

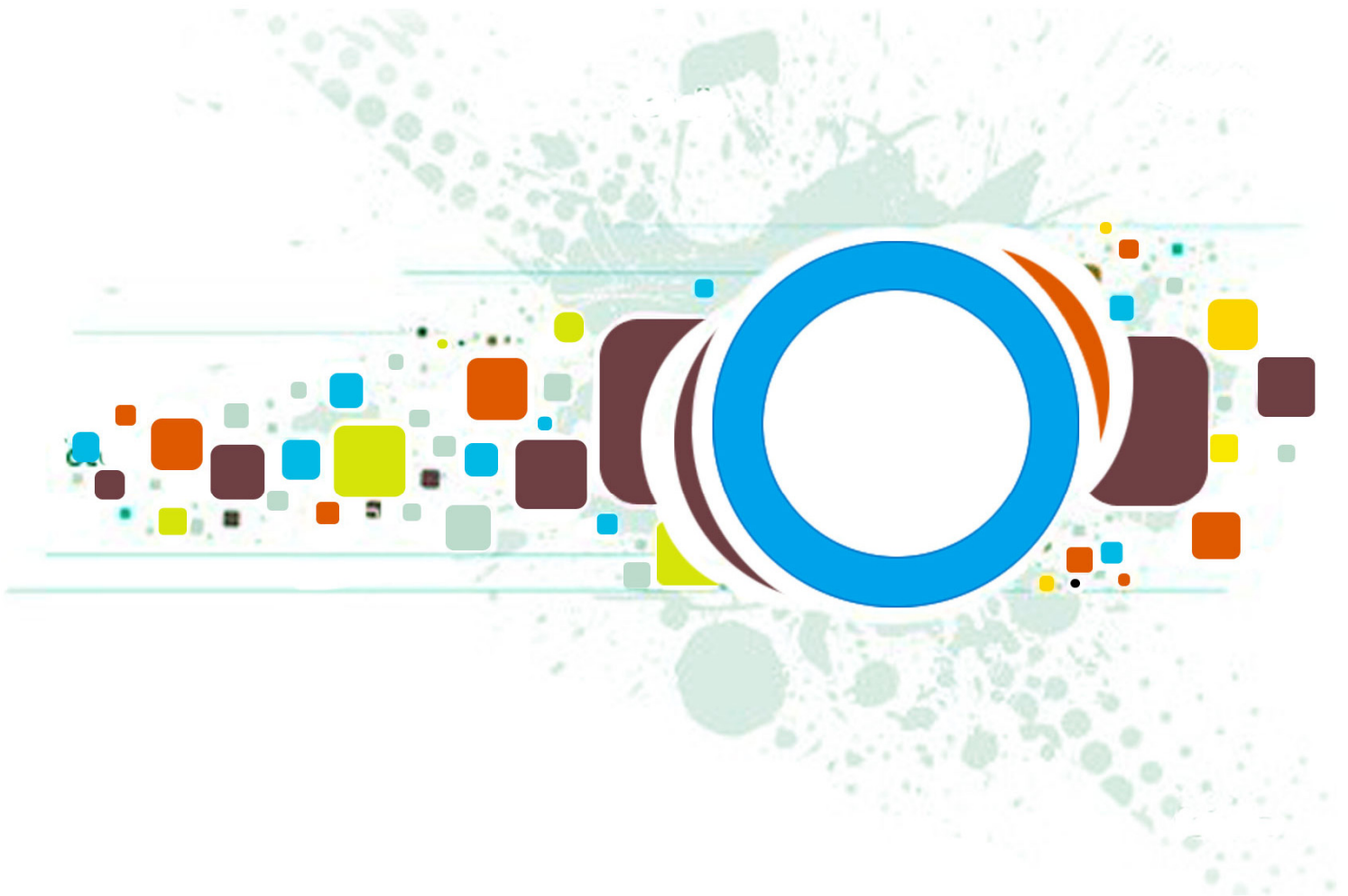
Volume 7 • Issue 1 • February 2013

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 7, ISSUE 1, 2013

**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 7, Issue 1, February 2013

Publishing Date: 28-02- 2013

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, 2013

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 first issue of volume seven 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 appears 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. Iftekharruddin
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

Associate Professor 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 7, Issue 1, February 2013

Pages

- 1 - 16 Pedestrian Counting in Video Sequences based on Optical Flow Clustering
Sizuka Fujisawa, Go Hasegawa, Yoshiaki Taniguchi, Hirotaka Nakano
- 17 - 30 Performance Evaluation of 2D Adaptive Bilateral Filter For Removal of Noise From Robust Images
B.Sridhar, Dr.K.V.V.S.Reddy
- 31 - 50 Speeded-up and Compact Visual Codebook for Object Recognition
Barathy Mayurathan, Amirthalingam Ramanan, Sinnathamby Mahesan, U.A.J. Piniidiyaarachchi
- 51 - 61 Image Processing Technique for Brain Abnormality Detection
Ashraf Anwar, Arsalan Iqbal
- 62 - 71 Digital Image Watermarking Using Different Levels of Intermediate Significant Bits with Zig-zag Embedding Approach
Ali Sharifara, Ghazali Bin Sulong, Mehran Ranjbar Seraydashti
- 72 - 80 MultiModal Identification System in Monozygotic Twins
Dinakar das CN, S.Peumal Sankar, Nisha George
- 81 - 89 Image Steganography Technique By Using Braille Method of Blind People (LSBraille)
Abdelmgeid Amin Ali, Al – Hussien Seddik Saad

- 90 - 100 Establishment of an Efficient Color Model from Existing Models for Better Gamma Encoding
In Image Processing
*T. M. Shahriar Sazzad, Sabrin Islam, Mohammad Mahbubur Rahman Khan Mamun, Md.
Zahid Hasan*
- 101 - 108 Selective Median Switching Filter for Noise Suppression in Microstructure Images of
Material
P.S. Hiremath, Anita Sadashivappa

Pedestrian Counting in Video Sequences based on Optical Flow Clustering

Shizuka Fujisawa

*Graduate School of Information Science and Technology
Osaka University
Suita, Osaka, 565-0871, Japan*

s-fujisw@ist.osaka-u.ac.jp

Go Hasegawa

*Cybermedia Center
Osaka University
Toyonaka, Osaka, 560-0043, Japan*

hasegawa@cmc.osaka-u.ac.jp

Yoshiaki Taniguchi

*Cybermedia Center
Osaka University
Toyonaka, Osaka, 560-0043, Japan*

y-tanigu@cmc.osaka-u.ac.jp

Hirotaka Nakano

*Cybermedia Center
Osaka University
Toyonaka, Osaka, 560-0043, Japan*

nakano@cmc.osaka-u.ac.jp

Abstract

The demand for automatic counting of pedestrians at event sites, buildings, or streets has been increased. Existing systems for counting pedestrians in video sequences have a problem that counting accuracy degrades when many pedestrians coexist and occlusion occurs frequently. In this paper, we introduce a method of clustering optical flows extracted from pedestrians in video frames to improve the counting accuracy. The proposed method counts the number of pedestrians by using pre-learned statistics, based on the strong correlation between the number of optical flow clusters and the actual number of pedestrians. We evaluate the accuracy of the proposed method using several video sequences, focusing in particular on the effect of parameters for optical flow clustering. We find that the proposed method improves the counting accuracy by up to 25% as compared with a non-clustering method. We also report that using a clustering threshold of angles less than 1 degree is effective for enhancing counting accuracy. Furthermore, we compare the performance of two algorithms that use feature points and lattice points when optical flows are detected. We confirm that the counting accuracy using feature points is higher than that using lattice points especially when the number of occluded pedestrians increases.

Keywords: Pedestrian Counting, Video Sequence, Optical Flow, Clustering.

1. INTRODUCTION

Pedestrian counting at event sites, buildings, or streets is an essential function for pedestrian traffic control, marketing surveys, surveillance in public places, and so on. For example, counting customers who enter and exit from each shop in shopping mall can be used for measuring business performance of the shops. Counting pedestrians at streets is effective to decide when to do road construction and maintenance. Traditionally, pedestrian counting is conducted manually, that has some disadvantages. For example, personnel expenses are high, simple work for long time is a cause of stress, it may introduce human errors, and so on. Therefore, automatic pedestrian counting has received much attention [1], [2].

Numerous approaches to automatically counting moving pedestrians have been reported in the literature. The most widely deployed methods utilize laser sensors [3], [4], [5], [6] and infrared sensors [7], [8], [9], [10], [11], [12], [13]. For example, [6] proposes the counting pedestrians in real time using a network of single row laser range scanners. Such sensor-based methods have advantages in terms of low cost and robustness against changes in the illumination conditions due to fluctuations of lighting and weather. However, these methods sometimes fail to count pedestrians correctly when the heights of pedestrians walking together are similar or when the heights do not fall within the presumed range, because such methods depend on the difference in propagation delays of reflected laser pulses or infrared light. Although the methods using multiple infrared sensors [7], [8], [9] can count pedestrians moving various directions, the counting accuracy degrades considerably when the street has much traffic and occlusion occurs frequently [14]. An occlusion is caused by pedestrians interacting with each other when many pedestrians are present [15].

On the other hand, pedestrian counting methods using video processing technologies have some advantages. For instance, the camera position is flexible such that it is vertically and obliquely above the pedestrians. Additionally, the video sequences can be retrieved at the distanced position from the target area where pedestrians are to be counted. Furthermore, information such as the height and gender of pedestrians can be obtained, in addition to the number of pedestrians. However, the methods using video processing technologies are affected by changes in the illumination conditions [16]. Furthermore, the counting accuracy of these methods degrades due to occlusion effect [17].

The counting pedestrians by video processing technologies is being actively researched, and a contest for counting pedestrians by using video processing technologies is held at the annual IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS) [18]. Texture-based methods [19], [20], [21], [22] are robust to the occlusion of moving objects whose texture pattern is uniform. However, people generally have non-uniformity in their appearance. Model-based methods [15], [23], [24], [25], [26] use a shape model to extract moving objects, and have the high counting accuracy in the presence of occlusions. However, the performance of these methods becomes considerably worse when the moving object is a non-rigid object such as a person. In [27], [28], a camera is set directly above pedestrians to avoid occlusion. However, the camera position is restricted to locations such as the entrance of a building. The methods in [29], [30], [31] use multiple cameras which give the three-dimensional positions of pedestrians to identify their locations more accurately. However, computing three-dimensional positions from point correspondences over multiple images leads to high computational complexity.

In this paper, we propose a method for counting the number of pedestrians in the video sequences retrieved by a single stationary camera. The proposed method is based on clustering optical flows of moving objects in video frames. Clustering optical flows is important for distinguishing multiple pedestrians who are walking together. Since optical flows detected from a single pedestrian have similar lengths, angles, and since their starting points exist within a certain range, we utilize such parameters in clustering. The proposed method estimates the number of pedestrians based on the strong correlation between the number of optical flow clusters detected by proposed method and the actual number of pedestrians. Since the degree of correlation depends on camera position and various other factors, we utilize statistical data to estimate the number of pedestrians.

The effectiveness of the proposed method is evaluated through implementation experiments with several video sequences, including the one used at the PETS contest. We compare the performance of the proposed method with a non-clustering method. Additionally, the performance of two kinds of optical flow extraction algorithms is evaluated. We also focus on the relationships between the optimal threshold of optical flow angles in clustering and the upper-limit of angle differentiation resolution of the method.

The rest of this paper is organized as follows. In Section 2, we introduce the overview of the pedestrian counting method in the video sequences. Section 3 explains the detail of optical flow clustering of the method. The pedestrian counting method is evaluated by using several video sequences in Section 4. Finally, Section 5 concludes the paper and describes future work.

2. SYSTEM DESCRIPTION

2.1 Overview

The overview of pedestrian counting method is depicted in Figure 1. The method places a stationary camera at an arbitrary position and records video of the target area (e.g., a street or building entrance) where pedestrians are to be counted. Retrieved video sequences are sent to a server, which computes the estimate of the number of pedestrians.

On the server, the method first uses the background difference method to identify the moving parts in each video frame. Next, optical flows of moving parts are detected by using a well-known method, followed by clustering them based on the lengths, angles, and source locations of detected optical flows. Finally, the number of pedestrians is estimated, based on the pre-learning correlation between the number of optical flow clusters and the actual number of pedestrians. In what follows, each step of the method proposed is explained in turn.

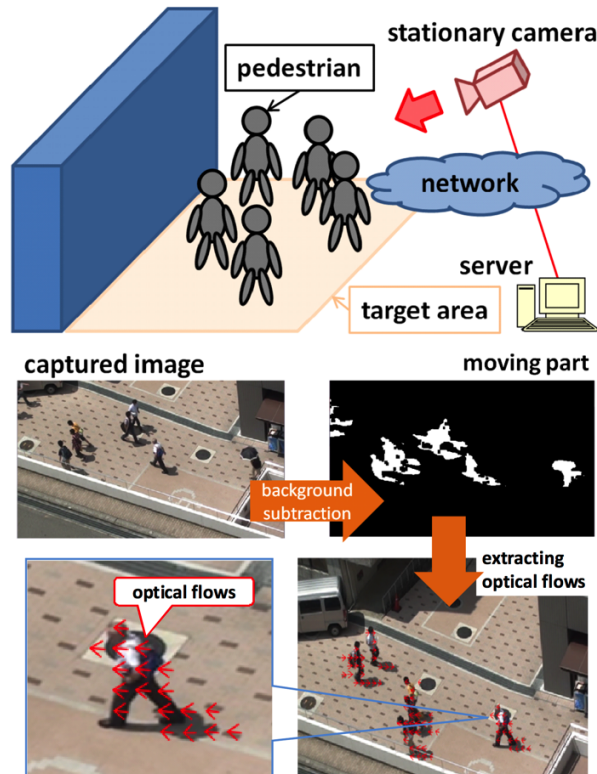


FIGURE 1: Overview of Pedestrian Counting Method.

2.2 Extracting Optical Flows

For the i -th frame ($2 \leq i \leq F$, where F is the total number of frames in the video sequence), denoted by f_i , the method identifies the moving parts in the frame by using the background difference method [32], [33], [34], [35]. To mitigate the adverse effects of changes in the illumination conditions, the method updates the background frame dynamically by using the exponential weighted moving average calculation as follows [36]:

$$\bar{B}_{i,x,y} = (1 - \alpha)\bar{B}_{i-1,x,y} + \alpha B_{i,x,y} \quad (1)$$

Here, $B_{i,x,y}$ ($B \in \{R, G, B\}$) denotes the RGB pixel value at coordinate (x, y) of frame f_i , $\bar{B}_{i,x,y}$ ($\bar{B} \in \{R, G, B\}$) denotes the RGB pixel value at coordinate (x, y) of the background frame, and α is a weight parameter. The moving parts are then identified in the frame by evaluating the difference between the RGB pixel value of the current frame and that of the background frame.

$$O_{i,x,y} = \begin{cases} 1 & (d_{i,x,y} > p_d) \\ 0 & (d_{i,x,y} \leq p_d) \end{cases} \quad (2)$$

$$d_{i,x,y} = |B_{i,x,y} - \bar{B}_{i-1,x,y}| \quad (3)$$

where $O_{i,x,y}$ indicates whether (x, y) is in moving parts of the frame. When $O_{i,x,y} = 1$, (x, y) is in moving parts of the frame. p_d is the threshold value.

Next, the method places multiple points on the moving parts for extracting optical flows. We consider the following two algorithms and compare them in Subsection 4.4. One uses the *feature points* detected by Harris method [37]. In Harris method, the corners of moving parts, which are defined as the intersection of two edges [38], are used as the points for detecting optical flows, as depicted in Figure 2. The other uses the lattice points, which are placed equally spaced in moving parts, as depicted in Figure 3. Figures 4 and 5 are snapshots where the feature points and the lattice points are detected in a video sequence, respectively.

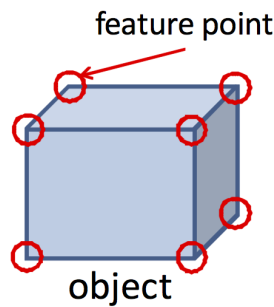


FIGURE 2: Feature Points Detected By Harris Method.

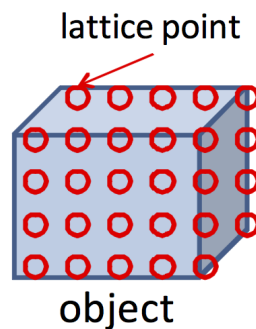


FIGURE 3: Lattice Points.



FIGURE 4: Feature Points Detected In A Video Sequence.



FIGURE 5: Lattice Points in A Video Sequence.

The method then extracts optical flows at multiple points placed using each method in each moving part of the frame. Optical flows are calculated by the Lucas-Kanade method [39]. Here, we denote the coordinate of the j -th point as (x_j, y_j) . The Lucas-Kanade method requires the following assumptions.

- The brightness of each moving part remains unchanged in frames f_{i-1} and f_i .
- The distance traveled by a moving part between frames f_{i-1} and f_i is within a certain range.
- Neighboring points, such as (x_j, y_j) and (x_{j+1}, y_{j+1}) of frame f_{i-1} , which belong to the same moving part, retain the same relative positions in frame f_i .

Figure 6 depicts that a point of frame f_{i-1} which corresponds to a point (x_k, y_k) of frame f_i as a result of movement is detected. By denoting the coordinate of such a point as (x_j, y_j) , the method determines that the optical flow whose starting point is (x_j, y_j) and whose ending point is (x_k, y_k) of frame f_i .

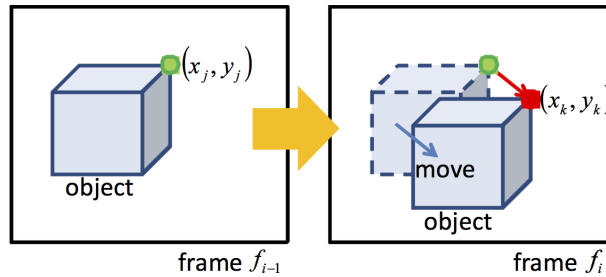


FIGURE 6: Detecting Optical Flows By Lucas-Kanade Method.

To alleviate the effects of noises and changes in the illumination conditions, the method deletes optical flows whose lengths do not fall into a certain range. That is, the method utilizes only the optical flows that satisfy the following condition, where $l_{i,j}$ is the length of an optical flow whose starting point is (x_j, y_j) of frame f_i .

$$l_{min} < l_{i,j} < l_{max} \quad (4)$$

l_{min} and l_{max} are determined empirically by pre-learning.

2.3 Clustering Optical Flows

Next, in the proposed method, optical flows are clustered based on the lengths, angles, and source locations of the detected optical flows, as shown in Figure 7. Clustering optical flows is important for distinguishing multiple pedestrians who are walking together. In Section 3, we explain the detail of clustering optical flows.

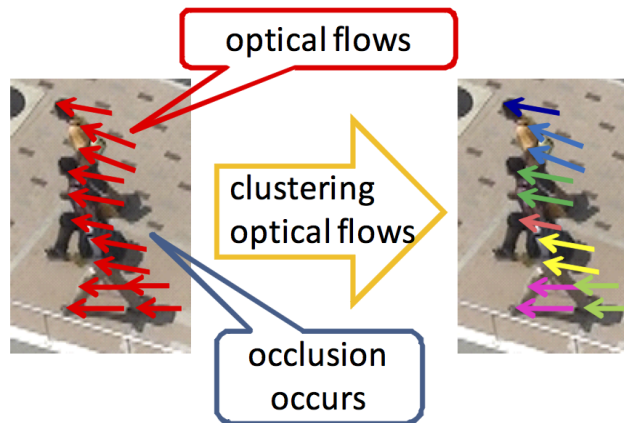


FIGURE 7: Optical Flow Clustering.

2.4 Estimating the Number of Pedestrians

We assume that the number of clusters detected in a frame and the actual number of the pedestrians in the frame have a strong correlation. Since optical flows are extracted from moving parts which indicate pedestrians in video frames, and optical flows detected from a single pedestrian tend to belong to the same cluster. The degree of correlation depends on the camera position, the size of pedestrians in the frames, and various other factors. The method utilizes a

part of frames in the video sequence for pre-learning and estimates the number of pedestrians for all frames in the video sequence. Here, the number of clusters in frame f_i is denoted by C_i . The method calculates the average number of clusters in the frames for pre-learning as follows:

$$\bar{C} = \frac{\sum_{q=1}^{F_l} C_q}{F_l} \quad (5)$$

where $F_l (< F)$ is the number of frames for pre-learning. The actual number of pedestrians in frame f_i is denoted by P_i . The method obtains the average number of pedestrians in the frames for pre-learning as follows.

$$\bar{P} = \frac{\sum_{q=1}^{F_l} P_q}{F_l} \quad (6)$$

Note that to calculate \bar{P} , the method must count the actual number of pedestrians in the part of the video sequence, which means pre-learning.

The average number of clusters per person in pre-learned frames, denoted by \bar{C}_p , is calculated as follows.

$$\bar{C}_p = \frac{\bar{C}}{\bar{P}} \quad (7)$$

The estimated number of pedestrians in frame f_i , denoted by P_i^e , is calculated as

$$P_i^e = \frac{C_i}{\bar{C}_p} \quad (8)$$

3. OPTICAL FLOW CLUSTERING

It is assumed that optical flows detected from a single pedestrian have similar lengths, angles, and since their starting points exist within a certain range. Therefore, optical flows are clustered based on the degree of similarity of their lengths, angles, and starting points. Here, a set of detected optical flows in frame f_i is denoted by D_i and the l -th optical flow in D_i is denoted by $o_{i,l} (\in D_i)$.

The proposed method conducts the clustering all optical flows in D_i . At first, a new cluster is created, denoted by $G_{i,1}$, and the optical flow $o_{i,1} (\in D_i)$ is added to $G_{i,1}$. For the optical flow $o_{i,l} (l \geq 2)$, $o_{i,l}$ is compared with all optical flows in each cluster $G_{i,m}$, in terms of length, angle, and starting point. We explain each step of optical flow clustering using two optical flows $o_{i,l}$ and $o_{i,s}$.

The lengths of two optical flows $o_{i,l}$ and $o_{i,s}$ are first compared where the following equation is satisfied or not:

$$|l_{i,l} - l_{i,s}| \leq l_{th} \quad (9)$$

where $l_{i,l}$ and $l_{i,s}$ are the lengths of $o_{i,l}$ and $o_{i,s}$, which are represented by following equations (Eq.(10) and Eq.(11)), and l_{th} is the threshold of for clustering optical flows.

$$l_{i,l} = \sqrt{(x_{i,l}^e - x_{i,l}^s)^2 + (y_{i,l}^e - y_{i,l}^s)^2} \quad (10)$$

$$l_{i,s} = \sqrt{(x_{i,s}^e - x_{i,s}^s)^2 + (y_{i,s}^e - y_{i,s}^s)^2} \quad (11)$$

where $(x_{i,l}^s, y_{i,l}^s)$ and $(x_{i,l}^e, y_{i,l}^e)$ are the starting point and ending point of optical flow $o_{i,l}$, respectively.

Next, the angles of two optical flows $o_{i,l}$ and $o_{i,s}$ are next compared where the following equation is satisfied or not:

$$|\theta_{i,l} - \theta_{i,s}| \leq \theta_{th} \quad (12)$$

where $\theta_{i,l}$ and $\theta_{i,s}$ are the angles of $o_{i,l}$ and $o_{i,s}$, which are represented by following equations (Eq.(13) and Eq.(14)), and θ_{th} is the threshold of angle.

$$\theta_{i,l} = \arctan\left(\frac{y_{i,l}^e - y_{i,l}^s}{x_{i,l}^e - x_{i,l}^s}\right) \quad (13)$$

$$\theta_{i,s} = \arctan\left(\frac{y_{i,s}^e - y_{i,s}^s}{x_{i,s}^e - x_{i,s}^s}\right) \quad (14)$$

Finally, the starting points of two optical flows $o_{i,l}$ and $o_{i,s}$ are compared.

$$|x_{i,l}^s - x_{i,s}^s| \leq x_{th} \quad (15)$$

$$|y_{i,l}^s - y_{i,s}^s| \leq y_{th} \quad (16)$$

where x_{th} and y_{th} are respectively the thresholds of x-coordinate and y-coordinate of optical flows.

When $o_{i,l}$ and all optical flows in $G_{i,m}$ satisfy Eq.(9)–(16), the proposed method adds $o_{i,l}$ to $G_{i,m}$. When $o_{i,l}$ does not belong to any existing cluster, a new cluster for $o_{i,l}$ is created. After clustering, the number of optical flows in each cluster is assessed. When the number is smaller than the threshold N_d , the cluster is discarded since we assume the cluster is caused by noise such as changes in the illumination conditions.

Figures 8(a) and 8(b) show the clustering results when θ_{th} is small and large, respectively. In these figures, the colored lines represent detected optical flows, and the optical flows that have the same color belong to the same cluster. When θ_{th} is small, since the number of optical flows in each cluster is small, the number of clusters is large. On the other hand, when θ_{th} is large, since the number of optical flows in each cluster is large, the number of clusters is small. Such difference with difference value of θ_{th} would affect the counting accuracy of the proposed method. In Subsection 4.3, the effect of θ_{th} on the counting accuracy is evaluated.

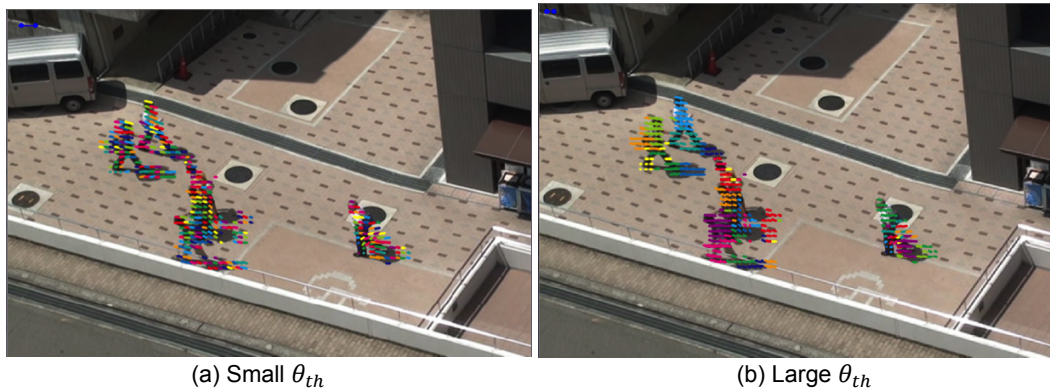


FIGURE 8: Results of Optical Flow Clustering.

4. EXPERIMENTAL EVALUATION

4.1 Experimental Environment

The effectiveness of the proposed method was evaluated through implementation experiments with three video sequences that have different characteristics. One was recorded a 30 [s] video (720×480 [pixels]) at the Toyonaka Campus of Osaka University, Japan at 29 [frames per second (fps)]. The shadows of pedestrians lay across the pavement in this video. Second video was recorded at a pedestrian crossing in Tokyo, Japan. In the video, many pedestrians move in different directions. The video (320×240 [pixels]) is 20 [s] long and was recorded at 29 [fps]. There is no shadow and the size of pedestrians is large in this video. The other is from the PETS contest, which was obtained from [40]. The PETS contest video (768×576 [pixels]) is 34 [s] long and was recorded at 7 [fps]. In this video, many pedestrians are present and occlusion occurs frequently. A snapshot from each video sequence is shown in Figures 9(a), 9(b), and 9(c), respectively. The proposed method was implemented on a PC with a 2.93 [GHz] Intel Core i7 CPU and 3.00 [GB] of memory, running Microsoft Windows 7 Enterprise. The parameters of the method for all video sequences are set as follows: $\alpha = 0.05$, $l_{min} = 0.2$, $l_{max} = 25.0$, $F_l =$, $\theta_{th} = 0.2$, $l_{th} = 3.0$, $x_{th} = 38.0$, $y_{th} = 46.0$, and $N_d = 13$. An example of the execution of the proposed method is shown in Figure 10.

