to edge signals but also prone to noise, whereas large-scale structure element is robust to noise but could filter out fine details. In this paper, a novel mathematical morphology edge detection algorithm is proposed to detect edges in remote sensing images that detects the edges as well as removes the noise better than the traditional methods for edge detection.

The paper is organized as follows. Section-II reviews the previous work; section-III describes some operators for boundary detection. Section-IV follows the proposed method and the result follows it in section-V. In the last section-VI, concludes the work.

2. REVIEW OF LITERATURE

Reconstruction is a very useful operator provided by mathematical morphology. The reconstruction transformation is relatively well-known in the binary case, where it simply extracts the connected components of an image which are "marked" by another image. In [1], the paper has three major goals: the first one is to provide a formal definition of grayscale reconstruction in the discrete case. In fact, they proposed two equivalent definitions:

The first one is based on the threshold superposition principle and the second one relies on grayscale geodesic dilations. The second part of the paper illustrates the use of binary and especially grayscale reconstruction in image analysis applications: examples proving the interest of grayscale reconstruction for such tasks as image filtering, extrema, domes and basins extraction in grayscale images, "top-hat" by reconstruction, binary and grayscale segmentation, etc., is discussed.

Most of the information about the image can be obtained from the boundary. The function of boundary detection is to identify the edges of homogeneous regions in an image based on properties such as intensity and texture. A lot of work has been done in this field. Several algorithms have been developed based on computation of the intensity gradient vector, which, in general, is sensitive to noise in the image. In order to suppress the noise, some spatial averaging may be combined with differentiation such as the Laplacian of Gaussian operator and the detection of zero crossing. Traditional edge detection like gradient operator, Robert operator, the Sobel operator, the Prewitt operator are the evaluation of derivatives of the image intensity. In [2], Raman Maini and J. S. Sobel evaluated the performance of the Prewitt edge detector for noisy image and demonstrated that the Prewitt edge detector works quite well for digital image corrupted with Poisson noise whereas its performance decreases sharply for other kind of noise. Davis, L. S. in [3], has suggested Gaussian pre-convolution for this purpose. However, all the Gaussian and Gaussian-like smoothing filters, while smoothing out the noise, also remove genuine high frequency edge features, degrade localization and degrade the detection of low contrast edges. The classical operators emphasize the high frequency components in the image and therefore act poorly in cases of moderate low SNR and/or low spatial resolution of the imaging device. Shin, M.C et al. in [4] presented an evaluation of edge detector performance using a structure from motion task. They found that the Canny detector had the best test performance and the best robustness in convergence and is one of the faster executing detectors. Rital, S. et al. in [5] proposed a new algorithm of edge detection based on properties of hyper graph theory and showed this algorithm is accurate, robust on both synthetic and real image corrupted by noise. Li Dong Zhang and Du Yan Bi in [6] presented an edge detection algorithm that the gradient image is segmented in two orthogonal orientations and local maxima are derived from the section curves. They showed that their algorithm can improve the edge resolution and insensitivity to noise. Fesharaki, M.N. and Hellestrand, G.R [7] presented a new edge detection algorithm based on a statistical approach using the student t-test. They selected a 5x5 window and partitioned into eight different orientations in order to detect edges. One of the partitioning matched with the direction of the edge in the image shows the highest values for the defined statistic in that algorithm. They show that this method suppresses noise significantly with preserving edges without a prior knowledge about the power of noise in the image. Canny [8] derived analytically optimal step edge operators and showed that the first derivative of Gaussian filter is a good approximation of such operators. An alternative to gradient techniques is based on statistical approaches. The idea is to examine the distribution of intensity values in the neighborhood of a given pixel and determine if the pixel is to be classified as an edge. In comparison with the differential approaches, less attention has been paid to statistical approaches. In [9], the authors have used PERCLOS algorithm in order to detect boundary of the objects in the image. In [10], the method of morphology has been applied for biomedical images. As per the authors it was shown an excellent performance.

3. STANDARD METHODS FOR BOUNDARY DETECTION

An operator for boundary detection is determined as the neighborhood operation that determines the extent to which each pixel's neighborhood can be partitioned by a simple arc passing through the pixel where pixels in the neighborhood on one side of the arc have one predominant value and pixels in the neighborhood on the other side of the arc have a different predominant value. Usually gradient operators, Laplacian operators, zero-crossing operators are used for edge detection. Mathematical morphology is developed from set theory. It was introduced by Matheron [11] as a technique for analyzing geometric structure of metallic and geologic samples. It was extended to image analysis by Serra [12]. As it is based on set theory, its operation is defined by set arithmetic and is to be represented by the set.

Some of the techniques for edge detection are mentioned below and are the traditional operators:

1. Sobel operator
2. Canny edge detection
3. Prewitt operator
4. Laplacian of Gaussian
5. Roberts edge detection

Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical and is therefore relatively inexpensive in terms of computations. Canny uses a multistage algorithm to detect a wide range of edges in images. Prewitt operator masks are the one of the oldest and best understood methods of detecting edges in images. Multiple masks are used in this method. One for detecting image derivatives in X and another is for detecting image derivative in Y. To find edges, a user convolves an image with both masks, producing two derivative images (dx and dy). The strength of the edge at given location is then the square root of the sum of the squares of these two derivatives. Roberts method is used frequently for hardware implementations for simplicity and speed are dominant factors [13, 14].

4. PROPOSED METHOD

The remote sensing image of island is collected from [15]. In this method, the image is represented using structuring element (SE). SE is characteristic of certain structure and feature, to measure the shape of image and then carry out the processing steps. The aim of this transformation is to search the special set structure of original set. The transformed set includes the information of the special set structure and the transformation is realized by special structuring element. Therefore, the result is correlative to some characteristics of structuring element.

The basic mathematical morphological operators are dilation and erosion and the other morphological operations are the synthesization of the two basic operations.

Let $A(x, y)$ denote a grey-scale two dimensional image, $B(s, t)$ denote structuring element.

Proposed Algorithm:

The following steps are stated for the algorithm of the proposed method:

1. Acquire the image [15].
2. Convert the image into grayscale image.
The process of conversion to grayscale image is performed as

$$\text{Grayscale image matrix } A(x, y) = \text{Red component} * 0.3 + \text{Green component} * 0.59 + \text{Blue component} * 0.11 \quad (1)$$

3. Create the suitable structuring elements (SE). The shape of all structuring elements may be line based flat, linear or both.

Different structuring elements were selected for the erosion and dilation operations. In order to have a basic link between both the operations a difference angle = 90° between the dilation angle

and the erosion angle is considered. A division angle of n has been considered which is used for generating the number of images i.e. $(180^\circ/n)$ which is to be used for the quantitative analysis.

4. Erode the image. Erosion of image $A(x, y)$ by a grey-scale structuring element $B(s, t)$ can be performed by

$$A \ominus B = \min_{[i,j] \in B} \{a[m-j, n-k] + b[j, k]\} \quad (2)$$

5. Dilate the image. Dilation of a grey-scale image $A(x, y)$ by a grey-scale structuring element $B(s, t)$ can be performed by

$$A \oplus B = \max_{[i,j] \in B} \{a[m-j, n-k] + b[j, k]\} \quad (3)$$

6. Find the edges using morphological operator for different structuring elements.

$$Edge(A) = (A \oplus B) - (A \ominus B) \quad (4)$$

7. Then MSE and PSNR were evaluated for different structuring elements as

$$MSE = \sum_{i=1}^m \sum_{j=1}^n (f1(i, j) - f2(i, j))^2 \quad (5)$$

where $f1$ is output image and $f2$ is input image.

$$PSNR = 10 \log(255^2 / MSE) \quad (6)$$

Opening and closing of grey-scale image $A(x, y)$ by grey-scale structuring element $B(s, t)$ are denoted respectively by

$$A \circ B = (A \ominus B) \oplus B \quad (7)$$

$$A \cdot B = (A \oplus B) \ominus B \quad (8)$$

Erosion is a transformation of shrinking, which decreases the grey-scale value of the image, while dilation is a transformation of expanding, which increases the grey-scale value of the image. But both of them are sensitive to the image edge whose grey-scale value changes obviously. Erosion filters the inner image while dilation filters the outer image. Opening is erosion followed by dilation and closing is dilation followed by erosion. Opening generally smoothes the contour of an image, breaks narrow gaps. As opposed to opening, closing tends to fuse narrow breaks, eliminates small holes, and fills gaps in the contours. Therefore, morphological operation is used to detect image boundary, and at the same time, noise can be eliminated from the image.

5. RESULTS



FIGURE 1: Original Image

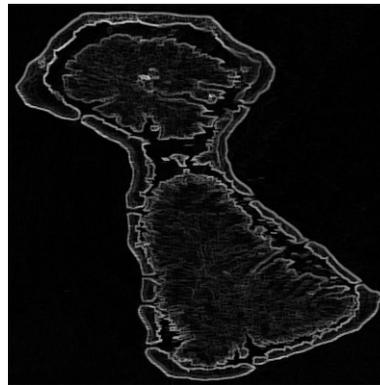


FIGURE.2: Result with 180° dilation angle & 90° erosion angle based structuring elements

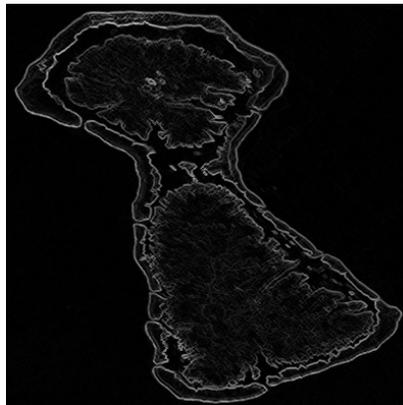


FIGURE3: Result with 135° dilation angle & 45° erosion angle based structuring elements

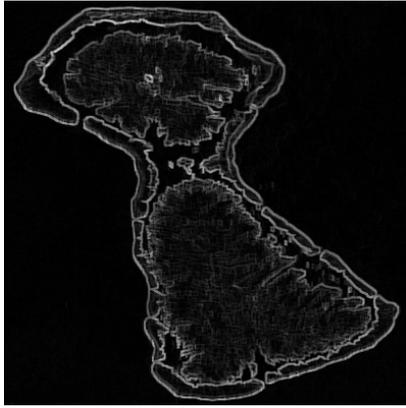


FIGURE 4: Result with 90° dilation angle & 0° erosion angle based structuring elements

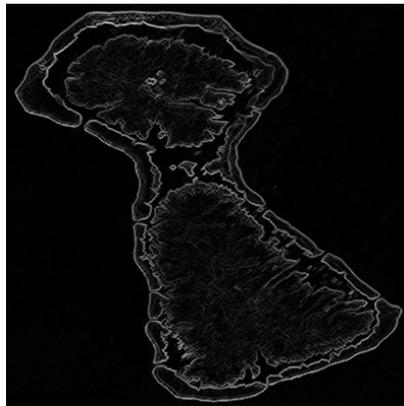


FIGURE 5: Result with 45° dilation angle & -45° erosion angle based structuring elements

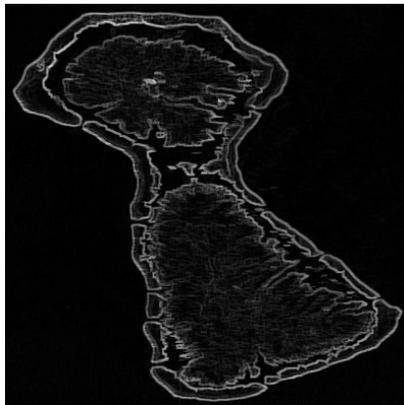


FIGURE 6: Result with 0° dilation angle & -90° erosion angle based structuring elements

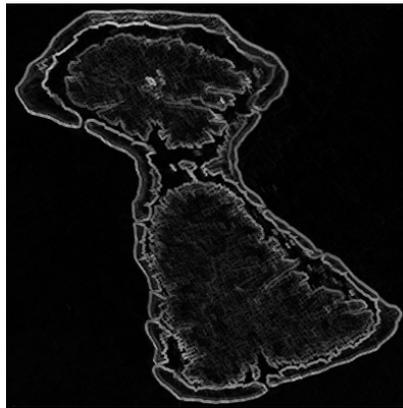


FIGURE 7: Result with 120° dilation angle & 30° erosion angle based structuring elements

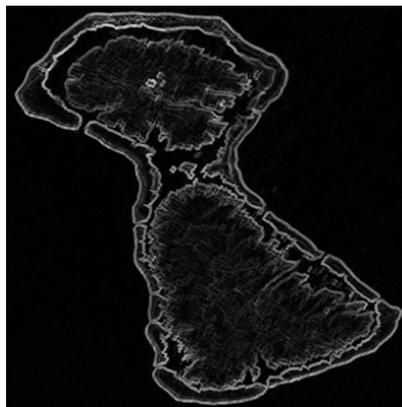


FIGURE 8: Result with 60° dilation angle & -30° erosion angle based structuring elements

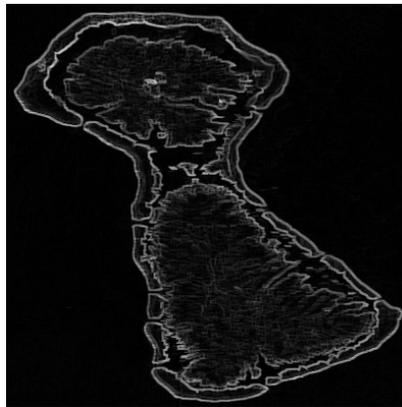


FIGURE 9: Result with 0° dilation angle & -90° erosion angle based structuring elements

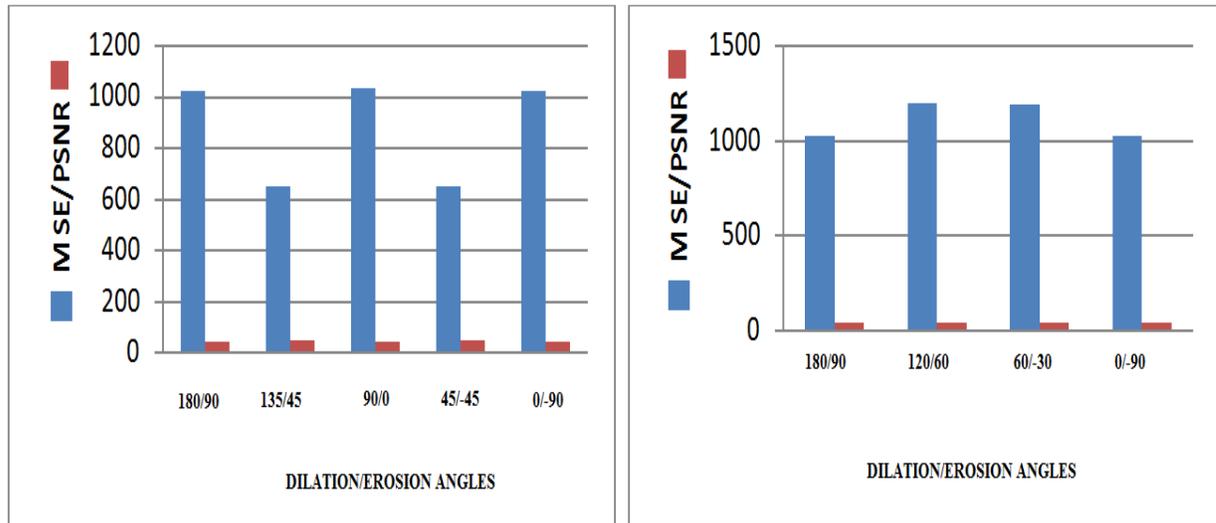


FIGURE 10: PSNR and MSE analysis for 45° division angle and 90° difference angle. **FIGURE 11:** PSNR and MSE analysis for 60° division angle and 90° difference angle.

6. CONCLUSION

The conclusion can be drawn as the boundary detection using mathematical morphology as proposed, is more efficient than the traditional methods. Also the method is simple and easy to implement. Also it can conclude that the method is most important for initial process in boundary detection for noisy images.

REFERENCES

- [1] Luc Vincent "Morphological Grayscale Reconstruction in Image Analysis: Applications and Efficient Algorithms", IEEE Transactions on Image Processing, Vol. 2, No. 2, pp. 176-201, April 1993.
- [2] Raman Maini and J. S. Sobel, "Performance Evaluation of Prewitt Edge Detector for Noisy Images", GVIP-ICGST Journal, Vol. 6, Issue 3, December 2006.
- [3] Davis, L. S., "Edge detection techniques", Computer Graphics Image Process. (4), pp.248-270, 1995.
- [4] Shin, M.C.; Goldgof, D.B.; Bowyer, K.W.; Nikiforou, S.; " Comparison of edge detection algorithms using a structure from motion task", Systems, Man and Cybernetics, Part B, IEEE Transactions on Volume 31, Issue 4, Page(s):589-601, Aug. 2001.
- [5] Rital, S.; Bretto, A.; Cherifi, H.; Aboutajdine, D.; "A combinatorial edge detection algorithm on noisy images", Video/Image Processing and Multimedia Communications 4th EURASIP/IEEE Region 8 International Symposium on VIPromCom, Page(s):351 – 355, 16-19 June 2002.
- [6] Li Dong Zhang; Du Yan Bi; "An improved morphological gradient edge detection algorithm", Communications and Information Technology, ISCIT 2005. IEEE International Symposium on Volume 2, 12-14 Oct. 2005, Page(s):1280 – 1283.
- [7] Fesharaki, M.N.; Hellestrand, G.R.; "A new edge detection algorithm based on a statistical approach", Speech, Image Processing and Neural Networks, Proceedings, ISSIPNN '94., International Symposium, vol.1, 13-16 April 1994, Page(s):21 - 24.
- [8] J. Canny, "A computational approach to edge detection," IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (1986), pp. 679–714.

- [9] Mihir Mohanty, Arabinda Mishra, Aurobinda Routray , "A non-rigid motion estimation algorithm for yawn detection in human drivers," International Journal of Computational Vision and Robotics 2009 - Vol. 1, No.1 pp. 89 - 109,doi: 10.1504/IJCVR.2009.
- [10] S. Behera, Mihir N. Mohanty, and A. Mishra, "Cyst Detection: An Image Processing Approach", IEEE Conf., ICECT, Apr. 2012,pp.55-59.
- [11] G.Matheron and J.Serra, "The birth of Mathematical Morphology", in Mathematical Morphology, H. Talbot, R. Beare, Edts Proc. ISMM 2002, CSIRO Sydney 2002 pp. 1-16.
- [12] J.Serra, "Image analysis and mathematical morphology" ,Academic Press Inc.1982.
- [13] Rafael C. Gonzalez and Richard E. Woods, "*Digital Image Processing*", 3e, PHI, 2008.
- [14] S.Behera; M.N.Mohanty, S. Patnaik" A Comparative Analysis on Edge Detection of Colloid Cyst: A Medical Imaging Approach " , Book Chapter in Soft Computing Techniques in Vision Sci., SCI 395, Springer-Verlag Berlin Heidelberg 2012, pp. 63–85.
- [15] <http://www.flickr.com/photos/digitalglobe-imagery/6539227229/in/photostream>

Active Shape Model based on a spatio-temporal a priori knowledge: applied to left ventricle tracking in scintigraphic sequences

Said Ettaieb

*Research laboratory of image, signal and information
Technology - University of Tunis El Manar
ROMMANA 1068, Tunis-B.P. n° 94, Tunisia*

settaieb@gmail.com

Kamel Hamrouni

*Research laboratory of image, signal and information
Technology - University of Tunis El Manar
ROMMANA 1068, Tunis-B.P. n° 94, Tunisia*

kamel.hamrouni@enit.rnu.tn

Su Ruan

*LITIS-Quantif
University of Rouen
22 boulevard Gambetta 76183 Rouen, France*

su.ruan@univ-rouen.fr

Abstract

The Active Shape Model – ASM is a class of deformable models that relies on a statistical a priori knowledge of shape for the segmentation of structures of interest [5]. The main contribution of this work is to integrate a new a priori knowledge about the spatio-temporal shape variation in this model. The aim is to define a new more stable method, allowing the reliable detection of structures whose shape changes considerably in time. The proposed method is based on two types of a priori knowledge: spatial and temporal variation of the shape of the studied structure. It was applied first on synthetic sequences then on scintigraphic sequences for tracking the left ventricle of the heart. The results were encouraging.

Keywords: Active shape model, a priori knowledge, spatio-temporal shape variation, scintigraphic sequences.

1. INTRODUCTION

The deformable models [1-6] are certainly the most popular approach in the field of medical images segmentation, due to their flexibility and ability to integrate a priori knowledge about the anatomical structures. The basic idea is to start with an initial coarse segmentation that will evolve gradually, according to several constraints, towards the target contours. These models have the advantage of segmenting an image by integrating a global vision of the shape of the structure to be extracted. They are widely studied and applied to the static segmentation of rigid structures, whether in the 2D case or the 3D case [7]. However, in some medical applications, it is sometimes necessary to follow up the spatio-temporal variation of non-rigid structures, whose shape varies over time. In this aim, several extensions of deformable models were proposed. For example, in [8], the authors propose to track anatomical structures in sequences of images by active contour [1] whose initialization in the image i is deduced automatically from the previous result in the image $i - 1$. In several other works [9-13], the sequence of images is treated in a global way and the studied shape variation is described by a single model that evolves over time. The majority of these works is focused mainly on the spatio-temporal tracking of cellular structures [7, 9] and the left ventricle of the heart [10, 11, 14, 15].

However, despite the success obtained in some cases, the quality of the results depends on the initialization step and the choice of propagation parameters. In addition, the used a priori knowledge has generally a global criterion.

The active shape model (ASM) is a particular class of deformable models, introduced by Cootes et al. [5] in order to extract complex and non-rigid objects. This model has two major advantages compared to the other classes of deformable models. On one hand, the initialization is a mean shape of the structure to be segmented. Thus, it will be very close to the target structure during the localization step, which affects advantageously the accuracy of the result. On the other hand, the progressive evolution of the initialization is guided by a statistical shape model that describes the geometry and the authorized deformation modes of the aimed structure. This reduces the solutions space and leads always to acceptable shapes. However, if the structure to be segmented changes considerably over time, these two advantages lose much of their interest. Because, if the shape variation is very important, the mean shape becomes more general and the statistical shape model becomes less precise. Thus, there might be a generation of shapes that is far from the target structures. In order to improve the precision of the active shape model in the case of segmentation of structures whose shape changes significantly over time, we suggest incorporating a new a priori knowledge about the spatio-temporal variation of shape into this model. Indeed, we propose to model the spatial variation of the studied structure over time in order to define a statistical spatio-temporal shape model. This model, which has to describe precisely the shape and the deformation modes of the studied structure at every moment, will be then used to guide a spatio-temporal localization stage to segment a sequence of images.

In this paper, we will explain first the steps of the proposed method. Then we will show its application on synthetic sequences and on real sequences of scintigraphic images. A comparative study between a ground truth drawn by an expert, the ASM and the ASMT, will also be established in order to deduce the interest from the integration of an a priori knowledge on the spatio-temporal shapes variation.

2. PROPOSED METHOD

Given a structure whose shape changes over time. At each instant t , it may take a different shape from that taken at an instant t_1 . We suppose that this variation according to time is represented by a sequence of images. The aim is the automatic localization, in the most reliable way, of this structure in all images of the sequence at the same time.

The proposed method can be described by figure 1.

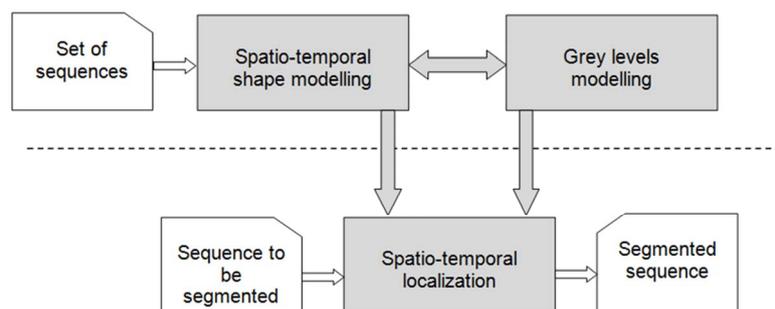


FIGURE 1: Proposed method.

This method requires three main stages: a stage of spatio-temporal shape modelling, a stage of grey levels modelling and a stage of spatio-temporal localization which is based on the results of both first ones to locate the target structure in a new sequence.

2.1 Stage of spatio-temporal shape modelling

The objective of this stage is to build a statistical spatio-temporal shapes model which describes exactly the variation over time of the non-rigid structure to be segmented.

It requires, first of all, the preparation of a spatio-temporal training set, which includes all the possible configurations of this structure.

Step 1: Preparation of a spatio-temporal training set

First, we have to collect a set of sequences of different images, reflecting the possible variations of the studied structure. Every sequence must contain the same structure in the same period of time and with the same number of images. Then, we have to extract the spatio-temporal shapes by putting, carefully, on the contour of every sequence, a sufficient number of landmark points on the wished contour. Each sequence is well modelled by a spatio-temporal shape which describes both the spatial and temporal variation of the studied structure. Given F the number of images by sequence and L the number of landmark points put on each image, the spatio-temporal shape that models a sequence i can then be represented by a vector S_i , constructed by concatenating the coordinates of the points defined on the contours of the studied structure through all the images in the sequence:

$$S_i = [u_{i1}, u_{i2}, u_{i3}, \dots \dots u_{iF}] \quad (1)$$

with $u_{ij} = [x_{ij1}, y_{ij1}, x_{ij2}, y_{ij2}, \dots \dots x_{ijL}, y_{ijL}]$ is a vector that models the J^{th} shape in the i^{th} sequence. Thus, the spatio-temporal training set will be represented by a set of spatio-temporal shapes: $\{S_i \text{ with } i = 1 \dots N \text{ (N number of sequences)}\}$

Step 2: Aligning spatio-temporal shapes

After extracting spatio-temporal shapes from samples of sequences, an alignment step of shapes is required in order to put the corresponding vectors $\{S_i\}$ at a centered position. This allows to eliminate the problem of variation in position and in size and to study only the most important variation in shape between the various configurations of the studied structure. The alignment procedure of the spatio-temporal shapes has the same idea of shapes alignment in the ASM [16]. First, it consists in taking, randomly, a spatio-temporal shape on which are aligned all the others. Then, in every iteration, a mean spatio-temporal shape is calculated, normalized and on which the others will be realigned. This process is stopped when the mean spatio-temporal shape reach some stability.

Step 3: Generation of the statistical spatio-temporal shapes model

The aligned vectors $\{S_i\}$, resulting from the two previous steps, can be arranged in an observation matrix whose size is $(2LF, N)$. This matrix describes both the spatial and temporal variation of the shape of the studied structure. The columns represent the temporal variation while the lines represent the spatial variation at each instant t . The aim is to deduce from this matrix, the modes and the amplitudes of the spatio-temporal variation of the studied structure. Using the same principle of the ASM, this can be done by applying principal component analysis (PCA) on the raw data. Indeed, the main modes of spatio-temporal variation of the studied structure will be represented by the principal components deduced from the covariance matrix C_s associated with the observation matrix (equation 2).

$$C_s = \frac{1}{N} \sum_{i=1}^N dS_i dS_i^t \quad (2)$$

where $dS_i = S_i - \bar{S}$ is the deviation of the i^{th} spatio-temporal shape S_i compared to a mean spatio-temporal shape, that is calculated:

$$\bar{S} = \frac{1}{N} \sum_{i=1}^N S_i \quad (3)$$

These principal components are given by the eigenvectors of the matrix C_s , such as:

$$C_s P_k = \lambda_k P_k \quad (4)$$

P_k is the K^{th} eigenvector of C_s and λ_k is the corresponding eigenvalue. Each vector represents a variability percentage of the variables used to build the covariance matrix.