In performance evaluations, we compared the counting accuracy of the proposed method with a non-clustering method that estimates the number of pedestrians based on a pre-learned correlation between the total number of optical flows, not clusters, and the actual number of pedestrians in each frame. The other parts of the non-clustering method are the same as in the proposed method.



FIGURE 9: Snapshots of Video Sequences.

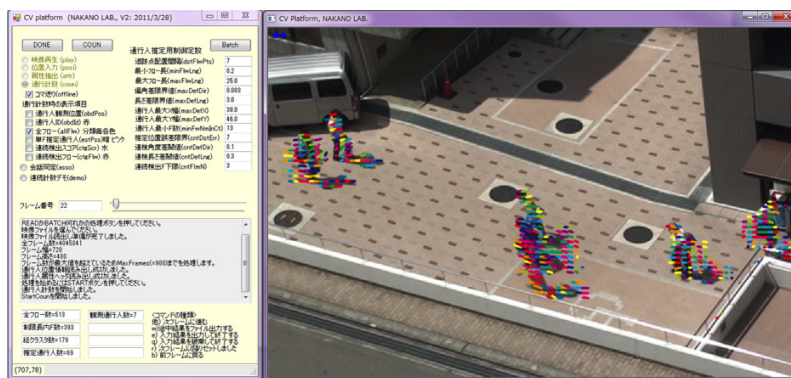


FIGURE 10: Snapshot of the Proposed Method.

4.2 Results and Discussions

Figures 11(a), 11(b), and 11(c) show the relationship between the number of clusters detected by the proposed method and the actual number of pedestrians for three video sequences. Each plot

shows the result for each frame. We observe a linear correlation between the number of clusters and the actual number of pedestrians regardless of the difference of the video sequence. This is because we utilize the simple pre-learning algorithm described in Subsection 2.4. From figures 11(a), 11(b), and 11(c), the degrees of correlation between the number of clusters and the actual number of pedestrians are different, since the camera position, the size of pedestrians, and other factors are different. Therefore, the proposed method utilizes the correlation in the video sequence for pre-learning.

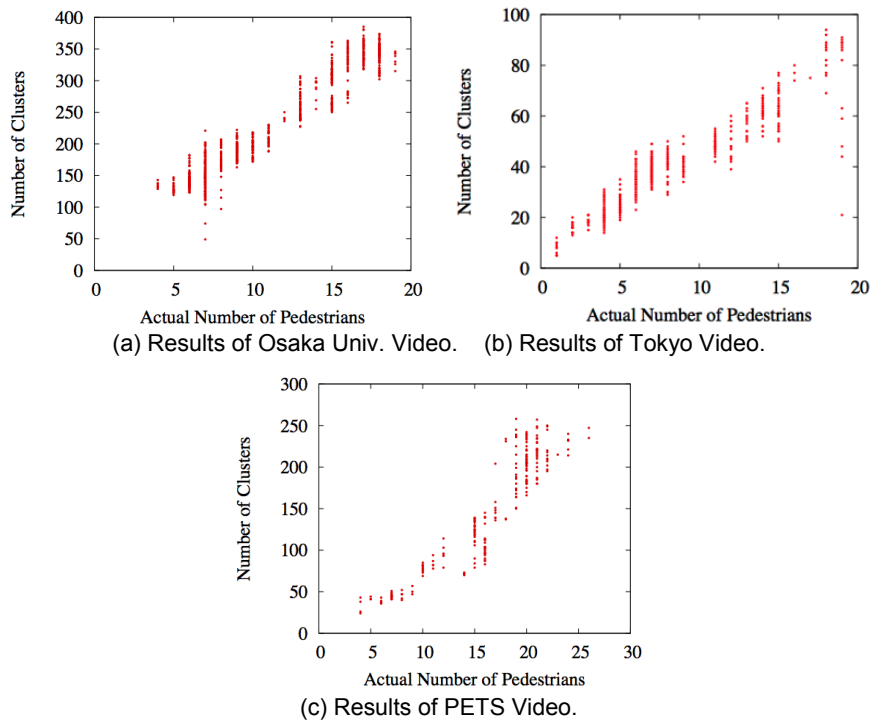


FIGURE 10: Correlations Between the Number of Clusters and the Actual Number of Pedestrians.

Tables 1(a), 1(b), and 1(c) summarize the evaluation results for each video sequence, which show the overall counting accuracy in terms of mean absolute error (MAE), mean relative error (MRE), and variance of the absolute error for the entire video sequence. For all video sequences, the proposed method outperforms the non-clustering method, indicating the effectiveness of the optical flow clustering proposed in this paper. For the video sequence taken in Tokyo in particular, the proposed method had counting accuracy that was about 25% higher than that of the non-clustering method. The proposed method had higher counting accuracy than that of the non-clustering method for another video sequences.

Method	MAE	MRE	Variance of absolute error
Proposed	0.952	0.103	1.211
Non-clustering	0.989	0.110	1.284

(a) Results of Osaka Univ. Video.

Method	MAE	MRE	Variance of absolute error
Proposed	1.123	0.170	1.603
Non-clustering	1.502	0.201	1.932

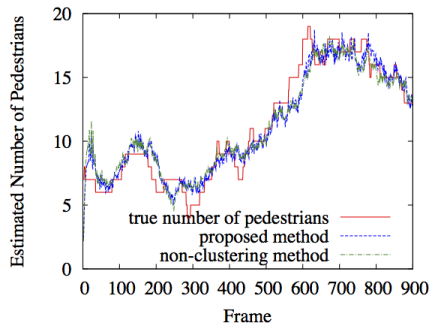
(b) Results of Tokyo Video.

Method	MAE	MRE	Variance of absolute error
Proposed	2.528	0.174	3.164
Non-clustering	3.085	0.220	3.738

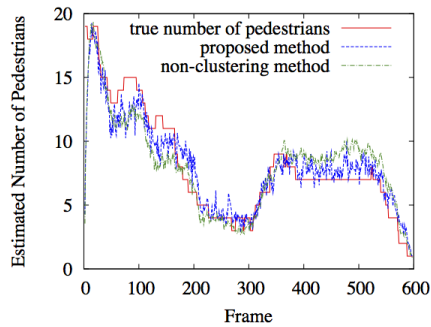
(c) Results of PETS Video.

TABLE 1: Overall Counting Accuracies.

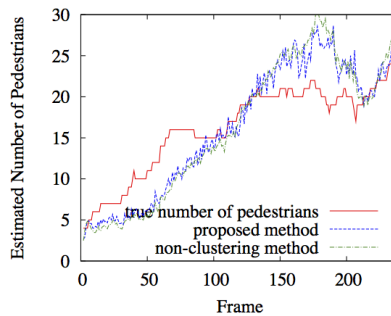
Figures 12(a), 12(b), and 12(c) show the temporal changes in the counting accuracy of the proposed method and that of the non-clustering method for each video sequence. The counting accuracy of both methods becomes worse especially when the actual number of pedestrians increases or decreases largely. One possible reason is that the method described in Subsection 2.2 fails to detect optical flows correctly when pedestrians enter or leave the target area. In such situations, a part of a pedestrian detected at the edge of the target area, which degrades the estimation accuracy based on pre-learning method. Other reasons for the lower counting accuracy include pedestrians that stop moving in the frame and the effect of shadow.



(a) Results of Osaka Univ. Video.



(b) Results of Tokyo Video.



(c) Results of PETS video.

FIGURE 12: Temporal Changes in Counting Accuracy.

It should be also noted that the counting accuracy of both methods significantly deteriorates for the video sequence from the PETS contest from Figure 12(c). In the PETS video sequence, tall and short pedestrians are walking together and the whole bodies of short pedestrians are sometimes completely occluded. Therefore, both methods cannot extract optical flows of occluded pedestrians. Additionally, since the frame rate of the video sequence from the PETS contest is extremely low, the pedestrians move largely between two successive frames. Therefore, the method described in Subsection 2.2 fails to detect optical flows.

Another reason for this large error is observed in Figure 11(c). Compared with Figures 11(a) and 11(b), the linear relationships between the number of pedestrians and the number of clusters is weak for PETS video sequence. Finding more appropriate fitting function is one of our future research.

4.3 Effect of Angle Threshold on Optical Flow Clustering

In this subsection, we focus on the angle threshold for optical flow clustering, since the angle threshold is important for distinguishing multiple pedestrians who are walking together. Figure 13 shows the relationship between the angle threshold and the MRE of the number of pedestrians estimated by the proposed method for the video sequence at Osaka University.

Note that the similar results have been obtained for the video sequence taken in Tokyo and that from the PETS contest. The finest angle resolution of the proposed method is approximately 0.2 degree, which is governed by the frame resolution, the average size, and the velocity of pedestrians. From this figure, it is found that the counting accuracy is high when the angle threshold is around 0.1 degree, which is roughly the same as the method's upper limit of angle resolution. We recall Figures 8(a) and 8(b) for small and large angle thresholds, where the angle threshold in Figure 8(a) is 0.1 degree and that in Figure 8(b) is 5.0 degree. By comparing these figures, it is observed that the proposed method detects more clusters for each pedestrian when small threshold is utilized. However, by utilizing the linear correlation exhibited in Fig. 11(a), the proposed method can increase the counting accuracy.

Figure 14 shows the relationship between the actual number of pedestrians and the MRE of the number of pedestrians estimated by the proposed method, for various angle threshold values in video sequence at Osaka University. The error bars in the graph indicate the 95% confidence interval. We can see from this figure that for detecting a small number of pedestrians, a larger threshold is desirable, whereas a smaller threshold is preferable for detecting a larger number of pedestrians. These characteristics are caused by the effect of the pedestrians' legs. When the number of pedestrians is small, less occlusion occurs and many optical flows are detected from the legs. Since the movement of the pedestrians' legs is more complicated than that of other body parts, a large angle threshold is better for avoiding redundant and fluctuating segmentation of optical flow clusters. On the other hand, when there are many pedestrians, a substantial amount of occlusion can be observed and the legs of the pedestrian are often hidden. In such a case, we should utilize a small angle threshold to distinguish multiple pedestrians who move together causing occlusion. This is also confirmed by Figures 8(a) and 8(b).

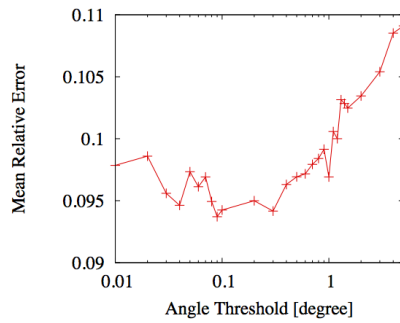


FIGURE 13: Effect of Angle Threshold on Counting Accuracy With Osaka Univ. Video.

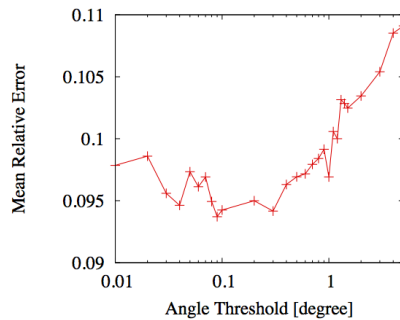


FIGURE 14: Effect of True Number of Pedestrians on Counting Accuracy.

4.4 Comparison of Two Algorithms of Extracting Optical Flows

In this subsection, we compared two algorithms of extracting optical flows, which utilize feature points and lattice points, explained in Subsection 2.2.

Figure 15 shows the relationship between the number of points set for extracting optical flows and the MRE of the number of pedestrians estimated by the proposed method using each algorithm in the video sequence at Osaka University. For both methods, the number of points is regulated by setting lower bound of pixels between two points. It is observed from this figure that the counting accuracy using feature points is higher than that using lattice points. The reason is that the algorithm using feature points detects the points almost from pedestrians' heads or shoulders. Since pedestrians' head and shoulders are hardly occluded even when severe occlusion occurs, the feature points are effective in extracting optical flows for distinguishing occluded pedestrians. On the other hand, the algorithm using lattice points utilizes the points equally separated. Therefore, the number of points utilized for extracting optical flows is only dependent on the size of moving part, not on the size of pedestrians included in the moving part. Then, the counting accuracy becomes worse especially when the number of occluded pedestrians increases.

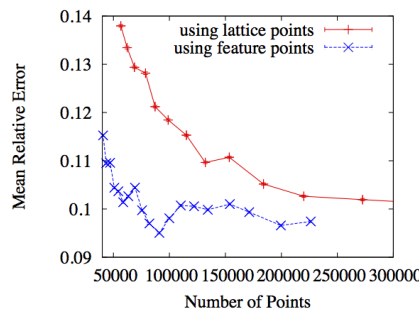


FIGURE 15: Performance Comparison of Two Algorithms for Extracting Optical Flows.

4 CONCLUSIONS

In this paper, we proposed an optical flow clustering method to improve the accuracy of counting pedestrians in the video sequences. In the proposed method, optical flows are clustered based on the lengths, angles, and source locations of optical flows. The proposed method counts the number of pedestrians using pre-learned statistics, based on the number of clusters detected by our method and the number of pedestrians having a strong and linear correlation. We evaluated the accuracy of the proposed method using several video sequences and showed that the counting accuracy using proposed method is up to 25% better than that of a non-clustering method. We also found that the clustering threshold of angles less than 1 degree is effective for enhancing counting accuracy. Additionally, in compared the performance of two algorithms which are use feature points and lattice points when optical flows are detected, the counting accuracy using feature points is higher than that using lattice points especially when the number of occluded pedestrians increases.

The most prioritized task on the present research work in the future is to compare the performance of the proposed method with other existing methods, to confirm the effectiveness of the proposed method. Especially, the comparison with method in latest PETS contest is important.

To emphasize wide applicability of the proposed method, we also plan to evaluate the proposed method using other video sequences to confirm the robustness of the proposed method. The improvement of the accuracy of the proposed method is necessary by resolving the problems explained in Subsections 4.2 and 4.3. By completing the above-mentioned tasks, we can completely confirm the advantage of the proposed method.

REFERENCES

- [1] F. Bu, R. Greene-Roesel, M. Diogenes, and D. Ragland, "Estimating pedestrian accident exposure: Automated pedestrian counting devices report," UC Berkeley Traffic Safety Center, Tech. Rep., Mar. 2007.
- [2] X. Liu, P. H. Tu, J. Rittscher, A. G. A. Perera, and N. Krahnstoever, "Detecting and counting people in surveillance applications," in Proc. AVSS 2005, pp. 306–311, Sep. 2005.
- [3] H. Zhao and R. Shibasaki, "A novel system for tracking pedestrians using multiple single-row laser-range scanners," *Systems, Man and Cybernetics*, vol. 35, pp. 283–291, Mar. 2005.
- [4] A. Fod, A. Howard, and M. J. Mataric, "Laser-based people tracking," in Proc. IEEE ICRA 2002, pp. 3024–3029, May 2002.
- [5] J. Cui, H. Zha, H. Zhao, and R. Shibasaki, "Laser-based detection and tracking of multiple people in crowds," *Computer Vision and Image Understanding*, vol. 106, pp. 300–312, May 2007.
- [6] C. Harris and M. Stephen, "A real-time system for monitoring pedestrians," in Proc. WACV-MOTION 2005, vol. 1, pp. 378–385, Jan. 2005.
- [7] A. Leykin and R. Hammoud, "Robust multi-pedestrian tracking in thermal-visible surveillance videos," in Proc. Conference on CVPR Workshop 2006, p. 136, Jun. 2006.
- [8] E. Goubet, J. Katz, and F. Porikli, "Pedestrian tracking using thermal infrared imaging," in Proc. SPIE 2006, no. 6206, pp. 797–808, Apr. 2006.
- [9] K. Hashimoto, M. Yoshinomoto, S. Matsueda, K. Morinaka, and N. Yoshiike, "Development of people-counting system with human-information sensor using multi-element pyroelectric infrared array detector," *Sensors and Actuators A: Physical*, no. 58, pp. 165–171, Feb. 1998.

- [10] SenSource Inc, "PCW-2BX03 Directional People Counter." Internet: www.sensourceinc.com/PDF/PCW-Wired-Directional%20-%20Brochure.pdf. [Nov. 22, 2012].
- [11] Eco counter, "People counters - Eco-counter." Internet: www.eco-compteur.com. [Nov. 22, 2012].
- [12] InfraRed Integrated Systems Ltd, "IRISYS people counter," Internet: irisys.co.uk/people-counting. [Nov. 22, 2012].
- [13] K. Hashimoto, M. Yoshinmoto, S. Matsueda, K. Morinaka, and N. Yoshiike, "Development of people-counting system with human-information sensor using multielement pyroelectric infrared array detector," *Sensors and Actuators*, vol. 58, no. 2, pp. 165–171, Feb. 1997.
- [14] Q. Chen, M. Gao, J. Ma, D. Zhang, L. M. Ni, and Y. Liu, "Mocus: Moving object counting using ultrasonic sensor networks," *International Journal of Sensor Networks*, vol. 3, no. 1, pp. 55–65, Dec. 2007.
- [15] T. Zhao and R. Nevatia, "Bayesian human segmentation in crowded situations," in *Proc. CVPR 2003*, pp. 459–466, Jun. 2003.
- [16] D. Conte, P. Foggia, G. Percannella, and M. Vento, "Performance evaluation of a people tracking system on pets2009 database," in *Proc. ACSS 2010*, pp. 119–126, Aug. 2010.
- [17] P. Kilambi, E. Ribnick, A. J. Joshi, O. Masoud, and N. Papanikolopoulos, "Estimating pedestrian counts in groups," *Computer Vision and Image Understanding*, vol. 110, pp. 43–59, Apr. 2008.
- [18] James Ferryman, "PETS 2010." Internet: www.cvg.rdg.ac.uk/PETS2010. [Nov. 22, 2012].
- [19] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Neural Networks*, vol. 24, no. 7, pp. 971–989, Jul. 2002.
- [20] S. Bileschi and L. Wolf, "A unified system for object detection, texture recognition, and context analysis based on the standard model feature set," in *Proc. BMVC 2005*, Sep. 2005.
- [21] M. Heikkila and M. Pietikainen, "A texture-based method for modeling the background and detection moving objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 4, pp. 657–662, Apr. 2006.
- [22] V. Mahadevan and N. Vasconcelos, "Background subtraction in highly dynamic scenes," in *Proc. CVPR 2008*, pp. 1–6, Jun. 2008.
- [23] T. Zhao and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 7, pp. 1198–1211, Jul. 2008.
- [24] H. Kitazawa, Z. Li, and K. Yabuta, "Moving object extraction and tracking based on the exclusive block matching," *IEICE technical report*, vol. 108, no. 4, pp. 49–54, Apr. 2008.
- [25] N. Baba, T. Ejima, and H. Matsuo, "Headfinder: A real-time robust head detection and tracking system," in *Proc. SPIE Electronic Imaging Conference on Real-Time Imaging 2002*, vol. 4666, pp. 42–51, Feb. 2002.

[26] A. Ess, B. Leibe, K. Schindler, and L. V. Gool, "A mobile vision system for robust multi-person tracking," in Proc. CVPR 2008, pp. 1–8, Jun. 2008.

[27] R. Eshel and Y. Moses, "Homography based multiple camera detection and tracking of people in a dense crowd," in Proc. CVPR 2008, pp. 1–8, Jun. 2008.

[28] T. Teixeira and A. Savvide, "Lightweight people counting and localizing in indoor spaces using camera sensor nodes," in Proc. ICDSC 2007, pp. 36–43, Sep. 2007.

[29] V. Kettner and R. Zabih, "Counting people from multiple cameras," in Proc. MCS 1999, vol. 2, pp. 267–271, Jul. 1999.

[30] A. Mittal and L. S. Davis, "M2tracker: A multi-view approach to segmenting and tracking people in a cluttered scene using region-based stereo," Computer Vision, vol. 2350, pp. 189–203, May 2002.

[31] J. P. Batista, "Tracking pedestrians under occlusion using multiple cameras," in Proc. ICIAR 2004, vol. 3212, pp. 555–562, Oct. 2004.

[32] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," in Proc. UAI 1997, Aug. 1997.

[33] A. Elegammal, D. Harwood, and L. Davis, "Non-parametric model for background subtraction," in Proc. ECCV 2000, vol. 2, pp. 751–767, Jun. 2000.

[34] T. Horprasert, D. Harwood, and L. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," in Proc. IEEE ECCV 1999 Frame-Rate Workshop, Sep. 1999.

[35] J.-S. Hu, T.-M. Su, and S.-C. Jeng, "Robust background subtraction with shadow and highlight removal for indoor surveillance," in Proc. IROS 2006, pp. 4545–4550, Oct. 2006.

[36] N. Friedman and S. Russell, "Image segmentation in video sequences: A probabilistic approach," in Proc. UAI 1997, pp. 175–181, Aug. 1997.

[37] C. Harris and M. Stephen, "A combined corner and edge detector," in Proc. AVC 1988, pp. 147–152, Aug. 1988.

[38] R. Mehrotra and S. Nichani, "Corner detection," Pattern Recognition, vol. 23, pp. 1223–1233, Mar. 1990.

[39] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in Proc. the 1981 DARPA Image Understanding Workshop, pp. 121–130, Apr. 1981.

[40] James Ferryman, "Video sequence take from PETS workshop," available Internet: www.cvg.rdg.ac.uk/PETS2010/a.html. [Nov. 22, 2012].

Performance Evaluation of 2D Adaptive Bilateral Filter For Removal of Noise From Robust Images

B.Sridhar

Faculty, Department of ECE
Lendi Engineering college
Vizianagaram 535220, India

srib105@gmail.com

Dr.K.V.V.S.Reddy

Professor, Department of ECE
Andhra University
Visakhapatnam, 531003, India

kvvsreddy@gmail.com

Abstract

In this paper, we present the performance analysis of adaptive bilateral filter by pixel to noise ratio and mean square errors. It was evaluate changing the parameters of the adaptive filter half width values and standard deviations. In adaptive bilateral filter, the edge slope is enhanced by transforming the histogram via a range filter with adaptive offset and width. The variance of range filter can also be adaptive. The filter is applied to improve the sharpens of a gray level and color image by increasing the slope of the edges without producing overshoot or undershoots. The related graphs were plotted and the best filter parameters are obtained

Keywords: Image Restoration, Adaptive Filter, Laplacian, Gaussian, Pixel To Noise Ratio, Mean Square Error, Half Width Factor, Standard Deviation Factor

1. INTRODUCTION

Image restoration is the process to construct the image from a blurred and noise image. It used to perform the operation inverse convolution methods. In basic methods the noises are to be estimated. But in practical situations, we unable to get the information regarding blurring and other effect directly. As specially if the robustness increase of an image it is difficult to restored. So to improve the quality we may apply adaptive filtering methods. Bilateral filters shows the prominent results now days, In the first step to develop a sharpening method that is fundamentally different from the unsharp mask filter which sharpens an image by enhancing the high frequency components of the image [4]. In the spatial domain, the boosted high-frequency components lead to overshoot and undershoot around edges, which causes objectionable ringing or halo artefacts. The second aspect of the problem we wish to address is noise[14]. In terms of noise removal, conventional linear filters work well for removing additive Gaussian noise, but they also significantly blur the edge structures of an image. Therefore, a great deal of research has been done on edge-preserving noise reduction[5,6]. The bilateral filter is essentially a smoothing filter; it does not restore the sharpness of a degraded image. The idea of bilateral filtering has since found its way into many applications not only in the area of image de-noising, but also computer graphics, video processing, image interpolation, dynamic range compression, and several others.

This paper examines the performance of the bilateral filtering, a recent approach proposed in [3] that represents the optimum parameter to chosen to get restored image form the robust image. The quality parameters are PSNR and MSE. Applications of bilateral filtering is varied (see, for example, the mean shift filtering [2] applications, mean shift and bilateral filtering, are closely related the repaid advancements of computing technology, any use of the computer-based technologies.

2. INTRODUCTION TO BILATERAL FILTER

The Bilateral filtering was proposed by Tomasi and Manduchi in 1998 as a non-iterative method for edge-preserving, smoothing and noise reducing smoothing filter[3]. The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels. This weight is based on a Gaussian distribution. Crucially the weights depend not only on Euclidean distance but also on the radiometric differences (differences in the range, e.g. color intensity). This preserves sharp edges by systematically looping through each pixel and according weights to the adjacent pixels accordingly.

$$f(m, n) = \sum_k \sum_l h[m, n; k, l] g[k, l]$$

performs 2-D bilateral filtering for the grayscale or color image A. A should be a double precision matrix of size NxMx1 or NxMx3 (i.e., grayscale or color images, respectively) with normalized values in the closed interval [0, 1].

$$h[m, n; m_0, n_0] = I(\Omega_{m_0, n_0}) \frac{1}{m_0, n_0} e^{-\frac{(m-m_0)^2 + (n-n_0)^2}{2\sigma_x^2}} \cdot e^{-\frac{1}{2} \left(\frac{g[m, n] - g[m_0, n_0] - f[m_0, n_0]}{\sigma_r[m_0, n_0]} \right)^2}$$

as shown in the fig.1 the response of bilateral filter. The intensity value at each pixel in an image is replaced by a weighted average of intensity values from nearby pixels. This weight is based on a Gaussian distribution. Crucially the weights depend not only on Euclidean distance but also on the radiometric differences (differences in the range, e.g. color intensity). This preserves sharp edges by systematically looping through each pixel and according weights to the adjacent pixels accordingly

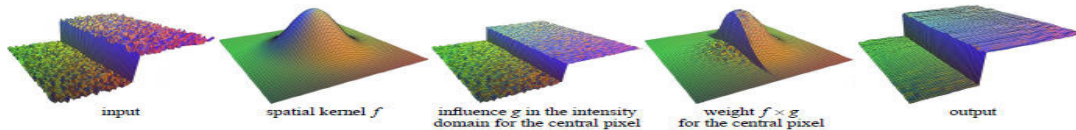


FIGURE 1: Response of Bilateral Filtering.

3. ADAPTIVE BILATERAL FILTERS

In adaptive filtering process here is using a combination of filtering process laplacian and Gaussian to improve the sharpness of edges of the image. the functional and graphical representation is shown in the figure 2,3.

The Laplacian is a 2-D isotropic measure of the 2nd spatial derivative of an image. The Laplacian of an image highlights regions of rapid intensity change and is therefore often used for edge detection (see zero crossing edge detectors). The Laplacian is often applied to an image that has first been smoothed with something approximating a Gaussian smoothing filter in order to reduce its sensitivity to noise, and hence the two variants will be described together here. The operator normally takes a single graylevel image as input and produces another graylevel image as output.[3].In fact, since the convolution operation is associative, we can convolve the Gaussian smoothing filter with the Laplacian filter first of all, and then convolve this hybrid filter with the image to achieve the required result. Doing things this way has two advantages:[7,8]

- Since both the Gaussian and the Laplacian kernels are usually much smaller than the image, this method usually requires far fewer arithmetic operations.
- The LoG ('Laplacian of Gaussian') kernel can be precalculated in advance one convolution needs to be performed at run-time on the image. The 2-D LoG function

centered on zero and with Gaussian standard deviation σ has the form:

$$\text{LoG}(x,y) = -\frac{1}{\pi\sigma^4} \left[1 - \frac{x^2+y^2}{2\sigma^2} \right] e^{-\frac{x^2+y^2}{2\sigma^2}}$$

as shown in the figure 2

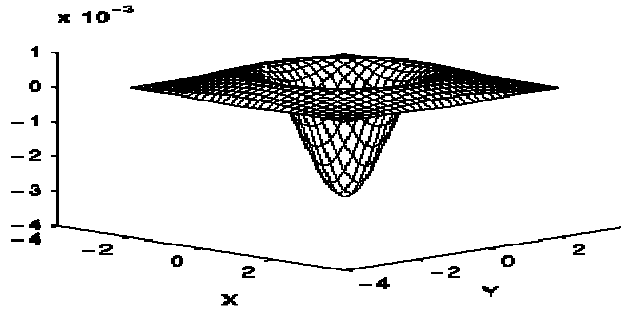


FIGURE 2: The 2-D Laplacian of Gaussian (LoG) Function.