The variability percentage represented by each vector is equal to its corresponding eigenvalue. In general, we can notice a very fast decreasing of the eigenvalues, which is used to classify the corresponding vectors in decreasing order.

Therefore, we can choose the first t eigenvectors, which represent the important variability percentage, as principal components. Every spatio-temporal shape S can be simply represented by the mean spatio-temporal shape and a linear combination of principal components (main deformation modes):

$$S = \bar{S} + Pb \tag{5}$$

where \bar{S} is the mean spatio-temporal shape, $P = (p_1, p_2, p_3, \dots, p_t)$ is the base of t principal components and $b = (b_1, b_2, b_3, \dots, b_t)^t$ is a weight vector representing the projection of the spatio-temporal shape S in the base P . Generally, the amplitude of the allowable deformation following a principal component P_k is limited as follows:

$$-3\sqrt{\lambda_k} \leq b_k \leq 3\sqrt{\lambda_k} \tag{6}$$

As a result, from the basic equation (Equation 5) we can deduce infinity of shapes describing the spatio-temporal studied structure by choosing correctly the b_k values (equation 6). Equation 5 defines, then, the statistical spatio-temporal shapes model, which defines an allowable deformation space for spatio-temporal studied structure.

This model will be used in the spatio-temporal localization stage to guide the evolution in such a way that it is only in the allowable space. Finally, this first step can be defined by the following functional algorithm:

Algorithm 1 : Spatio-temporal shape modelling

1. Enter the initial parameters:
 - Number of sequences of the training set : N
 - Number of images by sequence: F
 - Number of landmark points : L
 - Variability percentage to be represented: P
 2. Extract manually the spatio-temporal shapes: $\{S_i\} / i=1... N$
 4. Align of the spatio-temporal shapes
 5. Generate of the statistical spatio-temporal shapes model by PCA : $S = \bar{S} + Pb$
-

2.2 Stage of grey levels modelling

As for the ASM, in the stage of spatio-temporal localization, the proposed method is based on intensities information of the treated sequence. It is about finding an optimal correspondence between the properties of luminance of the treated sequence with information of luminance collected from sequences samples. For that purpose, in addition to the spatio-temporal shape modelling, it is necessary to model the grey levels information from the training sequences. For example, k is a sequence in the training set that is composed of three images. This sequence represents the variation over time of a structure. We suppose that ten landmark points are sufficient to extract the shape of the structure presented in every image (Figure 2).

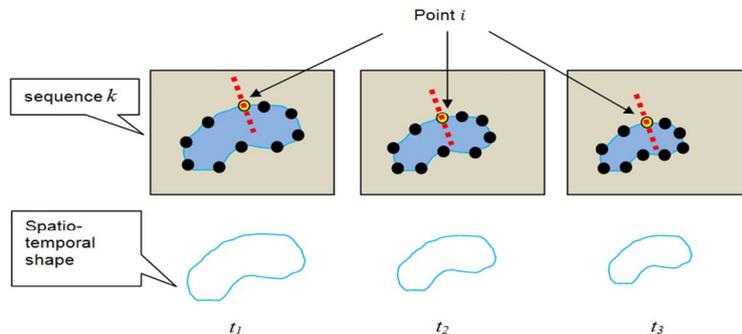


FIGURE 2: Grey levels modelling

The grey levels modelling consists in extracting, for each landmark point i (yellow point) on each image j of the sequence k and then through all the training sequences, the grey levels profile g_{ijk} from a segment of length n (red segment), centered in this point i and carried by its normal:

$$g_{ijk} = [g_{ijk0}, g_{ijk1}, g_{ijk2}, \dots, g_{ijkn-1}] \quad (7)$$

g_{ijkt} ($t = 0 \dots n - 1$) is the grey level of the t^{th} pixel of the examined segment. The derivative of this grey levels profile is defined by the expression 8:

$$dg_{ijk} = [g_{ijk1} - g_{ijk0}, g_{ijk2} - g_{ijk1}, \dots, g_{ijkn-1} - g_{ijkn-2}] \quad (8)$$

dg_{ijk} is a vector of size $(n - 1)$, including the differences in grey levels between two successive points of the examined segment. The normal derivative of this profile is defined by the expression 9:

$$y_{ijk} = \frac{dg_{ijk}}{\sum_{k=0}^{n-2} |dg_{ijkt}|} \quad (9)$$

Through all the images in a sequence and then through all the training sequences, we can define for each landmark point i , a mean normal derivative of the grey levels given by the expression 10:

$$\bar{y}_i = \frac{1}{FN} \sum_{j=1}^F \sum_{k=1}^N y_{ijk} \quad (10)$$

This mean normal derivative related to the point i , will be used in the stage of spatio-temporal localization to move the same point towards a better position. The stage of grey levels modelling can be summarized in by the following functional algorithm:

Algorithm 2 : Grey levels modelling

Enter the initial parameters :

Length (in points) of the grey levels profile: n

Number of sequences of the training set : N

Number of images by sequence: F

Number of landmark points: I

Calculate the mean normal derivative for each landmark point :

For i from 1 to I do

 For j from 1 to F do

 For k from 1 to N do

 Extract the profile : g_{ijk}

 Calculate the derivative : dg_{ijk}

 Calculate the mean derivative : y_{ijk}

 Add $X_i = X_i + y_{ijk}$

 End for

 End for

 Calculate the mean normal derivative : $\bar{y}_i = \frac{X_i}{FN}$

End for

2.3 Stage of spatio-temporal localization

The objective now is to bound the studied structure in a new sequence. A way to achieve this is to start with an initial spatio-temporal shape, which will gradually evolve towards the contours of the studied structure in all images simultaneously. This idea can provide a procedure of spatio-temporal localization, which consists in repeating iteratively the following four steps:

Step 1: Initialization

This step consists in putting an initial spatio-temporal shape S_i on the treated sequence. This shape can be built from a spatio-temporal shape S_a belonging to the training set:

$$S_i = M(k_i, \theta_i)[S_a] + t_i \quad (11)$$

with $M(k_i, \theta_i) = \begin{bmatrix} k_i \cos \theta_i & -k_i \sin \theta_i \\ k_i \sin \theta_i & k_i \cos \theta_i \end{bmatrix}$ a matrix (2*2)

$t_i = (t_{xi}, t_{yi}, t_{xi}, t_{yi}, t_{xi}, t_{yi} \dots t_{xi}, t_{yi})$ a translation vector of size $2 * F * N$

k_i, θ_i and t_i are respectively the homothety, rotation and translation to be applied to every point of S_a in order to build the initialization S_i .

Step 2: Search for the elementary movement

Having fixed an initial spatio-temporal shape S_i , the objective is to determine the elementary movement dS in order to slide the landmark points towards a better position; by using the grey levels characteristics. We will first address the problem of moving a single landmark point. Then, we will show how to calculate the elementary movement of the initial estimate dS_i .

Indeed, A is a particular landmark point of S_i (Figure 3).

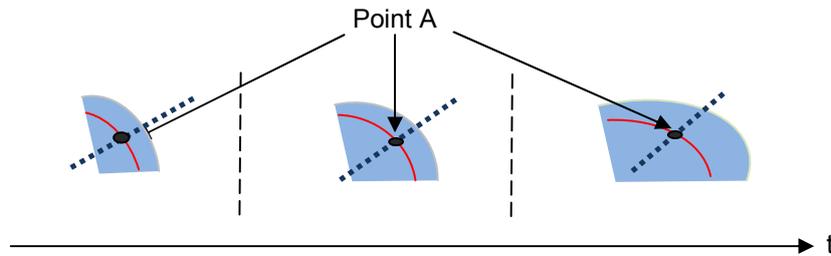


FIGURE 3: Movement of a landmark point A. The structure to be located is in blue. The initial spatio-temporal shape is the red curves.

To move the point A to the borders of the studied structure, the idea is to extract from each image j of the processed sequence, a search grey levels profile of length m pixels (with $m \gg (n - 1)$) which is centered in A and supported by the normal to the edge passing through this point (black segment).

Then, the point A will be represented by a matrix H_A , defined as follows:

$$H_A = \begin{matrix} & \mathbf{1} & \mathbf{2} & \dots & \mathbf{F} \\ \mathbf{1} & g_{A11} & g_{A21} & \dots & g_{AF1} \\ \mathbf{2} & g_{A12} & g_{A22} & \dots & g_{AF2} \\ \mathbf{3} & g_{A13} & g_{A23} & \dots & g_{AF3} \\ \vdots & \vdots & \vdots & (g_{Ajk}) & \vdots \\ \mathbf{m} & g_{A1m} & g_{A2m} & \dots & g_{AFm} \end{matrix}$$

H_A is a matrix of $m * F$ combinations where each column represents the search profile on the image j . (g_{Ajk} is the level grey of the k^{th} pixel on the segment passing by the point A in the j^{th} image). Knowing that each landmark point A is defined by a mean normal derivative of the grey levels \bar{y}_A (information calculated from sequences samples during the stage of grey levels modelling), we can calculate, from the matrix H_A , a new matrix $H'_A(j, l)$ which represents the difference between the grey level information surrounding the current point A and that related to the same point during the grey levels modelling.

This matrix can be defined by the expression (12):

$$H_A'(j, l) = (g_{Aj}(l) - \bar{y}_A)^t I_{n-1}^t I_{n-1} (g_{Aj}(l) - \bar{y}_A) \quad (12)$$

with j from 1 to F , l from 1 to m and I_{n-1} is the identity matrix ($n - 1$). $g_{Aj}(l)$ is the sub-profile of length $(n - 1)$ centered at the l^{th} position of the search profile g_{Aj} that contains the normal derivative of the intensities. (It is necessary to remind that $g_{Aj}(l)$ and \bar{y}_A have the same size $(n - 1)$). The best positions to which has to slide the point A on the treated sequence are given by the expression (13):

$$\{ p_{Aj} = [\min(H_A'(j, l))] \text{ with } j \text{ from } 1 \text{ to } F \text{ and } l \text{ from } 1 \text{ to } m \} \quad (13)$$

p_{Aj} : is the position to which has to slide the point A on the j^{th} image.
Therefore, we can calculate the elementary movements of the particular point A in all the images of the sequence, such as:

$$\{ dp_{Aj} = \mathbf{distance}(A, p_{Aj}) \text{ with } j \text{ from } 1 \text{ to } F \} \quad (14)$$

dp_{Aj} : is the elementary movement of point A in the j^{th} image.
distance : is the Euclidean distance between the point A and the position p_{Aj} .

By applying the same principle for the other landmark points of the initial spatio-temporal shape S_i , we can deduce finally the elementary movement dS_i :

$$dS_i = [dp_{Aj}] \text{ avec } A \text{ from } 1 \text{ to } L \text{ and } j \text{ from } 1 \text{ to } F \quad (15)$$

where:
L : Number of landmark points.
F : Number of images by sequence.

Step 3: Determining the parameters of position and shape
After determining the elementary movement dS_i , we must now determine the parameters of position and shape to make this movement, while respecting the constraints of spatio-temporal deformation imposed by the modelling stage.

- **Determining the position parameters:**
We suppose that the initial estimate S_i is centered in a position (x_c, y_c) with an orientation θ and an homothety k . Determining the position parameters means determining the parameters of geometric operations $1 + dk, d\theta$ and $dt = (dx_c, dy_c)$ to be applied to each point of S_i in order to reach the new position $(S_i + dS_i)$. A simple way to determine these parameters is to align the two vectors S_i and $(S_i + dS_i)$ [2].

- **Determining the shape parameters:**
Once the position parameters $(1 + dk, d\theta$ and $dt)$ are known, it remains to determine the shape parameters. That is to say, if we suppose that the initial estimate S_i is defined in the base of the principal components by a weight vector b , we seek to determine the variation db in order to trace $(S_i + dS_i)$ in the same base. Given that the initial estimate is built from a spatio-temporal shape S_a belonging to the training set $(S_i = M(k_i, \theta_i)[S_a] + t_i)$, determining the shape parameters db is to solve first in dx the following equation:

$$M(k_i(1 + dk), \theta_i + d\theta)[S_a + dx] + t_i + dt = S_i + dS_i \quad (16)$$

which means

$$M(k_i(1 + dk), \theta_i + d\theta)[S_a + dx] = S_i + dS_i - (t_i + dt) \quad (17)$$

but we have $S_i = M(k_i, \theta_i)[S_a] + t_i$

If we replace S_i by its value in the equation (17), we find

$$M(k_i(1 + dk), \theta_i + d\theta)[S_a + dx] = M(k_i, \theta_i)[S_a] + \cancel{t_i} + dS_i - \cancel{(t_i + dt)} \quad (18)$$

But we know that

$$M^{-1}(k, \theta)[...] = M(k^{-1}, -\theta)[...] \quad (19)$$

By applying this rule to the equation (18), we obtain

$$S_a + dx = M((k_i(1 + dk))^{-1}, -(\theta_i + d\theta)) [M(k_i, \theta_i)[S_a] + dS_i - dt] \quad (20)$$

what means that

$$dx = M((k_i(1 + dk))^{-1}, -(\theta_i + d\theta)) [M(k_i, \theta_i)[S_a] + dS_i - dt] - S_a \quad (21)$$

dx is determined in $2 * L * F$ size. However, we have t modes of variation. Then, we have to calculate dx' , the projection of dx in the base of principal components P . This can be done by adopting the approach of least squares [17]. Indeed, $dx' = wdx$ with $w = P(P^tP)^{-1}P^t$ is a projection matrix. However, the principal components of P are pairwise orthogonal, meaning that $P^tP = I$. This, then, gives $dx' = PP^tdx$. We know that $dx' = Pdb$, if we multiply both sides of this equation by P^t , we can deduce finally the shape parameters $db = P^tdx'$.

$db = (db_1, db_2, db_3, \dots, db_t)$ is a weight vector allowing to build and to limit the new vector $(S_i + dS_i)$ in the base of principal components (main modes of deformation).

- Movement of the spatio-temporal shape and the limitation of the shape parameters:

This last step consists in moving S_i to the new position $(S_i + dS_i = S_i^1)$, by using the already calculated parameters.

We obtain

$$S_i^1 = M(k_i(1 + dk), \theta_i + d\theta)[S_a + Pdb] + t_i + dt \quad (22)$$

We should note that the shape parameters $db = (db_1, db_2, db_3, \dots, db_t)$ must be limited in the allowable intervals of variation defined by the equation (6), to produce acceptable spatio-temporal shapes. Indeed, if for example a value db_k ($1 \leq k \leq t$) exceeds the maximum value in a component k , it will be limited as follows:

$$\left\{ \begin{array}{l} \text{if } db_k > v_{\max_k} \\ \text{then } db_k = v_{\max_k} \\ \text{if } db_k < -v_{\max_k} \\ \text{then } db_k = -v_{\max_k} \end{array} \right. \quad (23)$$

with $v_{\max_k} = 3\sqrt{|\lambda_k|}$ is the maximum value of allowable variation following the component k . λ_k is the eigenvalue related to the component k . Now, from S_i^1 , we will repeat the same steps to build S_i^2 then $S_i^3 \dots$ and so on, until no significant change is detected or the maximum number of iterations is reached. The stage of spatio-temporal localization can be described by the functional algorithm 3.

Algorithm 3 : Stage of spatio-temporal localization

```

Initialize of a spatio-temporal shape :  $S_i$ 
While (convergence==false and  $i < \text{nbr\_max\_iterations}$ )
    Search of elementary movement:  $dS_i$ 
    Determine the parameters of position and shape :
         $1 + dk, d\theta, dt$  and  $db$ 
    Movement of the spatio-temporal shape and limitation of the shape
        parameters:
             $-v_{\max_k} \leq db_k \leq v_{\max_k}$ 
             $S_{i+1} = M(k_i(1 + dk), \theta_i + d\theta)[S_i + Pdb] + t_i + dt$ 
    Convergence=compare ( $S_i, S_{i+1}$ )
     $i=i+1$ 
End While
    
```

3. EXPERIMENTAL RESULTS

The proposed method is designed mainly for tracking the spatio-temporal variation of the left ventricle in scintigraphic sequences of images of the heart. But before moving on to this real application, we chose to test the performance of our method on synthetic sequences, in order to deduce its effectiveness in an ideal case.

3.1 Validation on synthetic data

First, we built a database of synthetic sequences, which will serve as a spatio-temporal training set. We synthesized ten sequences disturbed by Gaussian noise which parameters are: $m = 0.2$ and $v = 0.1$. Each sequence contains six images of $256 * 256$ pixels, simulating the variation over time of a simple shape (figure 4.a). During the stage of spatio-temporal shape modelling, thirty landmark points were put on each image to extract the studied shape. Each sequence is modelled by a spatio-temporal shape which size is $2*30*6$ (figure 4.b). The variability percentage to be represented is fixed to 95%. The length of the grey levels profile in the modelling stage is 7 pixels, and in the localization stage, the length of search profile is 19 pixels. The maximum number of iterations is fixed to 40. Figure 5 shows an example of a result of spatio-temporal localization obtained on a synthetic test sequence.

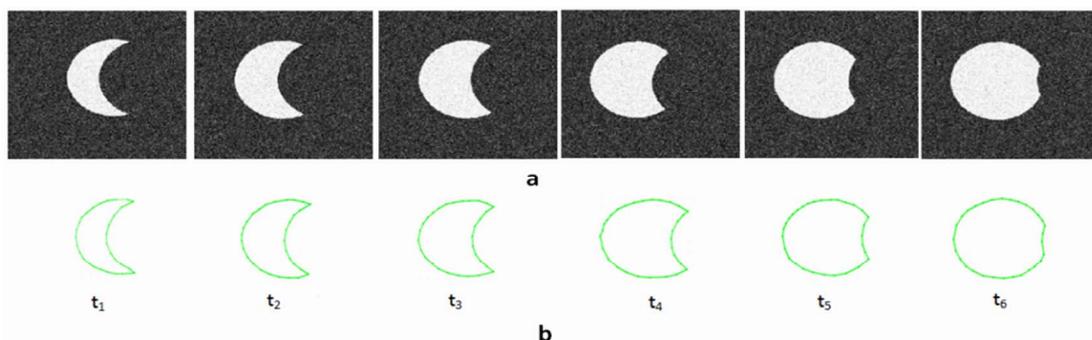


FIGURE 4: (a) Example of a sequence of the synthetic set simulating the variation over time of a simple shape (moon). (b) Corresponding spatio-temporal shape.

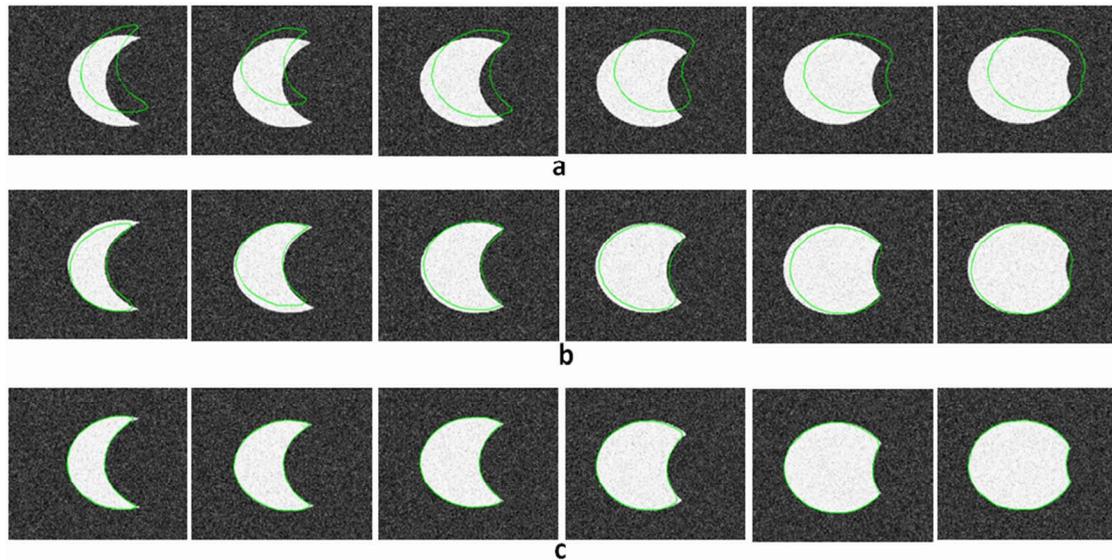


FIGURE 5: Result of spatio-temporal localization obtained on a synthetic sequence. (a) Initialization of the mean spatio-temporal shape. (b) Result obtained after 10 iterations. (c) Final localization result after 15 iterations.

Not surprisingly, we note that the spatio-temporal shape arrived to correctly locate the target shape in all images of the sequence. The accuracy of this result can be explained by two points. On the one hand, the target contours are quite clear and thus easily detectable.

On the other hand, we can say that the spatio-temporal shape modelling provided very accurate information about the studied shape at each moment. The mean spatio-temporal shape used as initialization is very close (in terms of shape) to the target structure on each image, which improves the accuracy of the result. This result shows clearly that our method works in a simple synthetic case. Further, we will apply it in a real case where even the manual tracing of the target contours is difficult.

3.2 Tracking of the left ventricle in scintigraphic sequences of images of the heart

- Background:

The heart is a hollow muscle in the middle of the chest, whose role is to circulate cyclically the blood in the body. In particular, the left ventricle is considered as the main pumping chamber of the heart, because of its great pushing force of the blood through the body against the body pressure. In clinical practice, the study of the function of the heart pump then requires necessarily to follow up the ventricle contraction, during the cardiac cycle in order to estimate the quantity of blood pumped during the corresponding time interval. However, this task is not easy to accomplish, especially for the scintigraphic images. This medical imaging modality is characterized by a low contrast and a low resolution, where even the manual tracing of the target contours is difficult. In this context, several methods are proposed to mark out the left ventricle on scintigraphic images. We can distinguish two types of methods: region-based segmentation [18, 19, 20, 21, 22] and contour-based segmentation [23, 24, 25, 26, 27, 28, 29]. All these works show that the marking out of the left ventricle in scintigraphic images is a difficult task. The results often depend on the parameters of the used method. We can conclude the importance of using a strong a priori knowledge about the studied organ's physiology and anatomy, in order to develop an effective segmentation method. Another important finding is that the developed methods treat the scintigraphic images sequences that represent the variation of left ventricle over time image by image. This fact is severely affecting the quality of the overall result on an entire sequence. For that reason, we thought to exploit a priori knowledge about the spatio-temporal shape variation for the marking out of the left ventricle.

- Experimentation:

The application of our contribution for tracking the left ventricle requires first the preparation of a spatio-temporal training set, which includes all possible configurations. The image database used in this work was provided by the department of nuclear medicine at the Institute Salah Azeiz in Tunis. This database contains 25 scintigraphic sequences of images of the heart from 25 different patients. Each sequence shows a heart beat cycle, represented by 16 images of $128 * 128$ pixels. After showing the collected sequences to a specialist doctor, we concluded that the details of the left ventricle can be represented by 20 landmark points. We noticed that if we work on the 25 sequences, the doctor has to put manually $20 * 16 * 25 = 8000$ landmark points! This is a tedious task. So, we have to reduce then the number of sequences in the database without losing the concept of variability.

To solve this problem, we propose to apply a selecting strategy of sequences to build an optimal spatio-temporal training set. We thought then to apply a classification of all the collected sequences. This classification aims to group sequences according to the shape of the left ventricle. To do this, we selected the same image of each sequence. Each of these images represents then a sequence. The entire population can be then classified into a small number of classes. Each class represents a mode of shape variability. To achieve this goal, we propose to move on from the field of parametric curves (contours of objects) to a field of invariants, where the invariance by translation, rotation and scaling factor is maintained. Thus, every shape will be represented by a set of invariants. The difference between the invariants of two different instances represents the difference between the natures of the shapes themselves. Therefore, to classify the population, it is sufficient to classify the vectors of invariants corresponding to each shape. For the invariants, we chose the Fourier descriptors known for their performance in the field. And for the classifiers, we chose the K-Means classifier, both for its simplicity and its performance. The results of the classification of the 25 collected sequences are given in the table 1:

Sequence	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12	S13
Class	2	2	2	2	2	2	2	3	1	4	3	2	3
Sequence	S14	S15	S16	S17	S18	S19	S20	S21	S22	S23	S24	S25	
Class	4	2	3	4	3	2	1	2	2	3	4	1	

TABLE 1: Results of the classification of 25 database sequences

The population was thus reduced to four classes. Then, we have just to choose arbitrarily two sequences from each class. Therefore, the used spatio-temporal training set is composed of 8 sequences.

The next step consists in extracting from each sequence, the spatio-temporal shape which represents both the spatial and temporal variation of the left ventricle of a heart cycle. This is done by placing 20 landmark points on the left ventricle contour through all images in all sequences. The result of this step is to obtain 8 vectors, each of size $2 \times 20 \times 16 = 640$. After aligning these vectors and using 95% as a variability percentage parameter, the application of PCA on these data provided five main modes of variability. The length of the grey levels profile in the modelling stage is 7 pixels, and in the localization stage, the length of profile search is 19 pixels. The maximum number of iterations is fixed at 60 iterations. The used test sequences are selected from the 17 remaining sequences in the original database. Figure 6 shows an example of the result of the spatio-temporal localization that is obtained on a test sequence.

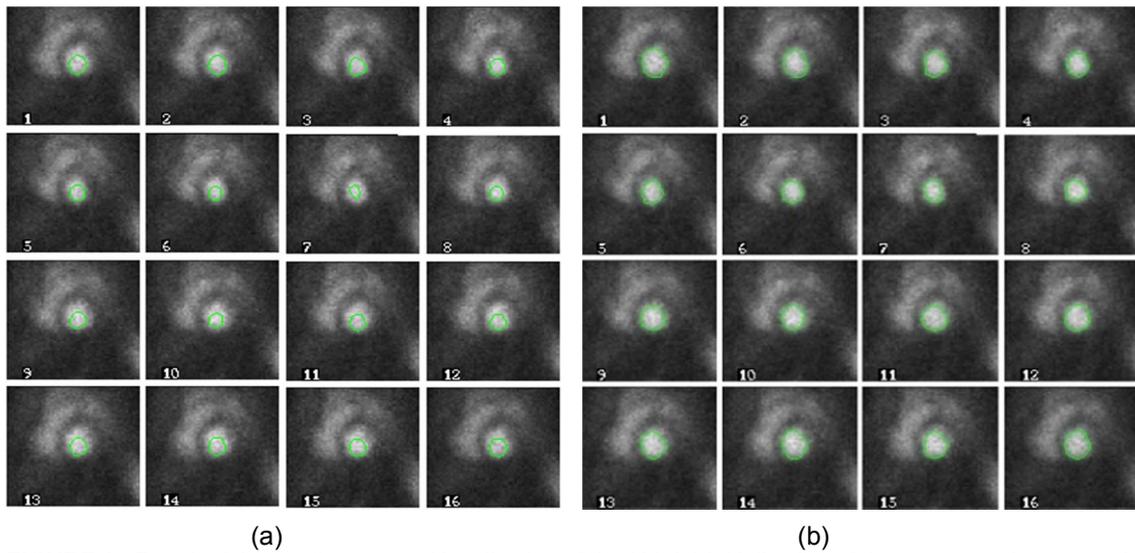
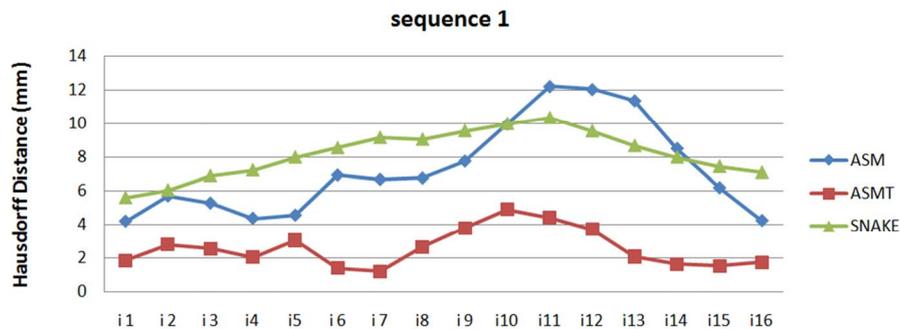


FIGURE 6: Result of the spatio-temporal localization of the LV. (a) Initialization of the mean spatio-temporal shape on a treated sequence. (b) Final result of the localization.