The x and y axes are marked in standard deviations (σ).

A discrete kernel that approximates this function (for a Gaussian $\sigma = 1.4$) is shown in Fig.

0	1	1	2	2	2	1	1	0
1	2	4	5	5	5	4	2	1
1	4	5	3	0	3	5	4	1
2	5	3	-	-	-	3	5	2
			12	24	12			
2	5	0	-	-	-	0	5	2
			12	40	24			
2	5	3	-	-	-	3	5	2
			12	24	12			
1	4	5	3	0	3	5	4	1
1	2	4	5	5	5	4	2	1
0	1	1	2	2	2	1	1	0

FIGURE 3: Discrete Approximation To LoG Function With Gaussian $\sigma = 1$.

Note that as the Gaussian is made increasingly narrow, the LoG kernel becomes the same as the simple Laplacian kernels shown in Figure 3. This is because smoothing with a very narrow

Gaussian ($\sigma < 0.5$ pixels) on a discrete grid has no effect. Hence on a discrete grid, the simple Laplacian can be seen as a limiting case of the LoG for narrow Gaussians.[12,13]

4. EXPERIMENTAL DETAILS

In this section we analyze the performance of bilateral filters on gray level and color images. The proposed algorithm shows in the figure 4. The two most common forms of degradation an image suffers are loss of sharpness or blur and noise. The degradation model consists of a linear shift-invariant blur followed by additive noise . in this paper a common process can choose at time

both gray and color images. It is simulated in MATLAB, the results and explanations as explained in the section –5.

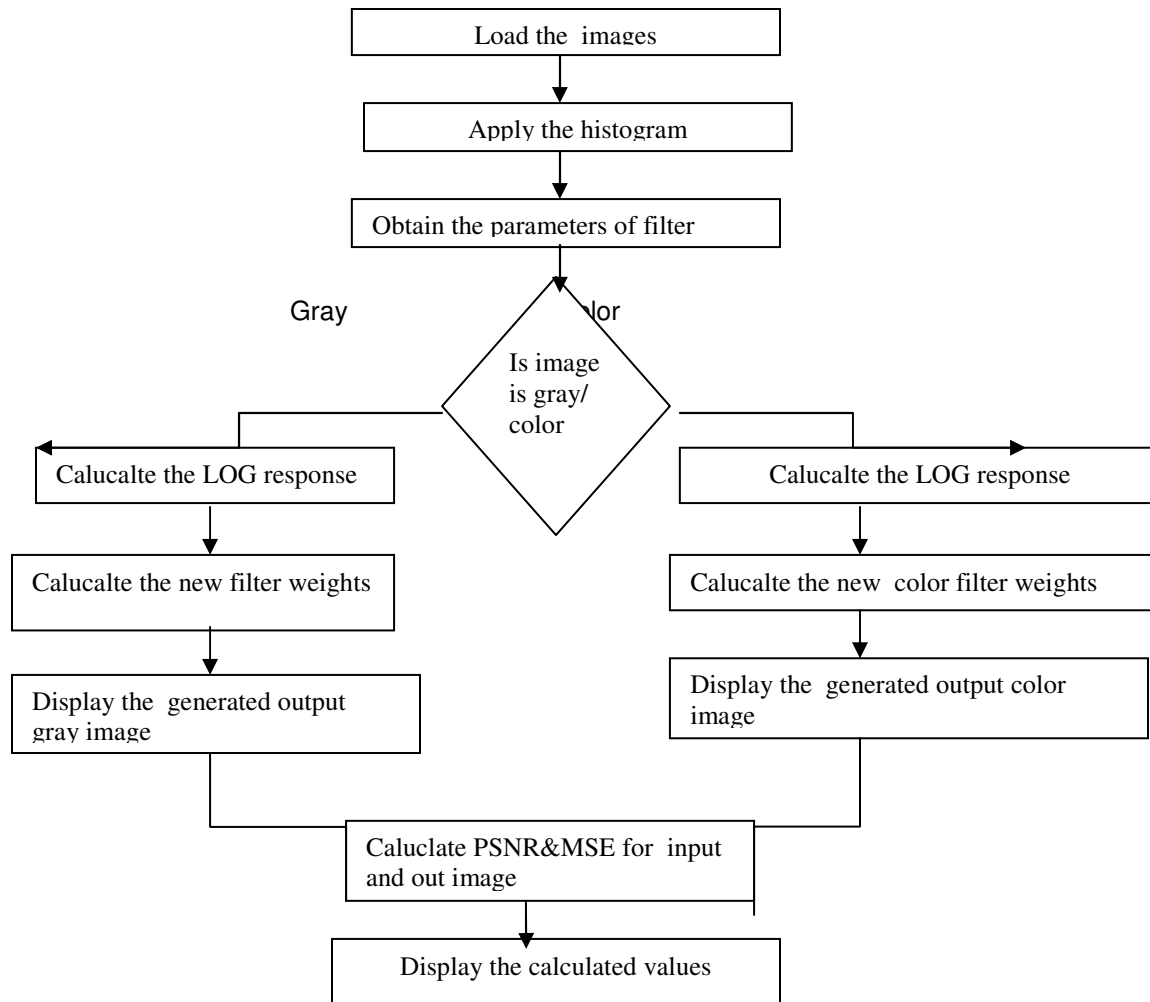


FIGURE 4: Flowchart For Proposed Method.

4.1 Mean Square Error

MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the squares of the "errors." The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate.[9].

The MSE is the second moment (about the origin) of the error, and thus incorporates both the variance of the estimator and its bias. For an unbiased estimator, the MSE is the variance. Like the variance, MSE has the same units of measurement as the square of the quantity being estimated.

The MSE of an estimator $\hat{\theta}$ with respect to the estimated parameter θ is defined as

$$MSE(\hat{\theta}) = E [(\hat{\theta} - \theta)^2]$$

The MSE is equal to the sum of the variance and the squared bias of the estimator

$$\text{MSE}(\hat{\theta}) = \text{Var}(\hat{\theta}) + (\text{Bias}(\hat{\theta}, \theta))^2$$

Thus assesses the quality of an estimator in terms of its variation and unbiasedness. The MSE is not equivalent to the expected value of the absolute error. Since MSE is an expectation, it is not a random variable. It may be a function of the unknown parameter θ , but it does not depend on any random quantities. However, when MSE is computed for a particular estimator of θ the true value of which is not known, it will be subject to estimation error. In a Bayesian sense, there are cases in which it may be treated as a random variable.

4.2 Peak Signal to Noise Ratio

The peak signal-to-noise ratio (PSNR) is the ratio between a signal's maximum power and the power of the signal's noise. Signals can have a wide dynamic range, so PSNR is usually expressed in decibels, which is a logarithmic scale[10,11]. It is most easily defined via the mean squared error (MSE) which for two $m \times n$ monochrome images I and K where one of the images is considered a noisy approximation of the other is defined as:

$$\text{MSE} = \frac{1}{m \cdot n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i,j) - K(i,j)]^2$$

The PSNR is defined as:

$$\text{PSNR} = 10 \cdot \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right)$$

$$= 20 \cdot \log_{10} \left(\frac{\text{MAX}_I}{\sqrt{\text{MSE}}} \right)$$

Here, MAX_I is the maximum possible pixel value of the image. Typical values for the PSNR in lossy image and video compression are between 30 and 50 dB, where higher is better. Acceptable values for wireless transmission quality loss are considered to be about 20 dB to 25 dB.

5. RESULTS DISCUSSIONS

In this chapter we will discuss the results of gray and color images plotted for different values of bilateral filter width and MSEs and PSNRs and taken the best values of half width(w) and standard deviation.

5.1 Input Figures

The figures 5 and 6 are taken as the input images for the implementation of Adaptive Bilateral Filter i.e., these images are loaded into MATLAB environment. The adaptive bilateral filtering is applied to both gray scale image and color image.

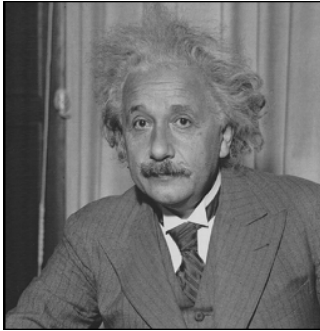


FIGURE 5: Input Gray Image.

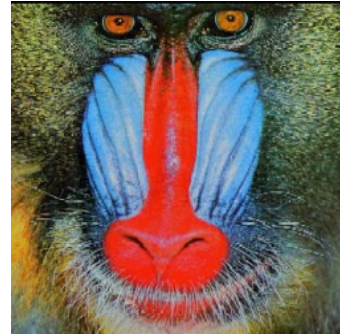


FIGURE 6: Input Color Image.

The dimensions of the grayscale image are 256x256 and the dimension of the color image is 256x256x3.

5.2 Histogram

An image histogram is a graphical representation of the tonal distribution in a digital image. It plots the number of pixels for each tonal value. The histogram for a specific image allows a viewer to judge the entire tonal distribution at a glance. The function “hist” displays the histogram of the images in Cartesian coordinate system.

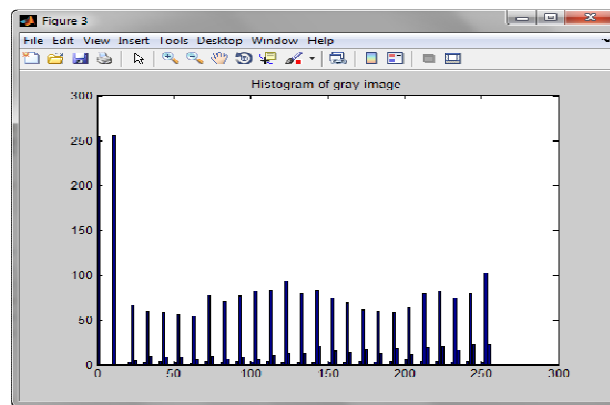


FIGURE 7: Histogram Function of Gray Image.

The Fig: 7 shows the histogram of the input gray image which represents the intensity distribution of different pixels in the image.

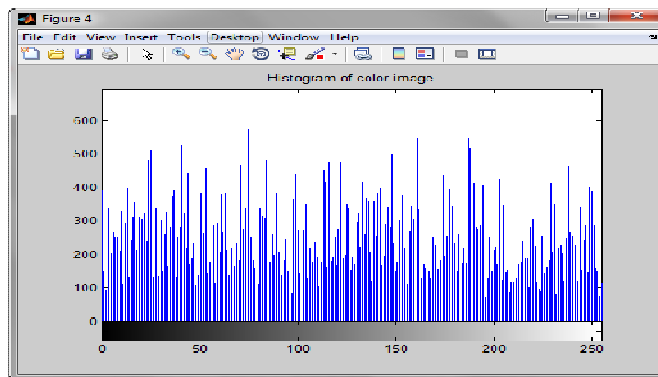


FIGURE 8: Histogram Function of Color Image.

The Figure 8 shows the histogram of the input colour image which represents the intensity distribution of different pixels in the image.

5.3 Output of Adaptive Bilateral Filter

After setting the bilateral parameters like bilateral filter width, domain and range filter standard deviations adaptive bilateral filtering is applied to both the images using the function

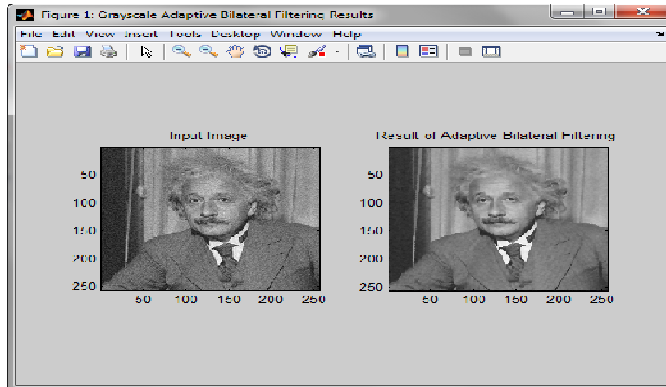


FIGURE 9: Adaptive Bilateral Filter Result for Gray Image.

The figure 9 shows the input gray image which is affected by some kind of noise and the output image after adaptive bilateral filtering. The input gray image appears blurred with unsharpened edges. As adaptive bilateral filter is the most effective one in removing the noise it enhances the slope without generating overshoot and undershoot around the edges. The output image is much clear compared to the input image because of the noise removal and edge preserving.

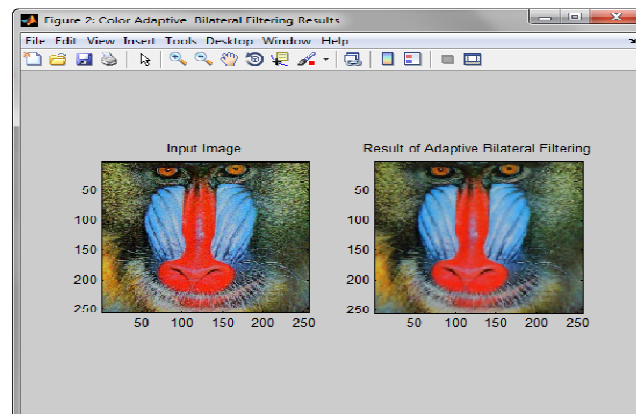


FIGURE 10: Adaptive Bilateral Filter Result for Color Image.

The figure 10 shows the input color image which is affected by some kind of noise and the output image after adaptive bilateral filtering. The input color image appears blurred with unsharpened edges. As adaptive bilateral filter is the most effective one in removing the noise it enhances the slope without generating overshoot and undershoot around the edges. The output image is much clear compared to the input image because of the noise removal and edge preserving.

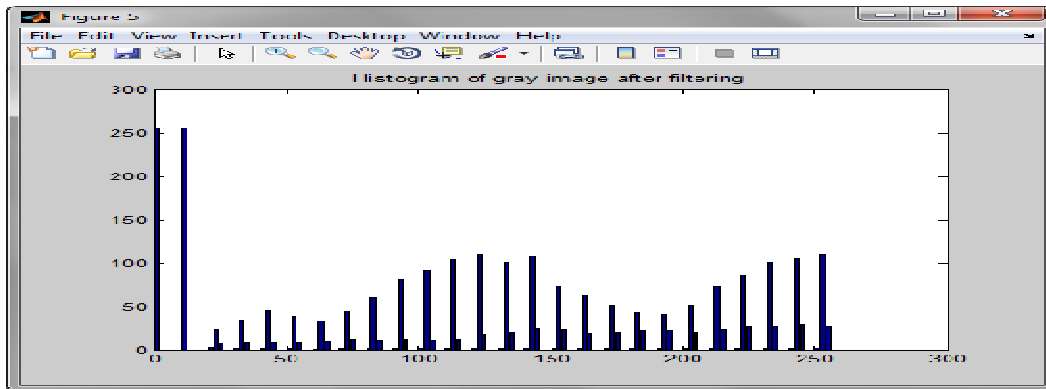


FIGURE 11: Histogram Function of Gray Image After Filtering.

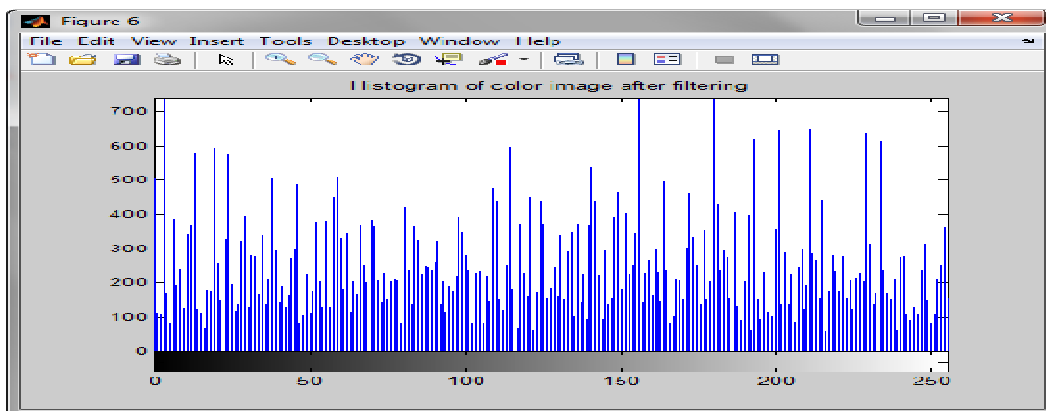


Figure: 12: Histogram Function of Color Image After Filtering.

Figures 11 and 12 represents the intensity levels of gray image and color image respectively after applying the adaptive bilateral filter.

5.5 Calculation of MSE And PSNR

By practically varying different values of the bilateral half width and by taking range filter standard deviation as $\sigma_R = 0.1$ the mean squared errors for gray image is measured and are tabulated as follows:

w	σ_R	Mean Squared Error		Peak signal to noise ratio	
		For Gray image	For Color image	For Gray image	For Color image
2	0.1	0.000501083	0.0016723	32.9944	27.7668
3	0.1	0.00050981	0.0019067	32.9259	27.1973
4	0.1	0.00050983	0.0020941	32.9257	26.7901
5	0.1	0.0005276	0.002253	32.7781	26.4724
2	0.2	0.00049998	0.0016711	33.0105	27.7701
3	0.2	0.00050168	0.0018949	32.9957	27.2241
4	0.2	0.00051173	0.002093	32.9096	26.7922
5	0.2	0.00052602	0.0022611	32.79	26.4569
2	0.3	0.00049903	0.0016618	33.0188	27.7942
3	0.3	0.00050871	0.0018971	32.9353	27.2193
4	0.3	0.00051518	0.0016618	32.8804	26.7983
5	0.3	0.00052855	0.0020901	32.7691	26.4465
2	0.4	0.00049568	0.001602	33.048	27.7828
3	0.4	0.00050479	0.0019082	32.9689	27.1938
4	0.4	0.00050963	0.0020929	32.9275	26.7925
5	0.4	0.00053133	0.0022544	32.7463	26.4697

TABLE 1: MSE And PSNR Values For Gray And Color Images For Fixed Values of σ_R .

The above Table .1 shows the vlaues for peak signal to noise ratio and mean squarred error for both gray and color images taken at fixed values of standard range deviation (σ_R) against varying the values of half width (W)

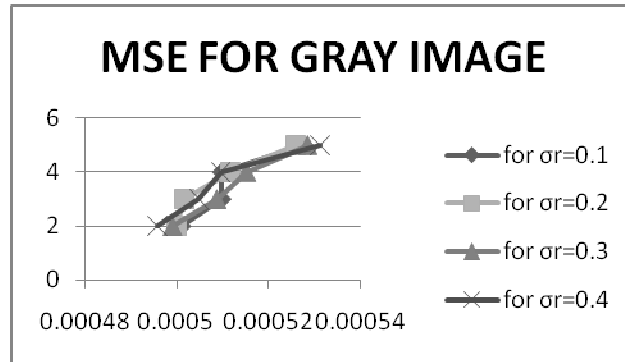


FIGURE 13: MSE Plot For Gray Image For Different Values of σ_R .

The above figure 13 shows the graph of a gray image plotted between MSE on x-axis and bilateral filter width on y-axis. For different values of bilateral filter standard range deviation σ_R , the values are taken between 0.1 to 0.4 and the respective graphs are being plotted.

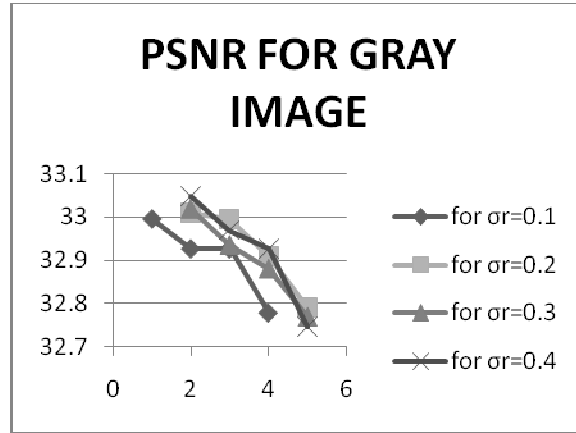


FIGURE 14: PSNR Plot for Gray Image For Different Values of σ_R .

The above figure 14 shows the graph of a gray image plotted between PSNR on y-axis and bilateral filter width on x-axis. For different values of bilateral filter standard range deviation σ_R , the values are taken between 0.1 to 0.4 and the respective graphs are being plotted.

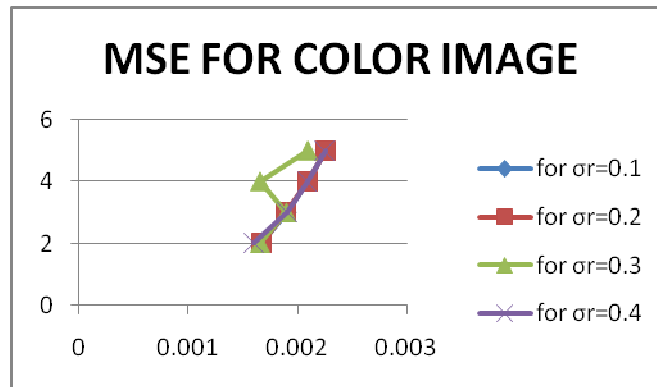


FIGURE 15: MSE Plot for Color Image For Different Values of σ_R .

The above figure 15 shows the graph of a color image plotted between MSE on x-axis and bilateral filter width on y-axis. For different values of bilateral filter standard range deviation σ_R , the values are taken between 0.1 to 0.4 and the respective graphs are being plotted.

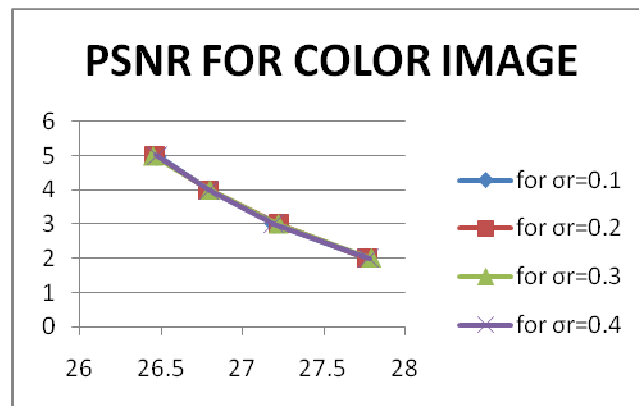


FIGURE 16: PSNR Plot for Color Image For Different Values of σ_R .

The above figure 16 shows the graph of a color image plotted between PSNR on y-axis and bilateral filter width on x-axis. For different values of bilateral filter standard range deviation σ_R , the values are taken between 0.1 to 0.4 and the respective graphs are being plotted.

w	σ_R	Mean Squared Error		Peak signal to noise ratio	
		For Gray image	For Color image	For Gray image	For Color image
2	0.1	0.00049729	0.0016573	33.0339	27.806
2	0.2	0.00050549	0.0016601	32.9629	27.7985
2	0.3	0.00049736	0.0016753	33.0333	27.759
2	0.4	0.00051037	0.0016589	32.9211	27.8019
3	0.1	0.00050324	0.0018911	33.001	27.2089
3	0.2	0.00050741	0.0019097	32.9464	27.1903
3	0.3	0.00050201	0.0018941	32.9929	27.226
3	0.4	0.00050843	0.0019041	32.9372	27.2036
4	0.1	0.00050887	0.0021106	32.934	26.756
4	0.2	0.00051846	0.002081	32.859	26.8173
4	0.3	0.00051271	0.002081	32.9013	26.8173
4	0.4	0.00050983	0.0020941	32.9257	26.7901
5	0.1	0.00053307	0.0022733	32.7322	26.4335
5	0.2	0.00052872	0.0022879	32.7677	24.4056
5	0.3	0.00052521	0.0022728	32.7967	26.4343
5	0.4	0.00052528	0.0022591	32.7961	26.4607

TALBE 2: MSE and PSNR Values For Gray And Color Images For Fixed Values of W.

The above table 5.2 shows the vlaues for peak signal to noise ratio and mean squarred error for both gray and color images taken at fixed values of half width (W) against varying the values of standard range deviation(σ_R)

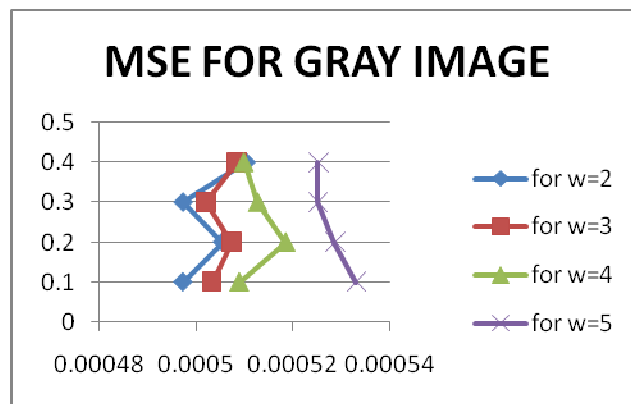


FIGURE 17: MSE Plot for Gray Image For Different Values of w.

The above figure 17 shows the graph of a gray image plotted between MSE on x-axis and bilateral filter standard range deviation σ_R on y-axis. For different values of bilateral filter half width(W), the values are taken between 2 to 5 and the respective graphs are being plotted.

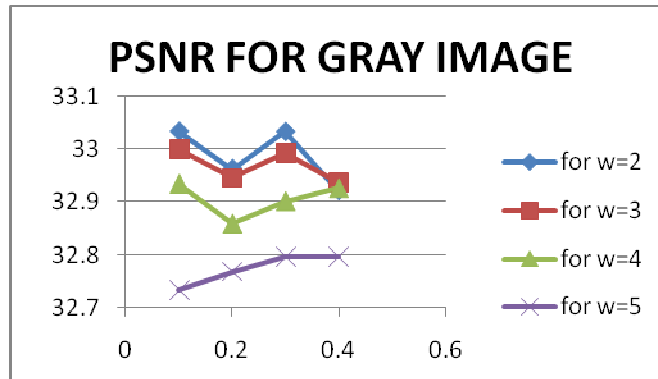


FIGURE 18: PSNR Plot for Gray Image For Different Values of w.

The above figure: 18 shows the graph of a gray image plotted between PSNR on x-axis and bilateral filter standard range deviation σ_R on y-axis. For different values of bilateral filter half width(W), the values are taken between 2 to 5 and the respective graphs are being plotted.