We note that the final spatio-temporal shape succeeded generally in locating the shape of the left ventricle. This can show the performance of the method even in the presence of contours that are difficult to identify. This result is qualitatively considered satisfactory by the medical specialists. However, we should establish a quantitative precise evaluation of the results. In our case, we have four sequences that are manually segmented by a radiologist. In order to deduce the interest from the integration of an a priori knowledge about the spatio-temporal variation of shape, we chose to compare our method ASMT with the ground truth, the basic model ASM and with another method that is proposed by Fekir and al. [8]. This method allows the tracking of non-rigid objects in sequences of images using active contour SNAKE [1] whose initialization in the image i is automatically deduced from the result in the image $i - 1$. Since the compared methods are contour-based methods, we chose the Hausdorff distance as a measure of segmentation quality [30]. This metric is widely used in multiple applications of the medical field. In our case, we use this distance to measure the similarity between two shapes.

Figure 7 shows the Hausdorff distance between each method (ASM, ASMT and SNAKE) and the reference segmentation of the four sequences.



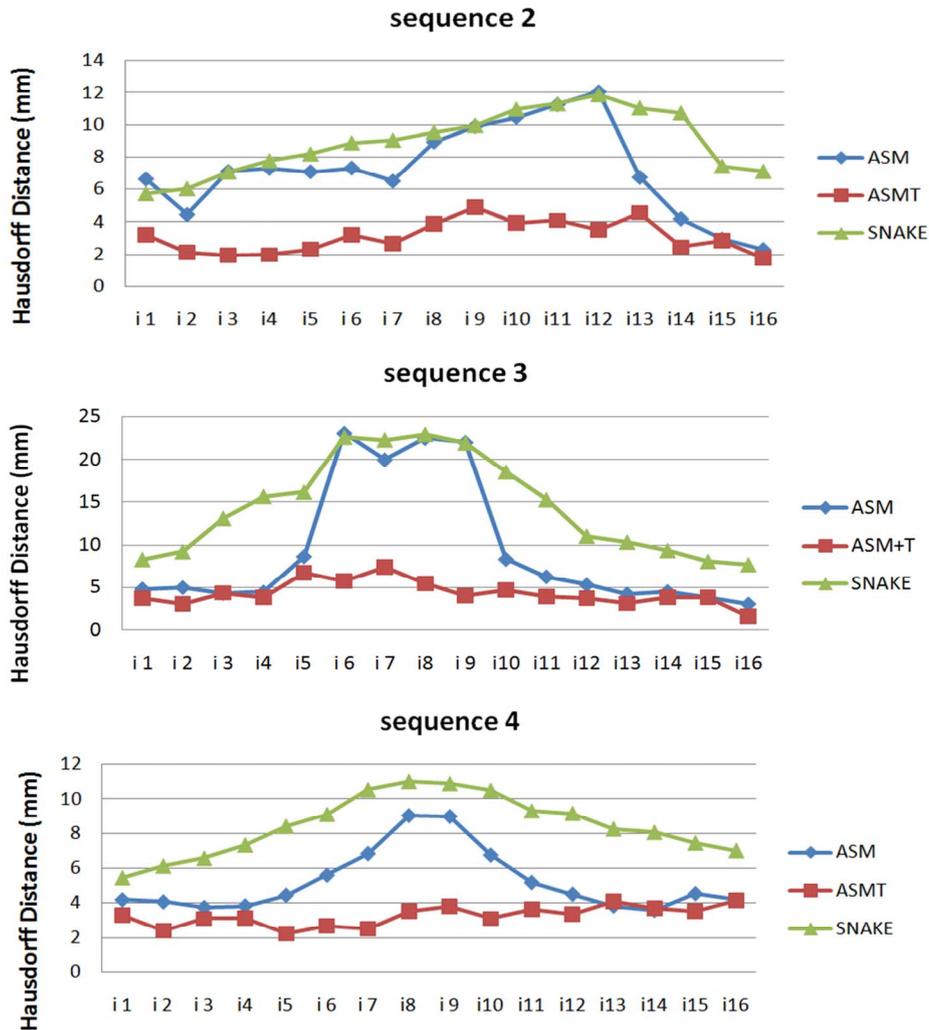


FIGURE 7: Hausdorff distance between each method (ASM, ASMT and SNAKE) and the reference segmentation of the four sequences.

On each diagram of figure 7, the horizontal axis represents the images of the sequence and the vertical axis represents the values of the Hausdorff distance. The red curve represents the values of the Hausdorff distance between the manual segmentation and the automatic segmentation obtained by our method ASMT. The blue curve represents the values of the Hausdorff distance between the manual segmentation and the automatic segmentation obtained by ASM. The green curve represents the values of the Hausdorff distance between the manual segmentation and the automatic segmentation obtained by the method based on SNAKE [8].

Looking at the four diagrams, we can see clearly that the red curve has some stability compared to the other curves (blue and green). Indeed, for the red curve and through the four diagrams, the values of the Hausdorff distance are between 1.18 and 7.38 (mm). By cons, for both blue and green curves, the values of the Hausdorff distance often represent great variations, which rose from 2.3 (mm) and reach 23.06 (mm). Through these measures, and although that in some cases the ASM and SNAKE provide acceptable results (especially in diastole images), we can deduce that our method provides for all images in each sequence an overall result that is more stable and closer to manual segmentation.

This proves the effectiveness of the integration of a priori knowledge about the spatio-temporal shape variation of the left ventricle. Indeed, the stage of spatio-temporal shape modelling provides more precise information on the spatial shape variation of the left ventricle at every moment of the cardiac cycle. This is what influences, consistently and in each image of the sequence, the accuracy of the results of the localization stage. The poor results obtained by ASM and SNAKE may be explained by the imprecision of the initialization in some images of the sequences and the generality of the a priori information about the shape. Besides, these results are mainly obtained in systole images (contraction stage) where the size of the left ventricle becomes very small and difficult to detect.

Another interesting finding is that the value of the Hausdorff distance, whether for ASM, SNAKE or ASMT compared to manual segmentation varies considerably from one sequence to another. For example, in sequence 3, this distance exceeds 23 (mm) for the ASM and SNAKE, while in sequence 4, this distance doesn't exceed 11 (mm). This may be related to the quality of the processed sequence, which affects then the result. That makes us ask: should we start with a pre-treatment stage to improve the quality of sequences before moving on to the segmentation stage?

In conclusion, it is clear that the integration of a priori knowledge about the spatio-temporal shape variation of the left ventricle improved significantly the results of segmentation. This increases the reliability of diagnostic parameters such as the activity-time curve and the ventricular ejection fraction, whose calculation is based on these results. However, we should know that these findings and results may be enriched to include more sequences in the process of quantitative validation. We should also note that the most delicate stage in our approach is the spatio-temporal shape modelling. This stage which is based on a manual process, affects widely the quality of results. It must be therefore made carefully with the help of an expert.

Figures 8, 9, 10 and 11 illustrate a qualitative comparison of the obtained results. Table 2 presents a report on the execution time of our approach. This time is divided into three stages: the spatio-temporal shape modelling, the grey levels modelling and the spatio-temporal localization.

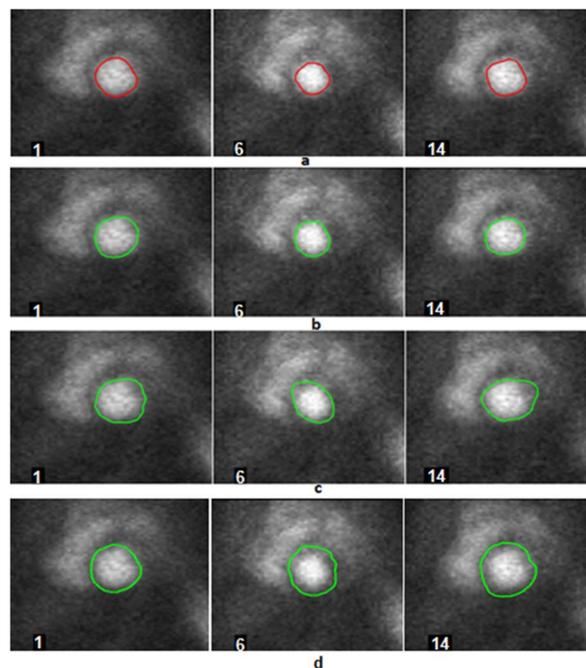


FIGURE 8: Qualitative comparison of results on sequence 1 (Images 1, 6, 14). (a) Manual segmentation, (b) Segmentation by ASMT, (c) Segmentation by ASM and (d) Segmentation by SNAKE.

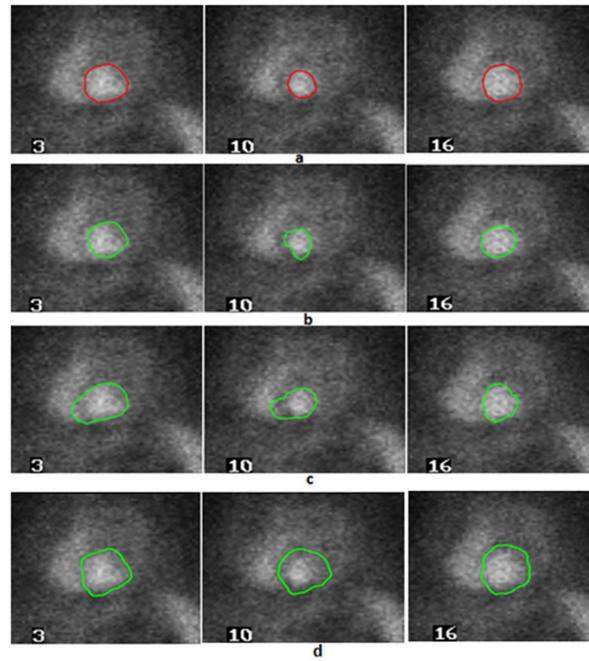


FIGURE 9: Qualitative comparison of results on sequence 2 (Images 3, 10, 16). (a) Manual segmentation, (b) Segmentation by ASMT, (c) Segmentation by ASM and (d) Segmentation by SNAKE.

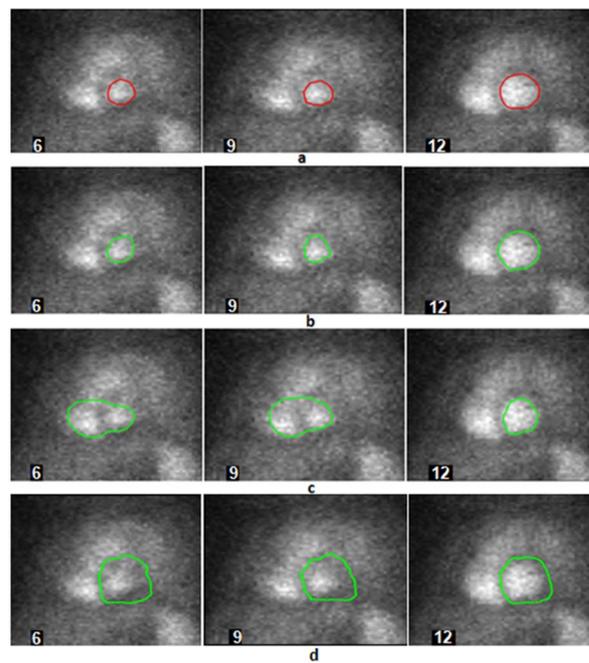


FIGURE 10: Qualitative comparison of results on sequence 3 (Images 6, 9, 12). (a) Manual segmentation, (b) Segmentation by ASMT, (c) Segmentation by ASM and (d) Segmentation by SNAKE.

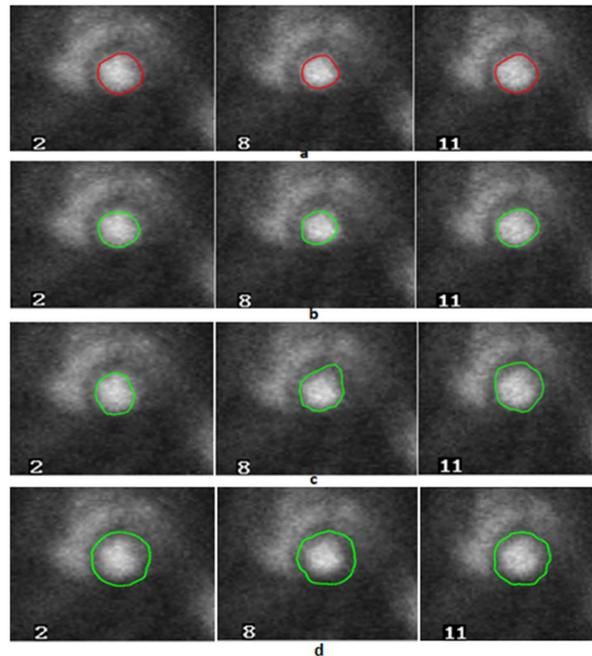


FIGURE 11: Qualitative comparison of results on sequence 4 (Images 2, 8, 11). (a) Manual segmentation, (b) Segmentation by ASMT, (c) Segmentation by ASM and (d) Segmentation by SNAKE.

Spatio-temporal shape modelling	Grey levels modelling	Spatio-temporal shape localization			
		Seq 1	Seq 2	Seq 3	Seq 4
26.27(s)	5.24(s)	29.12(s)	29.21(s)	29.56(s)	29.09(s)

Table 2: Report on the execution time of our method (Matlab 7.0.1, Processor: Intel®, Core™, i3, 2.53 GHz × 2.53 GHz and RAM: 4 GO).

4. CONCLUSION

In this paper, we proposed to incorporate a new a priori knowledge about the spatio-temporal shape variation in the active shape model in order to define a new simple and more stable method for detecting structures whose shape change over time. The proposed method is based on two types of a priori knowledge: the spatial and temporal variation of the studied structure. It has also the advantage of being applicable on sequences of images. The experimental validation of this method, whether it is on simple synthetic sequences or on scintigraphic sequences for the left ventricle tracking, shows the interest of integrating a priori knowledge of the spatio-temporal shape variation. Indeed, having accurate information (geometry and deformation modes) about the shape of the studied structure at every moment provides more stable results, uniformly on all images of the processed sequence. In the training stage, the proposed optimization step, which is based on Fourier descriptors and K-Means classifier, helped to reduce the labelling step without losing the concept of variability.

We are convinced of the relevance of the used method, however, some improvements can be added and the validation should be pursued. Indeed, the most difficult step in our approach is the labelling step. It consists in manually extracting the spatio-temporal shapes from training sequences. That is why, it is usually performed by an expert.

The complexity of this step is in function of the number of training sequences, the number of images by sequence and the number of landmark points needed to represent the target structure details. Once, these parameters become important, this task becomes tedious and time consuming. Then, we should think to make this task semi-automatic or fully automatic. A way to make it semi-automatic is to consider that the shape of the studied structure at instants $t - 1$, t and $t + 1$ has a low variation. The manual training can be thus done on a reduced number of images which correspond to well chosen moments of the sequence. Then, the result of this training will be used for the automatic segmentation of the remaining images. This segmentation is then considered as training. Thus, the complexity of the labelling task can be reduced at least 70%. Moreover, it is possible to enrich and further validate this approach for other types of applications. For example, if we replace the temporal component by the third spatial axis (z), this method can be effectively used for volume segmentation that is based on an important a priori knowledge of shape. In this case, we must solve some additional issues such as correspondence between the slices of training volumes and the slices of the volume to be segmented as well as the automatic determination of the slices that contain the studied structure during the segmentation.

5. REFERENCES

- [1] M. Kass, A. Witkin and D. Terzopoulos. "Snakes: Active contour models". International Journal of Computer vision, pp. 321 – 331, 1988
- [2] L. D. Cohen. "Active contour models and balloons". CVGIP: Image Understanding, vol.53, pp. 211– 218, 1991
- [3] S. Osher and J. A. Sethian. "Fronts propagating with Curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations". J. Comput. Phys., vol.79, pp 12–49, 1988
- [4] V. Caselles, R. Kimmel and G. Sapiro. "Geodesic Active Contours". International Journal of Computer Vision, vol.22, pp 61–79, 1997
- [5] T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham. "Active Shape Models - Their Training and Application". Computer Vision and Image Understanding, vol.61, pp 38–59, 1995
- [6] T. F. Cootes, G. J. Edwards and C. J. Taylor. "Active Appearance Models". IEEE Trans. Pattern Anal. Mach. Intell, vol.23, pp 681–685, 2001
- [7] A. Singh, D. Goldgof and D. Terzopoulos. "Deformable Models in Medical Image Analysis". IEEE Computer Society, Silver Spring, MD, 1998
- [8] A. Fekir, N. Benamrane and A. Taleb-Ahmed. "Détection et Suivi d'Objets dans une Séquence d'Images par Contours Actifs". In Proc. CIIA, 2009
- [9] F. Leymarie and M. Levine. "Tracking deformable objects in the plane using an active contour model". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp 617 – 634, 1993
- [10] T. McInerney and D. Terzopoulos. "A dynamic finite element surface model for segmentation and tracking in multidimensional medical images with application to cardiac 4D image analysis". Computerized Medical Imaging and Graphics, Vol.19, pp 69 – 83, 1995

- [11] W. Niessen, J. Duncan, M. Viergever and B. Romeny. "Spatio-temporal Analysis of Left Ventricular Motion". SPIE Medical Imaging, SPIE Press, pp 250 – 261, 1995,
- [12] S. Sclaroff, and J. Isidoro. "Active blobs". Proceedings of the Sixth IEEE International Conference on Computer Vision, pp 1146 – 1153, 1998
- [13] G. Hamarneh and G. Tomas. "Deformable spatio- temporal shape models: extending active shape models to 2D+time". Journal of Image Vision Computing, vol. 22, pp 461-470, 2004
- [14] B. Lelieveldt, S. Mitchell, J. Bosch, R. van der Geest, M. Sonka, J. Reiber. "Time-continuous segmentation of cardiac image sequences using active appearance motion models". Proceedings of the Information Processing in Medical Imaging, pp 446 – 452, 2001
- [15] A. Singh, L. Von Kurowski and M. Chiu. "Cardiac MR image segmentation using deformable models". SPIE Proceedings of the Biomedical Image Processing and Biomedical Visualization, pp 8 – 28, 1993
- [16] G. Hamarneh. "Active Shape Models, Modeling Shape Variations and Gray Level Information and an Application to Image Search and Classification". Technical report, Department of Signals and Systems, School of Electrical and Computer Engineering, Chalmers University of Technology, Goteborg, Sweden, 1998
- [17] G. Strang. "Linear Algebra and Its Applications". Saunders 1988
- [18] J. Pladellorenst, J. Serraa, A. Castell and M. Yzuelll. "Using mathematical morphology to determine left ventricular contours". Phys. Med. Biol., vol. 37, pp 3877-1894, 1992
- [19] W. Gao and Y. Jin. "Auto threshold region-growing method for edge detection of nuclear medicine images". Proceedings of SPIE, Medical Image Acquisition and Processing, pp. 126-130, 2001
- [20] Damien, L. Itti, P. Egroizard and R. Itti. "Left ventricle detection in radionuclide ventriculography by a model of neural network". Proc. 14th Annual International Conference of the IEEE Engineering in Medicine and Biology Society: IEEE-EMBS, pp 994-995, Paris, France, 1992
- [21] N. Khelifa, N. Hraiech and K. Hamrouni. "Segmentation d'images par l'algorithme EM et les ondelettes. Premier Congrès International de Signaux, Circuits & systèmes : SCS'2004, pp 208-211, Monastir, Tunisie, 2004
- [22] K. Hamrouni and N. Khelifa. "Two Methods for Analysis of Dynamic Scintigraphic Images of the Heart". The International Arab Journal of Information Technology, vol. 3, pp 118-125. 2006
- [23] A. Jouan, J. Verdenet and J. Cardot. "Extraction and analysis of left ventricular contours in cardiac radionuclide angiographies". Journal of Optics, pp 239-246, 1991
- [24] L. Sajn, M. Kukar, I. Kononenko and M. Milcinski. "Automatic segmentation of whole-body bone scintigrams as a preprocessing step for computer assisted diagnostics". In lecture Notes in Computer Science, vol. 3581, pp 363-372, 2005

- [25] N. Khelifa, T. Kraiem and K. Hamrouni. "Traitement et Analyse d'une Séquence d'Images Scintigraphiques du Cœur ". La 2ème Conférence Internationale : JTEA'2002, pp 236-244, Sousse, Tunisie, 2002
- [26] N. Khelifa, K. Hamrouni and T. Kraiem. "Etude de l'activité cardiaque par analyse d'images". International Conference on Image and Signal Processing: ICISP'2003, pp 324-232, Agadir, Maroc, 2003
- [27] N. Khelifa, A. Malek and K. Hamrouni. "Segmentation d'images par contours actifs : Application à la détection du ventricule gauche dans les images de scintigraphie cardiaque". International Conference Sciences of Electronic, Technologies of Information and Telecommunication: SETIT'2005, pp 44, Sousse, Tunisie, 2005
- [28] H. Najah, W. Daniel, K. Hamrouni. "An Active Contour Model Based on Splines and Separating Forces to Detect the Left Ventricle in Scintigraphic Images". In 2nd International Conference on Machine Intelligence, Tozeur, Tunisie, 2005
- [29] N. Khelifa, S. Ettaieb, Y. Wahabi and K. Hamrouni. "Left Ventricle Tracking in Isotopic Ventriculography Using Statistical Deformable Model". The International Arab Journal of Information Technology, Vol. 7, pp 213–222, 2010
- [30] D. Huttenlocher, D. Klanderman and A. Rucklidge. "Comparing images using the Hausdorff distance". IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, pp 850-863, 1993

Survey of The Problem of Object Detection In Real Images

Dilip K. Prasad

*School of Computer Engineering
Nanyang Technological University
Singapore, 639798*

dilipprasad@gmail.com

Abstract

Object detection and recognition are important problems in computer vision. Since these problems are meta-heuristic, despite a lot of research, practically usable, intelligent, real-time, and dynamic object detection/recognition methods are still unavailable. The accuracy level of any algorithm or even Google glass project is below 16% for over 22,000 object categories. With this accuracy, it's practically unusable. This paper reviews the various aspects of object detection and the challenges involved. The aspects addressed are feature types, learning model, object templates, matching schemes, and boosting methods. Most current research works are highlighted and discussed. Decision making tips are included with extensive discussion of the merits and demerits of each scheme. The survey presented in this paper can be useful as a quick overview and a beginner's guide for the object detection field. Based on the study presented here, researchers can choose a framework suitable for their own specific object detection problems and further optimize the chosen framework for better accuracy in object detection.

Keywords: Boosting, Object Detection, Machine learning, Survey.

1. INTRODUCTION

Object detection is a technologically challenging and practically useful problem in the field of computer vision. Object detection deals with identifying the presence of various individual objects in an image. Great success has been achieved in controlled environment for object detection/recognition problem but the problem remains unsolved in uncontrolled places, in particular, when objects are placed in arbitrary poses in cluttered and occluded environment. As an example, it might be easy to train a domestic help robot to recognize the presence of coffee machine with nothing else in the image. On the other hand imagine the difficulty of such robot in detecting the machine on a kitchen slab that is cluttered by other utensils, gadgets, tools, etc. The searching or recognition process in such scenario is very difficult. So far, no effective solution has been found for this problem.

A lot of research is being done in the area of object recognition and detection during the last two decades. The research on object detection is multi-disciplinary and often involves the fields of image processing, machine learning, linear algebra, topology, statistics/probability, optimization, etc. The research innovations in this field have become so diverse that getting a primary first hand summary of most state-of-the-art approaches is quite difficult and time consuming,

This paper is an effort to briefly summarize the various aspects of object detection and the main steps involved for most object detection algorithm or system. Section 2 provides brief introduction about the generic object detection framework and the importance of this study. Section 3 discusses various types of features used as key points for learning and subsequent object detection. Section 4 elaborates on generative and discriminative learning and comparison among them. Section 5 briefly discuss about the various types of representation used for storing the features after the machine learning stage. Various types of matching schemes used by various algorithms for object detection have been discussed in Section 6. Section 7 elaborates about boosting steps of object detection framework. The paper is concluded in Section 8.

2. PURPOSE AND SCOPE OF THE STUDY

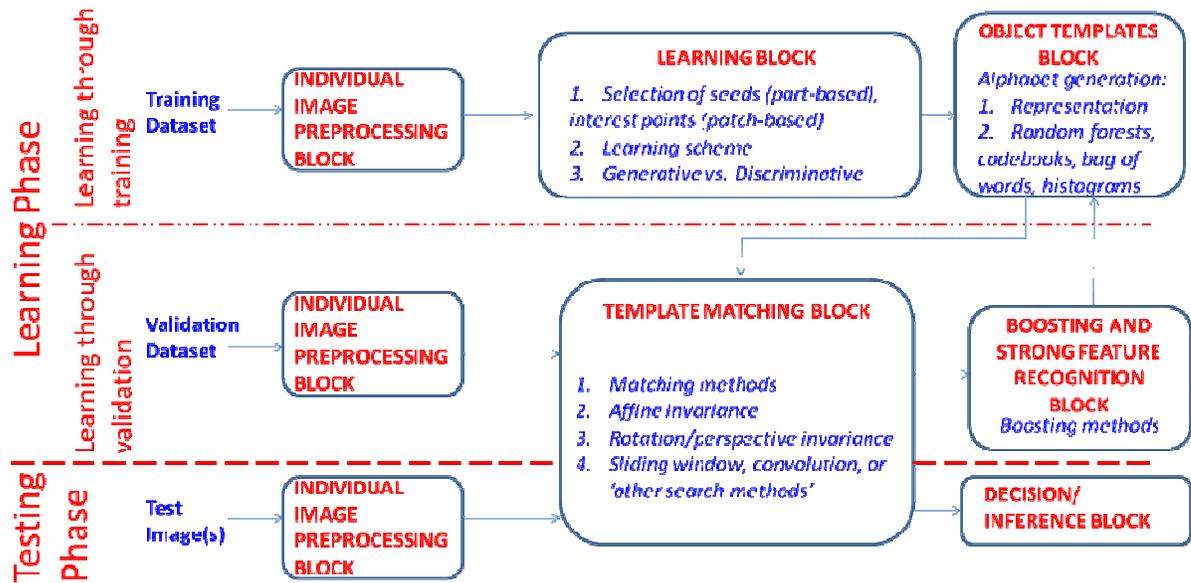


FIGURE 1: Basic block diagram of a typical object detection/recognition system.

In order to facilitate the discussion about the methods and ideas of various research works, we first present a general block diagram applicable to any object detection/recognition method in FIGURE 1. Specific methods proposed by various researchers may vary slightly from this generic block diagram.

Any such algorithm can be divided into two different phases, viz. learning phase and testing phase. In the learning phases, the machine uses a set of images which contains objects belonging to specific pre-determined class(es) in order to learn to identify the objects belonging to those classes. Once the algorithm has been trained for identifying the objects belonging to the specified classes, in the testing phase, the algorithm uses its knowledge to identify the specified class objects from the test image(s).