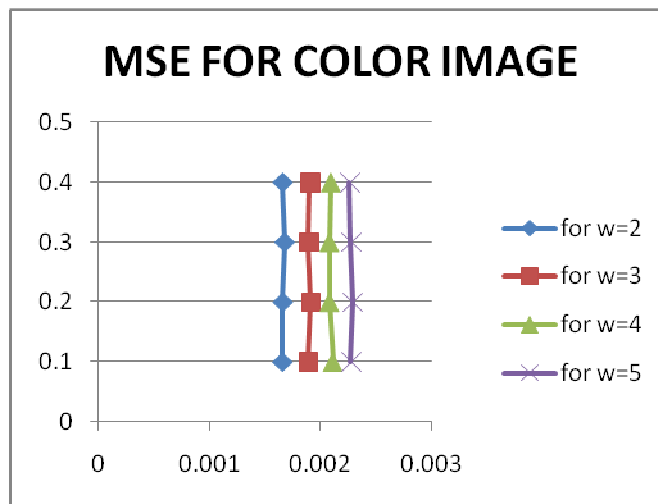


FIGURE 19: MSE Plot for Color Image For Different Values of w.

The above figure: 19. shows the graph of a color image plotted between MSE on x-axis and bilateral filter standard range deviation σ_R on y-axis. For different values of bilateral filter half width(W), the values are taken between 2 to 5 and the respective graphs are being plotted.

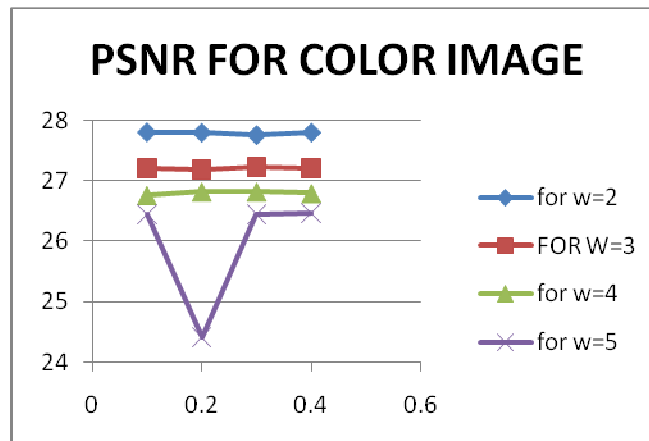


FIGURE 20: PSNR Plot for Color Image For Different Values of w .

The above figure: 20 shows the graph of a color image plotted between PSNR on x-axis and bilateral filter standard range deviation σ_R on y-axis. For different values of bilateral filter half width (W), the values are taken between 2 to 5 and the respective graphs are being plotted.

The proposed method was applied to gray level and color with different standard deviation noises. The PSNR of the restored image was measured for the proposed method as well as the original bilateral filtering method and the state-of-the-art image denoising method using Gaussian scale mixtures proposed by wang. [6] for comparison. A summary of the results is states that the average of PSNR is around 31. The proposed method achieves noticeable PSNR gains over 33 the original bilateral filtering method for all of the gray level. Also the color image signal filtering shows prominent results. It can be observed that the proposed method produced a restored image signal with noticeably improved perceptual quality compared to the Wang and the original bilateral filtering method.

6. CONCLUSIONS

In this paper, we present the adaptive bilateral filter for sharpness enhancement and noise removal. The adaptive bilateral filter sharpens an image by increasing the slope of the edges without producing overshoot or undershoots. In the adaptive bilateral filter, the edge slope is enhanced by transforming the histogram via a range filter with adaptive offset and width. The variance of range filter can also be adaptive. The adaptive bilateral filter is able to smooth the noise, while enhancing edges and textures in the image. Adaptive bilateral filter restored images are to be significantly sharper than those restored by the bilateral filter. The mean square error and peak signal to noise ratios are also calculated for different values of bilateral filter width and range filter standard deviation and corresponding graphs are plotted for both gray and color images. Adaptive bilateral filter works well for both gray images and color images. For future development of the adaptive bilateral filter, we would suggest that the following issues be addressed. First, the adaptive bilateral filter tends to resize the image, due to its fundamental mechanism of sharpening an image by pulling up or pushing down pixels along the edge slope. Second, the adaptive bilateral filter does not perform as well at corners as it does on lines and spatially slow-varying curves, since the adaptive bilateral filter is primarily based on transforming the histogram of the local data, which cannot effectively represent 2-D structures. Finally, in the current design of the adaptive bilateral filter, a fixed domain Gaussian filter is used. Future developments this proposed method does not work efficiently at corner edges, by choose a addition of transformation operation along with proposed method the problem will be solved.

7. REFERENCES

- [1] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," Proc. 1998 IEEE Int'l Conf. Computer Vision, 1998
- [2] D. Comaniciu and P. Meer, "Mean Shift Analysis and Applications," Proc. 1999 IEEE Int'l Conf. Computer Vision, 1999.
- [3] Buyue Zhang and Jan P. Allebach, "Adaptive Bilateral Filter for Sharpness Enhancement and Noise Removal" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 17, NO. 5, MAY 2008
- [4].Das and R. M. Rangayyan, "Enhancement of image edge sharpness and acutance," in Proc. SPIE, 1997, vol. 3026, pp. 133 – 142.
- [5]S. Kim and J. P. Allebach, "Optimal unsharp mask for image sharpening and noise removal," J Electron Imaging, vol. 14, no. 2, pp. 023007–1 – 023007–13, 2005.
- [6].Alexander Wong_ "Adaptive bilateral filtering of image signals using local phase characteristics" Journal of Elsevier Signal Processing 88 1615–1619(2008)
- [7].Rafael C. Gonzalez, Richard E. Woods and Steven L. Eddins (2004). Digital Image
- [8].Processing using MATLAB. Pearson Education. ISBN 978-81-7758-898-9.
Gilat, Amos (2004). MATLAB: An Introduction with Applications 2nd Edition. John Wiley & Sons. ISBN 978-0-471-69420-5.
- [9] <http://www4.comp.polyu.edu.hk/~cslzhang/code.htm>
- [10] <http://www4.comp.polyu.edu.hk/~cslzhang/>
- [11] <http://sse.tongji.edu.cn/linzhang/>
- [12] <http://sse.tongji.edu.cn/linzhang/UsefulLinks/links.htm>
- [13] <http://www.csse.uwa.edu.au/~ajmal/code.html>
- [14]<http://vlm1.uta.edu/~athitsos/vlm/>
- [15]<http://users.soe.ucsc.edu/~milanfar/students/>
- [16].susan approach-----<http://users.fmrib.ox.ac.uk/~steve/susan/index.html>

Speeded-up and Compact Visual Codebook for Object Recognition

B. Mayurathan

*Postgraduate Institute of Science,
University of Peradeniya, Sri Lanka.*

barathy@jfn.ac.lk

A. Ramanan, S. Mahesan

*Department of Computer Science,
University of Jaffna, Sri Lanka.*

a.ramanan, mahesans@jfn.ac.lk

U.A.J. Pinidiyaarachchi

*Department of Statistics and Computer Science,
University of Peradeniya, Sri Lanka.*

ajp@pdn.ac.lk

Abstract

The well known framework in the object recognition literature uses local information extracted at several patches in images which are then clustered by a suitable clustering technique. A visual codebook maps the patch-based descriptors into a fixed-length vector in histogram space to which standard classifiers can be directly applied. Thus, the construction of a codebook is an important step which is usually done by cluster analysis. However, it is still difficult to construct a compact codebook with reduced computational cost. This paper evaluates the effectiveness and generalisation performance of the Resource-Allocating Codebook (RAC) approach that overcomes the problem of constructing fixed size codebooks that can be used at any time in the learning process and the learning patterns do not have to be repeated. It either allocates a new codeword based on the novelty of a newly seen pattern, or adapts the codebook to fit that observation. Furthermore, we improve RAC to yield codebooks that are more compact. We compare and contrast the recognition performance of RAC evaluated with two distinctive feature descriptors: SIFT and SURF and two clustering techniques: K-means and Fast Reciprocal Nearest Neighbours (fast-RNN) algorithms. SVM is used in classifying the image signatures. The entire visual object recognition pipeline has been tested on three benchmark datasets: PASCAL visual object classes challenge 2007, UIUC texture, and MPEG-7 Part-B silhouette image datasets. Experimental results show that RAC is suitable for constructing codebooks due to its wider span of the feature space. Moreover, RAC takes only one-pass through the entire data that slightly outperforms traditional approaches at drastically reduced computing times. The modified RAC performs slightly better than RAC and gives more compact codebook. Future research should focus on designing more discriminative and compact codebooks such as RAC rather than focusing on methods tuned to achieve high performance in classification.

Keywords: Object Recognition, Codebook, K-means, RAC, fast-RNN, SIFT, SURF

1. INTRODUCTION

Object recognition is one of the major challenges in computer vision. Researchers in computer vision have been trying to understand this for many years. Individuals can look around them and recognise familiar objects without much trouble. The ability to generalise from examples and categorise objects, events, scenes, and places is one of the core capabilities of the human visual system. However, this is not an easy task for computers. Local features are used in many computer vision tasks including visual object categorisation, content-based image retrieval, and object recognition. Local features can be thought of as patterns in images that differ from the immediate neighbourhood. Such a pattern can be a corner, blob or a region. The term interest points (or keypoints) usually refer to the set of points that are used to describe these patterns. Local feature detectors are used to find areas of interest in the images.

In the state-of-the-art visual object recognition systems, the visual codebook model has shown excellent categorisation performance in large evaluations such as the PASCAL Visual Object Classes (VOC) Challenges [8] and Caltech object categories [11]. Desirable properties of a visual codebook are compactness, low computational complexity, and high accuracy of subsequent categorisation. Discriminative power of a visual codebook determines the quality of the codebook model, whereas the size of a codebook controls the complexity of the model. Thus, the construction of a codebook plays a central role that affects the model complexity. In general, there are two types of codebook that are widely used in the literature: global and category-specific codebooks. A global codebook may not be sufficient in its discriminative power but it is category-independent, whereas a category-specific codebook may be too sensitive to noise. The codebook itself is constructed by clustering a large number of local feature descriptors extracted from training data. Based on the choice of a clustering algorithm, one might obtain different clustering solutions, some of which might be more suitable than others for object class recognition.

The popular approach to constructing a visual codebook is usually undertaken by applying the traditional K -means method. However, clustering is a process that retains regions of high density in a distribution and it follows that the resulting codebook need not have discriminant properties. This is also recognised as a computational bottleneck of such systems. The resource-allocating codebook (RAC) approach [26] that we compare in this paper slightly outperforms more traditional approaches due to its tendency to spread out the cluster centres over a broader range of the feature space thereby including rare local features in the codebook than density-preserving clustering-based codebooks such as K -means and fast-RNN [20].

The objective of this paper is to study the performance of discriminative clustering techniques for object class recognition. Here a scale-invariant feature descriptors, SIFT and e-SURF have been included in order to study the performance of recognising objects. Consequently K -means, fast-RNN and RAC methods are used to cluster the extracted descriptors and these clustering techniques performances are compared.

Following the introductory section, the rest of this paper is organised as follows. In section 2, we summarise the background information that are closely related to this paper. This includes visual descriptors that are widely used in codebook model-based object recognition and various clustering techniques that have been used in constructing a codebook. Section 3 provides a summary of previous work on object recognition that has used a codebook model-based approach. Section 4 provides the experimental setup and testing results of our work. In section 5, we discuss the extension of RAC in constructing more compact codebooks for object recognition. Finally, section 6 concludes our work.

2. BACKGROUND

Several combinations of image patch descriptors, different features, matching strategies, various clustering methods and classification techniques have been proposed for visual object recognition. Assessing the overall performance of the individual components in such systems is difficult, since the computational requirements and the fine tuning of the different parts become crucial. The well-known framework in the literature uses the SIFT descriptors to describe the patches and cluster them using the standard K-means algorithm, in order to encode the images as a histogram of visual codewords. This section briefly describes several descriptors, clustering techniques and the well known classification method that were used in comparing different codebook models.

2.1 Local Invariant Features

Usually images are composed of different sets of colours, a mosaic of different texture regions, and different local features. Most previous studies have focused on using global visual features such as edge orientation, colour histogram and frequency distribution. Recent studies use local features that are more robust to occlusions and spatial variations. This new way of looking at local features has opened up a whole new range of applications and has brought us a step closer to cognitive level image understanding. Even though many different methods for detecting and describing local image regions have been developed, the simplest descriptor is a vector of image pixels. In this subsection we summarise the well known patch-based scale-invariant feature transform (SIFT) descriptors [21] and its follow up technique, the Speeded-Up Robust Features (SURF) descriptors [2].

2.1.1 Scale-Invariant Feature Transform (SIFT)

SIFT is a method to extract distinctive features from gray-value images, by filtering images at multiple scales and patches of interest that have sharp changes in local image intensities. The SIFT algorithm consists of four major stages: Scale-space extrema detection, keypoint localisation, orientation assignment, and representation of a keypoint descriptor. The features are located at maxima and minima of a difference of Gaussian (DoG) functions applied in scale space. Next, the descriptors are computed based on eight orientation histograms at a 4×4 sub region around the interest point, resulting in a 128 dimensional vector. In PCA-SIFT [15], the principal component analysis (PCA) is used instead of weighted histograms at the final stage of the SIFT. The dimensionality of the feature space is reduced from 128 to 20 which require less storage and increases speed in matching images. Although the size of the feature vector is significantly smaller than the standard SIFT feature vector, it has been reported that PCA-SIFT is less distinctive than SIFT [13].

2.1.2 Speeded-Up Robust Feature (SURF)

SURF is partly inspired by SIFT that makes use of integral images. The scale space is analysed by up-scaling the integral image-based filter sizes in combination with a fast Hessian matrix-based approach. The detection of interest points is selected by relying on the determinant of the Hessian matrix where the determinant is maximum. Next, the descriptors are computed based on orientation using 2D Haar wavelet responses calculated in a 4×4 sub region around each interest point, resulting in a 32 dimensional vector. When information about the polarity of the intensity changes is considered, this in turn results in a 64 dimensional vector. The extended version of SURF has the same dimension as SIFT. SURF features can be extracted faster than SIFT using the gain of integral images and yields a lower dimensional feature descriptor (i.e. 64 dimensions) resulting in faster matching and less storage space but it is not stable to rotation and illumination changes. In [13], e-SURF (i.e. 128 dimensions) is proved to have better performance than SURF.

2.2 Codebook Construction

A simple nearest neighbour design for patch-based visual object recognition is a possible way forward, but is computationally not feasible for large scale data. Hence, a way to cope with the enormous amount of the patch-based descriptors and their higher dimensionality is to cluster them by using an appropriate clustering method that captures the span of the feature space. Instead of the features themselves, the cluster centroids or representative points are used for the

different cluster members. Interest points are detected in training images and a visual codebook is constructed by a vector quantization technique that groups similar features together. Each group is represented by the learnt cluster centres referred to as 'codewords'. The size of the codebook is the number of clusters obtained from the clustering technique. Each interest point of an image in the dataset is then quantized to its closest codeword in the codebook, such that it maps the entire patches of an image in to a fixed-length feature vector of frequency histograms, i.e. the visual codebook model treats an image as a distribution of local features. In this subsection we describe the traditional K-means method and two other techniques known as RAC and fast-RNN that mainly focuses in constructing compact codebooks.

2.2.1 K-means

K-means is one of the simplest unsupervised learning algorithm that solves the well known clustering problem. Given a matrix $X \in \mathbb{R}^{N \times d}$ (representing N points described with respect to d features), then K-means clustering aims to partition the N points into K disjoint sets or clusters by minimizing an objective function, which is the squared error function, that minimizes the within-group sum of squared errors. K-means is a Gaussian mixture model with isotropic covariance matrix the algorithm is an expectation-maximization (EM) algorithm for maximum likelihood estimation.

There are several known difficulties with the use of K-means clustering, including the choice of a suitable value for K , and the computational cost of clustering when the dataset is large. It is also significantly sensitive to the initial randomly selected cluster centres. The time complexity of the traditional K-means method is $O(NdKm)$, where the symbols in parentheses represent number of data, dimensionality of features, the number of desired clusters and the number of iterations of the EM algorithm.

2.2.2 fast-RNN

In [20] the authors have presented a novel approach for accelerating the popular Reciprocal Nearest Neighbours (RNN) clustering algorithm and named as fast-RNN. A novel dynamic slicing strategy is used to speed up the nearest neighbour chain construction. When building nearest neighbour (NN) chains, it finds all the points that lie within a slice of the d -dimensional space of width 2ϵ centred at query point. To determine the nearest neighbour of x_i in S , i.e. $x_j = \text{NN}(x_i)$ where S is a set of N points $S = \{x_1, x_2, \dots, x_N\}$, it builds the first slice of width 2ϵ centred at x_i . Then, it performs a search for the NN of x_i considering only the points inside this slice. Once x_j is identified, it searches for its NN via slicing again, and so on.

Thereafter, agglomerative clustering builds the codebook by initially assigning each data point to its own cluster after that repeatedly selecting and merging pairs of clusters. Thus, it builds a hierarchical tree merging from the bottom (leaves) towards the top (root). The authors in [17] have improved the agglomerative clustering method based on the construction of RNN pairs.

2.2.3 Resource-Allocating Codebook

In [26], a Resource-Allocating Codebook (RAC) has been proposed for constructing a discriminate codebook. RAC is a much simplified algorithm that constructs a codebook in a one-pass process which simultaneously achieves increased discrimination and a drastic reduction in the computational needs. It is initialised by a random seed point selected from the set of visual descriptors. When a subsequent data item is processed, its minimum distance to all entries in the current codebook is computed using an appropriate distance metric. If this distance is smaller than a predefined threshold r (radius of the hypersphere) the current codebook is retained and no action is taken. If the threshold is exceeded by the smallest distance to codewords (i.e. it is a group represented by the learnt cluster centres), a new entry in the codebook is created by including the current data item as the additional entry. This process is continued until all data items are seen only once. The pseudocode of this approach is given in Algorithm 1. The RAC partitions the feature space into a set of overlapping hyperspheres when the distance metric used is the Euclidean norm.

Algorithm 1: Resource-Allocating Codebook

Input: Visual descriptors (**D**) and radius (r) of the hyperspheres.Output: Centres of the hyperspheres (**C**)Step 1: $C_1 \leftarrow D_1$ $i \leftarrow 2$ // to the next descriptor $j \leftarrow 1$ // size of the present **C**Step 2: Repeat Steps 3 to 4 while $i \leq \text{size}(\mathbf{D})$ Step 3: If $\min_j \|D_i - C_j\|^2 \geq r^2 \quad \forall_j$ then create a new hypersphere of r such that $C_j \leftarrow D_i$ $j \leftarrow j + 1$

endif

Step 4: $i \leftarrow i + 1$ Step 5: return **C**

2.3 Classification Using Support Vector Machine (SVM)

SVM is a supervised learning technique based on a statistical learning theory that can be used for pattern classification. In general SVMs outperform other classifiers in their generalisation performance [3]. A linear SVM finds the hyperplane leaving the largest possible fraction of points of the same class on the same side, while maximizing the distance of either class from the hyperplane. SVMs were originally developed for solving binary classification problems [5] and then binary SVMs have also been extended to solve the problem of multi-class pattern classification. There are four standard techniques frequently employed by SVMs to tackle multi-class problems, namely One-Versus-One (OVO) [7], One-Versus-All (OVA) [29], Directed Acyclic Graph (DAG) [25], and Unbalanced Decision Tree (UDT) [28].

OVO method is implemented using a “Max-Wins” voting strategy. This method constructs one binary classifier for every pair of distinct classes and in total it constructs $N(N-1)/2$ binary classifiers, where N is the number of classes. The binary classifier C_{ij} is trained with examples from the i^{th} class and the j^{th} class only. The max-wins strategy then assigns a test data X to the class receiving the highest voting score.

OVA method is implemented using a “Winner-Takes-All” strategy. It constructs N binary classifier models. The i^{th} binary classifier is trained with all the examples in the i^{th} class with positive labels, and the examples from all other classes with negative labels. For a test example X , the winner-takes-all strategy assigns it to the class with the highest classification boundary function value.

DAG-SVMs are implemented using a “Leave-One-Out” strategy. The training phase of the DAG is the same as the OVO method, solving $N(N-1)/2$ binary classifiers. In the testing phase it uses a rooted binary directed acyclic graph. Each node is a classifier C_{ij} from OVO. A test example X is evaluated at the root node and then it moves either to the left or the right depending on the output value.

UDT-SVMs are implemented using a “knock-out” strategy with at most $(N - 1)$ classifiers to make a decision on any input pattern. Each decision node of UDT is an OVA-based optimal classification model. Starting at the root node, one selected class is evaluated against the rest by the optimal model. Then the UDT proceeds to the next level by eliminating the selected class

from the previous level of the decision tree. UDT terminates when it returns an output pattern at a level of the decision node.

3. PREVIOUS WORK

In [6], the authors used the Harris affine region detector to identify the interest points in the images which are then described by SIFT descriptors. A visual codebook was constructed by clustering the extracted features using K-means method. Images are then described by histograms over the learnt codebook. K-means were repeated several times over a selected size of K and different sets of initial cluster centres. The reported results were the clusters that gave them the lowest empirical risk in classification. The size of the codebook used in reporting the results is 1000. The authors compared Naive Bayes and SVM classifiers in the learning task and found that the OVA SVM with linear kernel gives a significantly (i.e. 13%) better performance. The proposed framework was mainly evaluated on their 'in-house' database that is currently known as 'Xerox7' image set containing 1,776 images in seven object categories. The overall error rate of the classification is 15% using SVMs. It has been reported that RAC approach in [26] performs slightly better than the method in [6] but was achieved in a tiny fraction of computation time.

In [14], the authors proposed a mean-shift based clustering approach to construct codebooks in an under sampling framework. The authors sub-sample patches randomly from the feature set and allocate a new cluster centroid for a fixed-radius hypersphere by running a mean-shift estimator [4] on the subset. The mean-shift procedure is achieved by successively computing the mean-shift vector of the sample keypoints and translating a Gaussian kernel on them. In the next stage, visual descriptors that fall within the cluster are filtered out. This process is continued by monitoring the informativeness of the clusters or until a desired number of clusters is achieved. The features used in their experiments are the gray level patches sampled densely from multi-scale pyramids with ten layers. The size of the codebook was 2,500. The proposed method was evaluated on three datasets: Side views of cars from [28], Xerox7 image dataset [6] and the ETH-80 dataset [17]. Naive Bayes and linear SVM classifiers were compared in all their experiments. The authors' mean-shift based clustering method is computationally intensive in determining the cluster centroid by mean-shift iterations at each of the sub samples. The convergence of such a recursive mean-shift procedure greatly depends on the nearest stationary point of the underlying density function and its utility in detecting the modes of the density. In contrast, the RAC approach pursued in [26] has a single threshold that takes only one-pass through the entire data, making it computationally efficient.

In [34], the authors optimized codebooks by hierarchically merging visual words in a pair-wise manner using the information bottleneck principle [1] from an initially constructed large codebook. Training images were convolved with different filter-banks made of Gaussians and Gabor kernels. The resulting filter responses were clustered by the K-means method with a large value of K in the order of thousands. Mahalanobis distance between features is used during the clustering step. The learnt cluster centres and their associated covariances define a universal visual codebook. Classification results were obtained on photographs acquired by the authors, images from the web and a subset of 587 images in total that were selected from the PASCAL VOC challenge 2005 dataset. The initial codebook construction of this method which is based on hierarchically merging visual words in a pair-wise manner is extremely faster than K-means clustering on large number of features. However, if two distinct visual words are initially grouped in the same cluster, they cannot be separated later. Also the vocabulary is tailored according to the categories under consideration, but it would require fully retraining the framework on the arrival of new object categories.

In [22], local features are found by extracting edges with a multi-scale Canny edge detector with Laplacian-based automatic scale selection. For every feature, a geometry term gets determined, coding the distance and relative angle of the object centre to the interest point, according to the dominant gradient orientation and the scale of the interest point. These regions are then described with SIFT features that are reduced to 40-dimension via PCA. The visual codebook is constructed by means of a hierarchical K-means clustering. Given a test image, the features were

extracted and a tree structure is built using the hierarchical K-means clustering method in order to compare with the learnt model tree. Classification is done in a Bayesian manner computing the likelihood ratio. Experiments were performed on a five class problem taken from the PASCAL VOC 2005 image dataset.

In [33], a novel method is proposed for constructing a compact visual codebook using a sparse reconstruction technique. Initially a large codebook is generated by K-means method. Then it is reformulated in a sparse manner, and weight of each word in the old visual codebook is learnt from the sparse representation of the entire action features. Finally, a new visual codebook is generated through six different algorithms. These algorithms are mainly based on L_0 and L_1 distances. The authors approach has been tested on the Weizmann action database [10]. As a result, the obtained codebook which is half the size of the old visual codebook has the same performance as the one built by K-means.

In [24], images are characterised by using a set of category-specific histograms generated one per object category. Each histogram describes whether the content can be best modelled by a universal codebook or by its corresponding category-specific codebook. Category-specific codebooks are obtained by adapting the universal codebook using the class training data and a form of Bayesian adaptation based on the maximum a posteriori criterion. The maximum number of Gaussians in the universal codebook was set to 2048. An image is then characterised by a set of histograms called bipartite as they can be split into two equal parts. Each part describes how well one codebook accounts for an image compared to the other codebook. Local patches were described by SIFT and colour features. PCA was applied to reduce the dimensionality of SIFT from 128 to 50, and the RGB colour channels from 96 to 50. Evaluations were performed on their own in-house database containing 19 classes of object categories and scenes, and the PASCAL VOC 2006 dataset. Classification was performed using linear SVMs and a logistic regression with a Laplacian prior. However, if two visual object classes are visually close, there is no guarantee that a distinctive visual word will be obtained. The process that generates bipartite histograms is also computationally expensive.

4. EMPIRICAL EVALUATION

4.1 Datasets

We evaluate the combination of descriptors and codebook construction techniques on three benchmark datasets: PASCAL visual object classes challenge 2007 [8], UIUC texture [16], and MPEG-7 Part-B silhouette image [12] datasets.

4.1.1 PASCAL VOC Challenge 2007

The dataset includes data with all possible viewing conditions such as different viewpoints, scales, illumination conditions, occluded/truncated and poor quality. The goal of this challenge is to recognize objects from a number of visual object classes in realistic scenes (i.e. not pre-segmented objects).

There are twenty object classes: person, bird, cat, cow, dog, horse, sheep, aeroplane, bicycle, boat, bus, car, motorbike, train, bottle, chair, dining table, potted plant, sofa, and TV/monitor. The database contains a total of 9,963 annotated images. The dataset has been split nearly into 50 percent for training/validation and 50 percent for testing dataset. Figure1 shows the object categories that are in the PASCAL VOC 2007 dataset.

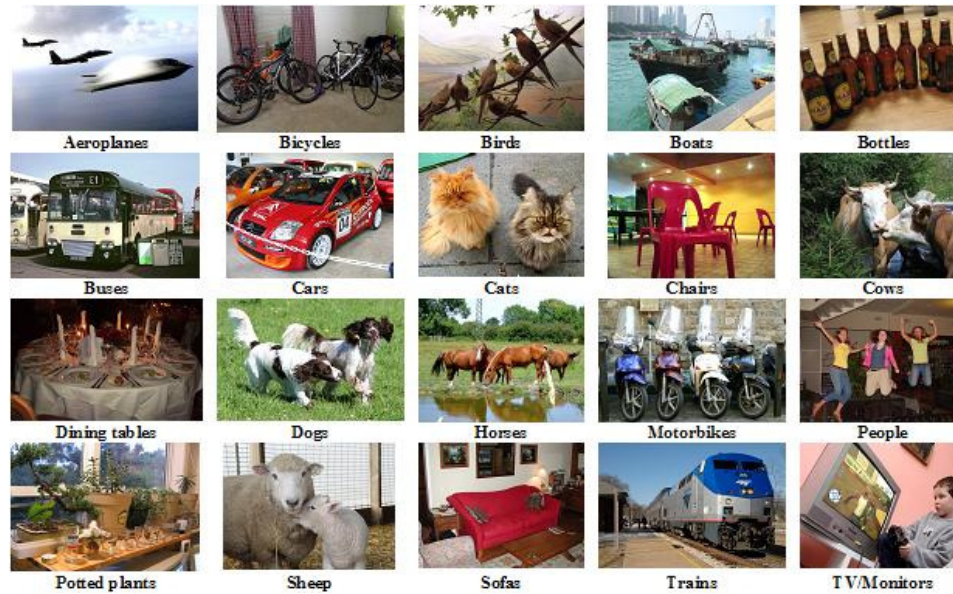


FIGURE 1: One Image from Each of the Object Categories In PASCAL VOC Challenge 2007 Image Dataset.