The algorithm for learning phase can be further subdivided into two parts, viz. learning through training and learning through validation. A set of images containing objects of the specified classes, called the training dataset, is used to learn the basic object templates for the specified classes. Depending upon the type of features (edge based features or patch based features), the training images are pre-processed and passed into the learning block. The learning block then learns the features that characterize each class. The learnt object features are then stored as object templates. This phase is referred to as 'learning through training'. The object templates learnt in this stage are termed as weak classifiers. The learnt object templates are tested against the validation dataset in order to evaluate the existing object templates. By using boosting techniques, the learnt object templates are refined in order to achieve greater accuracy while testing. This phase is referred to as 'learning through validation' and the classifiers obtained after this stage are called strong classifiers.

The researchers have worked upon many specific aspects of the above mentioned system. Some examples include the choice of feature type (edge based or patch based features), the method of generating the features, the method of learning the consistent features of an object class, the specificity of the learning scheme (does it concentrate on inter-class variability or intra-class variability), the representation of the templates, the schemes to find a match between a test/validation image and an object template (even though the size and orientation of an object in

the test image may be different from the learnt template), and so on. The following discussion considers one aspect at a time and provides details upon the work done in that aspect.

3. FEATURE TYPES

Most object detection and recognition methods can be classified into two categories based on the feature type they use in their methods. The two categories are edge-based feature type and patch based feature type. It is notable that some researchers have used a combination of both the edge-based and patch-based features for object detection [1-5]. In our opinion, using a combination of these two features shall become more and more prevalent in future because such scheme would yield a system that derives the advantages of both the feature types. A good scheme along with the advances in computational systems should make it feasible to use both feature types in efficient and semi-real time manner.

3.1 Edge-based features

The methods that use edge-based feature type extract the edge map of the image and identify the features of the object in terms of edges. Some examples include [1, 2, 6-22]. Using edges as features is advantageous over other features due to various reasons. As discussed in [6], they are largely invariant to illumination conditions and variations in objects' colors and textures. They also represent the object boundaries well and represent the data efficiently in the large spatial extent of the images.

In this category, there are two main variations: use of the complete contour (shape) of the object as the feature [7-12, 14, 17] and use of collection of contour fragments as the feature of the object [1, 2, 6, 13-20, 23, 24]. FIGURE 2 shows an example of complete contour and collection of contours for an image.

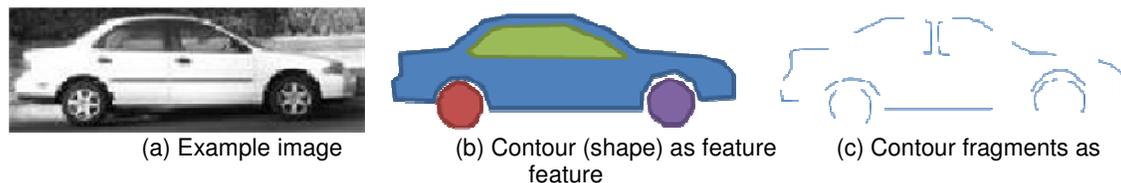


FIGURE 2: Edge-based feature types for an example image

The main motivation of using the complete contours as features is the robustness of such features to the presence of clutter [6, 11, 17, 25]. One of the major concerns regarding such feature type is the method of obtaining the complete contours (especially for training images). In real images, typically incomplete contours are inevitable due to occlusion and noise. Various researchers have tried to solve this problem to some extent [7, 11, 12, 14, 17, 20]. Hamsici [7] identified a set of landmark points from the edges and connected them to obtain a complete shape contour. Schindler [11] used segmenting approaches [26, 27] to obtain closed contours from the very beginning (he called the areas enclosed by such closed contours as super pixels). Ferrari [17, 20] used a sophisticated edge detection method that provides better edges than contemporary methods for object detection. These edges were then connected across the small gaps between them to form a network of closed contours. Ren [14] used a triangulation to complete the contours of the objects in natural images, which are significantly difficult due to the presence of background clutter. Hidden state shape model was used by Wang [28] in order to detect the contours of articulate and flexible/polymorphic objects. It is noticeable that all of these methods require additional computation intensive processing and are typically sensitive to the choice of various empirical contour parameters. The other problem involving such feature is that in the test and validation images, the available contours are also incomplete and therefore the degree of match with the complete contour is typically low [11]. Though some measures, like kernel based [7, 29] and histogram based methods [8, 9], can be taken to alleviate this problem, the detection of the severely occluded objects is still very difficult and unguaranteed [30]. Further, such features are less capable of incorporating the pose or viewpoint changes, large intra-class variability, articulate objects (like horses) and flexible/polymorphic objects (like cars) [11, 17, 20].

This can be explained as follows. Since this feature type deals with complete contours, even though the actual impact of these situations is only on some portions of the contour, the complete contour has to be trained.

On the other hand, the contour fragment features are substantially robust to occlusion if the learnt features are good in characterizing the object [1, 6, 8, 16, 17, 20, 31]. They are less demanding in computation as well as memory as the contour completion methods need not be applied and relatively less data needs to be stored for the features. The matching is also expected to be less sensitive to occlusion [6, 32]. Further, special cases like viewpoint changes, large intra-class variability, articulate objects and flexible/polymorphic objects can be handled efficiently by training the fragments (instead of the complete contour) [2, 6, 8, 17, 20, 32].

However, the performance of the methods based on contour fragment features significantly depends upon the learning techniques. While using these features, it is important to derive good feature templates that represent the object categories well (in terms of both inter-class and intra-class variations) [1, 33]. Learning methods like boosting [31, 33-54] become very important for such feature types.

The selection of the contour fragments for characterizing the objects is an important factor and can affect the performance of the object detection/recognition method. While all the contour fragments in an image cannot be chosen for this purpose, it has to be ensured that the most representative edge fragments are indeed present and sufficient local variation is considered for each representative fragment. In order to look for such fragments, Opelt [1] used large number of random seeds that are used to find the candidate fragments and finally derives only two most representative fragments as features. Shotton [6] on the other hand generated up to 100 randomly sized rectangular units in the bounding box of the object to look for the candidate fragments. It is worth noting that the method proposed in [1] becomes computationally very expensive if more than two edge fragments are used as features for an object category. While the method proposed by Shotton [6] is computationally efficient and expected to be more reliable as it used numerous small fragments (as compared to two most representative fragments), it is still limited by the randomness of choosing the rectangular units. Other computationally efficient way of approximating the contour fragments is by using dominant points or key points of the contours [55-59], guideline to choose suitable dominant point detection method has been given in [57, 60].

On the other hand, Chia [15] used some geometrical shape support (ellipses and quadrangles) in addition to the fragment features for obtaining more reliable features. Use of geometrical structure, relationship between arcs and lines, and study of structural properties like symmetry, similarity and continuity for object retrieval were proposed in [61]. Though the use of geometrical shape (or structure) [62-65] for estimating the structure of the object is a good idea, there are two major problems with the methods in [15, 61]. First problem is that some object categories may not have strong geometrical (elliptic [66, 67] and quadrangle) structure (example horses) and the use of weak geometrical structure may not lead to robust descriptors of such objects. Though [15] demonstrates the applicability for animals, the geometrical structure derived for animals is very generic and applicable to many classes. Thus, the inter-class variance is poor. The classes considered in [15], viz., cars, bikes and four-legged animals (four-legged animals is considered a single class) are very different from each other. Similarly, [61] concentrates on logos and the images considered in [61] have white background, with no natural background clutter and noise. Its performance may degrade significantly in the presence of noise and natural clutter. The second problem is that sometimes occlusion or flexibility of the object may result in complete absence of the components of geometrical structure. For example, if the structural features learnt in [61] are occluded, the probability of detecting the object is very low. Similarly, if the line features learnt in [15], used for forming the quadrangle are absent, the detection capability may reduce significantly.

Though we strongly endorse the idea of using geometric shapes for object detection [68], we suggest that such information should not be used as the only features for object detection. In

addition, they can be used to derive good fragment features and reduce the randomness of selection of the fragments.

3.2 Patch-based features

The other prevalent feature type is the patch based feature type, which uses appearance as cues. This feature has been in use since more than two decades [69], and edge-based features are relatively new in comparison to it. Moravec [69] looked for local maxima of minimum intensity gradients, which he called corners and selected a patch around these corners. His work was improved by Harris [70], which made the new detector less sensitive to noise, edges, and anisotropic nature of the corners proposed in [69].

In this feature type, there are two main variations:

- 1) Patches of rectangular shapes that contain the characteristic boundaries describing the features of the objects [1, 71-76]. Usually, these features are referred to as the local features.
- 2) Irregular patches in which, each patch is homogeneous in terms of intensity or texture and the change in these features are characterized by the boundary of the patches. These features are commonly called the region-based features.

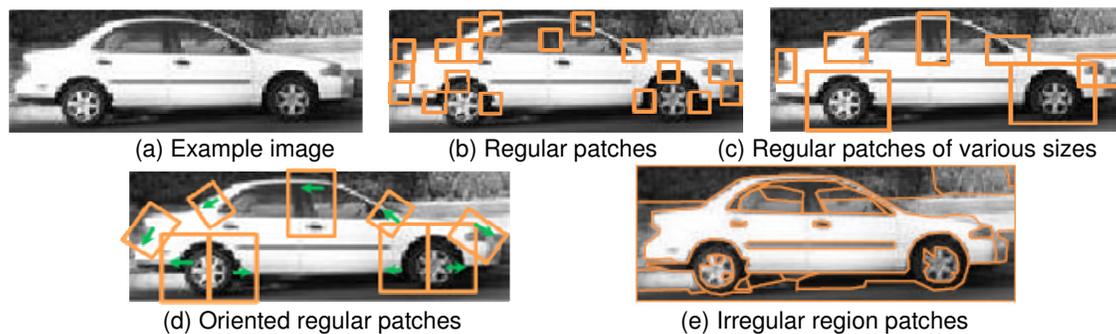


FIGURE 3: Patch-based feature types for an example image. Feature types shown in (b)-(d) are called local features, while the feature type shown in (e) is called region-based features.

FIGURE 3 shows these features for an example image. Subfigures (b)-(d) show local features while subfigure (e) shows region based features (intensity is used here for extracting the region features). As shown in FIGURE 3(b)-(d), the local features may be of various kinds. The simplest form of such features use various rectangular or square local regions of the same size in order to derive the object templates [77]. Such features cannot deal with multi-scaling (appearance of the object in various sizes) effectively. A fixed patch size may not be suitable because of the following reason. If the patch size is small, it may not cover a large but important local feature. Information of such feature may be lost in the smaller patch. On the other hand, if the patch size is large, it may cover more than one independent feature, which may or may not be present simultaneously in other images. Further, there is no way to determine the size that is optimal for all the images and various classes. Another shortcoming is that many small rectangular patches need to be learnt as features and stored in order to represent the object well. This is both computationally expensive and memory intensive.

A better scheme is to use features that may be small or big in order to appropriately cover the size of the local feature such that the features are more robust across various images, learning is better and faster, and less storage is required [78].

A pioneering work was done by Lowe [74], which enabled the use of appropriately oriented variable sized features for describing the object. He proposed a scale invariant feature transformation (SIFT) method. Lowe describes his method of feature extraction in three stages.

He first identified potential corners (key points) using difference of Gaussian function, such that these feature points were invariant to scale and rotation. Next, he identified and selected the corners that are most stable and determined their scale (size of rectangular feature). Finally, he computed the local image gradients at the feature points and used them to assign orientations to the patches. The use of oriented features also enhanced the features' robustness to small rotations. With the use of orientation and scale, the features were transformed (rotated along the suitable orientation and scaled to a fixed size) in order to achieve scale and rotational invariance. In order to incorporate the robustness to illumination and pose/perspective changes, the features were additionally described using the Gaussian weighing function along various orientations.

One of the major concerns in all the above schemes is the identification of good corner points (or key-points) that are indeed representative of the data. This issue has been studied by many researchers [4, 56, 57, 60, 74, 79-81]. Lowe [74] studied the stability of the feature points. However, his proposal would apply to his schema of features only. Carneiro [80] and Comer [82] proposed stability measures that could be applied to wide range and varieties of algorithms.

Another major concern is to describe these local features. Though the features can be directly described and stored by saving the pixel data of the local features, such method is naive and inefficient. Researchers have used many efficient methods for describing these local features. These include PCA vectors of the local feature (like PCA-SIFT) [21, 83], Fischer components [84, 85], wavelets and Gabor filters [13], Eigen spaces [86], kernels [7, 21, 29, 87, 88], dominant points [56-59], etc. It is important to note that though these methods use different tools for describing the features, the main mathematical concept behind all of them is the same except for the dominant points. The concept is to choose sufficient (and yet not many) linearly independent vectors to represent the data in a compressed and efficient manner [13]. Another advantage of using such methods is that each linearly independent vector describes a certain property of the local feature (depending on the mathematical tool used). For example, a Gabor wavelet effectively describes an oriented stroke in the image region [13]. Yet another advantage of such features is that while matching the features in the test images, properties of linear algebra (like linear dependence, orthogonality, null spaces, rank, etc.) can be used to design efficient matching techniques [13].

The region-based features are inspired by segmentation approaches and are mostly used in algorithms whose goal is to combine localization, segmentation, and/or categorization. While intensity is the most commonly used cue for generating region based features [51, 79, 89], texture [2, 89-92], color [91-93], and minimum energy/entropy [94, 95] have also been used for generating these features. It is notable that conceptually these are similar to the complete contours discussed in edge-based features. Such features are very sensitive to lighting conditions and are generally difficult from the perspective of scale and rotation invariance. However, when edge and region based features are combined efficiently, in order to represent the outer boundary and inner common features of the objects respectively, they can serve as powerful tools [2]. Some good reviews of feature types can also be found in [71, 96, 97].

In our opinion, SIFT features provide a very strong scheme for generating robust object templates [74, 98]. It is worth mentioning that though SIFT and its variants were proposed for patch-based features, they can be adapted to edge-fragments based features too. Such adaptation can use the orientation of edges to make the matching more efficient and less sensitive to rotational changes. Further, such scheme can be used to incorporate articulate and flexible/polymorphic objects in a robust manner.

It has been argued correctly by many researchers that a robust object detection and characterization scheme shall typically require more than one feature types to obtain good performance over large number of classes [1, 2, 5, 17, 18, 20, 50, 99-104]. Thus, we shall use region features along with contour fragments. As compared to [1], which has used only one kind of object template for making the final decision, we shall use a combined object template that stores edge, shape, and region features and assigns a strength value to each feature so that

combined probabilistic decision can be made while testing. Such scheme shall ensure that potential objects are identified more often, though the trust (likelihood) may vary and the decision can be made by choosing appropriate threshold. This shall be especially useful in severely occluded or noisy images.

4. GENERATIVE MODEL VS. DISCRIMINATIVE MODEL OF LEARNING

The relationship (mapping) between the images and the object classes is typically non-linear and non-analytic (no definite mathematical model applicable for all the images and all the object classes is available). Thus, typically this relationship is modeled using probabilistic models [105]. The images are considered as the observable variables, the object classes are considered as the state variables, and the features are considered as intermediate (sometimes hidden) variables. Such modeling has various advantages. First, it provides a generic framework which is useful for both the problems of object detection and recognition (and many other problems in machine vision and outside it). Second, such framework can be useful in evaluating the nature and extent of information available while training, which subsequently helps us to design suitable training strategies.

The probabilistic models for our problems can be generally classified into two categories, viz. discriminative models and generative models [106-110]. It shall be helpful to develop a basic mathematical framework for understanding and comparing the two models. Let the observable variables (images) be denoted by \mathbf{x}_i , $i = 1$ to N , where N is the number of training images. Let the corresponding state variables (class labels) be denoted as c_i and the intermediate variables (features/ feature descriptors) be denoted as θ_i . Accordingly, a simplistic graphical representation [107] of the discriminative and generative models is presented in FIGURE 4.

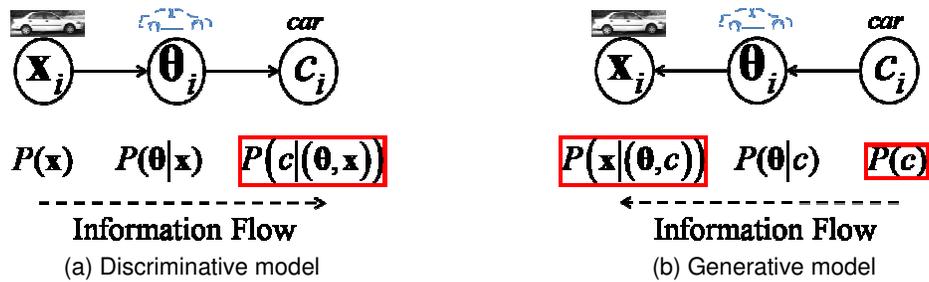


FIGURE 4: Graphical illustration of the discriminative and generative models. The probabilities in boxes are the model defining probabilities for the respective models.

As seen in the FIGURE 4, the discriminative model uses a map from the images to the class labels, and thus the flow of information is from the observables (images) to the state variables (class labels) [107]. Considering the joint probability $P(c, \theta, \mathbf{x})$, discriminative models expand $P(c, \theta, \mathbf{x})$ as $P(c, \theta, \mathbf{x}) = P(c | (\theta, \mathbf{x})) P(\theta | \mathbf{x}) P(\mathbf{x})$. Thus, $P(c | (\theta, \mathbf{x}))$ is the model defining probability [106] and the training goal is:

$$P(c | (\theta, \mathbf{x})) = \begin{cases} \alpha & \text{if } \mathbf{x} \text{ contains object of class } c \\ \beta & \text{otherwise} \end{cases} \quad (1)$$

Ideally, $\alpha = 1$ and $\beta = 0$. Indeed, practically this is almost impossible to achieve, and values between $[0, 1]$ are chosen for α and β .

In contrast, the generative model uses a map from the class labels to the images, and thus the flow of information is from the state variables (class labels) to the observables (images) [107]. Generative models use the expansion of the joint probability $P(c, \theta, \mathbf{x}) = P(\mathbf{x} | (\theta, c)) P(\theta | c) P(c)$.

Thus, $P(\mathbf{x} | (\theta, c))$ and $P(c)$ are the model defining probabilities [106] and the training goal is:

$$P(\mathbf{x} | (\theta, c)) P(c) = \begin{cases} \alpha & \text{if } \mathbf{x} \text{ contains object of class } c \\ \beta & \text{otherwise} \end{cases} \quad (2)$$

Ideally, $\alpha = 1$ and $\beta = 0$. Indeed, practically this is almost impossible to achieve, and some realistic values are chosen for α and β . It is important to note that in unsupervised methods, the prior probability of classes, $P(c)$ is also unknown.

Further mathematical details can be found in [106, 107]. The other popular model is the descriptive model, in which every node is observable and is interconnected to every other node. It is obvious that the applicability of this model to the considered problem is limited. Therefore, we do not discuss this model any further. It shall suffice to make a note that such models are sometimes used in the form of conditional random fields/forests [12, 51, 90].

With the above mentioned mathematical structure as a reference, we can now compare the discriminative and generative models from various aspects, in the following sub-sections.

4.1 Comparison of their functions

As the name indicates, the main function of the discriminative models is that for a given image, it should be able to discriminate the possibility of occurrence of one class from the rest. This is evident by considering the fact that the probability $P(c | (\theta, \mathbf{x}))$ is the probability of discriminating the class labels c for a given instance of image \mathbf{x} . On the other hand, the main function of generative models is to be able to predict the possibility of generating the object features θ in an image \mathbf{x} if the occurrence of the class c is known. In other words, the probabilities $P(\mathbf{x} | (\theta, c)) P(c)$ together represent the probability of generating random instances of \mathbf{x} conditioned to class c . In this context, it is evident that while discriminative models are expected to perform better for object detection purposes, generative models are expected to perform better for object recognition purposes [18]. This can alternatively be understood as the generative models are used to learn class models (and be useful even in large intra-class variation) [50, 75, 111, 112] while discriminative models are useful for providing maximum inter-class variability [112].

4.2 Comparison of the conditional probabilities of the intermediate variables

In the discriminative models, the intermediate conditional probability is $P(\theta | \mathbf{x})$, while in the generative models, the intermediate conditional probability is $P(\theta | c)$. Since we are interested in the joint probability $P(c, \theta, \mathbf{x})$, the probabilities $P(\theta | \mathbf{x})$ and $P(\theta | c)$ play an important role, though they do not appear in the training goals. In the discriminative models, $P(\theta | \mathbf{x})$ represents the strength of the features θ in representing the image well [17, 20], while in the generative models, $P(\theta | c)$ represent the strength of features in representing the class well. Though ideally we would like to maximize both, depending upon the type of feature and the problem, the maximum value of these probabilities is typically less than one. Further, it is difficult to quantitatively measure these probabilities in practice. In our opinion, while the shape features (closed contours) and region features (irregular patches) are more representative of the class (the object's 3-dimensional or 2-dimensional model), the edge fragments and local features are more

representative of the images [1, 50]. Thus, while shape and region features are widely used for segmentation and recognition, local features and edge fragments have been used more often for object detection [17, 18, 20, 50, 101]. Considering this argument, though most methods that use multiple feature types choose these feature types randomly, we recommend to choose a combination of two feature types where one feature is robust for characterizing the image, while the other is good in characterizing the class. In this regard, combining edge fragments and region features is the combination that is easiest to handle practically. Due to this many new methods have used a combination of these features [2, 5, 102-104].

4.3 Training data size and supervision

Mathematically, the training data size required for generative model is very large (at least more than the maximum dimension of the observation vector \mathbf{x}). On the other hand, discriminative models perform well even if the training dataset is very small (more than a few images for each class type). This is expected because the discriminative models invariably use supervised training dataset (the class label is specifically mentioned for each image). On the other hand, generative models are unsupervised (semi-supervised, at best) [113]. Not only the posterior probability $P(\mathbf{x}|\theta, c)$ is unknown, the prior probability of the classes $P(c)$ is also unknown for the generative models [106]. Another point in this regard is that since generative models do not require supervision and the training dataset can be appended incrementally [18, 106, 111] as vision system encounters more and more scenarios, generative models are an important tool for expanding the knowledge base, learning new classes, and keeping the overall system scalable in its capabilities.

4.4 Comparison of accuracy and convergence

The discriminative models usually converge fast and correctly (explained by supervised dataset). If the size of training dataset is asymptotically large, the convergence is guaranteed for the generative models as well. However, such convergence may be correct convergence or misconvergence. If the generative models converge correctly, then the accuracy of generative models is comparable to the accuracy of the discriminative models. But, if there has been a misconvergence, then the accuracy of the generative models is typically poorer than the discriminative models [114]. Since the dataset is typically finite, and in most cases small, it is important to compare the accuracy of these models when the dataset is finite. Mathematical analysis has shown that in such cases, the accuracy of the generative models is always lower than the discriminative methods [114]. It is notable that due to their basic nature, generative models provide good recall but poor precision, while discriminative models provide poorer recall but good precision. The restrictive nature of generative models has prompted more and more researchers to consider discriminative models [1, 17, 20, 93, 115-121]. On the other hand, considering the scalability, generalization properties, and non-supervised nature of generative models, other researchers are trying to improve the performance of generative models by using partial supervision or coupling the generative models and discriminative models in various forms [4, 18, 31, 75, 93, 111, 113, 122].

4.5 Learning methods

Generative models use methods like Bayesian classifiers/networks [18, 31, 75, 111], likelihood maximization [111, 122], and expectation maximization [4, 93, 113, 122]. Discriminative models typically use methods like logistic regression, support vector machines [17, 20, 93, 115-119], and k-nearest neighbors [93, 120, 121]. The k-nearest neighbors scheme can also be used for multi-class problems [109, 123-141] directly, as demonstrated in [120]. Boosting schemes are also examples of methods for learning discriminative models [1], though they are typically applied on already learnt weak features (they shall be discussed later in greater detail). In the schemes where generative and discriminative models are combined [93, 142], there are two main variations: generative models with discriminative learning [4, 102, 106], and discriminative models with generative learning [107]. In the former, typically maximum likelihood or Bayesian approaches are combined with boosting schemes or incremental learning schemes [4, 50, 102, 106, 111], while in the latter, usual discriminative schemes are augmented by 'generate and test'

schemes in the feedback loop [107, 143]. Learning scheme can be offline or online based on the demand of the application [144]. Online learning is now feasible due to advancement of cloud technology [145].

5. OBJECT TEMPLATES AND THEIR REPRESENTATION

The learning method has to learn a mapping between the features and the classes. Typically, the features are extracted first, which is followed by either the formation of class models (in generative models) or the most discriminative features for each class (in discriminative models) or random fields of features in which a cluster represents an object class (descriptive models, histogram based schemes, Hough transform based methods, etc). Based on them, the object templates suitable for each class are learnt and stored for the future use (testing). This section will discuss various forms of object templates used by researchers in computer vision.

While deciding on an object template, we need to consider factors like:

Is the template most representative form of the class (in terms of the aimed specificity, flexibility of the object, intra-class variation, etc)? For example, does it give the required intra-class and inter-class variability features? Does it need to consider some common features among various classes or instances of hierarchical class structure? Does it need to consider various poses and/or perspectives? Does it need to prioritize certain features (or kind of features)?

Is the model representation an efficient way of storing and using the template? Here, memory and computations are not the only important factors. We need to also consider if the representation enables good decision mechanisms.

The above factors will be the central theme in discussing the specific merits and demerits of the various existing object templates. We begin with the object templates that use the spatial location of the features. Such templates specifically represent the relative position of the features (edge fragments, patches, regions) in the image space. For this, researchers typically represent each feature using a single representative point (called the centroid) and specify a small region in which the location of the centroid may vary in various objects belonging to the same class [1, 6]. Then all the centroids are collected together using a graph topology. For example some researchers have used a cyclic/chain topology [11]. This simplistic topology is good to represent only the external continuous boundary of the object. Due to this, it is also used for complete contour representation, where the contour is defined using particular pivot points which are joined to form the contour [11]. Such a topology may fail if the object is occluded at one of the centroid locations, as the link between the chain is not found in such case and the remaining centroids are also not detected as a consequence. Further, if some of the characteristic features are inside the object boundary, deciding the most appropriate connecting link between the centroids of the external and internal boundaries may be an issue and may impact the performance of the overall algorithm. Other topology in use is the constellation topology [111, 146, 147], in which a connected graph is used to link all the centroids. A similar representation is being called multi-parts-tree model in [94], though the essentials are same. However, such topology requires extra computation in order to find an optimal (neither very deep nor very wide) representation. Again, if the centroids that are linked to more than one centroid are occluded, the performance degrades (though not as strongly as the chain topology). The most efficient method in this category is the star topology, in which a central (root) node is connected to all the centroids [1, 6, 8, 76]. The root node does not correspond to any feature or centroid and is just a virtual node (representing the virtual centroid of the complete object). Thus, this topology is able to deal with occlusion better than the other two topologies and does not need any extra computation for making the topology.

Other methods in which the features are described using transformation methods (like the kernel based methods, PCA, wavelets, etc., discussed in section 3), the independent features can be used to form the object templates. The object templates could be binary vectors that specify if a particular feature is present in an object or not. Such object templates are called bag-of-words, bag of visual words, or bag of features [1, 95, 115, 116, 119, 148-150]. All the possible features are analogous to visual words, and specific combinations of words (in no particular order) together represent the object classes. Such bag of words can also be used for features like colors, textures, intensity, shapes [95], physical features (like eyes, lips, nose for faces, and

wheels, headlights, mirrors for cars) etc. [93, 149, 151]. As evident, such bag of words is a simple yet powerful technique for object recognition and detection but may perform poorly for object localization and segmentation. Fusing the generic object template and visual saliency for salient object detection has been explored by Chang et. al. [152]. As opposed to them, spatial object templates are more powerful for image localization and segmentation.

In either of the above cases, the object templates can also be in the form of codebooks [1, 6, 17, 20, 75, 76, 150, 153]. A codebook contains a specific code of features for each object class. The code contains the various features that are present in the corresponding class, where the sequence of features may follow a specific order or not. An unordered codebook is in essence similar to the concept of bag of words, where the bag of words may have greater advantage in storing and recalling the features and the object templates. However, codebooks become more powerful if the features in the code are ordered. A code in the order of appearance of spatial templates can help in segmentation [6], while a code in the order of reliability or strength of a feature for a class shall make the object detection and recognition more robust.

Other hierarchical (tree like) object templates may be used to combine the strengths of both the codebooks and bag of words, and to efficiently combine various feature types [4, 18, 75, 84, 89, 92, 102, 113, 122, 147, 150, 154].

Another important method of representing the object templates is based on random forests/fields [90, 143, 155]. In such methods, no explicit object template is defined. Instead, in the feature space (where each feature represents one dimension), clusters of images belonging to same object class are identified [74, 75, 155]. These clusters in the feature space are used as the probabilistic object templates [84]. For every test image, its location in feature space and distance from these clusters determine the decision.

We prefer a hierarchical codebook, similar to the multi-parts-tree model [94, 113], which combines at least two feature types. We intend to place the strongest (most consistent and generic) features at the highest level and weaker features in subsequent nodes. Any single path in the hierarchy shall serve as a weak but sufficient object template and typically the hope is that more than one path are traversed if object of the class is present in an image. If all the paths are traversed, the image has a strong presence of the object class. The final inference will be based on the number and depth of the paths traversed. It is worth mentioning that while [94] used a minimization of the energy and Mahalanobis distance of the parts for generating the tree, we shall use the likelihood of each feature independently, and likelihood of each feature conditioned to the presence of higher level features in the tree. We might have considered another hierarchical structure where the strongest (but few) descriptors appear at the leaf nodes and the path towards the root incrementally confirms the presence of the object. But that would either require multiple bottom-up traversals (in order to reach the root) or a top-down traversal with very low initial confidence. On the other hand, the chosen top-down structure will ensure that we begin with a certain degree of confidence (due to the generic features with high likelihood at the highest level, details in section 6) in the presence of the object class and then tweak our confidence as we go further down the tree. If we cannot go further down the tree, we need not look for multiple other traversal paths beginning again from the top.

6. MATCHING SCHEMES AND DECISION MAKING

Once the object templates have been formed, the method should be capable of making decisions (like detecting or recognizing objects in images) for input images (validation and/or test images). We first discuss about the methods of finding a match between the object template and the input image and then discuss about the methods of making the final decision.

Discussion regarding matching schemes is important because of various reasons. While the training dataset can be chosen to meet certain requirements, it cannot be expected that the test images also adhere to those requirements. For example, we may choose that all the training images are of a particular size, illumination condition, contain only single object of interest viewed

from a fixed perspective, in uncluttered (white background), etc., such restrictions cannot be imposed on the real test images, which may be of varying size, may contain many objects of interest and may be severely cluttered and occluded and may be taken from various viewpoints. The problem of clutter and occlusion is largely a matter of feature selection and learning methods. Still, they may lead to wrong inferences if improper matching techniques are used. However, making the matching scheme scale invariant, rotation and pose invariant (at least to some degree), illumination independent, and capable of inferring multiple instances of multiple classes is important and has gained attention of many researchers [6, 68, 80, 82, 147, 156-185].

If the features in the object templates are pixel based (for example patches or edges), the Euclidean distance based measures like Hausdorff distance [174, 184, 186, 187] and Chamfer distance [1, 6, 17, 25, 94, 161, 170] provide quick and efficient matching tools. However, the original forms of both these distances were scale, rotation, and illumination dependent. Chamfer distance has become more popular in this field because of a lot of incremental improvement in Chamfer distance as a matching technique. These improvements include making it scale invariant, illumination independent, rotation invariant, and more robust to pose variations and occlusions [1, 6, 17, 25, 94, 161, 170]. Further, Chamfer distance has also been adapted for hierarchical codebooks [161]. In region based features, concepts like structure entropy [95, 188], mutual information [95, 154], and shape correlation have been used for matching and inference [157, 158]. Worth attention is the work by Wang [95] that proposed a combination of local and global matching scheme for region features. Such scheme can perform matching and similarity evaluation in an efficient manner (also capable of dealing with deformation or pose changes) by incorporating the spatial mutual information with the local entropy in the matching scheme. Amiri et. al. have proposed an improved SIFT based matching of potential interest points identified by searching for local peaks in Difference-of-Gaussian (DoG) images[189].

Another method of matching/inferring is to use the probabilistic model in order to evaluate the likelihood ratio [2, 4, 75] or expectation in generative models [105, 113]. Otherwise, correlation between the object template and the input image can be computed or probabilistic Hough transform can be used [77, 92, 93, 118]. Each of these measures is linked directly or indirectly with the defining ratio of the generative model, $P(\mathbf{x}|\theta, c)$, which can be computed for an input image and a given class through the learnt hidden variables θ [13]. For example, in the case of wavelet form of features, $P(\mathbf{x}|\theta, c)$ will depend upon the wavelet kernel response to the input image for a particular class [13]. Similarly, the posterior probability can be used for inference in the discriminative models. Or else, in the case of classifiers like SVM, k-nearest neighbors based method, binary classifiers, etc, the features are extracted for the input image and the posterior probability (based on the number of features voted into each class) can be used for inference [17, 20, 84]. If two or more classes have the high posterior probability, multiple objects may be inferred [75, 94]. However, if it is known that only one object is present in an image, refined methods based on feature reliability can be used.

If the object class is represented using the feature spaces, the distance of the image from the clusters in feature space is used for inference. Other methods include histograms corresponding to the features (the number of features that were detected) to decide the object category [74, 84, 149, 155].

7. BOOSTING METHODS - LEARNING WHILE VALIDATION

The weak object templates learnt during training can be made more class specific by using boosting mechanisms in the validation phase [190-211]. Boosting mechanisms typically consider an ensemble of weak features (in the object templates) and gives a boost to the stronger features corresponding to the object class. Technically, boosting method can be explained as follows. Suppose validation images \mathbf{x}_i , $i = 1$ to N contain the corresponding class labels $c_i = \pm 1$, where the value 1 indicates that the object of the considered class is present and -1 represents its absence. Let the weak classifier learnt while training be a combination of several individual

classifiers $h_j(\cdot)$, $j=1$ to J . Here, $h_j(\cdot)$ operates on the input image and gives an inference/decision regarding the presence/absence of class object. Evidently, $h_j(\cdot)$ is determined by the feature θ_j in the codebook and the inference mechanisms. Further, let us say that we want to extract maximum T strong classifiers. Then most boosting methods can be generally explained using the algorithm below:

Step 1:	Initialize the image weights $w_{i,t} = 1/N; \forall i$
Step 2:	For $t=1$ to T
Step 2.1:	Find the strongest classifier, $h_t(\cdot)$, using the current image weights. For this, first compute the error function for each classifier: $\varepsilon_j = \sum_i (w_{i,t} I_{j,i})$, where $I_{j,i} = 1$ if $c_i = h_j(\mathbf{x}_i)$, and 0 otherwise. Here, the index j is used to denote the j th classifier and the index i is used to denote the i th image. Find the classifier that resulted in minimum error (this is the strongest classifier for the weights $w_{i,t}$): $h_t(\cdot) = \arg(\min(\varepsilon_j)).$
Step 2.2:	Update the classifier weight for the chosen classifier: $\alpha_t = f(\varepsilon_t)$, where the function $f(\cdot)$ depends upon the chosen boosting technique and ε_t is the error corresponding to the current strongest classifier $h_t(\cdot)$.
Step 2.3:	If a termination condition is satisfied, then go to step 3. The termination condition depends upon the application or the boosting method used.
Step 2.4:	Update the weights $w_{i,t+1} = w_{i,t} g(\alpha_t I_i)$. Here, $g(\cdot)$ is the function that changes the weight distribution given to the validation images and is generally called the loss function. The general characteristic of $g(\cdot)$ is that it reduces the weight of the images that resulted in correct classification, so that in the next iteration, the method is less biased towards the current strong feature. Typically, $w_{i,t+1}$ is normalized after computation such that the sum of all the weights is 1.
Step 3:	The output of the boosting algorithm is typically specified as the strong classifier $h_{\text{strong}}(\cdot) = \sum_t \alpha_t h_t(\cdot).$

FIGURE 5: Generic algorithm for boosting

It is notable that some features may be repeatedly selected in step 2 of FIGURE 5, which indicates that though the method is getting lesser and lesser biased towards that feature, that feature is strong enough to be selected again and again.

There are many variations of boosting methods, which are typically differentiated based upon their loss function $g(\cdot)$ and the classifier update function $f(\cdot)$. We discuss some prominent methods used often in computer vision. The original boost used a constant value for the classifier update function $f(\cdot) = 1$ and an exponential loss function $g(\alpha_t I_i) = \exp(-\alpha_t c_i h_t(\mathbf{x}_i))$ [212, 213]. It was shown that such technique performed marginally better than the random techniques used for selecting the features from a codebook. However, the performance of boosting method was greatly enhanced by the introduction of adaptive boosting (Adaboost) [1, 212-215]. Here, the main difference is the classifier update function $f(\varepsilon_t) = 0.5 \ln((1 - \varepsilon_t) / \varepsilon_t)$. Since the value of $f(\cdot) = 0$ implies no further optimization, the termination condition is set as $\varepsilon_t \geq 0.5$. This boosting method was adapted extensively in the object detection and recognition field. Though it is efficient in avoiding the problem of over-fitting, it is typically very sensitive to noise and clutter.

A variation on the Ada-boost, Logit-boost [212, 213, 216] used similar scheme but a logistic regression function based loss function, $g(\alpha, I_i) = \ln(1 + \exp(-\alpha_i c_i h_i(\mathbf{x}_i)))$. As compared to the Ada-boost, it is more robust to the noisy and cluttered scenarios. This is because as compared to the Ada-boost, this loss function is flatter and provides a softer shift towards the noise images. Another variation on the Ada-boost is the GentleAda-boost [6, 8, 212, 213], which is similar to Ada-boost but uses a linear classifier update function $f(\varepsilon_i) = (1 - \varepsilon_i)$. The linear form of the classifier update function ensures that the overall update scheme is not severely prejudiced.

In order to understand and compare the four boosting schemes, we present the plots between the error ε and the loss function (which also incorporates the classifier update function through α) for the four boosting schemes in FIGURE 6. FIGURE 6(a) shows the value of loss function when the chosen classifier gives the correct inference for an image. If the classifier is weak (high error) and yet generates a correct inference for an image, that image is boosted so that the classifier gets boosted. Similarly, FIGURE 6(b) shows the plot when the chosen classifier generates incorrect inference for an image. If the classifier is strong (low error ε) and still generates an incorrect inference for an image, the classifier can be suppressed or weakened by boosting such image.

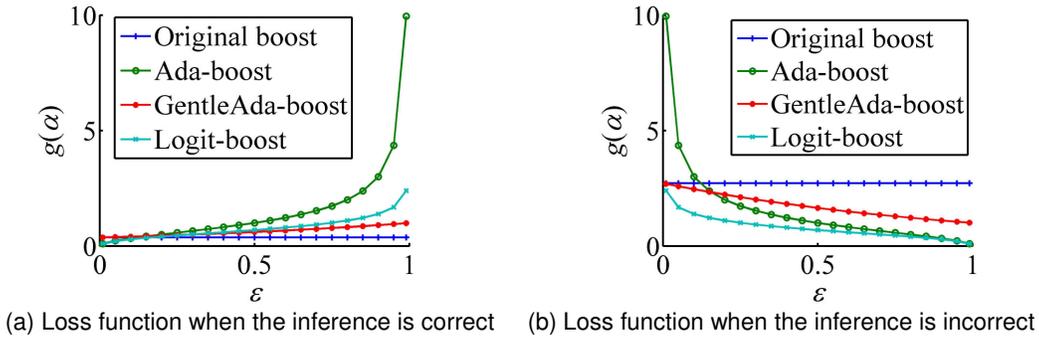


FIGURE 6: Comparison of boosting techniques. (a) Loss function when the inference is correct (b) loss function when the inference is incorrect.

It is evident that the desired property is not emulated well by the original boosting, which explains its slight (insignificant) improvement over the random selection of classifiers. On the other hand, Ada-boost is too strict in weakening or boosting the classifiers. Logit-boost and GentleAda-boost demonstrate a rather tempered performance, among whom, evidently Gentle-boost is the least non-linear and indeed the most gentle in weakening or boosting the classifiers. However, in our opinion, Logit-boost is the best among these methods precisely because of its combination of being gentle as well as non-linear. Due to the non-linearity, it is expected to converge faster than the GentleAda-boost and due to its gentle boosting characteristic, it is expected to be more robust than Ada-boost for noisy and cluttered images, where wrong inferences cannot be altogether eliminated.

The convergence of boosting techniques (except the original one) discussed above can be enhanced by using a gradient based approach for updating the weights of the images. Such approach is sometimes referred to as the Gradient-boost [123, 213, 217, 218]. However, this concept can be used within the framework of most boosting approaches. Similar contribution comes from the LP-boost (linear programming boost) methods [36, 212], where concepts of linear programming are used for computing the weights of the images. In both the schemes, the iteration (step 2 of

FIGURE 5) is cast as an optimization problem in terms of the loss function, such that the convergence direction and rate can be controlled. Such schemes also reduce the number of control parameters and make boosting less sensitive to them.

A recent work by Mallapragada [219], Semi-boost, is an interesting recent addition to the body of boosting algorithms. While the existing boosting methods assume that every image in the validation dataset is labeled, [219] considers a validation dataset in which only a few images need to be labeled. In this sense, it provides a framework for incorporating semi-supervised boosting. In each iteration (step 2 of

FIGURE 5), two major steps are done in addition to and before the mentioned steps. First, each unlabeled image is pseudo-labeled by computing the similarity of the unlabeled images with the labeled images, and a confidence value is assigned to each pseudo-label. Second, the pseudo-labeled images with high confidence values are pooled with the labeled images as the validation set to be used in the remaining steps of the iteration. As the strong features are identified iteratively, the pseudo-labeling becomes more accurate and the confidence of the set of unlabeled data increases. It has been shown in [219] that Semi-boost can be easily incorporated in the existing framework of many algorithms. This method provides three important advantages over the existing methods. First, it can accommodate scalable validation sets (where images may be added at any stage with or without labeling). Second, since semi-boost learns to increase the confidence of labeling the unlabeled images, and not just fitting the features to the labeled data, it is more efficient in avoiding over-fitting and providing better test performances. Third, though not discussed in [219], in our opinion, the similarity and pseudo-labeling schemes should help in identifying the presence of new (unknown) classes, and thus provide class-scalability as well. Although another recent work by Joshi [117] tries to attack the same problem as [219] by using a small seed training set that is completely labeled in order to learn from other unsupervised training dataset, his approach is mainly based on support vector machine (SVM) based learning. It may have its specific advantages, like suitability for multi-class data. However, semi-boost is an important improvement within the boosting algorithms, which have wider applicability than SVM based learning methods.

Another important method in the boosting techniques is the Joint-boost [1, 90], first proposed in [40, 220]. It can handle multi-class inferences directly (as opposed to other boosting techniques discussed above which use binary inference for one class at a time). The basis of joint boosting is that some features may be shared among more than one class [40, 220]. For this, the error metric is defined as $\varepsilon_j = \sum_{\kappa} \sum_i (\kappa w_{i,\kappa} \kappa I_i)$, where $\kappa = 1$ to K represents various classes, and the inference

κI_i is the binary inference for class κ . Thus, instead of learning the class-specific strong features, we can learn strong shared features. Such features are more generic over the classes and very few features are sufficient for representing the classes generically. Typically, the number of sufficient features is the logarithmic value of the number of classes [40, 220]. However, better inter-class distances can be achieved by increasing the number of features. Even then the number of features required for optimal generality and specificity is much lesser than boosting for one class at a time. Such scheme is indeed very beneficial if a bag of words is used for representing the object templates. Joint boost has also been combined with principal component analysis based system in [121] to further improve the speed of training.

8. CONCLUSION

This review paper addresses all the major aspects of an object detection framework. These include feature selection, learning model, object representation, matching features and object templates, and the boosting schemes. For each aspect, the technologies in use and the state-of-the-art research works are discussed. The merits and demerits of the works are discussed and key indicators helpful in choosing a suitable technique are also presented. Thus, the paper presents a concise summary of the state-of-the-art techniques in object detection for upcoming researchers. This study provides a preliminary, concise, but complete background of the object detection problem. Thus, based on this study, for a given problem environment and data availability, a proper framework can be chosen easily and quickly. Background codes/theories can be used from the works cited here relevant to the chosen framework and the focus of

research can be dedicated to improving or optimizing the chosen framework for better accuracy in the given problem environment.

9. REFERENCES

- [1] A. Opelt, A. Pinz, and A. Zisserman, "Learning an alphabet of shape and appearance for multi-class object detection," *International Journal of Computer Vision*, vol. 80, pp. 16-44, 2008.
- [2] Z. Si, H. Gong, Y. N. Wu, and S. C. Zhu, "Learning mixed templates for object recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 272-279.
- [3] R. Fergus, P. Perona, and A. Zisserman, "A sparse object category model for efficient learning and exhaustive recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 380-387.
- [4] Y. Chen, L. Zhu, A. Yuille, and H. J. Zhang, "Unsupervised learning of probabilistic object models (POMs) for object classification, segmentation, and recognition using knowledge propagation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1747-1774, 2009.
- [5] J. Shotton, "Contour and texture for visual recognition of object categories," Doctoral of Philosophy, Queen's College, University of Cambridge, Cambridge, 2007.
- [6] J. Shotton, A. Blake, and R. Cipolla, "Multiscale categorical object recognition using contour fragments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1270-1281, 2008.
- [7] O. C. Hamsici and A. M. Martinez, "Rotation invariant kernels and their application to shape analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1985-1999, 2009.
- [8] L. Szumilas and H. Wildenauer, "Spatial configuration of local shape features for discriminative object detection," in *Lecture Notes in Computer Science* vol. 5875, ed, 2009, pp. 22-33.
- [9] L. Szumilas, H. Wildenauer, and A. Hanbury, "Invariant shape matching for detection of semi-local image structures," in *Lecture Notes in Computer Science* vol. 5627, ed, 2009, pp. 551-562.
- [10] M. P. Kumar, P. H. S. Torr, and A. Zisserman, "OBJCUT: Efficient Segmentation Using Top-Down and Bottom-Up Cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 530-545, 2009.
- [11] K. Schindler and D. Suter, "Object detection by global contour shape," *Pattern Recognition*, vol. 41, pp. 3736-3748, 2008.
- [12] N. Alajlan, M. S. Kamel, and G. H. Freeman, "Geometry-based image retrieval in binary image databases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1003-1013, 2008.
- [13] Y. N. Wu, Z. Si, H. Gong, and S. C. Zhu, "Learning Active Basis Model for Object Detection and Recognition," *International Journal of Computer Vision*, pp. 1-38, 2009.
- [14] X. Ren, C. C. Fowlkes, and J. Malik, "Learning probabilistic models for contour completion in natural images," *International Journal of Computer Vision*, vol. 77, pp. 47-63, 2008.
- [15] A. Y. S. Chia, S. Rahardja, D. Rajan, and M. K. H. Leung, "Structural descriptors for category level object detection," *IEEE Transactions on Multimedia*, vol. 11, pp. 1407-1421, 2009.
- [16] J. Winn and J. Shotton, "The layout consistent random field for recognizing and segmenting partially occluded objects," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 37-44.
- [17] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Object detection by contour segment networks," in *Lecture Notes in Computer Science* vol. 3953, ed, 2006, pp. 14-28.
- [18] K. Mikolajczyk, B. Leibe, and B. Schiele, "Multiple object class detection with a generative model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 26-33.
- [19] R. C. Nelson and A. Selinger, "Cubist approach to object recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 1998, pp. 614-621.

- [20] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid, "Groups of adjacent contour segments for object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 36-51, 2008.
- [21] S. Ali and M. Shah, "A supervised learning framework for generic object detection in images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 1347-1354.
- [22] I. A. Rizvi and B. K. Mohan, "Improving the Accuracy of Object Based Supervised Image Classification using Cloud Basis Function Neural Network for High Resolution Satellite Images," *International Journal of Image Processing (IJIP)*, vol. 4, pp. 342-353, 2010.
- [23] D. K. Prasad and M. K. H. Leung, "A hybrid approach for ellipse detection in real images," in *2nd International Conference on Digital Image Processing*, Singapore, 2010, pp. 754601-6.
- [24] D. K. Prasad and M. K. H. Leung, "Methods for ellipse detection from edge maps of real images," in *Machine Vision - Applications and Systems*, F. Solari, M. Chessa, and S. Sabatini, Eds., ed: InTech, 2012, pp. 135-162.
- [25] P. F. Felzenszwalb, "Learning models for object recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2001, pp. 1056-1062.
- [26] E. Borenstein and S. Ullman, "Learning to segment," in *Lecture Notes in Computer Science* vol. 3023, ed, 2004, pp. 315-328.
- [27] E. Borenstein and J. Malik, "Shape guided object segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 969-976.
- [28] J. Wang, V. Athitsos, S. Sclaroff, and M. Betke, "Detecting objects of variable shape structure with Hidden State Shape Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 477-492, 2008.
- [29] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2006, pp. 13-13.
- [30] D. K. Prasad, C. Quek, and M. K. H. Leung, "Fast segmentation of sub-cellular organelles," *International Journal of Image Processing (IJIP)*, vol. 6, pp. 317-325, 2012.
- [31] Y. Amit, D. Geman, and X. Fan, "A coarse-to-fine strategy for multiclass shape detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1606-1621, 2004.
- [32] J. Shotton, A. Blake, and R. Cipolla, "Contour-based learning for object detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 503-510.
- [33] A. Opelt, A. Pinz, M. Fussenegger, and P. Auer, "Generic object recognition with boosting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 416-431, 2006.
- [34] Y. Freund, "Boosting a Weak Learning Algorithm by Majority," *Information and Computation*, vol. 121, pp. 256-285, 1995.
- [35] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-based object detection in images by components," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, pp. 349-361, 2001.
- [36] A. Demiriz, K. P. Bennett, and J. Shawe-Taylor, "Linear programming boosting via column generation," *Machine Learning*, vol. 46, pp. 225-254, 2002.
- [37] S. Agarwal, A. Awan, and D. Roth, "Learning to detect objects in images via a sparse, part-based representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1475-1490, 2004.
- [38] R. Fergus, P. Perona, and A. Zisserman, "A visual category filter for google images," in *Lecture Notes in Computer Science* vol. 3021, ed, 2004, pp. 242-256.
- [39] A. Opelt, M. Fussenegger, A. Pinz, and P. Auer, "Weak hypotheses and boosting for generic object detection and recognition," in *Lecture Notes in Computer Science* vol. 3022, ed, 2004, pp. 71-84.
- [40] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing features: Efficient boosting procedures for multiclass object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004, pp. 762-769.

- [41] A. Bar-Hillel, T. Hertz, and D. Weinshall, "Object class recognition by boosting a part-based model," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 702-709.
- [42] E. Bart and S. Ullman, "Cross-generalization: Learning novel classes from a single example by feature replacement," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 672-679.
- [43] R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from Google's image search," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 1816-1823.
- [44] F. Jurie and B. Triggs, "Creating efficient codebooks for visual recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 604-610.
- [45] Z. Tu, "Probabilistic boosting-tree: Learning discriminative models for classification, recognition, and clustering," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 1589-1596.
- [46] W. Zhang, B. Yu, G. J. Zelinsky, and D. Samaras, "Object class recognition using multiple layer boosting with heterogeneous features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 323-330.
- [47] P. Dollar, Z. Tu, and S. Belongie, "Supervised learning of edges and object boundaries," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 1964-1971.
- [48] A. Opelt, A. Pinz, and A. Zisserman, "Incremental learning of object detectors using a visual shape alphabet," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 3-10.
- [49] D. D. Le and S. Satoh, "Ent-Boost: Boosting using entropy measures for robust object detection," *Pattern Recognition Letters*, vol. 28, pp. 1083-1090, 2007.
- [50] A. Bar-Hillel and D. Weinshall, "Efficient learning of relational object class models," *International Journal of Computer Vision*, vol. 77, pp. 175-198, 2008.
- [51] P. Carbonetto, G. Dorko, C. Schmid, H. Kuck, and N. De Freitas, "Learning to recognize objects with little supervision," *International Journal of Computer Vision*, vol. 77, pp. 219-237, 2008.
- [52] L. Fuhrst, S. Fidler, and A. Leonardis, "Selecting features for object detection using an AdaBoost-compatible evaluation function," *Pattern Recognition Letters*, vol. 29, pp. 1603-1612, 2008.
- [53] X. Li, B. Yang, F. Zhu, and A. Men, "Real-time object detection based on the improved boosted features," in *Proceedings of SPIE - The International Society for Optical Engineering*, 2009.
- [54] J. J. Yokono and T. Poggio, "Object recognition using boosted oriented filter based local descriptors," *IEEJ Transactions on Electronics, Information and Systems*, vol. 129, 2009.
- [55] D. K. Prasad and M. K. H. Leung, "Reliability/Precision Uncertainty in Shape Fitting Problems," in *IEEE International Conference on Image Processing*, Hong Kong, 2010, pp. 4277-4280.
- [56] D. K. Prasad and M. K. H. Leung, "Polygonal representation of digital curves," in *Digital Image Processing*, S. G. Stanciu, Ed., ed: InTech, 2012, pp. 71-90.
- [57] D. K. Prasad, M. K. H. Leung, C. Quek, and S.-Y. Cho, "A novel framework for making dominant point detection methods non-parametric," *Image and Vision Computing*, vol. 30, pp. 843-859, 2012.
- [58] D. K. Prasad, C. Quek, and M. K. H. Leung, "A non-heuristic dominant point detection based on suppression of break points," in *Image Analysis and Recognition*. vol. 7324, A. Campilho and M. Kamel, Eds., ed Aveiro, Portugal: Springer Berlin Heidelberg, 2012, pp. 269-276.
- [59] D. K. Prasad, C. Quek, M. K. H. Leung, and S. Y. Cho, "A parameter independent line fitting method," in *Asian Conference on Pattern Recognition (ACPR)*, Beijing, China, 2011, pp. 441-445.
- [60] D. K. Prasad, "Assessing error bound for dominant point detection," *International Journal of Image Processing (IJIP)*, vol. 6, pp. 326-333, 2012.