4.1.2 UIUC Texture Dataset

The dataset contains 25 texture classes with 40 images per class. Each of the images is of size 640×480 pixels. This dataset has surfaces whose texture is mainly due to albedo variations (e.g. wood and marble), 3D shape (e.g. gravel and fur), as well as a mixture of both (e.g. carpet and brick). It also has significant viewpoint changes, uncontrolled illumination, arbitrary rotations, and scale differences within each class. Figure 2 shows some of the example images of the UIUCTex dataset.

4.1.3 Silhouette Images

The MPEG-7 Part B silhouette database is a popular database for shape matching evaluation consisting of 70 shape categories, where each category is represented by 20 different images with high intra-class variability. The shapes are defined by a binary mask outlining the objects. Figure 3 shows some example images of the dataset.

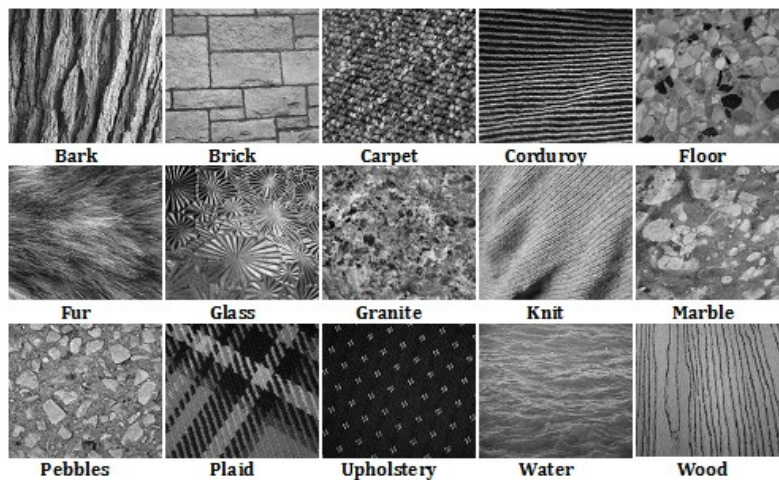


FIGURE 2: Example Images of the UIUCTex Dataset.

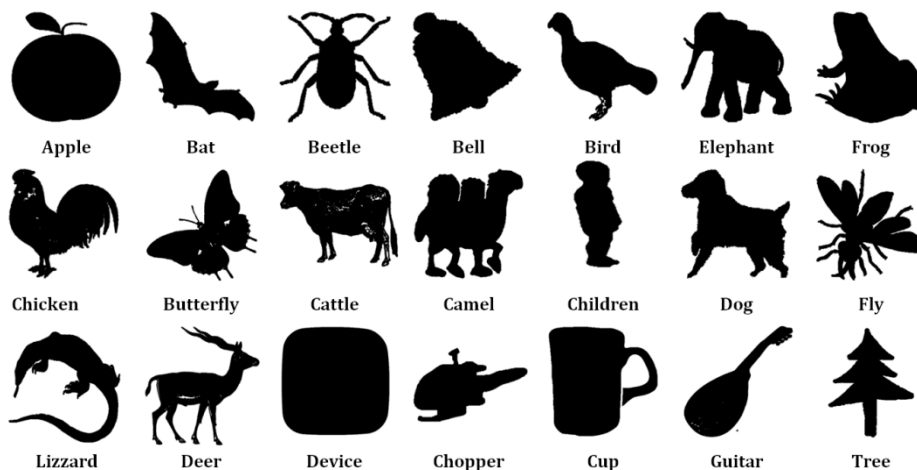


FIGURE 3: Example Images of the MPEG-7 Silhouette Dataset.

4.2 Vocabulary Construction

Popular approaches in the literature of codebook models [6, 9, 19, 31] have used K-means method with $K=1000$ to construct a codebook which shown better performance. For this reason when comparing the RAC and fast-RNN with K-means, we maintain the codebook size to be 1000. Moreover, K-means method was run three times with the same number of desired representative vectors and different sets of initial cluster centres.

The hyperparameter r of RAC is set to 0.8 in PASCAL VOC 2007 and 0.89 in silhouette and UIUCTex datasets. The choice of the radius r has the same set of difficulties associated with the choice of K in K-means. In [26], the approach to setting r is to take a small sample of the data, compute all pairwise distances between these samples and set the threshold, so that an approximate target codebook size is achieved.

When building NN chains in fast-RNN algorithm, we follow the authors [20] experimental setup in finding all the points that lie within a slice of the d -dimensional space of width 2ε centred at a query point instead of building a hypercube. The ε is set to 0.05.

4.3 Classification

In classification, we have used the OVA-based linear SVMs. The regularization parameter C was tuned with a range of values $[2^{-2}, 2^{-1}, \dots, 2^{11}, 2^{12}]$.

4.4 Evaluation Criterion

Average precision is used as performance evaluator for the binary object recognition tasks which has been widely used in recent PASCAL VOC challenges. Average precision is a single-valued measure that is proportional to the area under a precision-recall curve.

$$\text{Recall} = \frac{TP}{TP + FN} \quad \text{True positive rate} = \frac{TP}{TP + FN}$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{False positive rate} = \frac{FP}{FP + TN}$$

where TP, FP, TN and FN are true positive, false positive, true negative, and false negative respectively.

Receiver operating characteristics (ROC) curve shows how the number of correctly classified positive examples varies with the number of incorrectly classified negative examples. For multi-class classification, we report the classification rate as follows:

$$\text{rate} = \frac{\text{Number of correctly classified images}}{\text{Total number of testing images}} \times 100\%$$

4.5 Testing Results

Experiments in this work were mainly carried out to compare and contrast the RAC technique with traditional K-means method and the speeded-up RNN clustering algorithm in terms of classification rate, compactness of codebook, and time for constructing codebook. We have tested those algorithms on three benchmark datasets. Furthermore, we improve the standard RAC to yield more compact codebook.

For the PASCAL VOC 2007 dataset, we extracted features within the provided bounding box information from the combination of training and validation (trainval) set. Truncated objects were also included when extracting features. For evaluation purpose we have selected ten binary classes from PASCAL VOC 2007 image set (in Table 1). The classification results are shown (in Tables 1 and 2) as means of average precision and standard deviation.

In the UIUCTex dataset, we used ten-fold cross-validation. The dataset was split into ten partitions each containing four images from each of the 25 classes. Ten rounds of training and testing were performed in which nine partitions were used for training and the remaining partition was used for testing. According to [27], $2500 \times \mathbb{R}^{128}$ SIFT and e-SURF keypoints were randomly selected from each image of the training set and were individually clustered by K-means, RAC and fast-RNN techniques in order to form a locally merged global codebook. The random selection of a subset of SIFT and e-SURF descriptors was due to our previous experience of the prohibitive memory requirement by the traditional K-means in such a large number of keypoints (approx. 5000 keypoints per image). The K-means method in constructing 40 clusters per class was fed with a total of $36 \times 2500 \times \mathbb{R}^{128}$ SIFT and e-SURF descriptors. The resulting histogram of an image was of size 1000 in K-means method. The classification results are shown (in Tables 1 and 2) as means of average precision and standard deviation, over the ten runs.

In the Silhouette dataset, we used two-fold cross-validation. The Silhouette dataset was split into two partitions each containing ten images from each of the 25 classes. Two rounds of training and testing are performed in which one partition is used for training and the other was used for testing. We report the mean classification rate in Tables 1 and 2, together with the standard deviation, over the two runs.

All of our experiments were implemented in Matlab and executed on a desktop computer with an Intel Core 2 running at 2.4 GHz and 8GB of RAM.

4.5.1 Classification Rate

Table 1 details the classification rate of three independent runs of the proposed experiment using K-means, RAC and fast-RNN with SIFT descriptors respectively, whereas Table 2 details the rate for using SURF descriptors. Based on our testing results, when using SIFT descriptors, K-means performs better in three binary tasks, whereas RAC performs better in six binary tasks of the PASCAL VOC 2007 imageset. Fast-RNN performs better in only one task which can be observed in Table 1.

Dataset		SIFT		
		K-means	RAC	fast-RNN
PASCAL VOC 2007	Bird vs Aeroplane	81.86 \pm 1.27	82.10 \pm 0.80	75.29 \pm 1.24
	Aeroplane vs Horse	91.41 \pm 2.40	87.82 \pm 0.83	88.04 \pm 0.09
	Bicycle vs Motorbike	77.40 \pm 0.04	76.10 \pm 0.60	74.53 \pm 0.99
	Bus vs Train	79.80 \pm 0.55	76.00 \pm 0.60	74.31 \pm 0.69
	Dog vs Cat	68.22 \pm 1.51	70.73 \pm 0.65	71.10 \pm 1.40
	Cow vs Sheep	67.77 \pm 1.20	71.10 \pm 1.40	70.89 \pm 0.78
	Pottedplant vs Dining table	72.97 \pm 0.53	76.00 \pm 1.19	74.00 \pm 1.60
	Bottle vs Potted plant	59.69 \pm 2.15	67.72 \pm 0.31	62.54 \pm 1.30
	Boat vs TV/monitor	74.85 \pm 1.20	81.20 \pm 1.80	79.91 \pm 1.10
	Aeroplane vs Boat	80.00 \pm 0.90	87.80 \pm 0.80	80.16 \pm 0.81
UIUCTex		98.77 \pm 0.97	98.10 \pm 0.99	95.12 \pm 0.65
MPEG 7 Part B (Silhouette)		77.60 \pm 2.26	77.20 \pm 3.39	75.19 \pm 0.40

TABLE 1: Classification Rate as Mean Average Precision With Standard Deviation When Using SIFT Descriptors.

In the case of SURF descriptors, K-means performs better in eight binary tasks, whereas RAC performs better in three binary tasks of the PASCAL VOC 2007 imageset and fast-RNN performs better in only one task which can be observed in Table 2.

We carried out F-tests (one-way ANOVA) to compare the classification rates of K-means, RAC and fast-RNN techniques when applied on the PASCAL VOC 2007 imageset using SIFT and SURF descriptors. We may conclude that those three techniques are equally comparable in classification rates for the SIFT and SURF descriptors with the p-values 0.71 and 0.94, respectively.

Even though K-means slightly outperforms RAC in classification rate when applied on UIUCTex and silhouette image classification tasks, the negligible increase in performance was achieved at a huge computational time which can be observed in Table 3. Fast-RNN not only shows less classification rate compared to RAC but also higher codebook construction time to RAC. Even

though fast-RNN has some demerits to RAC it still constructs more compact codebooks than the other approaches.

Dataset		e-SURF		
		K-means	RAC	fast-RNN
PASCAL VOC 2007	Bird vs Aeroplane	86.22 ± 0.21	86.48 ± 1.02	85.36 ± 0.21
	Aeroplane vs Horse	91.39 ± 0.28	85.36 ± 0.85	85.58 ± 0.42
	Bicycle vs Motorbike	81.80 ± 0.22	79.47 ± 0.32	75.98 ± 0.79
	Bus vs Train	69.74 ± 1.66	79.61 ± 2.25	76.13 ± 1.38
	Dog vs Cat	70.31 ± 0.92	68.83 ± 0.80	64.70 ± 0.12
	Cow vs Sheep	67.26 ± 0.97	65.06 ± 1.65	66.24 ± 1.81
	Pottedplant vs Dining table	74.06 ± 0.31	72.18 ± 1.53	70.32 ± 0.41
	Bottle vs Potted plant	60.92 ± 0.67	61.04 ± 1.65	60.04 ± 2.41
	Boat vs TV/monitor	56.82 ± 1.51	71.96 ± 3.12	73.45 ± 1.37
	Aeroplane vs Boat	75.27 ± 1.80	71.20 ± 2.04	69.13 ± 1.30
UIUCTex		97.88 ± 0.92	97.05 ± 1.47	94.52 ± 0.03
MPEG 7 Part B (Silhouette)		75.03 ± 0.05	74.00 ± 0.56	70.71 ± 0.23

TABLE 2: Classification Rate as Mean Average Precision With Standard Deviation Using SURF Descriptors.

The classification performance for K-means, RAC and fast-RNN clustering techniques with SIFT and e-SURF descriptors is also represented using ROC curves. Figure 4 illustrates some example of the ROC curves which are randomly selected from our results. The ROC curve details how the number of correctly classified positive examples varies with the number of incorrectly classified negative examples. In these figures, Fig. 4(a), (d) and (e) lie in the upper-left-hand corner representing a good classification result and SIFT + RAC performs better than other combinations.

In addition to this, we carried out limited experiments to see the influence of the order of presentation of data to RAC. Based on the limited experiments, we found that RAC is slightly sensitive to the order of presentation of data, similar to the random initial selection of cluster centres in the K-means method.

Dataset		SIFT			e-SURF		
		K-means	RAC	fast-RNN	K-means	RAC	fast-RNN
PASCAL VOC 2007	Aeroplane	34712	495	34735	1845	12	1664.8
	Bird	89538	780	74695	6540	12	2117
	Bottle	17824	66	14097	536	2	124
	Boat	12652	123	11202	2670	6	709
	Bus	27657	489	27657	5927	12	2092
	Bicycle	34906	270	31140	10269	14	2951
	Cat	248810	334	173563	3809	14	2588
	Cow	27999	116	22440	1800	5	583
	Dining table	35544	252	33263	5338	10	1490
	Dog	271695	452	265264	7380	21	5027
	Horse	86639	793	79988	5587	18	3622
	Motorbike	80002	619	83496	10172	23	6236
	Potted plant	32137	195	27274	8398	9	1481
	Sheep	10489	50	9256	900	2	264
	Train	78621	969	75324	16497	22	5629
TV/Monitor	7608	69	24023	466	2	172	
UIUCTex	70585	10050	40100	60141	79	57125	
MPEG 7 Part B (Silhouette)	11597	20	10200	6135	4	5135	

TABLE 3: Estimated Time for Codebook Construction Using SIFT and SURF Descriptor.

4.5.2 Compactness of Codebook

A compact codebook may be achieved directly by reducing the codebook size or by carefully selecting the codebook elements. RAC sizes are slightly greater than others. Table 4 indicates average size of the codebooks by using K-means, RAC and fast-RNN with SIFT and e-SURF descriptors in that order. As shown in Table 4 the size of the codebook obtained by e-SURF with fast-RNN yields codebooks of an average size of 131 while SIFT with fast-RNN produces codebooks of an average of 280.

4.5.3 Time For Constructing Codebook

The construction of a visual codebook is often performed from thousands of images and each image contains hundreds or even one thousands of patch based interest points described in a higher dimensional space, in order to capture sufficient information for efficient classification.

The time complexity of the traditional K-means method is $O(NdKm)$, whereas for the fast-RNN technique it is $O(N \log N)$ where, N -number of descriptors, d -dimensionality of features, K -number of desired clusters and m -number of iterations of the expectation-maximization (EM) algorithm. The time complexity of RAC depends on the size of the candidate cluster set C , i.e., it compares each newly seen pattern to all existing clusters. Thus, RAC has far lower computational cost than K-means and fast-RNN clustering techniques.

In our experiments codebook construction time is quantified in two cases: SIFT and e-SURF and have been tested with those three algorithms. In all cases, e-SURF is faster than SIFT based on the outcomes of the results reported in Table 4 and as found by [2]. When commenting on the performance of descriptors, we conclude that SIFT performs better than SURF descriptors, which has also been proved by [13].

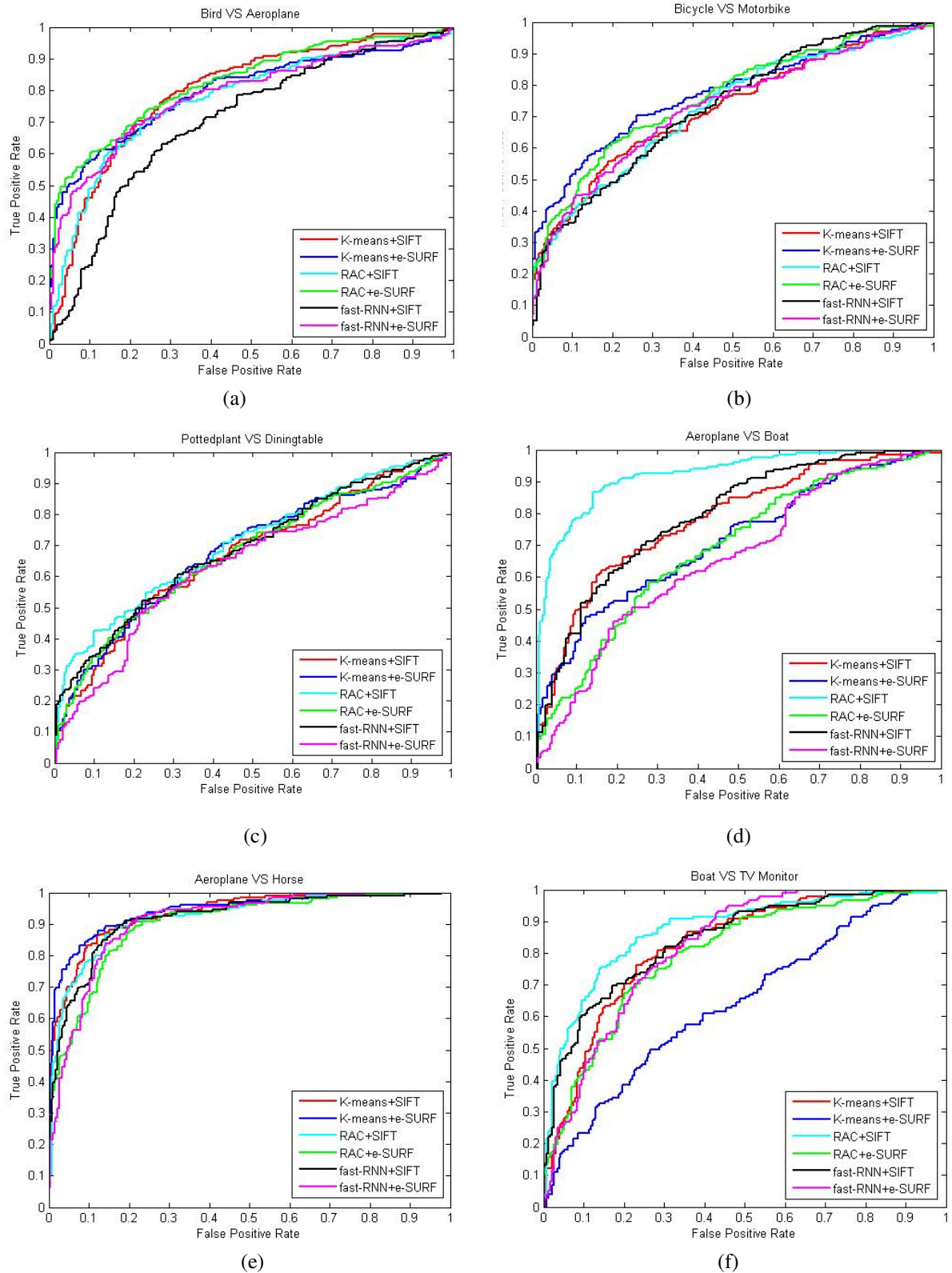


FIGURE 4: Classification of the Test Set in PASCAL VOC 2007 Represented Using the ROC Curves for Binary Classes.

Since the total number of interest points detected by e-SURF is much less than that of SIFT descriptors as shown in Figure 5, K-means contradicts with fast-RNN in both cases. Based on our computational time to construct a codebook, RAC seems particularly suitable for the codebook construction due to its speed than other methods compared in this paper.

Objects		SIFT			e-SURF		
		K-means	RAC	fast-RNN	K-means	RAC	fast-RNN
PASCAL VOC 2007	Aeroplane	1000	1325	342	1000	526	140
	Bird		1165	291		461	120
	Bottle		945	220		321	97
	Boat		1032	251		383	109
	Bus		1233	286		492	135
	Bicycle		1148	265		423	105
	Cat		1167	260		506	260
	Cow		932	210		369	100
	Dining table		1219	302		452	111
	Dog		1305	268		540	120
	Horse		1297	303		541	119
	Motorbike		1399	316		510	120
	Potted plant		1074	245		376	103
	Sheep		711	174		277	75
	Train		1410	347		531	147
	TV Monitor		943	248		358	129
	UIUCTex			1524		375	
MPEG 7 Part B		1242	340		1370	125	

TABLE 4: Average Size of the Codebook Using SIFT and e-SURF Descriptors.

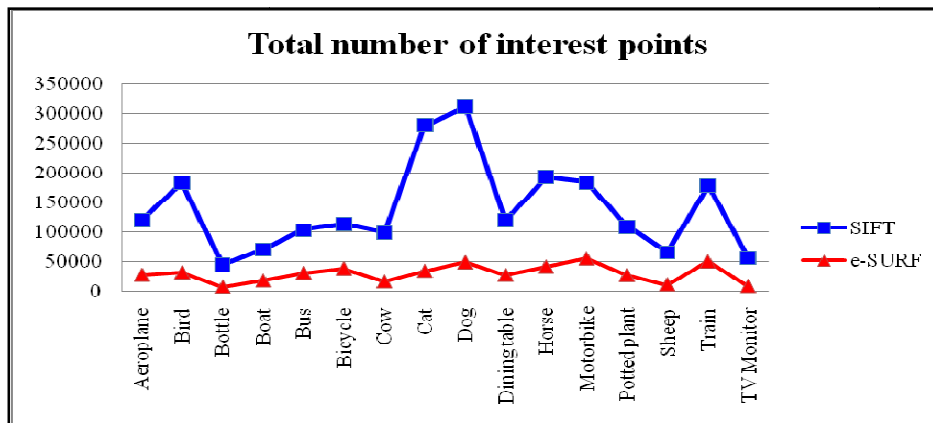


FIGURE 5: Total Number of Interest Points Detected in PASCAL VOC 2007 Dataset.

4.6 Further Improvement

Based on our testing results, we can easily conclude that RAC outperforms to the traditional K-means and fast-RNN but the codebook sizes constructed by RAC are slightly greater than others. To overcome this issue, we have reduced the codebook sizes of RAC by removing hyperspheres that contain less number of members. That is if any visual word in codebook contains very small number of local descriptors (i.e., $n \leq 10$) that hypersphere is removed from the codebook. We refer this method as compact RAC.

Table 5 details the classification rate of three independent runs of the proposed experiment using Compact-RAC that is compared with the standard RAC. Recognition results are shown as means of average precision with standard deviation.

Dataset		SIFT		e-SURF	
		RAC	Compact RAC	RAC	Compact RAC
PASCAL VOC 2007	Bird vs Aeroplane	82.10 ± 0.80	81.12 ± 0.42	86.48 ± 1.02	87.20 ± 0.50
	Aeroplane vs Horse	87.82 ± 0.83	87.00 ± 0.04	85.36 ± 0.85	85.18 ± 1.79
	Bicycle vs Motorbike	76.10 ± 0.60	77.37 ± 0.95	79.47 ± 0.32	79.27 ± 2.12
	Bus vs Train	76.00 ± 0.60	78.91 ± 0.16	79.61 ± 2.25	79.61 ± 2.36
	Dog vs Cat	70.73 ± 0.65	69.65 ± 1.42	68.83 ± 0.80	68.28 ± 2.04
	Cow vs Sheep	71.10 ± 1.40	71.33 ± 0.72	65.06 ± 1.65	65.27 ± 0.21
	Potted plant vs Dining table	76.00 ± 1.19	75.62 ± 0.97	72.18 ± 1.53	71.48 ± 0.99
	Bottle vs Potted plant	67.72 ± 0.31	63.37 ± 2.13	61.04 ± 1.65	60.64 ± 0.88
	Boat vs TV/monitor	81.20 ± 1.80	76.37 ± 3.70	71.96 ± 3.12	74.03 ± 2.99
	Aeroplane vs Boat	87.80 ± 0.80	73.19 ± 2.80	71.20 ± 2.04	71.34 ± 1.70
	UIUCTex	98.10 ± 0.99	97.04 ± 0.96	97.05 ± 1.47	96.00 ± 2.67
MPEG 7 Part B	77.20 ± 3.39	73.80 ± 0.84	74.00 ± 0.56	74.00 ± 1.13	

TABLE 5: RAC vs Compact-RAC: Recognition Results as Mean Average Precision With Standard Deviation.

Figure 6 shows the average size of the codebook by using Compact-RAC. Based on the results of Table 5 and Figure 6, recognition results using Compact-RAC are reduced or increased by one percentage compared with RAC. But the size of the codebook is significantly compact when using Compact-RAC.

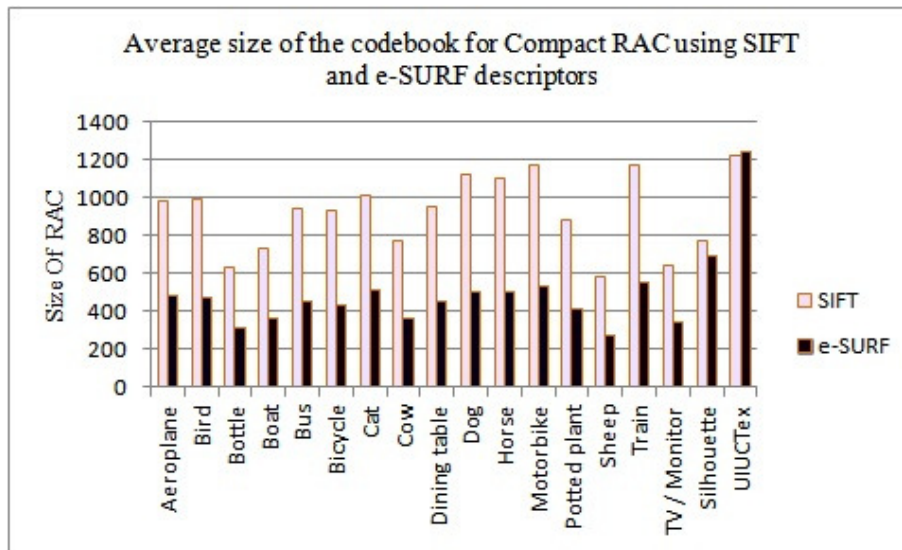


FIGURE 6: Average Size of the Codebook for *Compact-RAC* Using SIFT and e-SURF Descriptors.

5. DISCUSSION AND CONCLUSION

This paper mainly focuses on the evaluation of codebook construction techniques: K-means, fast-RNN and RAC algorithms that are used in patch-based visual object recognition. Our work suggests the need for an online codebook construction technique such as RAC in constructing discriminant and compact codebooks at a drastically reduced time. We also compare the well known patch-based descriptors SIFT and SURF in classification rate. In practice, the construction of a visual codebook is often performed from thousands of images and each image on average contains hundreds or even one thousand of patch-based interest points described in a higher dimensional space of one hundred, in order to capture sufficient information for efficient classification. While clustering algorithms and their performance characteristics have been studied extensively over recent years, a major bottleneck lies in handling the massive scale of the datasets. The Caltech and PASCAL VOC Challenge image datasets are becoming gold standard for measuring recognition performance in recent vision papers but the size of these datasets nearly grows exponentially over the years. The size of the codebooks that have been used in the literature ranges from 102 to 104, resulting in very high-dimensional histograms. A larger size of codebook increases the computational needs in terms of memory usage, storage requirements, and the computational time to construct the codebook and to train a classifier. On the other hand, a smaller size of codebook lacks good representation of true distribution of features. Thus, the choice of the size of a codebook should be balanced between the recognition rate and computational needs.