- [61] Y. Chi and M. K. H. Leung, "Part-based object retrieval in cluttered environment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 890-895, May 2007.
- [62] D. K. Prasad and M. K. H. Leung, "An ellipse detection method for real images," in *25th International Conference of Image and Vision Computing New Zealand (IVCNZ 2010)*, Queenstown, New Zealand, 2010, pp. 1-8.
- [63] D. K. Prasad, M. K. H. Leung, and S. Y. Cho, "Edge curvature and convexity based ellipse detection method," *Pattern Recognition*, vol. 45, pp. 3204-3221, 2012.
- [64] D. K. Prasad, M. K. H. Leung, and C. Quek, "ElliFit: An unconstrained, non-iterative, least squares based geometric Ellipse Fitting method," *Pattern Recognition*, 2013.
- [65] D. K. Prasad, C. Quek, and M. K. H. Leung, "A precise ellipse fitting method for noisy data," in *Image Analysis and Recognition*. vol. 7324, A. Campilho and M. Kamel, Eds., ed Aveiro, Portugal: Springer Berlin Heidelberg, 2012, pp. 253-260.
- [66] D. K. Prasad and M. K. H. Leung, "Error analysis of geometric ellipse detection methods due to quantization," in *Fourth Pacific-Rim Symposium on Image and Video Technology (PSIVT 2010)*, Singapore, 2010, pp. 58 - 63.
- [67] D. K. Prasad, R. K. Gupta, and M. K. H. Leung, "An Error Bounded Tangent Estimator for Digitized Elliptic Curves," in *Discrete Geometry for Computer Imagery*. vol. 6607, ed: Springer Berlin / Heidelberg, 2011, pp. 272-283.
- [68] C. F. Olson, "A general method for geometric feature matching and model extraction," *International Journal of Computer Vision*, vol. 45, pp. 39-54, 2001.
- [69] H. P. Moravec, "Rover visual obstacle avoidance," in *Proceedings of the International Joint Conference on Artificial Intelligence*, Vancouver, CANADA, 1981, pp. 785-790.
- [70] C. Harris and M. Stephens, "A combined corner and edge detector," presented at the Alvey Vision Conference, 1988.
- [71] J. Li and N. M. Allinson, "A comprehensive review of current local features for computer vision," *Neurocomputing*, vol. 71, pp. 1771-1787, 2008.
- [72] K. Mikolajczyk and H. Uemura, "Action recognition with motion-appearance vocabulary forest," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1-8.
- [73] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2003, pp. 1615-1630.
- [74] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, pp. 91-110, 2004.
- [75] B. Ommer and J. Buhmann, "Learning the Compositional Nature of Visual Object Categories for Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.
- [76] B. Leibe, A. Leonardis, and B. Schiele, "Robust object detection with interleaved categorization and segmentation," *International Journal of Computer Vision*, vol. 77, pp. 259-289, 2008.
- [77] M. Varma and A. Zisserman, "A statistical approach to material classification using image patch exemplars," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2032-2047, 2009.
- [78] P. M. Roth, S. Sternig, H. Grabner, and H. Bischof, "Classifier grids for robust adaptive object detection," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 2727-2734.
- [79] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, pp. 761-767, 2004.
- [80] G. Carneiro and A. D. Jepson, "The quantitative characterization of the distinctiveness and robustness of local image descriptors," *Image and Vision Computing*, vol. 27, pp. 1143-1156, 2009.
- [81] W. T. Lee and H. T. Chen, "Histogram-based interest point detectors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1590-1596.
- [82] H. T. Comer and B. A. Draper, "Interest Point Stability Prediction," in *Proceedings of the International Conference on Computer Vision Systems*, Liege, 2009.

- [83] Y. Ke and R. Sukthankar, "PCA-SIFT: A more distinctive representation for local image descriptors," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2004, pp. 506-513.
- [84] H. Zhang, W. Gao, X. Chen, and D. Zhao, "Object detection using spatial histogram features," *Image and Vision Computing*, vol. 24, pp. 327-341, 2006.
- [85] R. Sandler and M. Lindenbaum, "Optimizing gabor filter design for texture edge detection and classification," *International Journal of Computer Vision*, vol. 84, pp. 308-324, 2009.
- [86] H. Bischof, H. Wildenauer, and A. Leonardis, "Illumination insensitive recognition using eigenspaces," *Computer Vision and Image Understanding*, vol. 95, pp. 86-104, 2004.
- [87] C. H. Lampert and J. Peters, "Active structured learning for high-speed object detection," in *Lecture Notes in Computer Science* vol. 5748, ed, 2009, pp. 221-231.
- [88] C. Wallraven, B. Caputo, and A. Graf, "Recognition with local features: The kernel recipe," in *Proceedings of the IEEE International Conference on Computer Vision*, 2003, pp. 257-264.
- [89] A. Zalesny, V. Ferrari, G. Caenen, and L. Van Gool, "Composite texture synthesis," *International Journal of Computer Vision*, vol. 62, pp. 161-176, 2005.
- [90] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context," *International Journal of Computer Vision*, vol. 81, pp. 2-23, 2009.
- [91] M. V. Rohith, G. Somanath, D. Metaxas, and C. Kambhamettu, "D - Clutter: Building object model library from unsupervised segmentation of cluttered scenes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2783-2789.
- [92] C. Gu, J. J. Lim, P. Arbeláez, and J. Malik, "Recognition using regions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1030-1037.
- [93] A. Bosch, A. Zisserman, and X. Muñoz, "Scene classification using a hybrid generative/discriminative approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 712-727, 2008.
- [94] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, pp. 55-79, 2005.
- [95] H. Wang and J. Oliensis, "Rigid shape matching by segmentation averaging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 619-635, 2010.
- [96] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1615-1630, 2005.
- [97] T. Tuytelaars and K. Mikolajczyk, "Local invariant feature detectors: A survey," *Foundations and Trends in Computer Graphics and Vision*, vol. 3, pp. 177-280, 2007.
- [98] L. Juan and O. Gwun, "A Comparison of SIFT, PCA-SIFT and SURF," *International Journal of Image Processing (IJIP)*, vol. 3, pp. 143-152, 2009.
- [99] N. Adluru and L. J. Latecki, "Contour grouping based on contour-skeleton duality," *International Journal of Computer Vision*, vol. 83, pp. 12-29, 2009.
- [100] D. Cailliere, F. Denis, D. Pele, and A. Baskurt, "3D mirror symmetry detection using Hough transform," in *Proceedings of the IEEE International Conference on Image Processing*, San Diego, CA, 2008, pp. 1772-1775.
- [101] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2003, pp. 409-415.
- [102] P. Schnitzspan, M. Fritz, S. Roth, and B. Schiele, "Discriminative structure learning of hierarchical representations for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2238-2245.
- [103] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation by image exploration," in *Lecture Notes in Computer Science* vol. 3021, ed, 2004, pp. 40-54.
- [104] V. Ferrari, T. Tuytelaars, and L. Van Gool, "Simultaneous object recognition and segmentation from single or multiple model views," *International Journal of Computer Vision*, vol. 67, pp. 159-188, 2006.
- [105] A. R. Pope and D. G. Lowe, "Probabilistic models of appearance for 3-D object recognition," *International Journal of Computer Vision*, vol. 40, pp. 149-167, 2000.

- [106] J. A. Lasserre, C. M. Bishop, and T. P. Minka, "Principled hybrids of generative and discriminative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 87-94.
- [107] A. E. C. Pece, "On the computational rationale for generative models," *Computer Vision and Image Understanding*, vol. 106, pp. 130-143, 2007.
- [108] M. Andriluka, S. Roth, and B. Schiele, "Discriminative Appearance Models for Pictorial Structures," *International Journal of Computer Vision*, vol. 99, pp. 259-280, Sep 2012.
- [109] C. Desai, D. Ramanan, and C. C. Fowlkes, "Discriminative Models for Multi-Class Object Layout," *International Journal of Computer Vision*, vol. 95, pp. 1-12, Oct 2011.
- [110] Y. Aytar and A. Zisserman, "Tabula Rasa: Model Transfer for Object Category Detection," in *IEEE International Conference on Computer Vision*, 2011, pp. 2252-2259.
- [111] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental Bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, pp. 59-70, 2007.
- [112] I. Ulusoy and C. M. Bishop, "Generative versus discriminative methods for object recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 258-265.
- [113] G. Bouchard and B. Triggs, "Hierarchical part-based visual object categorization," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 710-715.
- [114] T. M. Mitchell, *Machine Learning*: Mcgraw-Hill International Edition, 2010.
- [115] C. H. Lampert, H. Nickisch, and S. Harmeling, "Learning to detect unseen object classes by between-class attribute transfer," in *Proceedings of the IEEE Computer Vision and Pattern Recognition Workshop*, 2009, pp. 951-958.
- [116] T. Yeh, J. J. Lee, and T. Darrell, "Fast concurrent object localization and recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 280-287.
- [117] A. J. Joshi, F. Porikli, and N. Papanikolopoulos, "Multi-class active learning for image classification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2372-2379.
- [118] S. Maji and J. Malik, "Object detection using a max-margin hough transform," in *Proceedings of the IEEE Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 1038-1045.
- [119] L. Wu, Y. Hu, M. Li, N. Yu, and X. S. Hua, "Scale-invariant visual language modeling for object categorization," *IEEE Transactions on Multimedia*, vol. 11, pp. 286-294, 2009.
- [120] P. Jain and A. Kapoor, "Active learning for large Multi-class problems," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 762-769.
- [121] A. Stefan, V. Athitsos, Q. Yuan, and S. Sclaroff, "Reducing JointBoost-Based Multiclass Classification to Proximity Search," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 589-596.
- [122] D. Parikh, C. L. Zitnick, and T. Chen, "Unsupervised learning of hierarchical spatial structures in images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2743-2750.
- [123] S. Chen, J. Q. Wang, Y. Ouyang, B. Wang, C. S. Xu, and H. Q. Lu, "Boosting part-sense multi-feature learners toward effective object detection," *Computer Vision and Image Understanding*, vol. 115, pp. 364-374, Mar 2011.
- [124] C. Dubout and F. Fleuret, "Tasting Families of Features for Image Classification," in *IEEE International Conference on Computer Vision*, 2011, pp. 929-936.
- [125] P. Felzenszwalb, "Object Detection Grammars," in *IEEE International Conference on Computer Vision Workshops*, 2011.
- [126] G. X. Huang, H. F. Chen, Z. L. Zhou, F. Yin, and K. Guo, "Two-class support vector data description," *Pattern Recognition*, vol. 44, pp. 320-329, Feb 2011.
- [127] M. Jamieson, Y. Eskin, A. Fazly, S. Stevenson, and S. J. Dickinson, "Discovering hierarchical object models from captioned images," *Computer Vision and Image Understanding*, vol. 116, pp. 842-853, Jul 2012.

- [128] L. Jie, T. Tommasi, and B. Caputo, "Multiclass Transfer Learning from Unconstrained Priors," in *IEEE International Conference on Computer Vision*, 2011, pp. 1863-1870.
- [129] T. Y. Ma and L. J. Latecki, "From Partial Shape Matching through Local Deformation to Robust Global Shape Similarity for Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1441-1448.
- [130] M. Maire, S. X. Yu, and P. Perona, "Object Detection and Segmentation from Joint Embedding of Parts and Pixels," in *IEEE International Conference on Computer Vision*, 2011, pp. 2142-2149.
- [131] S. Nilufar, N. Ray, and H. Zhang, "Object Detection With DoG Scale-Space: A Multiple Kernel Learning Approach," *IEEE Transactions on Image Processing*, vol. 21, pp. 3744-3756, Aug 2012.
- [132] E. Rahtu, J. Kannala, and M. Blaschko, "Learning a Category Independent Object Detection Cascade," in *IEEE International Conference on Computer Vision*, 2011, pp. 1052-1059.
- [133] N. Razavi, J. Gall, and L. Van Gool, "Scalable Multi-class Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1505-1512.
- [134] S. Rivera and A. M. Martinez, "Learning deformable shape manifolds," *Pattern Recognition*, vol. 45, pp. 1792-1801, Apr 2012.
- [135] R. Salakhutdinov, A. Torralba, and J. Tenenbaum, "Learning to Share Visual Appearance for Multiclass Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1481-1488.
- [136] J. Stottinger, A. Hanbury, N. Sebe, and T. Gevers, "Sparse Color Interest Points for Image Retrieval and Object Categorization," *IEEE Transactions on Image Processing*, vol. 21, pp. 2681-2692, May 2012.
- [137] G. Toliás and Y. Avrithis, "Speeded-up, relaxed spatial matching," in *IEEE International Conference on Computer Vision*, 2011, pp. 1653-1660.
- [138] K. Tzevanidis and A. Argyros, "Unsupervised learning of background modeling parameters in multicamera systems," *Computer Vision and Image Understanding*, vol. 115, pp. 105-116, Jan 2011.
- [139] M. Villamizar, J. Andrade-Cetto, A. Sanfeliu, and F. Moreno-Noguer, "Bootstrapping Boosted Random Ferns for discriminative and efficient object classification," *Pattern Recognition*, vol. 45, pp. 3141-3153, Sep 2012.
- [140] X. G. Wang, X. Bai, W. Y. Liu, and L. J. Latecki, "Feature Context for Image Classification and Object Detection," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 961-968.
- [141] A. Zaharescu, E. Boyer, and R. Horaud, "Keypoints and Local Descriptors of Scalar Functions on 2D Manifolds," *International Journal of Computer Vision*, vol. 100, pp. 78-98, Oct 2012.
- [142] M. Fritz, B. Leibe, B. Caputo, and B. Schiele, "Integrating representative and discriminant models for object category detection," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 1363-1370.
- [143] Y. Li, L. Gu, and T. Kanade, "A robust shape model for multi-view car alignment," in *Proceedings of the IEEE Computer Vision and Pattern Recognition Workshop*, 2009, pp. 2466-2473.
- [144] D. K. Prasad, "Adaptive traffic signal control system with cloud computing based online learning," in *8th International Conference on Information, Communications, and Signal Processing (ICICS 2011)*, Singapore, 2011.
- [145] D. K. Prasad, "High Availability based Migration Analysis to Cloud Computing for High Growth Businesses," *International Journal of Computer Networks (IJCN)*, vol. 4, 2012.
- [146] M. F. Demirci, A. Shokoufandeh, and S. J. Dickinson, "Skeletal shape abstraction from examples," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 944-952, 2009.
- [147] M. Bergtholdt, J. Kappes, S. Schmidt, and C. Schnörr, "A study of parts-based object class detection using complete graphs," *International Journal of Computer Vision*, vol. 87, pp. 93-117, 2010.

- [148] J. R. R. Uijlings, A. W. M. Smeulders, and R. J. H. Scha, "What is the spatial extent of an object?," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 770-777.
- [149] F. Perronnin, "Universal and adapted vocabularies for generic visual categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 1243-1256, 2008.
- [150] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2006, pp. 2169-2178.
- [151] D. A. Ross and R. S. Zemel, "Learning parts-based representations of data," *Journal of Machine Learning Research*, vol. 7, pp. 2369-2397, 2006.
- [152] K. Y. Chang, T. L. Liu, H. T. Chen, and S. H. Lai, "Fusing Generic Objectness and Visual Saliency for Salient Object Detection," in *IEEE International Conference on Computer Vision*, 2011, pp. 914-921.
- [153] S. Lazebnik and M. Raginsky, "Supervised learning of quantizer codebooks by information loss minimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1294-1309, 2009.
- [154] B. Epshtein and S. Ullman, "Feature hierarchies for object classification," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 220-227.
- [155] J. Gall and V. Lempitsky, "Class-specific hough forests for object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2009, pp. 1022-1029.
- [156] S. Basalamah, A. Bharath, and D. McRobbie, "Contrast marginalised gradient template matching," in *Lecture Notes in Computer Science* vol. 3023, ed, 2004, pp. 417-429.
- [157] S. Belongie, J. Malik, and J. Puzicha, "Matching shapes," in *Proceedings of the IEEE International Conference on Computer Vision*, 2001, pp. 454-461.
- [158] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 509-522, 2002.
- [159] A. C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2005, pp. 26-33.
- [160] S. Biswas, G. Aggarwal, and R. Chellappa, "Robust estimation of albedo for illumination-invariant matching and shape recovery," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 884-899, 2009.
- [161] G. Borgefors, "Hierarchical Chamfer matching: A parametric edge matching algorithm," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 10, pp. 849-865, 1988.
- [162] A. M. Bronstein, M. M. Bronstein, A. M. Bruckstein, and R. Kimmel, "Partial similarity of objects, or how to compare a centaur to a horse," *International Journal of Computer Vision*, vol. 84, pp. 163-183, 2009.
- [163] R. Brunelli and T. Poggio, "Template matching: Matched spatial filters and beyond," *Pattern Recognition*, vol. 30, pp. 751-768, 1997.
- [164] G. J. Burghouts and J. M. Geusebroek, "Performance evaluation of local colour invariants," *Computer Vision and Image Understanding*, vol. 113, pp. 48-62, 2009.
- [165] J. R. Burrill, S. X. Wang, A. Barrow, M. Friedman, and M. Soffen, "Model-based matching using elliptical features," in *Proceedings of SPIE - The International Society for Optical Engineering*, 1996, pp. 87-97.
- [166] M. Ceccarelli and A. Petrosino, "The orientation matching approach to circular object detection," in *Proceedings of the IEEE International Conference on Image Processing*, 2001, pp. 712-715.
- [167] S. H. Chang, F. H. Cheng, W. H. Hsu, and G. Z. Wu, "Fast algorithm for point pattern matching: Invariant to translations, rotations and scale changes," *Pattern Recognition*, vol. 30, pp. 311-320, 1997.
- [168] F. H. Cheng, "Multi-stroke relaxation matching method for handwritten Chinese character recognition," *Pattern Recognition*, vol. 31, pp. 401-410, 1998.

- [169] Y. Chi and M. K. H. Leung, "A local structure matching approach for large image database retrieval," in *Proceedings of the International Conference on Image Analysis and Recognition*, Oporto, PORTUGAL, 2004, pp. 761-768.
- [170] T. H. Cho, "Object matching using generalized hough transform and chamfer matching," in *Lecture Notes in Computer Science* vol. 4099, ed, 2006, pp. 1253-1257.
- [171] O. Choi and I. S. Kweon, "Robust feature point matching by preserving local geometric consistency," *Computer Vision and Image Understanding*, vol. 113, pp. 726-742, 2009.
- [172] P. F. Felzenszwalb, "Representation and detection of deformable shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 208-220, 2005.
- [173] P. F. Felzenszwalb and J. D. Schwartz, "Hierarchical matching of deformable shapes," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.
- [174] Y. S. Gao and M. K. H. Leung, "Line segment Hausdorff distance on face matching," *Pattern Recognition*, vol. 35, pp. 361-371, Feb 2002.
- [175] H. Hakalahti, D. Harwood, and L. S. Davis, "Two-dimensional object recognition by matching local properties of contour points," *Pattern Recognition Letters*, vol. 2, pp. 227-234, 1984.
- [176] F. Li, M. K. H. Leung, and X. Z. Yu, "A two-level matching scheme for speedy and accurate palmprint identification," in *Proceedings of the International Multimedia Modeling Conference*, Singapore, SINGAPORE, 2007, pp. 323-332.
- [177] X. Lin, Z. Zhu, and W. Deng, "Stereo matching algorithm based on shape similarity for indoor environment model building," in *Proceedings of the IEEE International Conference on Robotics and Automation*, 1996, pp. 765-770.
- [178] Z. Lin and L. S. Davis, "Shape-based human detection and segmentation via hierarchical part-template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 604-618, 2010.
- [179] H.-C. Liu and M. D. Srinath, "Partial shape classification using contour matching in distance transformation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 1072-1079, 1990.
- [180] G. Mori, S. Belongie, and J. Malik, "Efficient shape matching using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 1832-1837, 2005.
- [181] C. F. Olson and D. P. Huttenlocher, "Automatic target recognition by matching oriented edge pixels," *IEEE Transactions on Image Processing*, vol. 6, pp. 103-113, 1997.
- [182] F. C. D. Tsai, "Robust affine invariant matching with application to line features," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 1993, pp. 393-399.
- [183] C. Xu, J. Liu, and X. Tang, "2D shape matching by contour flexibility," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 180-186, 2009.
- [184] X. Z. Yu and M. K. H. Leung, "Shape recognition using curve segment Hausdorff distance," in *Proceedings of the International Conference on Pattern Recognition*, Hong Kong, PEOPLES R CHINA, 2006, pp. 441-444.
- [185] X. Z. Yu, M. K. H. Leung, and Y. S. Gao, "Hausdorff distance for shape matching," in *Proceedings of the IASTED International Conference on Visualization, Imaging, and Image Processing*, Marbella, SPAIN, 2004, pp. 819-824.
- [186] F. Li and M. K. H. Leung, "Two-stage approach for palmprint identification using Hough transform and Hausdorff distance," in *Proceedings of the International Conference on Control, Automation, Robotics and Vision*, Singapore, SINGAPORE, 2006, pp. 1302-1307.
- [187] D. G. Sim and R. H. Park, "Two-dimensional object alignment based on the robust oriented Hausdorff similarity measure," *IEEE Transactions on Image Processing*, vol. 10, pp. 475-483, 2001.
- [188] S. Lazebnik, C. Schmid, and J. Ponce, "A maximum entropy framework for part-based texture and object recognition," in *Proceedings of the IEEE International Conference on Computer Vision*, 2005, pp. 832-838.
- [189] M. Amiri and H. R. Rabiee, "RASIM: A Novel Rotation and Scale Invariant Matching of Local Image Interest Points," *IEEE Transactions on Image Processing*, vol. 20, pp. 3580-3591, Dec 2011.

- [190] D. Gerónimo, A. D. Sappa, D. Ponsa, and A. M. López, "2D-3D-based on-board pedestrian detection system," *Computer Vision and Image Understanding*, p. In press, 2010.
- [191] P. Wang and Q. Ji, "Multi-view face and eye detection using discriminant features," *Computer Vision and Image Understanding*, vol. 105, pp. 99-111, 2007.
- [192] S. Avidan, "Ensemble tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 261-271, 2007.
- [193] M. Enzweiler and D. M. Gavrila, "Monocular pedestrian detection: Survey and experiments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2179-2195, 2009.
- [194] Z. He, T. Tan, Z. Sun, and X. Qiu, "Toward accurate and fast iris segmentation for iris biometrics," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1670-1684, 2009.
- [195] C. Huang, H. Ai, Y. Li, and S. Lao, "High-performance rotation invariant multiview face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 671-686, 2007.
- [196] J. J. LaViola Jr and R. C. Zeleznik, "A practical approach for writer-dependent symbol recognition using a writer-independent symbol recognizer," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 1917-1926, 2007.
- [197] S. Z. Li and Z. Q. Zhang, "FloatBoost learning and statistical face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1112-1123, 2004.
- [198] E. Makinen and R. Raisamo, "Evaluation of gender classification methods with automatically detected and aligned faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 541-547, 2008.
- [199] J. J. Rodríguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A New classifier ensemble method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 1619-1630, 2006.
- [200] J. Wu, S. C. Brubaker, M. D. Mullin, and J. M. Rehg, "Fast asymmetric learning for cascade face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 369-382, 2008.
- [201] S. Baluja and H. A. Rowley, "Boosting sex identification performance," *International Journal of Computer Vision*, vol. 71, pp. 111-119, 2007.
- [202] J. H. Elder, S. J. D. Prince, Y. Hou, M. Sizintsev, and E. Olevskiy, "Pre-attentive and attentive detection of humans in wide-field scenes," *International Journal of Computer Vision*, vol. 72, pp. 47-66, 2007.
- [203] Y. Liu, X. L. Wang, H. Y. Wang, H. Zha, and H. Qin, "Learning Robust Similarity Measures for 3D Partial Shape Retrieval," *International Journal of Computer Vision*, pp. 1-24, 2009.
- [204] J. Porway, Q. Wang, and S. C. Zhu, "A Hierarchical and Contextual Model for Aerial Image Parsing," *International Journal of Computer Vision*, pp. 1-30, 2009.
- [205] H. Schneiderman and T. Kanade, "Object detection using the statistics of parts," *International Journal of Computer Vision*, vol. 56, pp. 151-177, 2004.
- [206] J. Šochman and J. Matas, "Learning fast emulators of binary decision processes," *International Journal of Computer Vision*, vol. 83, pp. 149-163, 2009.
- [207] K. Tieu and P. Viola, "Boosting Image Retrieval," *International Journal of Computer Vision*, vol. 56, pp. 17-36, 2004.
- [208] Z. Tu, X. Chen, A. L. Yuille, and S. C. Zhu, "Image parsing: Unifying segmentation, detection, and recognition," *International Journal of Computer Vision*, vol. 63, pp. 113-140, 2005.
- [209] P. Viola and M. J. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision*, vol. 57, pp. 137-154, 2004.
- [210] P. Viola, M. J. Jones, and D. Snow, "Detecting pedestrians using patterns of motion and appearance," *International Journal of Computer Vision*, vol. 63, pp. 153-161, 2005.
- [211] B. Wu and R. Nevatia, "Detection and tracking of multiple, partially occluded humans by Bayesian combination of edgelet based part detectors," *International Journal of Computer Vision*, vol. 75, pp. 247-266, 2007.
- [212] P. J. Bickel, Y. Ritov, and A. Zakai, "Some theory for generalized boosting algorithms," *Journal of Machine Learning Research*, vol. 7, pp. 705-732, 2006.

- [213] M. Culp, K. Johnson, and G. Michailidis, "Ada: An R package for stochastic boosting," *Journal of Statistical Software*, vol. 17, pp. 1-27, 2006.
- [214] R. E. Schapire and Y. Singer, "Improved boosting algorithms using confidence-rated predictions," *Machine Learning*, vol. 37, pp. 297-336, 1999.
- [215] Y. Freund, "An adaptive version of the boost by majority algorithm," *Machine Learning*, vol. 43, pp. 293-318, 2001.
- [216] M. Collins, R. E. Schapire, and Y. Singer, "Logistic regression, AdaBoost and Bregman distances," *Machine Learning*, vol. 48, pp. 253-285, 2002.
- [217] J. H. Friedman, "Greedy function approximation: A gradient boosting machine," *Annals of Statistics*, vol. 29, pp. 1189-1232, 2001.
- [218] J. H. Friedman, "Stochastic gradient boosting," *Computational Statistics and Data Analysis*, vol. 38, pp. 367-378, 2002.
- [219] P. K. Mallapragada, R. Jin, A. K. Jain, and Y. Liu, "SemiBoost: Boosting for semi-supervised learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 2000-2014, 2009.
- [220] A. Torralba, K. P. Murphy, and W. T. Freeman, "Sharing visual features for multiclass and multiview object detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 854-869, 2007.

Hyperspectral Data Compression Using Spatial-Spectral Lossless Coding Technique

Ayman Ahmed

*Egyptian Space Program
National Authority for Remote Sensing and Space science
Cairo, Egypt*

a_ymn2002@yahoo.com

Salwa ElRamly

*Ain-Shams University
Department of Electronics and Communication Engineering
Cairo, Egypt*

sramlye@netscape.net

Mohamed El. Sharkawy

*E-JUST—Egypt-Japan University of science and technology
Alexandria, Egypt*

DSPSYS@aol.com

Abstract

Hyperspectral imaging is widely used in many applications; especially in vegetation, climate changes, and desert studies. Such kind of imaging has a huge amount of data, which requires transmission, processing, and storage resources especially for space borne imaging. Compression of hyperspectral data cubes is an effective solution for these problems. Lossless compression of the hyperspectral data usually results in low compression ratio, which may not meet the available resources; on the other hand, lossy compression may give the desired ratio, but with a significant degradation effect on object identification performance of the hyperspectral data. Moreover, most hyperspectral data compression techniques exploits the similarities in spectral dimensions; which requires bands reordering or regrouping, to make use of the spectral redundancy. In this paper, we analyze the spectral cross correlation between bands for Hyperion hyperspectral data; spectral cross correlation matrix is calculated, assessing the strength of the spectral matrix, and finally, we propose new technique to find highly correlated groups of bands in the hyperspectral data cube based on "inter band correlation square", from the resultant groups of bands we propose a new predictor that can predict efficiently the whole bands within data cube based on weighted combination of spectral and spatial prediction, the results are evaluated versus other state of the art predictor for lossless compression.

Keywords: Hyperspectral Compression; Band Regrouping; Edge Detection; Spectral correlation Matrix.

1. INTRODUCTION

Hyperspectral data contains a huge amount of spectral data distinctive in spectral resolution, which allows identification of each pixel based on its spectral footprint. On the other hand; this amount of data increases as spectral bands increase, usually satellite instrument that measures earth's illumination at specified spectral band, has more dynamic range than visual images; typically ranges from 10 up to 16 bits per pixel per band; additionally, considering the swath width of the satellite imagery; hyperspectral imaging session of a satellite may contain tremendous amount of digital data to be transmitted to ground station[1]; this limits the imaging session and spatial resolution.

Many researches have been conducted to efficiently "carefully" compress this amount of data without losing the main advantage of hyperspectral imaging which is spectral resolution; two known compression approaches are usually investigated, lossy and lossless techniques; lossless

compression is perfect for compression data and keeping the original information without distortion and in the same time allow further processing of the image to identify earth's objects accurately; unfortunately this approach of compression gives compression ratio ranges from 1 up to 3[2][3]; that means the compressed data will have smaller volume down to 3 times less the original one; this, in practical situations, is not sufficient; while compressed data still represents a significant issue for onboard satellite designer for transmission and storage[4].

On the other hand; lossy compression approach gives a great compression ratio, which may goes up to 40 times; this ratio, is achieved scarifying the low distortion rate; that means more losses will appear on the reconstructed data; this losses can and will affect the process of earth's object identification process[5]. Another approach of compression is known as near lossless; this approach achieves relatively higher compression ratio than achieved by lossless approach and smaller distortion less than resulted from lossy compression approach; that is controlled by losses threshold.

Meanwhile researches are continuing to find optimum solution that can fit onboard satellites[6]; most researches goes around exploitation of either spectral or spatial redundancy of hyperspectral data or both of them; spatial redundancy results from the fact that imaging certain territory will have similarity in spatial dimensions; these characteristics are exhaustively investigated during last decades; and as a result we have discrete cosine transformation and wavelet transformation; it exploits the spatial redundancy in the images in different ways of implementation either in ground image processing software or onboard satellite instrumentations. On the other hand; spectral redundancy is relatively a new dimension in hyperspectral imaging; many researches are trying to investigate the best way to deal with this redundancy[7], [8] and optimum techniques to exploit it; these facts lead to another activity of investigation and analysis of spectral structure of hyperspectral data [9],[10].

We introduce a new technique that can exploit spectral and spatial redundancies in the hyperspectral data cubes, based on the on modified median predictor[11] that can be extended to third dimension of prediction based on spectral analysis.

This paper is organized as follow; first section discusses the origin of the topic, followed by explaining similarity measurement and spectral cross correlation structure of the hyperspectral data cubes, third section proposes calculation of global reference band that will be used in prediction; fourth section discusses the proposed J-predictor; section five, explains the predictor model and finally we evaluate the results against state of the art lossless compression in section six.

2. INTER-BAND SPECTRAL CROSS CORRELATION AND SIMILARITY MEASUREMENTS

Hyperspectral data can be viewed as a "Data-Cube"; this data cube has two spatial dimensions and one spectral dimension, spectral dimension represent the captured image in different spectral bands, usually successive. Spectral redundancy is based on the similarity between bands and each other's; these similarity can be measured by spectral cross correlation[12], Conditional entropy, mutual information, Euclidian Distance, Maximum Absolute Distance, and Centered Euclidian Distance; these measures are well studied by researchers and compared to determine which one is best fit for regrouping the bands for prediction based compression techniques; correlation is found to be the best for similarity measurement [8][13][14]; this results is a good point to start the analysis of spectral structure of the hyperspectral data cube.

Correlation between spectral bands is named "spectral correlation" as it represents the correlation between two identical images in different spectral domains. The imaged piece of land by hyperspectral instrument is treated as unknown object, since the main objective of the satellite imagery is to provide data about these objects. The dependencies between bands should be reflectance of the material spectral response collected by imager; this may appear to be random

dependencies, but the average correlation between bands is the main measuring criteria of similarity, which is dependent on average material spectral response within the single band. Cross correlation is usually a standard measure of degree of similarity between two images (matrices); some techniques of cross correlation estimation was used to investigate the hyperspectral inter-band correlation; this process is time consuming and requires extensive computational power.

Cross correlation mainly depends on covariance calculation between the two bands; while normalized cross correlation uses variance of each band as divisor to remove values dependency on the variation of both brightness and contrast of the image.

Selection of estimation technique should be based on deterministic criteria; such as simplicity, speed, minimum resources of memory and computational power. Fast normalized cross correlation [15][16] is a very good technique for calculating the similarity between two images; it is fast and calculate how much similarity two images are independent of their individual brightness and contrast.

Using fast normalized cross correlation techniques, correlation between all hyperspectral bands in the data cube is estimated; for band i, correlation value is estimated with all other bands in the data cube j; using the Eq. (1).

Data cube is estimated; for band i, correlation value is estimated with all other bands in the data cube j; using the Eq. (1).

$$NCC(i, j) = \frac{\sum_{x,y} [(D_i(x,y) - \bar{D}_i) \times (D_j(x,y) - \bar{D}_j)]}{\sqrt{\sum_{x,y} (D_i(x,y) - \bar{D}_i)^2} \times \sqrt{\sum_{x,y} (D_j(x,y) - \bar{D}_j)^2}} \quad (1)$$

Where:

NCC(i,j): Normalized cross correlation between bands i. and j,

$D_i(x,y)$: intensity of pixel, (x, y) pixel indices within one band,

\bar{D}_i : mean of pixel intensity values of band i.

This function is implemented in fast, optimized way in Matlab image processing tool box [17], based on "Fast Normalized Cross-Correlation" [15].

We estimating the inter-band cross correlation matrix for each hyperspectral data cube; three spaceborne hyperspectral data samples are processes and used in this study; Table 1, illustrates the details of each data samples, associated figure that reflects the image view of spectral correlation matrix, and name of spectral correlation matrix (SCM) used in calculations the resulted spectral correlation matrix for three hyperspectral data samples, from both airborne and space borne instruments [18][19], are illustrated in FIGURE 1, FIGURE 2, FIGURE 3.

We can notice that for "corr_mtxer" there is a strong correlation between the groups of bands starting from approximately band number 10 till band number 55, and group of bands starting from approximately band number 180 till band number 220 (GOB (~10-55) and GOB (~180-220)). This correlation appears in a square manner, dashed square in, this is called "Inter Band Correlation Square" (IBCS); IBCS itself is not symmetric around spectral correlation matrix diagonal. Checking of the existence of the IBCS, allows determining the correlation between

bands in hyperspectral data cube, this helps in band regrouping techniques used in compression of hyperspectral data[20][21][22][15].

Hyperspectral data sample	Details	Figure number	Spectral correlation matrix name
"ErtAle"	Hyperion instrument: 3187 lines x 256 samples x 242 bands, instrument bit depth = 12 bits.	FIGURE 1	"corr_mtxer"
"LakeMonona"	Hyperion instrument: 3176 lines x 256 samples x 242 bands, instrument bit depth = 12 bits.	FIGURE 2	"corr_mtxl"
"MtStHelens"	Hyperion instrument: 3242 lines x 256 samples x 242 bands, instrument bit depth = 12 bits.	FIGURE 3	"corr_mtxhel"

TABLE 1 Summary of hyperspectral data samples, its corresponding figures of correlation matrix, and name of correlation matrix

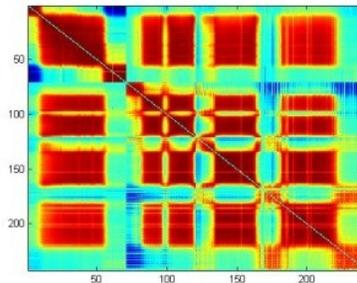


FIGURE 1: Image view of SCM for "corr_mtxer"

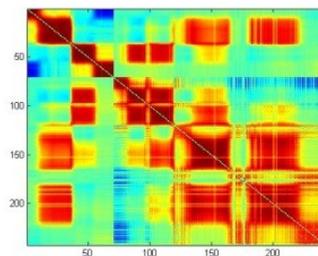


FIGURE 2: Image view of SCM for "corr_mtxl"

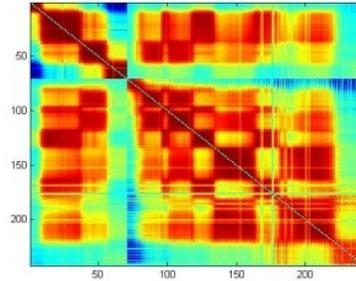


FIGURE 3: Image view of SCM for “corr_mtxhel”

3. GLOBAL REFERENCE BAND

Inter band correlation square is pattern of spectral correlation between bands in hyperspectral data; finding this square(s) refers to the location of the group(s) of bands that are highly correlated, and usually these GOBs are far away from each other.

Edge detection is concept of image processing that helps to locate edges in the processed image; some algorithms are used in this area; such as, Sobel Method, Prewitt Method, Roberts Method, Laplacian of Gaussian Method, Zero-Cross Method, and Canny Method.

The interest here to find the algorithm that determines the location(s) of the IBCS in the SCM; the algorithm should, at least, be able to determine the location of the biggest IBCS; or, if many similar exists, to determine at least one of them.

Sobel, Prewitt, and Roberts methods find edges using the corresponding approximation to the derivative, and return edges at those points of maximum gradient. The Laplacian of Gaussian method finds edges by looking for zero crossings after filtering the matrix with a Laplacian of Gaussian filter. Zero-cross method finds edges by looking for zero crossings after filtering matrix with a selected filter.

The Canny method finds edges by looking for local maxima of the gradient. The gradient is calculated using the derivative of a Gaussian filter. The method uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This method is therefore less likely than the others to be fooled by noise and more likely to detect true weak edges[23][24].

Comparative Study of these techniques[25] have been carried out; also performance analysis of each of them [26], recommends that Canny Method has better performance while detection of edges and less sensitivity for noise.

Using "Canny method" to estimate the edges of the IBCS[27]; we got the results in FIGURE 4

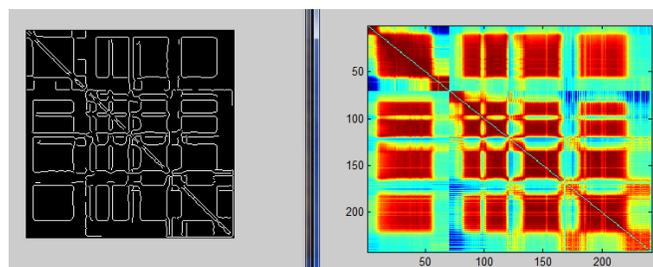


FIGURE 4: Canny Edge detection for SCM

4. BANDS REGROUPING TECHNIQUE

Finding IBCS in SCM will help in determining groups of bands that are highly correlated to each other's; detection threshold used in edge detection process of each IBCS is the correlation value for each GOB, which indicates the level of similarity between bands and each other.

We will take "corr_mtxer" as SCM, to investigate the principle of grouping based on IBCS. for example, there is a strong correlation between the groups of bands starting from approximately band number 10 till band number 55, and group of bands starting from approximately band number 180 till band number 220 (GOB (~10-55) and GOB (~180-220)).

Detection Thr.	GOB	GOB
0.7	9-53	9-53
0.65	81-120	9-53
0.6	181-221	101-120
0.45	181-221	81-96
0.4	125-165	9-53
0.3	181-221	9-53

TABLE 2: Group of bands and correlation threshold value of detection.

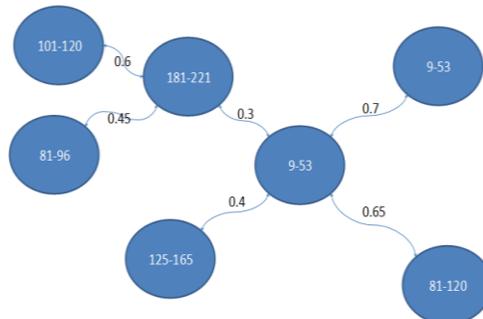


FIGURE 5: GOBs correlation with each other.

FIGURE 5 and

TABLE 2 illustrate the correlation between groups of bands; as we can see, the GOB (9-53) is much correlated to almost all bands in the data cube; we call it the central group of correlation.

From this central band we calculate what we call Global Reference Band (GRB)[28]; this band represents the average context of the whole correlated bands in the hyperspectral data cube; we use this band to exploit spectral redundancy[29].

We use simple exponential smoothing which assumes that the data fluctuates around a reasonably stable mean as the GOB is correlated to itself; forecasting the bands is based on equation (2):

$$\bar{X}_{t+1} = \alpha X_t + (1 - \alpha)\bar{X}_t \tag{2}$$

Where \bar{X}_{t+1} is the band under prediction; real data of the previous (reference) band, \bar{X}_t predicted values of the previous band, α is smoothing factor, α is empirically selected [30].

We apply simple exponential smoothing for GOB (9-53) to generate Global reference band (GRB); GRB is then used as a reference band represents the mean values of correlated bands.

5. J-PREDICTOR

JPEG –LS[31] introduces a simple and efficient prediction scheme for spatial redundancy called median predictor or Median Edge Detection; JPEG-LS was developed with the aim of providing a low-complexity lossless and near-lossless image compression standard that could offer better compression efficiency than lossless JPEG. The JPEG-LS predictor is shown in Equation (3).

$$\bar{Y} = \begin{cases} \text{Min}(N, W) & \text{if } NW \geq \text{Max}(N, W) \\ \text{Max}(N, W) & \text{if } NW \leq \text{Min}(N, W) \\ N + W - NW & \text{Otherwise} \end{cases} \quad (3)$$

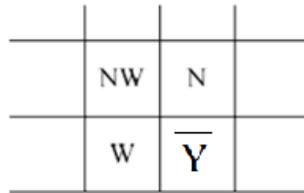


FIGURE 6: Neighboring pixels location relative to predicted pixel- Median predictor.

JPEG-LS predictor uses the Neighbor pixels for prediction FIGURE 6; in our modified model we add the corresponding pixel in the GRB for prediction, as it gives a considerable amount of information about the pixel in the same location in all correlated group of bands as in FIGURE 7; JPEG-LS predictor compares the upper left pixel with the upper and left pixels to select which one is more adequate for predicting the current pixel; this exploits the spatial redundancy in the image plane; as shown earlier.

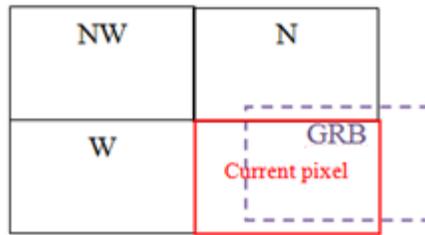


FIGURE 7: J-Predictor

The proposed predictor makes use of the collocated pixel in the previous band and predicts the current pixel based on similar model, each spectrally and spatially predicted values is assessed relative to surrounding pixel values and if spectrally predicted value is accepted it will be used in prediction process with its half weight .

$$\bar{Y} = w_a \bar{Y}_a + w_c \bar{Y}_c \quad (4)$$

Where, \bar{Y}_a spatially predicted value; \bar{Y}_c spectrally predicted value, w_a weight of spatially predicted value; w_c weight of spectrally predicted value; \bar{Y} predicted value of current pixel.

$$\bar{Y}_a = \begin{cases} \text{Min}(N, W) & \text{if } NW \geq \text{Max}(N, W) \\ \text{Max}(N, W) & \text{if } NW \leq \text{Min}(N, W) \\ N + W - NW & \text{Otherwise} \end{cases}$$

\bar{Y}_c = value of the co-located pixel in the GRB (Y_{GRB}).

$$w_a = 1 - w_c;$$

$$w_c = \begin{cases} 0 & \text{if } Y_{GRB} \geq \text{Max}(N, W, NW) \\ 0 & \text{if } Y_{GRB} \leq \text{Min}(N, W, NW) \\ 0.5 & \text{Otherwise} \end{cases}$$

The model combines both spatially and spectrally predicted values with empirically selected weighting factor; it rejects the spectrally predicted value if it is far away from the surrounding pixels; while it add the spectrally predicted value multiplied by weight, in case of the value within the range.

The prediction error is then encoded using JPEG2000 [33] in lossless mode.

6. RESULTS

The results of using the weighted prediction model combined with JPEG2000 in lossless compression mode are compared for other techniques for the same hyperspectral data samples shown in TABLE 3.

	Erta_Ale	LakeMonona	MtStHelen	Average
Original data	12	12	12	12
JPEG2000+ KLT[32]	6.02	5.75	6.06	5.94
JPEG2000+ KLT Static[32]	5.96	5.72	6.01	5.90
JPEG2000+ KLT Dynamic[32]	6.16	5.87	6.28	6.10
JPEG2000(IWT(CDF))	6.40	6.14	6.56	6.37
PRIM+FlossTree+LSCM[34]	5.007	4.870	5.019	4.905
PRIM+FlossTree+LSCM with band reorder [34]	4.992	4.859	4.995	4.831
JPEG2000 with J-predictor	4.623	4.489	4.601	4.398

TABLE 3: results of using J-predictor with JPEG2000 compared with other techniques

JPEG2000 is combined with Karhunen–Loève transform (KLT); where reversible integer wavelet transform is used for lossless transform.

The PRIM+FlossTree+LSCM algorithm is combination of reordering technique followed by predictive lossless compression algorithm; Prim’s algorithm is used for bands reordering; where fast lossless Free/Libre and Open Source Software (FLOSS) algorithm is used for predictive lossless coding, the reordering process is based on local/causal Spectral Correlation Mapper[34]. Our proposed predictor has performance better than the listed techniques in table an improvement of number of bit per pixel per band is reduced using the J-predictor.

7. CONCLUSIONS

As hyperspectral compression is requiring an efficient techniques for exploitation of both spatial and spectral redundancies; we have proposed a new modified predictor for lossless compression of hyperspectral data cubes; the techniques mainly depends on calculation of spectral correlation matrix to discover the correlated group of bands; and as a result we calculate the global reference band that represent a mean value of spectral information of the hyperspectral data bands; this

bands is then used in modified predictor to predict the whole bands and the residual prediction error is then encoded by JPEG2000 in lossless mode.

As a comparison this technique and other techniques showed that J-predictor can minimize the prediction error and consequently the number of bits required to encode the data.

The main disadvantage of this techniques is that it depends on the calculation of spectral correlation matrix; which is a time consuming process, this issue can be addressed in future research.

8. REFERENCES

- [1] X. Pan, R. Liu, and X. Lv, "Low-Complexity Compression Method for Hyperspectral Images Based on Distributed Source Coding," *IEEE Geoscience and Remote Sensing Letters*, 9(2), 224 – 227(2012).
- [2] M. Slyz and L. Zhang, "A block-based inter-band lossless hyperspectral image compressor," *Data Compression Conference, 2005. Proceedings*, 427-436(2005).
- [3] J. Zhang and G. Liu, "Hyperspectral images lossless compression by a novel three-dimensional wavelet coding," *15th international conference on Multimedia, New York, NY, USA*, 2007.
- [4] F. Sepehrband, P. Ghamisi, A. zadeh, M. Sahebi, and J. Choupan, "Efficient Adaptive Lossless Compression of Hyperspectral Data using Enhanced DPCM," *International Journal of Computer Applications*, 35(4), 6-11 (2011).
- [5] G. Vilchez, F. Muñoz-Marí, J. Zorteza, M. Blanes, I. González-Ruiz, V. Camps-Valls, G. Plaza, and A. Serra-Sagrìsta, "On the Impact of Lossy Compression on Hyperspectral Image Classification and Unmixing," *IEEE Geoscience and Remote Sensing Letters*, 8(2), 253 – 257(2011).
- [6] S. Bergeron, M. Cunningham, I. Gagnon, and L. Hollinger, "Near lossless data compression onboard a hyperspectral satellite," *IEEE Trans. on Aerospace and Electronic Systems*, 42(3), 851 – 866 (2006).
- [7] B. Aiazzi, L. Alparone, S. Baronti, and C. Lastrì, "Crisp and fuzzy adaptive spectral predictions for lossless and near-lossless compression of hyperspectral imagery," *IEEE Geoscience Remote Sens. Lett.* (4)4, 532-536 (2007).
- [8] W. Wang, Z. Zhao, and H. Zhu, "Hyperspectral Image Compression Method Based on Spectral Statistical Correlation," *2nd International Congress on Image and Signal Processing, Tianjin, china*, 1 – 5(2009).
- [9] D. Manolakis, R. Lockwood, and T. Cooley, "On the Spectral Correlation Structure of Hyperspectral Imaging Data," *IEEE International Geoscience and Remote Sensing Symposium, Boston, MA USA*, 581-584(2008).
- [10] G. Liu and F. Zhao, "Efficient compression algorithm for hyperspectral images based on correlation coefficients adaptive 3D zero tree coding," *Image Processing, IET*, 2(2), 72–82(2008).
- [11] M. J. Weinberger, G. Seroussi, and G. Sapiro, "LOCO-I: A low complexity, context-based, lossless image compression algorithm," in *Proc. Data Compression Conference, Snowbird, UT*, 140–149(1996).

- [12] Z. Zhou, Y. Tan, and J. Liu, "Satellite Hyperspectral Imagery Compression Algorithm Based on Adaptive Band Regrouping," International Conference on Wireless Communications, Networking and Mobile Computing, Wuhan, China, 1-4 (2006).
- [13] J. Gaucel, C. Thiebaut, R. Hugues, and R. Camarero, "On-board compression of hyperspectral satellite data using band-reordering," Satellite Data Compression, Communications, and Processing VII, SPIE Proc. 8157,1152-1160 (2011).
- [14] D. Cesmecci, Gullu, M.K. Erturk, "Segmentation of hyperspectral images using phase correlation based on adaptive thresholding," IEEE 16th Signal Processing, Communication and Applications Conference, Aydin, Turkey, 1-4(2008).
- [15] M. Mori and K. Kashino, "Fast Template Matching Based on Normalized Cross Correlation Using Adaptive Block Partitioning and Initial Threshold Estimation," IEEE International Symposium on Multimedia, Taichung, Taiwan, 196 – 203(2010).
- [16] J. P. Lewis, "Fast Template Matching", Canadian Image Processing and Pattern Recognition Society, Quebec, Canada, 15-19(1995).
- [17] Matlab www.matlab.com [March, 2012]
- [18] SPECTIR <http://www.spectir.com/download.html> [February, 2012]
- [19] AVIRIS, Hyperion <http://compression.jpl.nasa.gov/hyperspectral/> [October, 2011]
- [20] G. Liu¹, F. Zhao¹, and G. Qu, "An efficient compression algorithm for hyperspectral images based on a modified coding framework of H.264/AVC," IEEE International Conference on Image Processing, San Antonio, TX, USA, 341-344(2007).
- [21] H. Liu, Y. Li. Song, and X. C. Wu, "Distributed Compressive Hyperspectral Image Sensing," Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing 2010, Darmstadt, Germany. 15-17 (2003).
- [22] Q. Du, W. Zhu, H. Yang, and J.E. Fowler, "Segmented Principal Component Analysis for Parallel Compression of Hyperspectral Imagery," IEEE Geoscience and Remote Sensing Letters, 6(4), 713 – 717(2009).
- [23] J.CANNY, "A Computational Approach to Edge Detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(6), 679- 698(1986).
- [24] R. Maini and H. Aggarwal "Study and Comparison of Various Image Edge Detection Techniques International," international Journal of Image Processing,3(1),1-11(2009).
- [25] R. Maini, and H. Aggarwal, "Study and Comparison of Various Image Edge Detection Techniques," International Journal of Image Processing, 3(1), (1-11)2009.
- [26] M. Juneja, and P. Sandhu, "Performance Evaluation of Edge Detection Techniques for Images in Spatial Domain," International Journal of Computer Theory and Engineering, 1(5), 1793-8201(2009).
- [27] Ayman Mahmoud, S. Elramly, and M. Elsharkawy "Analysis of Inter-band Spectral Cross-Correlation Structure of Hyperspectral Data" WSEAS International Conference on Information Technology and Computer Networks (ITCN '12), Vienna, Austria, November 2012.
- [28] Ayman Mahmoud, S. Elramly, and M. Elsharkawy "Bands Regrouping of Hyperspectral Data Based on Spectral Correlation Matrix Analysis," International Journals of Engineering &

Sciences, 12(4), August 2012.

- [29] Ayman Mahmoud, S. Elramly, and M. Elsharkawy "Hyperspectral band referencing based on correlation structure," IEEE International Conference on Control System, Computing and Engineering Malaysia 2012.
- [30] Prajakta S. Kalekar. "Time series forecasting using Holt-Winters Exponential Smoothing," Kanwal Rekhi School of Information Technology, 1-13, (2004).
- [31] Marcelo J. Weinberger and Gadiel Seroussi. "The LOCO-I Lossless Image Compression Algorithm: Principles and Standardization into JPEG-LS" IEEE Transactions on Image Processing, 9(8), 1309 – 1324(2000).
- [32] Ian Blanes, and Joan Serra-Sagristà, "Cost and Scalability Improvements to the Karhunen–Loève Transform for Remote-Sensing Image Coding" IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, 48(7), 50-62 2010.
- [33] S. Singh, R.K. Sharma and M.K. Sharma "Use of Wavelet Transform Extension for Graphics Image Compression using JPEG2000 Standard," International Journal of Image Processing, 3(1),55-60(2009).
- [34] J, Gaucel, C. Thiebaut, R. Hugues, and R. Camarero, "On-board compression of hyperspectral satellite data using band–reordering," Satellite Data Compression, Communications, and Processing VII, SPIE Proc. 81-97 (2011).

Joint, Image-Adaptive Compression and Watermarking by GA-Based Wavelet Localization: Optimal Trade-Off between Transmission Time and Security

M. M. Ramya

*Department of Computer Applications
Hindustan University
Chennai, India*

mmramya@hindustanuniv.ac.in

R. Murugesan

*Director (Innovative Programmes)
ChettinadUniversity
Chennai, India*

rammurugesan@yahoo.com

Abstract

Teleradiology using internet can offer patients in remote locations the benefit of diagnosis and advice by a super specialist present in a metropolis. However, exchange of vital information such as the clinical images and textual facts in the public network poses challenges of transmission of large volume of data as well as prevention of the distortion of the images. In this paper, a novel application system to jointly compress and watermark the medical images in a near-lossless, image-adaptive fashion is proposed to address these challenges. The system design uses genetic algorithm for adaptive wavelet coding to generate compressed data and integration of dual watermarks to realize the security and authentication of the compressed data. The GA-based image adaptive compression provides feasible way to obtain optimal compression ratio without compromising the image fidelity upon subsequent watermarking. A multi-gene approach, with one gene coding for the embedding strength of the robust watermark and the other for the number of bits for embedding the semi-fragile watermark is used for optimal image-adaptive watermarking. A multi-parameter fitness function is designed to address the conflicting requirements of image compression, authenticity and integrity associated with teleradiology. Experimental results show the ability of the system to detect tampering and to limit the peak error between the original and the watermarked images. Moreover, as the watermarking is performed on the compressed image, the overhead for watermarking gets reduced.

Keywords: Adaptive Compression, Dual Watermarking, Multi-gene Genetic Algorithm, Multi-objective Fitness Function, Teleradiology.

1. INTRODUCTION

Within the expanding paradigm of medical imaging and wireless communications, there arises an ever increasing demand of fast transmission of diagnostic medical imagery over error-prone wireless communication channels such as those encountered in cellular phone technology. The mobile transmission of such images is prohibitive without the use of image compression to reduce the image size [1]. Therefore, medical images must be compressed to minimize transmission time, and robustly coded to ensure security [2]. This is especially favorable if the end application is teleradiology, because rural areas do not have high bandwidth communication network. The primary challenge of medical image compression is to reduce the data volume and to achieve a low bit rate in the digital representation of radiological images without perceived loss of image quality [3].