Based on our testing results the fast-RNN method constructs codebooks that are more compact but shows less classification rate than K-means and RAC algorithms. Even though K-means slightly performs better than RAC, it requires more computational resources such as memory, disk space and huge time in constructing a codebook. In contrast, RAC sequentially processes large number of descriptors in a higher dimensional feature space to constructing compact codebooks for reliable object categorisation performance at drastically reduced computational needs.

SIFT and SURF descriptors are invariant to common image transformations, such as scale changes, image rotation, and small changes in illumination. These descriptors are also invariant to translations as from the use of local features. SURF features can be extracted faster than SIFT using the gain of integral images and yield a lower dimensional feature descriptor resulting in faster matching and less storage space. SIFT descriptors have been found highly distinctive in performance evaluation [23] which has also been proved in our experiments.

The RAC discussed in this paper seems particularly suitable for the codebook construction owing to its simplicity, speed and performance and its one-pass strategy that requires relatively little memory. RAC is also fundamentally different from traditional approaches where it is not the density of detected patches one needs to retain in the codebook but the coverage across the feature space. We have demonstrated RAC with a computationally much simplified algorithm compared to what others have achieved.

6. REFERENCES

- [1]. 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.
- [2]. H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "SURF: Speeded Up Robust Features", In *Computer Vision and Image Understanding*, Vol. 110, pp. 346–359, 2008.
- [3]. C. J. Burgues, "A Tutorial on Support Vector Machines for Pattern Recognition", *Knowledge Discovery and Data Mining*, Vol. 2, pp. 121–167, 1998.
- [4]. D. Comaniciu and P. Meer, "Mean Shift: A Robust Approach toward Feature Space Analysis", In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, pp. 603–619, 2002.
- [5]. N. Cristianini and J. Shawe-Taylor, "An introduction to Support Vector Machines and other Kernel-based Learning Methods", Cambridge University Press, 2000.
- [6]. G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual Categorization with Bags of Keypoints", In *Workshop on Statistical Learning in Computer Vision, ECCV*, pp. 1–22, 2004.
- [7]. R. Debnath, N. Takahide, and H. Takahashi, "A Decision based One-Against-One Method for Multi-class Support Vector Machine", *Pattern Analysis Application*, Vol. 7, pp. 164–175, 2004.
- [8]. M. Everingham, L. Van-Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results".
- [9]. L. Fei-Fei, and P. Perona, "A Bayesian Hierarchical Model for Learning Natural Scene Categories", In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2, pp. 524–531, 2005.
- [10]. L. Gorelick, M. Blank, E. Shechtman, M. Irani, and R. Basri, "Actions as Space-Time Shapes", In *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, Vol. 29, pp. 2247–2253, 2007.
- [11]. G. Griffi, A. Holub and P. Perona, "The Caltech-256 Object Category Dataset", Technical Report, California Institute of Technology, 2007.
- [12]. L. Jan Latecki, R. Lakamper and U. Eckhardt, "Shape Descriptors for Non-rigid Shapes with a Single Closed Contour", In *proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 424–429, 2000.
- [13]. L. Juan and O. Gwun, "A Comparison of SIFT, PCA-SIFT and SURF", In *International Journal of Image Processing*, Vol. 3, pp. 143–152, 2009.
- [14]. F. Jurie and B. Triggs, "Creating Efficient Codebooks for Visual Recognition", In *Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)*, Vol. 01, pp. 604 – 610, 2005.
- [15]. Y. Ke and R. Sukthankar, "PCA-SIFT: A More Distinctive Representation for Local Image Descriptors", In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 511–517, 2004.

- [16]. S. Lazebnik, C. Schmid, and J. Ponce, "A Sparse Texture Representation using Local Affine Regions", In IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 27, pp. 1265–1278, 2005.
- [17]. B. Leibe and B. Schiele, "Interleaved Object Categorization and Segmentation", In Proceedings of the British Machine Vision Conference (BMVC'03), pp. 759–768, 2003.
- [18]. D. Li, L. Yang, X. Hua and H. Zhan, "Large-scale Robust Visual Codebook Construction", ACM International Conference on Multimedia (ACM-MM), pp. 1183–1186, 2010.
- [19]. T. Li, T. Mei and I.-S. Kweon, "Learning Optimal Compact Codebook for Efficient Object Categorization", In IEEE workshop on Applications of Computer Vision, pp. 1–6, 2008.
- [20]. R. J. Lopez-Sastre, D. Onoro-Rubio, P. Gil-Jimenez, and S. Maldonado-Bascon, "Fast Reciprocal Nearest Neighbours Clustering", Signal Processing, Vol. 92, pp. 270–275, 2012.
- [21]. D. Lowe, "Distinctive Image Features from Scale-invariant Keypoints", International Journal of Computer Vision, Vol. 60 (2), pp. 91–110, 2004.
- [22]. K. Mikolajczyk, B. Leibe, and B. Schiele, "Multiple Object Class Detection with a Generative Model", In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06), Vol. 1, pp. 26–36, 2006.
- [23]. K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors", In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 27, pp. 1615–1630, 2005.
- [24]. F. Perronnin, "Universal and Adapted Vocabularies for Generic Visual Categorization", In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, pp. 1243–1256, 2008.
- [25]. J. C. Platt, N. Cristianini, and J. Shawe-Taylor, "Large Margin DAGs for Multiclass Classification", In Advances in Neural Information Processing Systems (NIPS'00), Vol. 12, pp. 547–553, 2000.
- [26]. A. Ramanan and M. Niranjan, "A One-Pass Resource-Allocating Codebook for Patch-based Visual Object Recognition", In Proceedings of the IEEE International Workshop on Machine Learning for Signal Processing (MLSP'10), pp. 35 – 40, 2010.
- [27]. A. Ramanan, R. Paheerathy, and M. Niranjan, "Speeding Up Multi-Class Texture Classification by One-Pass Vocabulary Design and Decision Tree", In Proceedings of the Sixth IEEE International Conference on Industrial and Information Systems (ICIIS'11), pp. 255-260, 2011.
- [28]. A. Ramanan, S. Suppharangsarn, and M. Niranjan. "Unbalanced Decision Trees for Multi-class Classification", In Proceedings of the IEEE International Conference on Industrial and Information Systems (ICIIS'07), pp. 291–294, 2007.
- [29]. R. Rifkin and A. Klautau, "In Defense of One-vs-All Classification", Journal of Machine Learning Research, Vol. 5, pp. 101–141, 2004.
- [30]. E. B. Sudderth, A. Torralba, W. T. Freeman and A. S. Willsky, "Describing Visual Scenes using Transformed Objects and Parts", International Journal of Computer Vision, Vol. 77, pp. 291–330, 2008.
- [31]. N. Tishby, F. C. Pereira, and W. Bialek, "The Information Bottleneck Method", In the 37th Annual Allerton Conference on Communication, Control and Computing, pp. 368–377, 1999.
- [32]. Q. Wei, X. Zhang, Y. Kong, W. Hu and H. Ling, "Compact Visual Codebook for Action Recognition", In International Conference on Image Processing (ICIP), pp. 3805–3808, 2010.
- [33]. J. Winn, A. Criminisi, and T. Minka, "Object Categorization by Learned Universal Visual Dictionary", In Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 1800–1807, 2005.

- [34]. C. Zhang, J. Liu, Y. Ouyang, Q. Tian, H. Lu, S. Ma, "Category Sensitive Codebook Construction for Object Category Recognition", In International Conference on Image Processing (ICIP), pp. 329–332, 2009.
- [35]. H. Zhang, A. Berg, M. Maire, and J. Malik. "SVM-KNN Discriminative Nearest Neighbour Classification for Visual Category Recognition", In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition (CVPR), pp. 2126-2136, 2006.

Image Processing Technique for Brain Abnormality Detection

Ashraf Anwar

prof1aa@yahoo.com

*Associate Professor of Computer Science,
Chair of Online Technologies & National Advisory Council Member
University of Atlanta
Atlanta, GA, USA*

Arsalan Iqbal

arsalan.iqbal@email.uofa.edu

*School of Computer and Information Science
University of Atlanta
Atlanta, Georgia, USA*

Abstract

Medical imaging is expensive and very much sophisticated because of proprietary software and expert personalities. This paper introduces an inexpensive, user friendly general-purpose image processing tool and visualization program specifically designed in MATLAB to detect much of the brain disorders as early as possible. The application provides clinical and quantitative analysis of medical images. Minute structural difference of brain gradually results in major disorders such as schizophrenia, Epilepsy, inherited speech and language disorder, Alzheimer's dementia etc. Here the main focusing is given to diagnose the disease related to the brain and its psychic nature (Alzheimer's disease).

Keywords: Alzheimer, Bilinear Interpolation, Brain Imaging, Cerebral Cortex, Image Registration, Neuroinformatics.

1. INTRODUCTION

The human brain [1] is the center of the human nervous system and is the most complex organ in any creature on earth. Any abnormality in brain leads to the total collapse of entire vital functions of the body [2]. This paper introduces a concept of simple user friendly GUI application to process an image of brain and analyze its morphological abnormalities.

2. PROBLEM DEFINITION

Typical abnormality of brain is the shrinking of its cerebral cortex (as shown in *Figure 1*). The spaces in the folds of the brain (the sulci) are grossly enlarged. This type of abnormality leads to Alzheimer's disease. Here the major objective is to identify the cerebral cortex region of brain and measure the cavity area in the sulci (*Figure 2*) using software.

Currently, Brain MRI images could be checked for abnormality using more recent techniques [3], [4], [5].

Goyal et al. [3] showed a technique for automatic detection of some types of brain abnormalities, along with techniques for tumor segmentation in MRI sequences. They presented an automated and clinically-tested method for detection of brain abnormalities and tumor-edema segmentation using MRI sequences. Their method follows a Radiologist's approach to the brain diagnosis using multiple MRI sequences instead of any prior models or training phases. Their procedure consists of the following steps:

- i. Pre-processing of the MRI sequences, T2, T1, and T1 post-contrast (enhanced) for size standardization, contrast equalization and division into active cells.
- ii. Identification of the T2 MRI sequence as normal or abnormal by exploiting the vertical symmetry of the brain.
- iii. Determination of the region of abnormality using its hyper-intense nature.
- iv. Separation of tumor from edema using the T1 and its post-contrast (enhanced) sequences.
- v. Estimation of the volume of tumor found and generation of an anatomical differential of the possible disorders.

Lashkari [4] introduced an automatic brain tumor detection method to increase the accuracy and yield, and decrease the diagnosis time. The goal in his work was to classify the tissues into two classes: normal and abnormal. *MR* images that were used in his work were images from normal and abnormal brain tissues. He tried to give clear description from brain tissues using *Zernike* Moments, Geometric Moment Invariants, energy, entropy, contrast and some other statistic features such as mean, median, variance, correlation between corresponding points, and values of maximum and minimum intensity. He used a feature selection method to reduce the feature space as well. His method used neural networks to do that classification. The purpose was to classify the brain tissues into normal and abnormal classes automatically; which saves the radiologist time, and increases accuracy and yield of diagnosis.

Reddy et al. [5] showed an improvement to fuzzy clustering means (FCM). They introduced an earlier spatial constraint into FCM algorithm, in which the spatial information is encoded through mutual influences of neighboring positions. To detect the abnormalities of Brain MRI images, they used a new spatial FCM, and compared the results with k-means and FCM techniques.



FIGURE 1:
Cerebral Cortex.



FIGURE2:
Sulci.

3. PROJECT GOALS

A. Alzheimer- The Brain Disorder: Just like the rest of our bodies, our brains change as we age. Most of us notice some slowed thinking and occasional problems remembering certain things. However, serious memory loss, confusion and other major changes in the way our minds work are not a normal part of aging [6]. These may be the signs of brain cells failure [7].

The brain has 100 billion nerve cells (neurons). Each nerve cell communicates with many others to form network. Nerve cell networks have special jobs. Some are involved in thinking, learning and remembering. Others help us see, hear and smell. Still others tell our muscles when to move. In Alzheimer's disease, as in other types of dementia, increasing numbers of brain cells deteriorate and die [6]. *Figure 3* shows the gradual increase of Alzheimer's diseased people.

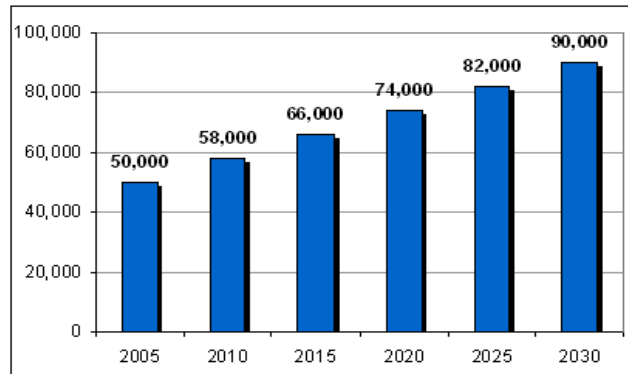


FIGURE 3: Alzheimer's Patients' Census.

B. Brain imaging methods: Functional imaging of electric brain activity requires specific models to transform the signals recorded at the surface of the human head into an image [8]. Two categories of model are available: 1) single-time-point and 2) spatio-temporal methods. The instantaneous methods rely only on few voltage differences measured at one sampling point. To create a spatial image from this limited information, they require strict assumptions that rarely conform to the underlying physiology. Spatio-temporal models create two kinds of images: first, a spatial image of discrete equivalent multiple dipoles or regional sources, and second, an image of source current waveforms that reflect the temporal dynamics of the brain activity in circumscribed areas. The accuracy of the spatial image is model dependent and limited, but it can be validated from the spatio-temporal data by the "regional source imaging" technique, introduced here. The source waveforms are linear combinations of the scalp waveforms, and thus, specific derivations which image local brain activities at a macroscopic level.

Brain source imaging of somatosensory evoked potentials revealed temporally overlapping activities from the brainstem, thalamus, and from multiple sources in the region of the contralateral somatosensory projection areas.

MRI (Magnetic Resonance Imaging): Magnetic Resonance Imaging (MRI), or nuclear magnetic resonance imaging (NMRI), is primarily a medical imaging technique most commonly used in radiology to visualize the internal structure and function of the body. MRI provides much greater contrast between the different soft tissues of the body than computed tomography (CT) does, making it especially useful in neurological (brain), musculoskeletal, cardiovascular, and oncological (cancer) imaging. Unlike CT, it uses no ionizing radiation, but uses a powerful magnetic field to align the nuclear magnetization of (usually) hydrogen atoms in water in the body. Radio frequency (RF) fields are used to systematically alter the alignment of this magnetization, causing the hydrogen nuclei to produce a rotating magnetic field detectable by the scanner. This signal can be manipulated by additional magnetic fields to build up enough information to construct an image of the body.

Magnetic Resonance Imaging is a relatively new technology. The first MR image was published in 1973 and the first study performed on a human took place on July 3, 1977. By comparison, the first human X-ray image was taken in 1895.

Magnetic Resonance Imaging was developed from knowledge gained in the study of nuclear magnetic resonance [8]. In its early years the technique was referred to as nuclear magnetic resonance imaging (NMRI). However, as the word nuclear was associated in the public mind with ionizing radiation exposure it is generally now referred to simply as MRI. Scientists still use the term NMRI when discussing non-medical devices operating on the same principles. The term Magnetic Resonance Tomography (MRT) is also sometimes used.

Functional MRI: Functional MRI or functional Magnetic Resonance Imaging (fMRI) is a type of specialized MRI scan. It measures the haemodynamic response related to neural activity in the brain or spinal cord of humans or other animals. It is one of the most recently developed forms of neuroimaging. Since the early 1990s, fMRI has come to dominate the brain mapping field due to its low invasiveness, lack of radiation exposure, and relatively wide availability.

Positron Emission Tomography (PET): Positron emission tomography (PET) is a nuclear medicine imaging technique which produces a three-dimensional image or picture of functional processes in the body. The system detects pairs of gamma rays emitted indirectly by a positron-emitting radionuclide (tracer), which is introduced into the body on a biologically active molecule. Images of tracer concentration in 3-dimensional space within the body are then reconstructed by computer analysis. In modern scanners, this reconstruction is often accomplished with the aid of a CT X-ray scan performed on the patient during the same session, in the same machine.

If the biologically active molecule chosen for PET is FDG, an analogue of glucose, the concentrations of tracer imaged then give tissue metabolic activity, in terms of regional glucose uptake. Although use of this tracer results in the most common type of PET scan, other tracer molecules are used in PET to image the tissue concentration of many other types of molecules of interest.

C. Image Registration: In this project, sets of data acquired by sampling the same scene or object at different times, or from different perspectives, will be in different coordinate systems. Image registration is the process of transforming the different sets of data into one coordinate system. [9] Registration is necessary in order to be able to compare or integrate the data obtained from different measurements.

Medical image registration (for data of the same patient taken at different points in time) often additionally involves elastic (also known as nonrigid) registration to cope with deformation of the subject (due to breathing, anatomical changes, and so forth). Nonrigid registration of medical images can also be used to register a patient's data to an anatomical atlas, such as the Talairach atlas for neuroimaging.

Image registration is the process of overlaying two or more images of the same scene taken at different times, from different viewpoints, and/or by different sensors. It geometrically aligns two images—the reference and sensed images. The present differences between images are introduced due to different imaging conditions. Image registration [10] is a crucial step in all image analysis tasks in which the final information is gained from the combination of various data sources like in image fusion, change detection, and multichannel image restoration. Typically, registration is required in remote sensing (multispectral classification, environmental monitoring, change detection, image mosaicing, weather forecasting, creating super-resolution images, integrating information into geographic information systems (GIS)), in medicine (combining computer tomography (CT) and NMR data to obtain more complete information about the patient, monitoring tumor growth, treatment verification, comparison of the patient's data with anatomical atlases), in cartography (map updating), and in computer vision (target localization, automatic quality control), to name a few. During the last decades, image acquisition devices have undergone rapid development and growing amount and diversity of obtained images invoked the research on automatic image registration.

D. Image registration methodology: Image registration, as it was mentioned above, is widely used in remote sensing, medical imaging, computer vision etc. In general, its applications can be divided into four main groups according to the manner of the image acquisition [9]:

1. Different viewpoints (multi-view analysis). Images of the same scene are acquired from different viewpoints. The aim is to gain larger a 2D view or a 3D representation of the scanned

scene. Examples of applications: Remote sensing—mosaicing of images of the surveyed area. Computer vision—shape recovery (shape from stereo).

2. Different times (multi-temporal analysis). Images of the same scene are acquired at different times, often on regular basis, and possibly under different conditions. The aim is to find and evaluate changes in the scene which appeared between the consecutive image acquisitions. Examples of applications: Remote sensing—monitoring of global land usage, landscape planning. Computer vision— automatic change detection for security monitoring, motion tracking. Medical imaging— monitoring of the healing therapy, monitoring of the tumor evolution.

3. Images of the same scene are acquired by different sensors. The aim is to integrate the information obtained from different source streams to gain more complex and detailed scene representation. Examples of applications: Remote sensing—fusion of information from sensors with different characteristics like panchromatic images, offering better spatial resolution, color/multispectral images with better spectral resolution, or radar images independent of cloud cover and solar illumination. Medical imaging—combination of sensors recording the anatomical body structure like magnetic resonance image (MRI), ultrasound or CT with sensors monitoring functional and metabolic body activities like positron emission tomography (PET), single photon emission computed tomography (SPECT) or magnetic resonance spectroscopy (MRS). Results can be applied, for instance, in radiotherapy and nuclear medicine.

4. Scene to model registration. Images of a scene and a model of the scene are registered. The model can be a computer representation of the scene, for instance maps or digital elevation models (DEM) in GIS, another scene with similar content (another patient), 'average' specimen, etc. The aim is to localize the acquired image in the scene/model and/or to compare them. Examples of applications: Remote sensing—registration of aerial or satellite data into maps or other GIS layers. Computer vision—target template matching with real-time images, automatic quality inspection. Medical imaging— comparison of the patient's image with digital anatomical atlases specimen classification.

Due to the diversity of images to be registered and due to various types of degradations it is impossible to design a universal method applicable to all registration tasks. Every method should take into account not only the assumed type of geometric deformation between the images but also radiometric deformations and noise corruption, required registration accuracy and application-dependent data characteristics. Nevertheless, the majority of the registration methods consist of the following four steps (closed-boundary regions, edges, contours, line intersections, corners, etc.) are manually or, preferably, automatically detected. For further processing, these features can be represented by their point representatives (centers of gravity, line endings, distinctive points), which are called control points (CPs) in the literature. Feature matching. In this step, the correspondence between the features detected in the sensed image and those detected in the reference image is established. Various feature descriptors and similarity measures along with spatial relationships among the features are used for that purpose. Transform model estimation. The type and parameters of the so-called mapping functions, aligning the sensed image with the reference image, are estimated. The parameters of the mapping functions are computed by means of the established feature correspondence. Image resampling and transformation. The sensed image is transformed by means of the mapping functions. Image values in non-integer coordinates are computed by the appropriate interpolation technique. The implementation of each registration step has its typical problems. First, we have to decide what kind of features is appropriate for the given task. The features should be distinctive objects, which are frequently spread over the images and which are easily detectable. Usually, the physical interpretability of the features is demanded. The detected feature sets in the reference and sensed images must have enough common elements, even in situations when the images do not cover exactly the same scene or when there are object occlusions or other unexpected changes. The detection methods should have good localization accuracy and should not be sensitive to the assumed image degradation. In an ideal case, the algorithm should be able to detect the same

features in all projections of the scene regardless of the particular image deformation. In the feature matching step, problems could be caused by an incorrect feature.

E. Mathematics of Image Registration: Given two images of the same object (possibly at different times, from different distances or angles or points of view, with distortion from a lens or other intervening medium, or using different imaging technologies), one would like to find a mapping that identifies points of one image with points in the other. An ideal mapping should have various features: in particular, it should be relatively smooth, and the data should fit the mapping well. (Criteria such as these can of course be made more precise using appropriate mathematical and statistical language.) Since an "ideal" mapping may be computationally infeasible (or may not even exist) the goal is to develop mathematical algorithms to find a mapping of a given specified type (e.g., definable by BÄ©zier transformations) and which is as close to ideal as various practical limitations admit.

F. Defining Project Goals: Fundamental problems in the analysis of functional and structural imaging data include data transport, boundary identification (including manual tracing, edge detection, and tissue segmentation), volume estimation, three-dimensional reconstruction and display, surface and volume rendering, shape analysis, and image overlay. These problems require that research investigators have access to suitable methods of image analysis, implemented on a set of software programs, in order to conduct neuroimaging research.

Different cost functions, different minimization methods, and various sampling, smoothing, and editing strategies were compared. Internal consistency measures were used to place limits on registration accuracy for MRI data, and absolute accuracy was measured using a brain phantom for PET data. All strategies were consistent with subvoxel accuracy for intrasubject, intramodality registration. Estimated accuracy of registration of structural MRI images was in the 75 to 150 µm range. Sparse data sampling strategies reduced registration times to minutes with only modest loss of accuracy.

The registration algorithm described is a robust and flexible tool that can be used to address a variety of image registration problems. Registration strategies can be tailored to meet different needs by optimizing tradeoffs between speed and accuracy. The main emphasis is given to the color combination of some brain tissues. ***The spaces in the folds of the brain (the sulci) can be easily detected by their color combination. Color intensity of that area will be measured; based on that measurement results, hypothesis will be generated. Cavity in cerebral cortex region stained in PET image is shown in figure 4.***

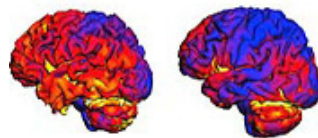


FIGURE 4: Cerebral Cortex with Color Stain (PET Image).

4. MODEL FORMULATION

1. Read and zoom the image using Bilinear Interpolation [11]


```
im1(1+(i-1)*fac,1+(j-1)*fac,:)=im(i,j,:)
```

Where im1 is the image array and fac is the zooming factor

$$im1(i, j, :) = b1 + b2 * dx + b3 * dy + b4 * dx * dy$$

Where b1, b2, b3, b4 bilinear interpolation constants and dx, dy localized pixels
2. Manage all Exceptions while loading the zoomed image: ExceptionHandling (Try Catch statement [12]).

3. Get the pixmap (Identified as 3D array of image in MATLAB [12]).
4. Extract cavity area of sulci (Apply imcrop function [12]).
5. Find the sum of each pixel value of three arrays; (RED+GREEN+BLUE) from cropped area
6. If value= =0 identify as black.
7. Put total identified pixels (black) into an array called as cavity array.
8. Else put those pixels (not black) into an array called as cortex array.
9. Compare the length of both cavity and cortex arrays (Apply length function).
10. If (length of cavity array is more), then trace as abnormality in the figure.
11. Else normal nature.
12. Set the graph of both cavity and cortex arrays (Apply plot function).

5. TESTED RESULTS USING MATLAB

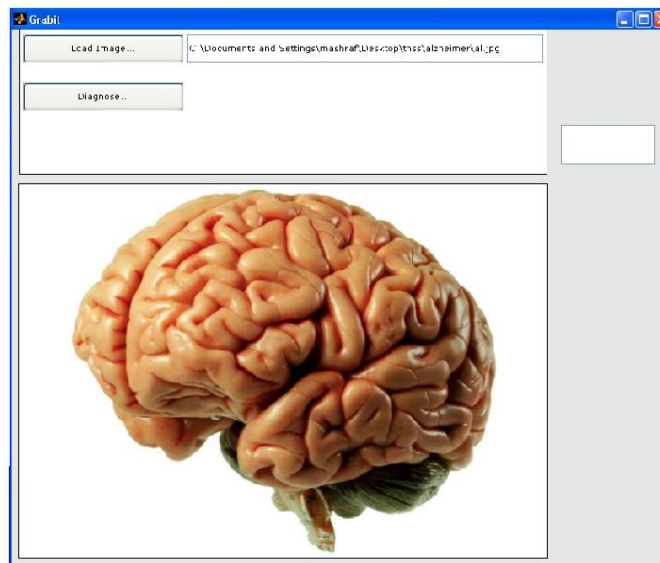


FIGURE 5: Simulated Image First Load.

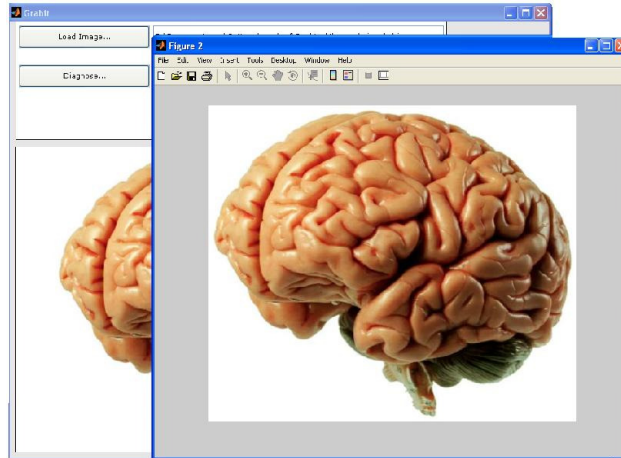


FIGURE 6: Image in Zoom.