To control the amount of information lost during the compression process, a class of algorithms capable of strictly controlling the compression loss has been devised and grouped under term

Near-Lossless Compression, whose main requirement is that of ensuring that the maximum error between the original and the compressed image does not exceed a fixed threshold. In the same line, the concept of near-lossless watermarking has been introduced recently to satisfy the strict requirements for medical image watermarking [4]. Moreover, these techniques do not adaptively arrive at an optimal compression ratio. A single compression technique might not be suitable for all medical images because of their differing noise characteristics. A high compression ratio is preferable for reducing transmission time. But, it is difficult to attain the same compression ratio for an image with low PSNR, because it might degrade the original image, making it difficult for clinical reading. Hence it is necessary to find an optimal trade-off between the image quality and compression ratio. In addition, it is also essential to ensure that the compressed image has sufficient bandwidth to accommodate the watermark payload. This work attempts to investigate, for the first time, the application of GA in achieving an optimal compression ratio for dual watermarking in wavelet domain without degrading the image.

2. GA-BASED IMAGE ADAPTIVE COMPRESSION

Discrete wavelet transform (DWT) has gained extensive interest as a method of information coding [5, 6], due to its ability to decorrelate data effectively. Due to their inherent multiresolution nature, the coefficients of DWT are localized in both spatial and frequency domains, which is highly desired because HVS functions as a bandpass filter with the localization property [7-10]. In contrast to the conventional wavelet transform, the lifting scheme (LS) allows faster implementation of the wavelet transform [11]. In addition, it is better matched to the HVS characteristics. In the context of image authentication through joint coding and watermarking is highly desirable, since otherwise the fragile nature of the watermark will identify image compression as an unwanted manipulation, and will eventually fail to distinguish between compression (allowed) and tampering (not allowed). On the other side, tying the watermarking system to a particular coding format limits the flexibility of the authentication scheme, since the watermark is likely not to survive lossless format changes. It is one of the goals of the algorithm developed in this paper to embed the watermark in the compressed image, while still allowing the recovery of the watermark.

Optimal compression can be measured as sufficient fidelity (~37 dB) of medical images given that an appropriate amount of compression is used [12]. To calculate the error value in the compressed image various parameters like mean square error (MSE), root mean square error (RMSE) may be used. These parameters help to measure the trade-off between image quality and compression ratio (CR), defined as $CR = \frac{I}{IC}$, is used as a relative measure. Here I and IC correspond to the input and compressed images, respectively. CR performs as a good measure for all images, independent of the way they are encoded, because it is only a ratio of the respective image sizes. The activity diagram of the novel joint GA algorithm for optimal compression and watermarking is presented in Fig. 1.

In a general wavelet compression algorithm, an image is decomposed using wavelet transform to obtain LL, HL, LH and HH sub-bands. In the wavelet quantization context, the decorrelation property suggests that processing the coefficients independent of each other and the sparseness (or "heavy-tailedness") property pave a way to use threshold estimators aimed at removing coefficients that are "small" relative to the noise. The classical choices for performing thresholding of lifting wavelet transform (LWT) coefficients are the hard and soft thresholding functions [6]. However, most of the wavelet thresholding methods suffer from the drawback that the chosen threshold may not match the specific distribution of signal and noise components in different scales.

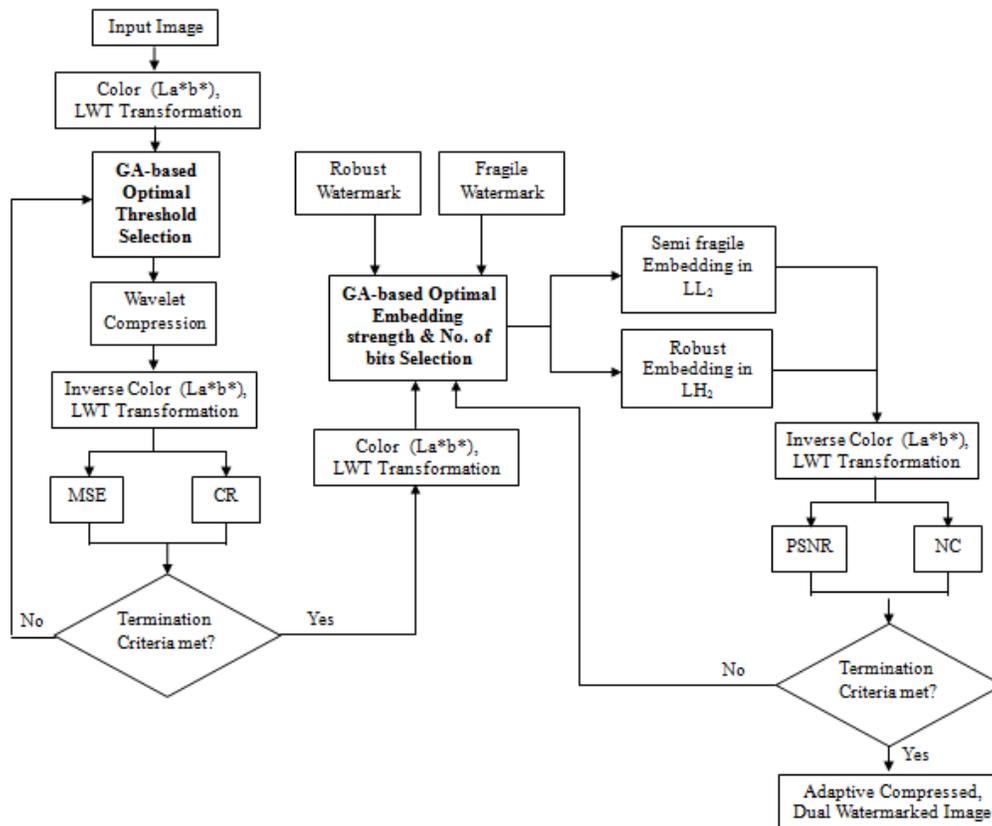


FIGURE 1:Activity Diagram

This is especially so, for medical images where both the PSNR and compression ratio are equally important. To address this problem, adaptive thresholding based on statistical priors of the noise models may be applied as it is done in denoising [13]. But, success of these methods depends upon the exact knowledge of the spectral SNR relationship, which varies with the type of imagery, leading to a case-by-case basis investigation. This motivates the development of robust and versatile compression methods that are capable of universal application, rather than being optimal under very specific condition. An ideal method should perform quantization by intelligently arriving at the level of compression from the images themselves without a priori input information. One such innovative approach is to use GA to optimize the threshold of each sub-band across different scales. Hence, an optimal trade-off between the MSE and compression ratio of medical images forms the basis for the fitness function used in the algorithm developed here..

2.1 Chromosome Encoding

The basic structure of GA revolves around the concept of encoding a solution and evolving successive solutions according to their fitness. In the present work, the genes of a chromosome represent the threshold for a given image. Wavelet threshold values were represented by real-coded chromosome to offer a number of advantages in numerical function optimization over the binary encoding. Efficiency of the GA is increased as there is no need to convert chromosomes to phenotypes before each function evaluation; less memory is required as efficient floating point internal computer representations can be used directly. There is no loss in precision by discretization to binary or other values; and there is greater freedom to use different genetic operators. The chromosome encoding adopted in this work is presented in Fig. 2.

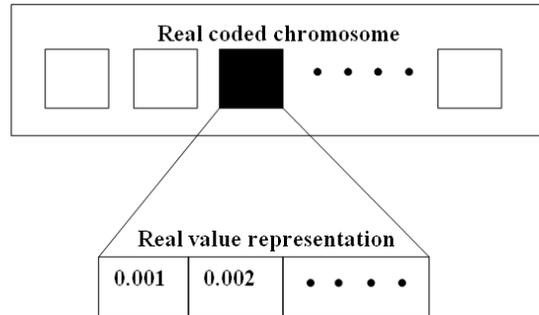


FIGURE 2:Real-coded chromosome for GA-based image adaptive compression.

2.2 Fitness Function

The desired compression of medical images not only aims at achieving a higher compression ratio but also to retain the image quality. Hence the correct choice of threshold is crucial to the performance of a LWT-based compression algorithm. Hence, a multi-objective fitness function may be designed. In the present study, the fitness function is designed to maximize CR while keeping the error metric, MSE low. Such a function is given in Eqn. 1.

$$f = CR - wt * MSE \quad (1)$$

Similarly a multi-objective fitness function for adaptive dual watermarking is designed to maximize PSNR while keeping the NC low. The function is given in Eqn. 2.

$$f = \max(PSNR(I, I_w) - (wt * avg(NC(W_r, W_r'), NC(W_f, W_f')))) \quad (2)$$

Here I_w refers to the watermarked image. Two watermarks W_r (robust) and W_f (semifragile) are embedded in I . The recovered watermarks are represented as W_r' and W_f' .

3. RESULTS AND DISCUSSION

For validation of the algorithm, the six test images (Lena and five medical images as shown in Column 1 of Fig. 5) were used. To provide data integrity, an image containing patient name, physician name, hospital logo, date and remark of diagnosis was used as a robust watermark. The size of the semi-fragile watermark was kept small. Hence, a binary image containing the physician’s signature was used as the semi-fragile watermark. The watermarks used in this work are presented in Table 1.

	Semi-fragile	Robust
Watermarks		Cardiologist: Dr. Rathinavel Patient Name: Sathish kumar Date: 29/07/07 Remarks: Rheumatic Heart Disease 
Size	64 x 65	56 x 28

TABLE 1:Watermarks Used

The efficiency of GA depends upon tuning of the GA-parameters for specific application. Hence several trials were performed to select optimal GA-parameters. Roulette wheel method was used for selection of chromosomes to the next generation. From extensive experiments, we found the proposed GA-based approach to work well with a single point crossover with a random crossover probability of $P_c = 0.4$ and a single point mutation with a mutation probability of $P_m = 0.05$.

3.1 Influence of Population Size on Convergence

To arrive at a suitable population size, experiments were performed using different population sizes with the above mentioned fixed crossover and mutation rates. Experiments were carried out with population sizes ranging from 10 to 150. There is a gradual increase in the fitness function

with increase of population size up to 90. Further increase in population size does not show any significant improvement in the fitness function. The computational overhead of different population sizes was also investigated. There is a fast increase in execution time when the population size increases above 50. Based on these results, a population size of 50 was selected as an optimal trade-off

3.2 Halting Condition

The termination condition adopted here is based on both evaluating the progress made by the algorithm in a predefined number of generations (100) or arriving at the average distance between the individuals to be less than 0.01. Similar experiments were done to estimate the GA parameters for dual watermarking of medical images. The various GA parameters thus estimated are collected in Table 2.

Genetic Operator	GA parameters for adaptive compression	GA parameters for adaptive watermarking
Initial population	90 chromosomes	60 chromosomes
Encoding	Real encoding scheme	9 bit encoding scheme
Fitness function	CR and MSE values as in Eqn. 1	PSNR and NC values as in Eqn. 2
Selection	Roulette wheel selection	Roulette wheel selection
Crossover	Single-point crossover with probability 0.4	Two-point crossover with probability 0.5
Mutation	Single-point mutation with probability of 0.05	Single-point mutation with probability of 0.062
Convergence	Convergence to single result or 100 generations	Convergence to single result or 100 generations

TABLE 2:GA Parameters

The compressed images obtained using the GA-based image-adaptive compression is presented for visual evaluation in column 2 of Fig. 3. The quantitative parameters computed are presented in Table 3. The threshold values for different images clearly bring out the scope of the adaptive nature of the proposed algorithm. The threshold value obtained by the proposed algorithm is larger for the standard Lena image. Due to its high image quality, the Lena image is able to withstand a high compression ratio of 80.2:1. Since, medical images are of relatively low PSNR, a higher threshold value may lead to more information loss. The GA-based image adaptive compression algorithm automatically selects a lower threshold when the PSNR of input image is low, thus finding an optimal solution for the two conflicting objectives of fidelity and compression.

3.3 Performance Evaluation

The compressed images should be evaluated for their fitness to watermarking, because our overall aim is speed as well as secure transmission of medical images. Hence, the compressed images were tested whether they can withstand the dual watermarking process without suffering loss of clinical diagnostic reading. Hence, the multi-gene, multi-objective GA-based dual watermarking algorithm was used to watermark the compressed images, given in column 2 of Fig. 3. The visual results (the watermarked images) presented in column 3 of Fig. 3 reveal that, the image fidelity is retained even after compression and dual watermarking.

Image	CR	MSE	RMSE	PSNR	Threshold
Lena	80.2 : 1	4.26	2.06	41.84	20.17
Echo-A	63.7 : 1	3.86	1.96	42.26	9.06
Echo-B	63.2 : 1	3.58	1.89	42.59	8.96
Echo-C	40.2 : 1	4.82	2.20	41.30	5.23
Fundus	62.1 : 1	3.27	1.81	42.99	7.83
fMRI	61.8 : 1	4.21	2.05	41.89	7.21

TABLE 3:Optimal threshold adaptively arrived by the GA-based algorithm.

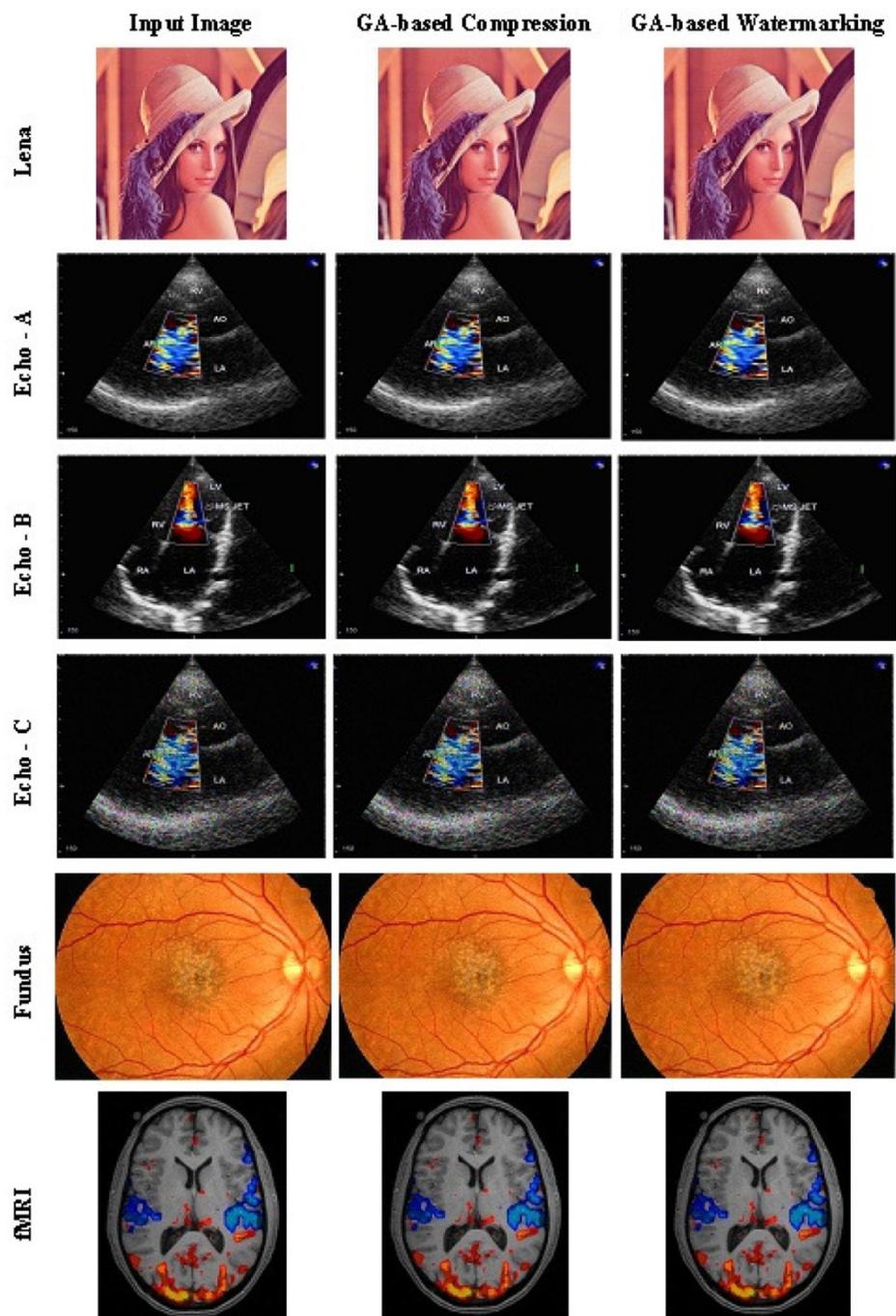


FIGURE 3: Visual comparison of input, compressed and watermarked images

The optimal embedding strength and the number of bits selected by the multi-gene, multi-objective GA are collected in Table 4. The CR values in Table 4 correspond to the compression ratio achieved by GA-based compression approach.

The PSNR values obtained for all images, given in column 3 of Fig. 3 are greater than 37 dB, satisfying the fidelity requirement of medical image watermarking. At the same time, the

recovered watermarks from these images result in correlation above 0.85, satisfying the conflicting requirements of fidelity, robustness and image size.

Image	α	No. of bits	PSNR	NC	CR
Lena	0.46	3	39.21	0.92	80.2 : 1
Echo-A	0.22	2	38.62	0.89	63.7 : 1
Echo-B	0.25	2	39.45	0.9	63.2 : 1
Echo-C	0.11	1	37.02	0.85	40.2 : 1
Fundus	0.17	2	39.45	0.89	62.1 : 1
fMRI	0.27	2	38.86	0.89	61.8 : 1

TABLE 4:Watermarking and image quality parameters.

The common approach followed in reducing the transmission time in teleradiology is to compress the watermarked image. But, in this paper, for the first time, a novel algorithm which initially compresses the image by multi-objective GA, and then embeds the watermarks in the compressed image by a multi-gene, multi-objective GA is implemented. It would be of interest to compare these approaches in terms of fidelity and robustness. Such a comparative evaluation is presented in Table 5. Row (a) corresponds to the recovered watermarks (robust and fragile) from the proposed GA-based joint, image-adaptive compression and watermarking algorithm developed in this chapter. Row (b) shows the recovered dual watermarks from the common approach of compressing the watermarked images. Row (a) reveals high NC value for the recovered watermarks and good image fidelity (PSNR > 37dB). But, for the same compression ratio (63.7:1), the recovery of the watermark fails, when the other widely used approach of compressing the watermarked image (Row b).

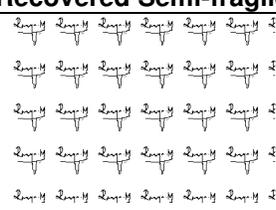
	CR	Recovered Robust	Recovered Semi-fragile	PSNR	NC
(a)	63.7:1	Cardiologist: Dr. Rathinavel Patient Name: Sathish kumar Date: 29/07/07 Remarks: Rheumatic Heart Disease 		38.62	0.96
(b)	63.7:1			27.64	0.61

TABLE 5:Recovered watermarks (robust and semi-fragile)

These results prove the superior performance and wide ranging significance of the algorithm presented here. There is also an added advantage. Since the watermarking is performed on the compressed image, the overhead for the watermarking process becomes less due to the compactness of the compressed image.

3.3.1 Robustness to Attacks

Five different types of attacks were used to evaluate the robustness of the GA-based algorithm for adaptive watermarking of the adaptively compressed image. The watermarks recovered after various attacks, shown in Table 6, reveal the robust watermark to withstand the attacks like copy, rotation and noise addition. However, the robust watermark does not withstand compression,

since further compression of a compact image leads to more information loss, and thus making it difficult for the detector to recover the watermark.

	Copy	Rotation	Noise Addition	Compression (20%)
Robust				
Fragile				

TABLE 6:Recovered watermarks after attacks

3.3.2 Tamper Localization

The semi-fragile watermark embedded in the compressed image can be used to spy the malicious tampering of the watermarked image. The results of tamper localization are shown in Table 7. The three tampering operations (Columns a-c) correspond to malicious drawing, copy paste and cropping attacks. The non-malicious compression (10%) is shown in (d). Rows 1, 2, 3 of columns a-d of Table 7 show the tampered images, recovered watermarks and the tamper localization results, indicating the efficiency of the algorithm for joint compression and dual watermarking. It is interesting to note that the image compression of 10% is not seen as tampering. Thus, the algorithm presented here satisfies many demands of medical image watermarking.

	(a)	(b)	(c)	(d)
Tampered images				
Recovered watermarks				
Tamper localization				

TABLE 7:Performance evaluation of semi-fragile watermarking

3.4 Computational Time

The computational time required for the joint, image-adaptive compression and watermarking algorithm was estimated on an Intel dual core 2 GHz processor with 1 GB RAM. The time taken for compression and watermarking is presented in Table 8. It is interesting to note that, the time required for GA-based dual watermarking algorithm to embed in the uncompressed image is 12.36 min. But, the time taken for the same dual watermarking on the same image after adaptive compression is only 8.58 min. Since the joint image-adaptive compression and watermarking algorithm embeds the dual watermarks in the compressed image, the computational time required for embedding is reduced by 4 min. Though, there is an additional computational overhead of 4.82 min for GA-based image compression, the reduction in size can help improving the transmission time.

Modules	Time taken (min)
GA-based dual watermarking	12.36
GA-based dual watermarking after adaptive compression	8.58

TABLE 8:Computational time

4. CONCLUSION

This paper presents a novel algorithm to enhance the potential of teleradiology by integrating GA-based image adaptive compression with GA-based color image dual watermarking. Two image processing applications, compression and watermarking are coupled to enhance the potential of watermarking in teleradiology, by exploiting the localization property of wavelet transform, using genetic algorithm. The algorithm permits to jointly compress and watermark medical images and at the same time to maintain fidelity of the image. The algorithm is designed in such a way that the compression ratio and watermarking error can be controlled (near lossless compression and optimal watermarking). The results show that the GA-based system can automatically and image-adaptively arrive at optimal compression and watermarking parameters. The application system can readily detect any tampering in the images. While the developed system is tested to work on telemedicine imagery, its use may be extended to other sensitive areas such as remote sensing, military application, video surveillance etc.

Acknowledgements

One of the authors, M. M. Ramya thanks Hindustan University for encouragement and support.

5. REFERENCES

1. J. Strom, P. Cosman. "Medical Image Compression with Lossless Regions of Interest". Signal Processing.vol. 3, pp.155-171, 1997.
2. M. J. Zukoski, T. Boulton, T. Lyriboz. "A Novel Approach to Medical Image Compression". *Bioinformatics research and applications*, vol. 2, pp. 89-103, 2006.
3. T. Lyriboz *et al.* "A Comparison of Wavelet and JPEG Lossy Compression Methods Applied to Medical Images". *Journal of Digital Imaging*, vol. 12, pp. 14-17, 1999.
4. K. Chen, T. V. Ramabadran. "Near-Lossless Compression of Medical Image through Entropy-Coded DPCM". *IEEE Transactions on Medical Imaging*.vol. 13, pp. 538-548, 1994.
5. J. Froment, S. Mallat. "Second Generation Compact Image Coding with Wavelets". *Wavelets: A Tutorial in Theory and Applications*, New York: Academic Press, 1992.

6. S. Lewis, G. Knowles. "Image Compression Using the 2D Wavelet Transform". *IEEE Transactions on Image Processing*, vol. 1, pp. 244-250, 1992.
7. S. Bhattacharjee, M. Kutter. "Compression Tolerant Image Authentication". In Proc. of ICIP, 1998, pp. 435-439.
8. M. M. Ramya, R. Bhaskran, R. Murugesan. "GA-based Image Adaptive Watermark Embedding for Optimal Fidelity and Robustness in Medical Images". In Proc. of ICICOT, 2007, pp. 47-52.
9. Y. Ester, T. Kai-Shiang. "HDWT-based Grayscale Watermark for Copyright Protection". *Expert Systems with Applications*, vol. 35, pp. 301-306, 2008.
10. W. Linet *et al.* "A blind watermarking method using maximum wavelet coefficient quantization". *Expert systems with Applications*, vol. 36, pp. 11509-11516, 2009.
11. W. Sweldens. "The Lifting Scheme: A Custom-Design Construction of Biorthogonal Wavelets". *Applied Computational Harmonic Analysis*, vol. 3, pp. 186-200, 1996.
12. F. A. P. Petitcolas. "Watermarking Schemes Evaluation". *IEEE Signal processing*, vol. 17, pp. 58-64, 2000.
13. M. Vetterli, C. Herley. "Wavelets and Filter Banks: Theory and Design". *IEEE Transactions on Signal Processing*, vol. 40, pp. 2207-2232, 1992.

INSTRUCTIONS TO CONTRIBUTORS

The *International Journal of Image Processing (IJIP)* aims to be an effective forum for interchange of high quality theoretical and applied research in the Image Processing domain from basic research to application development. It emphasizes on efficient and effective image technologies, and provides a central forum for a deeper understanding in the discipline by encouraging the quantitative comparison and performance evaluation of the emerging components of image processing.

We welcome scientists, researchers, engineers and vendors from different disciplines to exchange ideas, identify problems, investigate relevant issues, share common interests, explore new approaches, and initiate possible collaborative research and system development.

To build its International reputation, we are disseminating the publication information through Google Books, Google Scholar, Directory of Open Access Journals (DOAJ), Open J Gate, ScientificCommons, Docstoc and many more. Our International Editors are working on establishing ISI listing and a good impact factor for IJIP.

The initial efforts helped to shape the editorial policy and to sharpen the focus of the journal. Starting with volume 7, 2013, IJIP will be appearing in more focused issues. Besides normal publications, IJIP intends to organize special issues on more focused topics. Each special issue will have a designated editor (editors) – either member of the editorial board or another recognized specialist in the respective field.

We are open to contributions, proposals for any topic as well as for editors and reviewers. We understand that it is through the effort of volunteers that CSC Journals continues to grow and flourish.

LIST OF TOPICS

The realm of International Journal of Image Processing (IJIP) extends, but not limited, to the following:

- Architecture of imaging and vision systems
- Character and handwritten text recognition
- Chemistry of photosensitive materials
- Coding and transmission
- Color imaging
- Data fusion from multiple sensor inputs
- Document image understanding
- Holography
- Image capturing, databases
- Image processing applications
- Image representation, sensing
- Implementation and architectures
- Materials for electro-photography
- New visual services over ATM/packet network
- Object modeling and knowledge acquisition
- Photographic emulsions
- Prepress and printing technologies
- Remote image sensing
- Autonomous vehicles
- Chemical and spectral sensitization
- Coating technologies
- Cognitive aspects of image understanding
- Communication of visual data
- Display and printing
- Generation and display
- Image analysis and interpretation
- Image generation, manipulation, permanence
- Image processing: coding analysis and recognition
- Imaging systems and image scanning
- Latent image
- Network architecture for real-time video transport
- Non-impact printing technologies
- Photoconductors
- Photopolymers
- Protocols for packet video
- Retrieval and multimedia

- Storage and transmission

- Video coding algorithms and technologies for ATM

CALL FOR PAPERS

Volume: 7 - Issue: 4

i. Paper Submission: May 31, 2013

ii. Author Notification: July 15, 2013

iii. Issue Publication: August 2013

CONTACT INFORMATION

Computer Science Journals Sdn Bhd

B-5-8 Plaza Mont Kiara, Mont Kiara

50480, Kuala Lumpur, MALAYSIA

Phone: 006 03 6207 1607

006 03 2782 6991

Fax: 006 03 6207 1697

Email: cscpress@cscjournals.org

CSC PUBLISHERS © 2012
COMPUTER SCIENCE JOURNALS SDN BHD
M-3-19, PLAZA DAMAS
SRI HARTAMAS
50480, KUALA LUMPUR
MALAYSIA

PHONE: 006 03 6207 1607
006 03 2782 6991

FAX: 006 03 6207 1697
EMAIL: cscpress@cscjournals.org