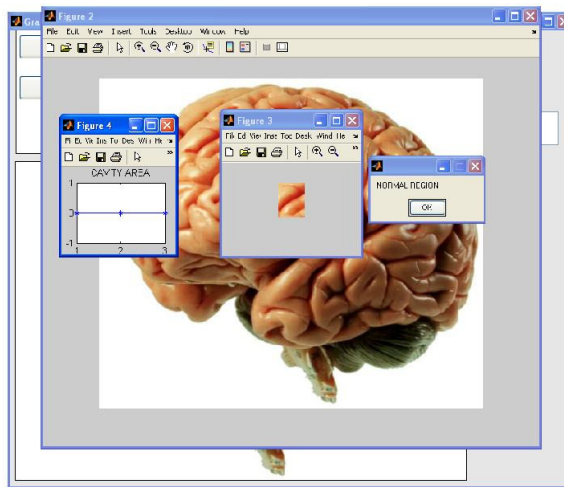


FIGURE 7: Normal Area.

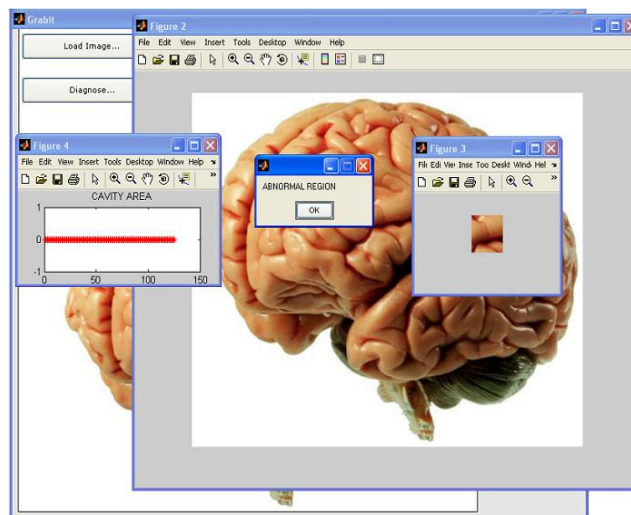


FIGURE 8: Abnormal Area.

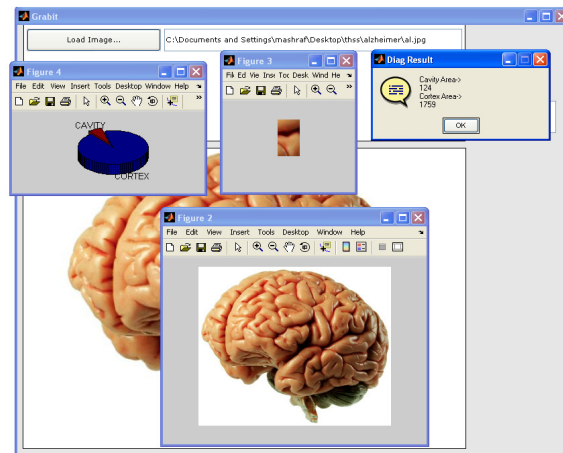


FIGURE 9: Measure of Abnormality.

6. SCOPE OF STUDY

The scope of this study can open an emerging trend in the field of computer and medical science; a new branch of science called Neuroinformatics, it may deal all the issues of neuroscience and computer science.

Neuroinformatics is a research field that encompasses the organization of neuroscience data and application of computational models and analytical tools. These areas of research are important for the integration and analysis of increasingly fine grain experimental data and for improving existing theories about nervous system and brain function. Neuroinformatics provides tools, creates databases and the possibilities for interoperability between and among databases, models, networks technologies and models for the clinical and research purposes in the neuroscience community and other fields.

Neuroinformatics stands at the intersection of neuroscience and information science. Other sciences, like genomics, have proved the effectiveness of data sharing through databases and by applying theoretical and computational models for solving complex problems in the field. Through Neuroinformatics facilities, researchers can share their data and contribute to other disciplines using available tools for the analysis and integration of data. Researchers can also more easily quantitatively confirm their working theories by means of computational modeling. Additionally, Neuroinformatics fosters more collaborative research of which one aim is to provide better possibilities to study the brain at multiple levels of brain structure. There are three main directions where Neuroinformatics has to be applied:

- a. The development of tools and databases for management and sharing of neuroscience data at all levels of analysis,
- b. The development of tools for analyzing and modeling,
- c. The development of computational models of the nervous system and neural processes.

In the recent decade; as vast amounts of diverse data about the brain were gathered by many research groups, the problem of how to integrate the data from thousands of publications in order to enable efficient tools for further research; was raised. The biological and neuroscience data are highly interconnected and complex, and by itself represents a great challenge for scientists.

Combining informatics research and brain research provides benefits for both fields of science. On one hand, informatics facilitates brain data processing and data handling, by providing new electronic and software technologies for arranging databases, modeling and communication in brain research. On the other hand, enhanced discoveries in the field of neuroscience will invoke the development of new methods in information technologies.

7. CONCLUSION

Due to high cost and requirement of professionalism much of medical imaging softwares are far from common man. Here the attempt was develop simple software in MATLAB to detect the structural abnormality of brain. The task almost fulfilled but it requires much perfection. The highlight of this software is simplicity, user friendliness and keen observation of the image at minute level.

8. FUTURE RESEARCH

In this research the measurement of cavity area is totally based on the color coding system. This measurement should be converted into some metric form. After that the abnormality can be categorized as mild, moderate and severe.

9. REFERENCES

- [1] Gilat, Amos (2008). *“Matlab: An Introduction with Applications, 3rd ed.”*, Wiley: USA. ISBN-10: 0470108770 | ISBN-13: 978-0-470-10877-2.
- [2] Sousa, David A. (2006). *“How the Special Needs Brain Learns, 2nd ed.”*, Crowin Press: USA. ISBN-10: 1-4129-4986-6 | ISBN-13: 978-1412949866.
- [3] Goyal, Soniya, Shekhar, Sudhanshu, and Biswas, K. (2011). *“Automatic Detection of Brain Abnormalities and Tumor Segmentation in MRI Sequences”*. Image and Vision Computing New Zealand Conference, IVCNZ 2011: New Zealand.
- [4] Lashkari, AmirEhsan (2010). *“A Neural Network Based Method for Brain Abnormality Detection in MR Images Using Zernike Moments and Geometric Moments”*. International Journal of Computer Applications, vol. 4, no. 7, July 2010. ISSN: 0975-8887.
- [5] Reddy, A. Ramaswamy, Prasad, E. V., and Reddy, L. S. (2012). *“Abnormality Detection of Brain MRI Images Using a New Spatial FCM Algorithm”*. International Journal of Engineering Science & Advanced Technology, vol. 2, no. 1, pp. 1-7. ISSN: 2250-3676.
- [6] Kuhn, Daniel, and Bennett, David (2003). *“Alzheimer’s Early Stages: First Steps for Family, Friends, and Caregivers, 2nd ed.”*, Hunter House: USA. ISBN-10: 0897933974 | ISBN-13: 978-0-89793-397-1.
- [7] Marshall, Louise H., and Magoun, Horace Winchell (1998). *“Discoveries in the Human Brain: Neuroscience Prehistory, Brain Structure, and Function, 1st ed.”*, Humana Press: USA. ISBN-10: 0896034356 | ISBN-13: 978-0896034358.
- [8] Shih, Frank Y. (2010), *“Image Processing and Pattern Recognition: Fundamentals and Techniques”*, Wiley-IEEE Press: USA. ISBN-10: 0-470-40461-2, ISBN-13: 978-0-470-40461-4.
- [9] Jain, Anil K (1988), *“Fundamentals of Digital Image Processing (Prentice Hall Information and System Sciences Series)”*, Prentice Hall: USA. ISBN-10: 0133361659.

- [10] Aubert, Gilles, and Kornprobst, Pierre (2006), "*Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations (Applied Mathematical Sciences)*, 2nd ed.", Springer: USA. ISBN-10: 0-387-32200-0 | ISBN-13: 978-0387-32200-1.
- [11] Debaggio, Thomas (2002). "*Losing My Mind, An Intimate Look At Life With Alzheimer's*, 1st edition", Free Press: USA. ISBN-10: 07432-05650 | ISBN-13: 978-0743205658.
- [12] Gonzalez, Rafael C., and Woods, Richard E. (2007). "*Digital Image Processing*, 3rd ed.", Prentice Hall: USA. ISBN-10: 013168728X | ISBN-13: 978-0131687288.

Digital Image Watermarking Using Different Levels of Intermediate Significant Bits with Zig-zag Embedding Approach

Ali Sharifara

*Dept. of Software engineering, faculty of computing
UTM – Malaysia – Skudai 81310*

a.sharifara@gmail.com

Ghazali Bin Sulong

*Dept. of Software engineering, faculty of computing
UTM – Malaysia – Skudai 81310*

ghazali@spaceutm.edu.my

Mehran Ranjbar Seraydashti

*Dept. of Software engineering, faculty of computing
UTM – Malaysia – Skudai 81310*

rsmehran2@live.utm.my

Abstract

The rapid growths of computer technologies have been increased over the last half century in terms of amount and complexity of data. Broadcasting of digital contents on the networks (especially Internet) has become more important and access to the data also has become much easier than before. Digital watermarking techniques are used to protect the copyrights of multimedia data by embedding secret information inside them. For example, embedding watermark in images, audios, and videos. Digital Image watermarking also has been using to detect original images against forged images by embedding an evidence of the owner of the digital image. Imperceptibility, on the other hand, is one of the problems in digital image watermarking which a repeated method in different bit planes of cover image has been presented to improve the imperceptibility of watermarking in both embedding and extracting processes. Moreover, embedding process aims to embed watermark in different bit planes by using a non-sequential method to improve security of image rather than simple sequential embedding.

Keywords: Digital Image Watermarking, Invisible Watermarking, Imperceptibility in Watermarking, Copyright Protection, Intermediate Significant Bits (ISB).

1. INTRODUCTION

Over the last half century the pace of change in the digital technologies has been widely increased. Digital images as one of the digital technologies also have been replaced with the analog images. Moreover, Internet also has become one of the most important tools to transfer digital images from one part to other parts of the world. For this reason, the security of digital documents became a challenging concern and digital image watermarking as a solution is use to decrease the number of digital forged documents.

There are lots of digital images on the Internet without having watermark which can be downloaded and modified by anyone illegally [1]. Furthermore, the ownership of the image cannot be traced without the watermarking [2]. Digital watermarking provides a solution in order to identify the owner of the digital images [3]. It has also been presented for purposes such as: copyright protection, data authentication, fingerprint, medical applications, and broadcast monitoring [4]. Image watermarking is the process of embedding an image into a host image [5]. For Instance, watermarks are embedded in bank cheque for preventing forgery. Consequently, unauthorized modification of data is the concern of researchers about copyright of documents and numerous image watermarking methods, which have been proposed with different

complexity and efficiency levels so far. Also all of image watermarking methods aimed to set up a balance between the quality and the robustness of the watermarked images.

One of the most important aspects about the digital image watermarking is to improve the robustness of the watermarked image [6]. Robustness is related to the ability of recovering the watermark after performing various processing attacks on watermarked image. The robustness must be sufficient to handle any kind of attacks occur. The watermarking scheme should be able to protect the watermark against any possible signal processing operations.

2. WHAT IS WATERMARKING

The growth of the Internet over the last several years has identified the techniques in order to protect the digital documents in the digital world [7]. Also the Internet has made it easy to distribute the digital documents such as images, video, and audio [8]. Hence, without having the powerful techniques for protecting the digital documents, it is impossible to identify the genuine owner and the plagiarist.

Therefore, the digital watermarking is one of the proper solutions to protect copyright of the digital documents against forgers. By using watermarking in digital documents, we can reduce the misusing against the documents such as Will, Cheque and etc. For instance, in the figure 1, an analog watermarking has been used to protect money against forgers.



FIGURE 1: Example of Analog Watermarking.

Unlike the analog images, all of the images on the Internet are provided as a digital content. Therefore, a desirable technique for protecting copyright is needed. Digital watermarking can authenticate the legal copyright and that cannot be removed or manipulated easily by forgers without having the secure key or other security restrictions. For instance, a duplication of a cheque can decrease confidence in its authenticity. For this reason, watermarks are employed in currencies to reduce the risk of forgery. However, the watermark technique is not the only technique that has been using for preventing forged and illegal uses.

2.1 Proposed Method

Least Significant Bit (LSB) is one of the first techniques for the watermarking purpose and it uses the lowest bit plane of image [9]. It can modify bits of both host and watermark images to embed the given watermark. The basic idea behind this technique is the substitution of the lowest bit plane of the host image with the watermark's bits. In other words, in the spatial domain, the watermark is embedded directly into the host image by altering the pixel values [10].

Although, LSB is extremely easy in term of implementation, it is difficult to detect watermark easily [11]. However, this method is imperceptible and there is no significant difference between watermarked image and original image, but it is not robustness enough to protect the authority of watermarked images. This means, in the spatial domain, embedding capacity can be large, but the watermark could be easily found by unauthorized forgers. The main primary concept of Most

Significant Bit (MSB) is the same as LSB that we have described. The only difference between them is that, LSB uses the least significant bit, meanwhile in the MSB is completely different and it uses most significant bits, instead.

When we can say that a watermark method is robust whereby the watermarked image suffers from different attacks and the embedded image (Watermark) still can be extracted from the watermarked image [12]. In other words, the attacks must not have any effects in the extraction process. For example, when a watermarked image is rotated, the algorithm still must identify the watermark that we have embedded inside the cover image and extract it as well.

2.2 The best quality of image

There is no doubt that the quality of the watermarked image is the most significant parameter in all invisible methods [13]. In all invisible watermarking methods, the watermark must be embedded with having the least effect on the quality of the host data. In order to improve the quality of the watermarked image, beside the security, an Intermediate Significant Bit (ISB) method has been presented. This means that only some bit(s) – between bit 2 and bit 7 – will be changed to guarantee the quality of watermarked image which will not have any effect on other bits.

2.3 The best robustness

Robustness as another concern in watermarking also must be improved [14]. In this research, the goal is achieved by repeating the embedded bits of watermark for certain times. This idea has tried to make the algorithm more robust against attacks. Hence, the watermark is encoded into main signals T times. For instance, in this case we have considered $T=3$ and image is divided into blocks with size of 3 pixels, moreover, all of the values of each block should be the same as it is demonstrated in the figure 2.

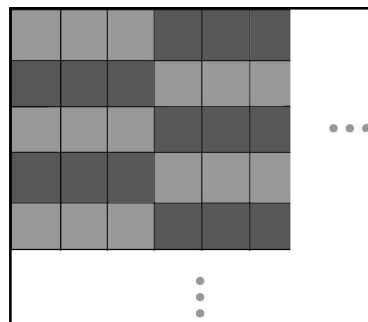


FIGURE 2: Divided Cover Image Into Blocks.

To clarify, figure 3 depicts how one pixel of watermark is embedded into the cover image; by this assumption that the first pixel of watermark contains (11011101), the pattern of embedding is sequential, and each bit of the watermark image is repeated into the cover image by three times. Figure 3 shows the result for one pixel of the watermarked image.

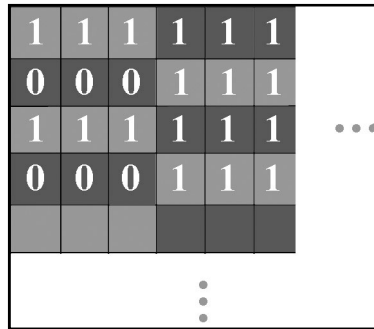


FIGURE 3: Repeating Bits in Each Block.

Extracting process must also be done in order to reverse this process. This procedure protects the watermark against attacks and it can decrease the risk of losing bits that can be damaged by the attacks. This process verifies the majority of bits while extracting watermark from the host image is to figure out the value of the watermark. In other words, the algorithms read all three bits for one pixel and makes decision between them. For example, if $T=3$, (000) represents 0, the (101) depicts 1, the (001) represents 0 and so on. To improve the robustness of the watermarked image, one bit in each pixel can be embedded in a block of pixels instead of embedding in only one pixel. This method has some advantages and disadvantages as well. The most important disadvantage of this method is the decreasing of the size of the watermark data which can be embedded in the host image. On the other hand, this approach makes an acceptable robustness by increasing the size of each block. Moreover, each block can be divided into 3, 5, 7, and 9 pixels as can be seen from the figure 4.

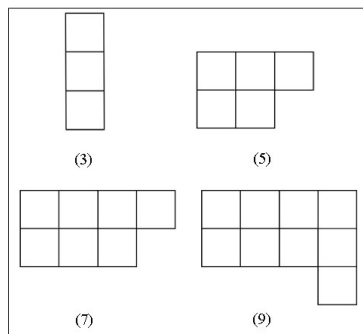


FIGURE 4: Repeating Bits in Each Block.

Beside the improvement of robustness in the watermarking system, other parameters also need to be improved such as security and capacity of image which can hide the information inside. Hence, a few steps are required to be developed which can provide guarantee that the proposed method is more robust and secure as well. Security beside the robustness is also another important point which may help to cut-off the attacks of malicious users. It is important to note that all of these stages can be executed separately. In the following section more details have been discussed.

2.4 Security in Watermarking

Security is a vital parameter in the watermarking system which must be considered as well [15]. The main aim of the attack in terms of security is to obtain the secret key of encoding process as well as decoding [16]. This means a strategy must be used to protect the watermark against malicious users and this strategy must be hidden. In other words, the watermark must be secured enough to be protected from the people those do not have the right to modify or manipulate the

Methods	PSNR
Proposed Method	45.71
HPDM	36.42
ST-SCS	36.42
color-based-encoding	30.48
Side match method	41.22
PVD	41.79
Conventional LSB	31.71
DCT	31.847
Vulnerability of PVD	45.1

Table 2: Comparison Between Different Methods [19].

2.7 Extracted Watermark after performing selected attacks

As it can be inferred from the table 2, the proposed method is more robust in comparison with conventional LSB and MSB methods. Nevertheless, most of the image attacks tend to destroy low bits (1, 2, and 3), but the proposed method can extract the watermark after the attacks by acceptable percentage of healthy pixels in the watermark even in the low bit-planes [18]. Surprisingly, as can be seen from the result of proposed method, the average PSNR in all of the 8 layers are approximately same by over 30 db. Hence, the above result proves that the watermark can embed into all bit-planes of cover image regardless of the visibility of the watermark.

The below figures also provides information about the PSNR of the watermark after extraction for proposed method compared to the conventional LSB method. The x-axis indicates 8 bit-planes and the PSNR is indicates on y-axis.

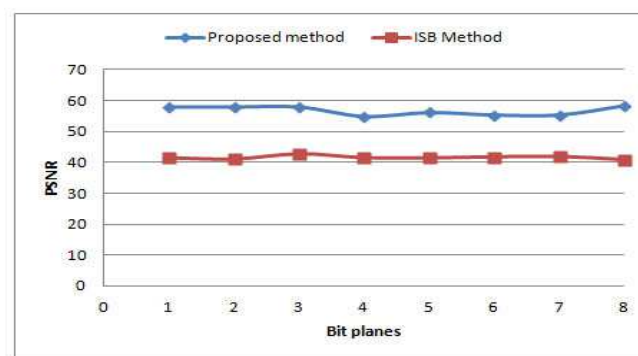


FIGURE 6: Applied Salt and Pepper Attacks for Both Proposed and Conventional LSB Method.

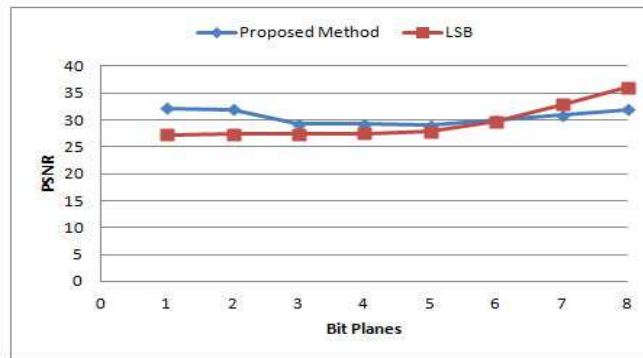


FIGURE 7: Applied Blurring Attacks for Both Proposed and Conventional LSB Method.

As can be seen from the figure 6 and 7, the selected attacks do not have any significant effect on the watermark and can be extracted by high PSNR (Peak signal-to-noise ratio). Furthermore, other two mentioned attacks also have less effect on the watermark image after extraction in all different 8 bit planes. Table 3 and 4 show the watermark images after extracting in both proposed method and the conventional LSB, respectively.

Bit- planes	Salt and Pepper	Gaussian	Speckle	Poisson	Blurring
1,2,3					
PSNR	57.8942	35.1767	35.1767	35.1211	35.1767
1,3,4					
PSNR	58.0278	33.1767	35.1767	31.9604	30.2204
2,3,4					
PSNR	58.0448	29.2292	28.8208	29.1933	28.9032
2,4,5					
PSNR	54.8024	29.1511	29.0914	29.2553	28.7651
3,4,5					
PSNR	57.1409	29.0285	29.7285	28.9973	28.2251
3,5,6					
PSNR	55.2978	28.9489	29.7255	29.9973	29.0686
4,5,6					
PSNR	55.2978	28.8878	31.5728	30.5728	28.9786

Table 3: Extracted Watermark After Attack in Proposed Method.

The result proves that the watermark can embed into all bit-planes of the cover image regardless of the visibility of the watermark. We also have performed similar tests to other images and we have achieved the same results.









































Bit- Planes	Salt and Pepper	Gaussian	Speckle	Poisson	Blurring
1					
PSNR	41.3402	27.2872	27.4521	27.2553	27.3459
2					
PSNR	40.9581	27.4718	27.5265	27.3376	27.4557
3					
PSNR	42.6408	27.2759	27.3816	27.3793	27.2623
4					
PSNR	41.3980	27.3839	27.5088	27.4279	27.7070
5					
PSNR	41.3206	27.4665	29.5288	27.8759	27.8265
6					
PSNR	41.6746	27.6354	32.2871	29.7421	28.2866
7					
PSNR	41.8143	29.4687	34.7684	32.8219	27.4516
8					
PSNR	40.7806	32.2926	39.5445	36.0818	27.4319

Table 4: Extracted Watermark After Attack in Conventional Method LSB.

3. CONCLUSION AND FUTURE WORK

Watermarking is a method that is being used to hide information or identify data within the digital documents. Moreover, digital documents can be divided into three categories: video, audio and image. In this research, we have focused primarily on the watermarking of digital images. Digital watermarking is becoming popular, mainly for embedding undetectable marks such as copyright information. In this research, gray scale images have been applied to embed as an evidence of the ownership for the digital images. The proposed method also used Intermediate Significant Bits (ISB) to improve the quality of the watermarked image.

In this research, we have proposed a new watermarking method to improve previous works in area of spatial domain in terms of robustness and imperceptibility. The proposed method has been aimed to improve robustness of the watermarked image by using repeated variable for each pixel of the watermark. Also, it improves the probability of extracting intact bits after attacks and the security guaranteed by using a secret key and none-sequential embedding method.

Although the proposed method is tried to achieve the best quality, robustness, and security for watermarking in grayscale images, but some other important parameters must be considered as well, such as: Capacity and Integrity. Also grayscale image has been used for this research, but the present study can also be extended for colored image (RGB).

4. REFERENCES

- [1] Chin-Shiuh Shieh, Hsiang-Cheh Huang, Feng-Hsing Wang, Jeng-Shyang Pan, "Genetic watermarking based on transform-domain techniques". *Pattern Recognition*,. 37(3): p. 555-565, 2004.
- [2] Golea, N.E.-H., "A blind RGB color image watermarking based on singular value decomposition", in *Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications - AICCSA 2010*, IEEE Computer Society. p. 1-5. 2010.
- [3] Yongjian Hu, Guangzhou, China Kwong, S. "Using invisible watermarks to protect visibly watermarked images" ,p. 584-587. 2004.
- [4] Potdar, V.M., Han, S., Chang, "A survey of digital image watermarking techniques". 2005. p. 709- 716.
- [5] Bamatraf, A. Ibrahim, R. Salleh, M.N.B.M, "Digital watermarking algorithm using LSB", in *Computer Applications and Industrial Electronics (ICCAIE)*, Kuala Lumpur, 2010.
- [6] Dharwadkar, N.V., Amberker, B.B. Gorai, A., "Non-blind watermarking scheme for color images in RGB space using DWT-SVD", in *Communications and Signal Processing (ICCSP)*, 2011 International Conference, Calicut, 2011.
- [7] Dehkordi, A.B., Esfahani, S.N., Avanaki, A.N. , "Robust LSB watermarking optimized for local structural similarity", in *Electrical Engineering (ICEE)*, 2011 19th Iranian Conference. Tehran, 2011.
- [8] Arya, D., "A Survey of Frequency and Wavelet Domain Digital Watermarking Techniques". 2010. 1(2).
- [9] Zhao Xingyang, "A novel color image fragile watermarking based on the extended channel", in *Broadband Network & Multimedia Technology*, 2009. IC-BNMT '09. 2nd IEEE International Conference, Beijing, Oct. 2009.

- [10] Gil-Je Lee, "A New LSB Based Digital Watermarking Scheme with Random Mapping Function", in Ubiquitous Multimedia Computing. UMC '08. International Symposium: Hobart, ACT, Oct 2008.
- [11] Ker, A., "Improved Detection of LSB Steganography in Grayscale Images Information Hiding", J. Fridrich, Editor, Springer Berlin / Heidelberg, 2005.
- [12] Kumar, N.M., Manikandan, T., Sathagirivasan, V. , "Non blind image watermarking based on similarity in contourlet domain", in Recent Trends in Information Technology (ICRTIT), 2011 International Conference, Chennai, Tamil Nadu, 2011.
- [13] Maity, S.P. and M.K. Kundu, "DHT domain digital watermarking with low loss in image informations". AEU - International Journal of Electronics and Communications. 64(3): p. 243-257, 2010.
- [14] MohammadReza Keyvanpour, Farnoosh Merrikh Bayat, "Blind image watermarking method based on chaotic key and dynamic coefficient quantization in the DWT domain", Mathematical and Computer Modelling, Available online 27 July 2012, ISSN 0895-7177, 2012.
- [15] Hu, M.-C., D.-C. Lou, and M.-C. Chang, "Dual-wrapped digital watermarking scheme for image copyright protection". Computers & Security 26 (4) , 319-330, 2007.
- [16] Jian Cao; Jiwu Huang; , "Controllable Secure Watermarking Technique for Tradeoff Between Robustness and Security," Information Forensics and Security, IEEE Transactions on , vol.7, no.2, pp.821-826, April 2012.
- [17] M.Padmaa, "ZIG-ZAG PVD – A Nontraditional Approach". International Journal of Computer Applications. Volume 5– No.7, August 2010.
- [18] Mir Shahriar Emami , Ghazali Bin Sulong, "Set Removal Attack: A New Geometric Watermarking Attack" in 2011 International Conference on Future Information Technology IPCSIT. 2011, IACSIT Press: Singapore. 2011.
- [19] Akram M. Zeki. Khedher, Azizah Abdul Manaf. "Digital watermarking and data hiding techniques". Proceedings of the Postgraduate Annual Research Seminar p. 79-84, 2006.

MultiModal Identification System in Monozygotic Twins

Dinakardas C.N.

*Research Scholar, Nims University, Jaipur
& Associate Professor, Lourdes Matha College of Science & Technology
Trivandrum, 695574, India*

dinakardas@gmail.com

Dr.S.Perumal Sankar

*Research Guide, Nims University, Jaipur
& Professor, CAPE Institute of Technology
Kanyakumari, 629001, India*

spsankar2004@yahoo.co.in

Nisha George

*Assistant Professor, Lourdes Matha College of Science & Technology
Trivandrum, 695574, India*

ddcnisha@gmail.com

Abstract

With the increase in the number of twin births in recent decades, there is a need to develop alternate approaches that can secure the biometric system. In this paper an effective fusion scheme is presented that combines information presented by multiple domain experts based on the rank-level fusion integration method. The developed multimodal biometric system possesses a number of unique qualities, starting from utilizing Fisher's Linear Discriminant methods for face matching, Principal Component Analysis for fingerprint matching and Local binary pattern features for iris matching and fused the information for effective recognition and authentication. The importance of considering these boundary conditions, such as twins, where the possibility of errors is maximum will lead us to design a more reliable and robust security system. The proposed approach is tested on a real database consisting of 50 pair of identical twin images and shows promising results compared to other techniques. The Receiver Operating Characteristics also shows that the proposed method is superior compared to other techniques under study.

Keywords: Fisher Faces, Principal Component Analysis, Local Binary Pattern Receiver Operating Characteristics Curve.

1. INTRODUCTION

A biometric identification (matching) system is an automatic pattern recognition system that recognizes a person by determining the authenticity of a specific physiological and/or behavioral characteristic (biometric) possessed by that person. Fertility treatments have resulted in an increase in the identical twin birth rate [1]. The identical twin birth rate is about twice as high for women who use fertility drugs. Identical twins are more precisely described by the term monozygotic, indicating that they come from the split of a single fertilized embryo thus, they have the same DNA. Fraternal, or dizygotic, twins are the result of two different fertilized embryos and have different DNA [2]. Identical twins has been considered to be a problem of only academic interest but due to consistent increase in twin births in recent decades. The extent of variation in a physical trait due to random development process differs from trait to trait.

Typically, most of the physical characteristics such as body type, voice, and face are very similar for identical twins and automatic identification based on face and hand geometry will fail to distinguish them [3],[4],[5]. However, a recent experiment demonstrates that iris biometric template aging can be detected after as little as two years. A significant number of twin pairs (206) have been studied for handwriting. These samples were processed with features extracted

and conclusions drawn by comparing verification performances with twins and non-twins. In that study, the conclusion was that twins are discriminable but less so than an arbitrary pair of individuals [5],[7]. Turk and Pentland popularized the use of PCA for face recognition [8]. They used PCA to compute a set of subspace basis vectors (which they called "eigenfaces") for a database of face images, and projected the images in the database into the compressed subspace. New test images were then matched to images in the database by projecting them onto the basis vectors and finding the nearest compressed image in the subspace (eigenspace). Kong *et al.* [9] observed that palmprints from identical twins have correlated features.

The same observation was made by Jain *et al.* [10],[14] for fingerprints also. Srihari *et al.*[4] analyzed the similarity between twins fingerprints in a study using fingerprint images from 298 pairs of twins. The authors analyzed this similarity based on the pattern of the ridge flow, and minutiae. They concluded that the similarity between twin fingers is higher than between two arbitrary fingers, but twins can still be distinguished using fingerprints .Kocaman *et al.*[11] proposed a study on PCA,FLDA,DCVA,and evaluate error and hit rates of four algorithms which were calculated by random subsampling and k-fold cross validation. . Chang *et al.*[12] compared PCA technique for both face and ear images and showed similar performance as biometrics . Gaurav *et al.*[16] proposed a method for distinguishing identical twins and the ROC shows the various comparisons for various set of facial marks in identical twins. Kodate *et al.* [17] experimented with 10 sets of identical twins using a 2D face recognition system. Recently, Sun *et al.* [6] presented a study of distinctiveness of biometric characteristics in identical twins using fingerprint, face and iris biometrics. They observed that though iris and fingerprints show little to no degradation in performance when dealing with identical twins, face matchers find it hard to distinguish between identical twins. It is believed that the texture of every iris is determined entirely at random. This implies that the iris textures of two identical twins are no more similar to each other than the iris textures of unrelated persons. The researchers compared the distribution of difference values between iris codes from the eyes of identical twins to that between iris codes of unrelated persons and found that the iris textures of identical twins are no more similar than those of unrelated persons. Among all biometric traits, the textural structure of the human iris has been observed to be robust and reliable. However, the performance of iris recognition systems is adversely affected by the quality of the acquired images.

Today technologies are well-studied, but research shows they have many drawbacks which decrease the success of the methods applied. The frequently used and most common biological traits in the field of biometrics are face, finger, and iris. Identifying identical twins is crucial for all biometric systems. The systems that cannot handle identical twins have a serious security hole.

The rest of this paper is organized as follows. Section 2 describes the Fisher linear Discriminant analysis, Principal component analysis and Local binary pattern and the results are reported to evaluate the performance of our proposed approach in Section 3. Finally, conclusions are given in Section 4.

2. BIMODAL BIOMETRIC SYSTEM DESIGN

In this section deals the development procedures of the proposed multimodal biometric system. Fisher face features are extracted from the face images and the PCA features from the fingerprints and the Local Binary pattern based texture pattern from the iris pattern are used for the enrollment and recognition of biometric traits. This system integrate multiple modalities in user verification and identification which will lead to higher performance A more detailed representation of the proposed system is shown in Fig. 1.

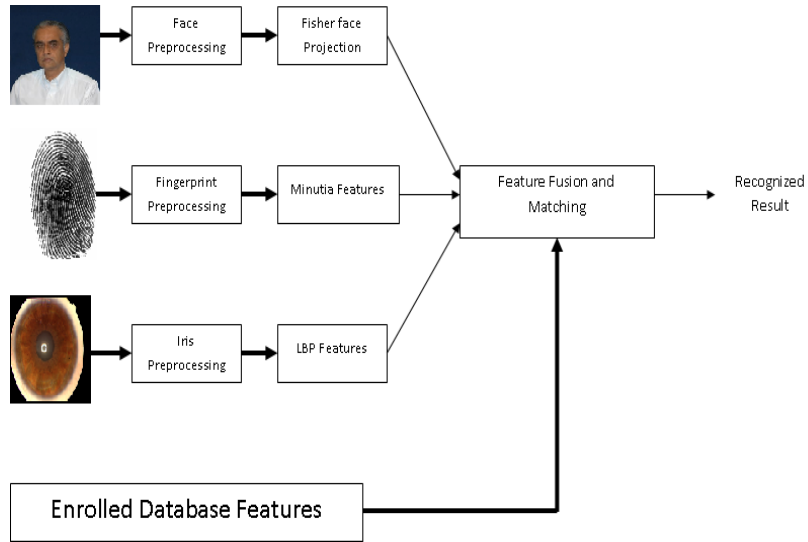


FIGURE 1: Block Diagram of the Proposed Bimodal Biometric Identification System.

2.1. Fisher linear discriminant analysis

Fisher linear discriminant analysis (LDA), a widely-used technique for pattern classification, finds a linear discriminant that yields optimal discrimination between two classes which can be identified with two random variables, say X and Y in R^n . PCA in the form of eigen space representation is very sensitive to image conditions such as background noise, image shift, occlusion of objects, scaling of the image, and illumination change. When substantial changes in illumination and expression are present in any image, much of the variation in data is due to these changes [13], and the eigenimage technique, in this case, cannot give highly reliable results. Due to certain illumination changes in the face images of the database used in this work, a fisherface based face recognition method [23] is developed to compare with the eigenface technique. The fisherface method uses both PCA and LDA to produce a subspace projection matrix, similar to that used in the eigenface method. The terms Fisher's linear discriminant and LDA are often used interchangeably, although Fisher's original article actually describes a slightly different discriminant, which does not make some of the assumptions of LDA such as normally distributed classes or equal class covariances. Suppose two classes of observations have means,

$$\vec{\mu}_y = 0, \vec{\mu}_y = 1$$

and covariance

$$\sum_y = 0, \sum_y = 1$$

Then the linear combination of features $\vec{\omega} \cdot x$ will have means

$$\vec{\omega} \cdot \vec{\mu}_y = i$$

and variances

$$\vec{\omega}^T \sum_y = i \cdot \vec{\omega}, \text{ for } i = 0, 1.$$

Fisher defined the separation between these two distributions to be the ratio of the variance between the classes to the variance within the classes as in eq(1)

$$\begin{aligned}
 S &= \frac{\sigma_{between}^2}{\sigma_{within}^2} = \\
 &= \frac{(\vec{\omega} \cdot \vec{\mu}_{y=1} - \vec{\omega} \cdot \vec{\mu}_{y=0})^2}{\vec{\omega}^T \sum_{y=1} \vec{\omega} + \vec{\omega}^T \sum_{y=0} \vec{\omega}} \\
 &= \frac{(\vec{\omega} \cdot (\vec{\mu}_{y=1} - \vec{\mu}_{y=0}))^2}{(\vec{\omega}^T \sum_{y=0} + \sum_{y=1}) \vec{\omega}} \tag{1}
 \end{aligned}$$

This measure is, in some sense, a measure of the signal-to-noise ratio for the class labeling. It can be shown that the maximum separation occurs when

$$\vec{\omega} = \left(\sum_{y=0} + \sum_{y=1} \right)^{-1} (\vec{\mu}_{y=1} - \vec{\mu}_{y=0})$$

When the assumptions of LDA are satisfied, the above equation is equivalent to LDA. The vector $\vec{\omega}$ is the normal to the discriminant hyperplane. As an example, in a two dimensional problem, the line that best divides the two groups is perpendicular to $\vec{\omega}$.

Generally, the data points to be discriminated are projected onto $\vec{\omega}$, then the threshold that best separates the data is chosen from analysis of the one-dimensional distribution. There is no general rule for the threshold. However, if projections of points from both classes exhibit approximately the same distributions, the good choice would be hyperplane in the middle between projections of the two means,

$$\vec{\omega} \cdot \vec{\mu}_{y=0} \text{ and } \vec{\omega} \cdot \vec{\mu}_{y=1} .$$

In this case the parameter c in threshold condition $\vec{\omega} \cdot \vec{x} < c$ can be found explicitly as in eq(2):

$$c = (\vec{\omega} \cdot (\vec{\mu}_{y=0} + \vec{\mu}_{y=1})) / 2 \tag{2}.$$

2.2. Principal Component Analysis

Principal Components Analysis is a method that reduces data dimensionality by performing a covariance analysis between factors. As such, it is suitable for data sets in multiple dimensions, such as a large experiment involving huge amount of data. PCA is an unsupervised technique and as such does not include label information of the data. Kirby and Sirovich[15] were among the first to apply principal component analysis (PCA) to face images, and showed that PCA is an

optimal compression scheme that minimizes the mean squared error between the original images and their reconstructions for any given level of compression

PCA, mathematically defined as an orthogonal linear transformation [23] that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

Define a data matrix, \mathbf{X}^T , with zero empirical mean, where each of the n rows represents a different repetition of the experiment, and each of the m columns gives a particular kind of datum.

The singular value decomposition of \mathbf{X} is $\mathbf{X} = \mathbf{W}\Sigma\mathbf{V}^T$, where $m \times m$ matrix, \mathbf{W} is the matrix of eigenvectors of $\mathbf{X}\mathbf{X}^T$, matrix Σ is an $m \times n$ rectangular diagonal matrix with nonnegative real numbers on the diagonal, and the $n \times n$ matrix \mathbf{V} is the matrix of eigenvectors of $\mathbf{X}^T\mathbf{X}$.

The PCA transformation that preserves dimensionality is then given by eq (3).

$$\mathbf{Y}^T = \mathbf{X}^T\mathbf{W} = \mathbf{V}\sum^T\mathbf{W}^T\mathbf{W} = \mathbf{V}\sum^T \quad (3).$$

\mathbf{V} is not uniquely defined in the usual case when $m < n - 1$, but \mathbf{Y} will usually still be uniquely defined. Since \mathbf{W} is an orthogonal matrix, each row of \mathbf{Y}^T is simply a rotation of the corresponding row of \mathbf{X}^T . The first column of \mathbf{Y}^T is made up of the "scores" of the cases with respect to the "principal" component; the next column has the scores with respect to the "second principal" component, and so on. For reduced-dimensionality representation, project \mathbf{X} down into the reduced space defined by only the first L singular vectors, \mathbf{W}_L .

$$\mathbf{Y} = \mathbf{W}_L^T \mathbf{X} = \sum_L \mathbf{V}^T$$

Where with $I_{L \times m}$ the $L \times m$ rectangular identity matrix. The matrix \mathbf{W} of singular vectors of \mathbf{X} is equivalently the matrix \mathbf{W} of eigenvectors of the matrix of observed covariances $\mathbf{C} = \mathbf{X}\mathbf{X}^T$,

$$\mathbf{X}\mathbf{X}^T = \mathbf{W}\sum\sum^T\mathbf{W}^T \quad (4).$$

Given a set of points in Euclidean space, the first principal component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the distances of the points from the line. The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted from the points. The singular values in Σ are the square roots of the eigenvalues of the matrix $\mathbf{X}\mathbf{X}^T$. Each eigenvalue is proportional to the portion of the "variance" that is correlated with each eigenvector. The sum of all the eigenvalues is equal to the sum of the squared distances of the points from their multidimensional mean. PCA essentially rotates the set of points around their mean in order to align with the principal components. This moves as much of the variance as possible (using an orthogonal transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be small and may be dropped with minimal loss of information. PCA is often used in this manner for dimensionality reduction. PCA has the distinction of being the optimal orthogonal transformation for keeping the subspace that has largest "variance". One characteristic of both PCA and LDA is that they produce spatially global feature vectors. In other words, the basis vectors produced by PCA and LDA are non-zero for almost all dimensions, implying that a change to a single input pixel will alter every dimension of its subspace projection.

2.3. Local Binary Pattern for Iris pattern

Local Binary Pattern (LBP) is an efficient method used for feature extraction and texture classification it was first introduced by Ojala et al in 1996 [19], this was the first article to describe LBP. The LBP operator was introduced as a complementary measure for local image

contrast, and it was developed as a grayscale invariant pattern measure adding complementary information to the amount of texture in images. LBP is ideally suited for applications requiring fast feature extraction and texture classification. Local Binary Pattern (LBP) is a very efficient texture operator which labels the pixels of an image by thresholding the neighborhood of each pixel and considers the result as a binary number. Due to its discriminative power and computational simplicity, LBP texture operator has become a popular approach in various applications. Local binary patterns are adopted for representing the textural characteristics of local sub-regions. It can be seen as a unifying approach to the traditionally divergent statistical and structural models of texture analysis. Each iris images can be considered as a composition of micro-patterns which can be effectively detected by the LBP operator. The LBP operator [18] forms labels for the image pixels by thresholding the 3 x 3 neighborhood of each pixel with the center value and considering the result as a binary number. The histogram of these $2^8 = 256$ different labels can then be used as a texture descriptor. The features of the iris pattern is extracted using the above procedure and is explained in Fig.2

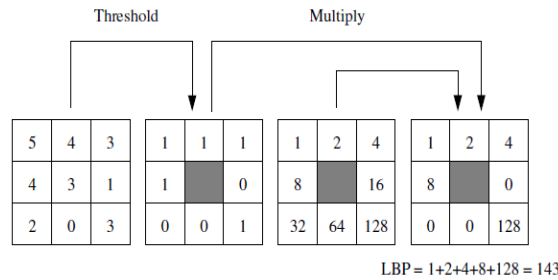


FIGURE 2: Calculating the Original LBP Code and A Contrast Measure.

3. Experiment and Results

In this section the performance of the proposed multibiometric recognition is tested on a real time database consisting of 50 pair of identical twins from whom the face, fingerprint and iris images of the persons are collected. The images are acquired in a resolution of 200x200 sizes. We have implemented our multibiometric system in MATLAB 7.10 on a Pentium-IV Windows XP workstation. To build our virtual multimodal database, we have chosen 100 images. Face images are randomly sampled as training samples, and the remaining are left as test samples. The technique is also applied for fingerprint and iris databases to collect training samples. Then, each sample of the face database is randomly combined with one sample of the fingerprint database and one sample of the iris database.

The performance of a biometric system can be shown as a Receiver Operating Characteristic (ROC) curve that plots the Genuine Accept Rate against the False Accept Rate (FAR) at different thresholds on the matching score. Fig. 5 shows the performance of the hybrid approach presented here.

We compare this performance with other approaches that does not utilize texture information for representing the fingerprint. As can be seen in the graph, the proposed hybrid approach outperforms over a wide range of FAR values.

The results obtained using various multibiometric systems were analyzed and the area under the ROC curve for each method using Real Time database are shown in Table 1., and it shows the area under the ROC curve (**Az**), Standard Deviation (**S.D**) and 95% Confidence Interval (**CI**) for each classifier. Results show that high performance was obtained by the proposed scheme when compared to other multibiometric systems.

	Single mode PCA	Multimodal PCA	Proposed
Az	0.94470	0.96036	0.96147
S.D	0.01710	0.01413	0.01370
95% CI	0.91119	0.93268	0.93419

TABLE 1: Classification Results.

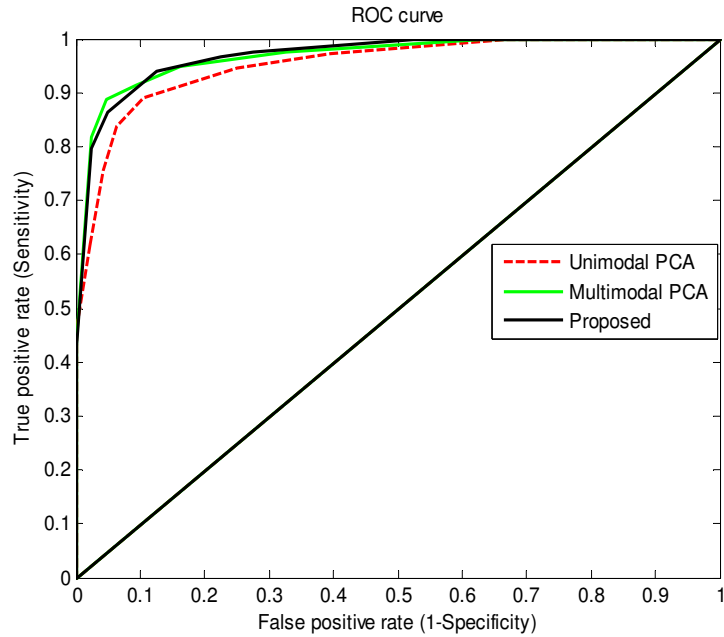


Figure 3. Receiver Operating Characteristics Curve.

4. CONCLUSION

This paper has presented a multimodal approach and comparison of three approaches in implementing a biometric system. Using a multimodal database we have investigated the relative merits of adopting a single classifier approach, an approach which uses a multimodal classifier configuration operating on a single modality and, finally, a multimodal biometric solution which combines different biometric samples in providing an identification decision. Our study has provided quantitative data to demonstrate the relative performance levels, in terms of ROC curve, attainable in each case, and we have shown how multimodal biometric solutions, while offering other additional advantages where appropriate, provide only modest improvements over an approach based on a multimodal classifier approach and a single modality, bringing some potentially significant benefits in terms of usability.

5. ACKNOWLEDGMENT

The author would like to express his sincere gratitude to his research guide Perumal Sankar and his co author Nisha George for all their helpful guidance and advice.

6. REFERENCES

[1] D. Costello, Families: The Perfect Deception: *Identical Twins*", Wall Street Journal, February 12, 1999.

- [2] J. Martin, H.-C. Kuang, T. J. Mathews, D. L. Hoyert, D. M. Strobino, B. Guyer, and S. R. Sutton, "Annual summary of vital statistics: 2006," in *Pediatrics*, pp. 788–801, 2008.
- [3] Anil K. Jain, Salil Prabhakar and Sharath Pankanti "Twin Test: On Discriminability of Fingerprints".
- [4] Sargur N. Srihari, Harish Srinivasan, and Gang Fang "Discriminability of Fingerprints of Twin", *Journal of Forensic Identification*, March 19, 2007, pp.109-121.
- [5] K. Nandakumar, A. Ross and A. K. Jain, *Handbook of Multibiometrics*. New York: Springer-Verlag, 2006.
- [6] Z. Sun, A. A. Paulino, J. Feng, Z. Chai, T. Tan, and A. K. Jain, "A study of multibiometric traits of identical twins," in *Biometric Technology for Human Identification VII*, B. V. K. V. Kumar, S. Prabhakar, and A. A. Ross, Eds., vol. Proc. SPIE 7667, 2010.
- [7] P. Jonathon Phillips, Patrick J. Flynn, Kevin W. Bowyer, Richa W. Vorder Bruegge, Patrick J. Grother, George W. Quinn and Matthew Pruitt, "Distinguishing identical twins by face recognition". *Automatic Face & Gesture Recognition and Workshops, IEEE Digital Explore* 21-25 March 2011, pp. 185 - 192
- [8] M. Turk, A. Pentland, Eigenfaces for recognition, *Journal of Cognitive Neuroscience* 3 (1991) 71–86.
- [9] A. Kong, D. Zhang, and G. Lu, "A study of identical twins palmprints for personal authentication."
- [10] K. Nandakumar, A. Ross and A. K. Jain, *Handbook of Multibiometrics*. New York: Springer-Verlag, 2006.
- [11] Kocaman, B., Kirci, M., Gunes, E.O., Cakir, Yand Ozbudak, O. "On Ear biometrics", *EUROCON, IEEE*, 2009, pp.327-332.
- [12] K. Chang, K. W. Bowyer, S. Sarkar, and B. Victor, "Comparison and combination of ear and face images in appearance-based biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25 no. 9, pp. 1160- 1165, Sept. 2003.
- [13] T. Heseltine, N. Pears, J. Austin, and Z. Chen, "Face recognition: A comparison of appearance-based approaches," in *Proc. 7th Digit. Image Comput.: Tech. Appl.*, C. Sun, H. Talbot, S. Ourselin, and T. Adriaansen, (Eds.), Sydney, Australia, 2003, pp. 59–68.
- [14] K. Jain, S. Prabhakar, and S. Pankanti, "On the similarity of identical twin fingerprints," 2002.
- [15] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, pp. 103-107, 1990.
- [16] Nisha Srinivas, Gaurav Aggarwal, Patrick J Flynn, and Richard W. Vorder Bruegge, "Facial Marks as Biometric Signatures to Distinguish between Identical Twins", pp.113-120
- [17] E. W. Kashiko Kodate, Rieko Inaba and T. Kamiya, "Facial recognition by a compact parallel optical correlator," *Measurement Science and Technology*, vol. 13, Nov 2002.
- [18] B.A. Draper, K. Baek, M.S. Bartlett and J.R. Beveridge, "Recognizing faces with PCA and ICA", *Computer Vision and Image Understanding*, 91 (1-2), pp. 115-137, 2003.
- [19] X. Jing, Y. Yao, D. Zhang, M. Li, Face and palmprint pixel level fusion and Kernel DCV-RBF classifier for small sample biometrics recognition, *Pattern Recognition* 40 (11), 2007, pp. 3209–3224.

- [20] S. Ribaric, I. Fratric, A biometric identification system based on eigenpalm and eigenfinger features, *IEEE Trans. Pattern Anal. Mach.Intell.* 27 (11), 2005, pp. 1698–1709.
- [21] S. C. Dass, K. Nandakumar & A. K. Jain, *A principal approach to scorelevel fusion in Multimodal Biometrics System*, Proceedings of ABVPA, 2005.
- [22] Sheetal Chaudhary, and Rajender Nath, *A Multimodal Biometric Recognition System Based on Fusion of Palmprint, Fingerprint and Face*, Proceedings of ICART in Communication and Computing, 2009.
- [23] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “*Eigenfaces vs. fisherfaces: Recognition using class specific linear projection*,” *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 19, No. 7, pp. 711–720, Jul. 1997.

Image Steganography Technique By Using Braille Method of Blind People (LSBraille)

Abdelmgeid Amin Ali

*Associate Professor, Dept. of Computer Science
Faculty of Science, Al – Minia University
Al – Minia, Egypt*

abdelmgeid@yahoo.com

Al – Hussien Seddik Saad

*Assistant Lecturer, High Institute for
Engineering and Technology, (H.I.E.T)
Al – Minia, Egypt*

al.hussien_seddik@yahoo.com

Abstract

Steganography is the art and science of writing hidden messages in such a way that no one, apart from the sender and intended recipient, suspects the existence of the message, a form of security through obscurity. Steganography is a Greek origin word which means “hidden writing”. Steganography word is classified into two parts: steganos which means “secret or covered” (where you want to hide the secret messages) and the graphic which means “writing” (text). It can be defined as the art of hiding the fact that communication is taking place, by hiding information in other information. Many different carrier file formats can be used, but digital images are the most popular. In this paper, a new image steganography method is proposed. The proposed method hides the secret message inside the cover image by representing the secret message characters using Braille method of reading and writing for blind people that can save more than one-fourth of the required space for embedding. From the experimental results it is seen that the proposed method achieves higher visual quality as indicated by the high peak signal-to-noise ratio (PSNR) in spite of hiding a large number of secret bits in the image.

Keywords: Steganography, Peak Signal-to-Noise Ratio (PSNR), Least Significant Bit (LSB).

1 INTRODUCTION

Data security or data privacy has become increasingly important as more and more systems are connected to the Internet [1]. In general, protecting the secret messages during transmission becomes an important research issue. To protect secret message from being stolen during transmission, there are two ways to solve this problem. One way is encryption, which refers to the process of encoding secret information in such a way that only the right person with a right key can decode and recover the original information successfully. Another way is steganography and this is a technique which hides secret information into a cover media or carrier so that it becomes unnoticed and less attractive [2].

Steganography is a branch of data hiding that allows the people to communicate secretly. As increasingly more material becomes available electronically, the influence of steganography on our lives will continue to grow [3]. In general, steganography is the art of hiding a message signal in a host signal without any perceptual distortion of the host signal. The composite signal is usually referred to as the stego signal. By using steganography, information can be hidden in carriers such as images, audio files, text files and videos [2].

The main terminologies used in the steganography systems are: the cover message, secret message, secret key and embedding algorithm. The cover message is the carrier of the message such as image, video, audio, text, or some other digital media. The secret message is the

information which is needed to be hidden in the suitable digital media. The secret key is usually used to embed the message depending on the hiding algorithms. The embedding algorithm is the way or the idea that usually use to embed the secret information in the cover message [4].

The most frequently used carriers are digital images. The use of digital images for steganography makes use of the weaknesses in the human visual system (HVS), which has a low sensitivity in random pattern changes and luminance. The human eye is incapable of discerning small changes in color or patterns. Because of this weakness the secret Message can be inserted into the cover image without being detected [5].

In this paper a new method that hides the secret message inside the cover image by representing the secret message characters using Braille method of reading and writing for blind people. Braille system uses six raised dots in a systematic arrangement with two columns of three dots, known as a Braille cell [8]; a 6-dot matrix which is the basis of Braille [7].

This paper is organized as follows, section 1 is an introduction. Section 2 briefly introduces the types of image steganography and least significant bit (LSB) technique. Section 3 introduces Braille method in details. Section 4 contains some previous LSB proposed methods. Section 5 presents our proposed LSBBraille method. Section 6 contains the experimental results of our proposed method. Finally, section 7 concludes the paper.

2 IMAGE STEGANOGRAPHY

As said before, steganography is the technique that allows people to communicate secretly. Actually, there are two kinds of image steganographic techniques, spatial domain and frequency domain based methods. The schemes of the first kind directly embed the secret data or secret message within the pixels of the cover image. One of the most known examples of spatial domain method is LSB (Least Significant Bit) insertion. The schemes of the second kind embed the secret data within the cover image that has been transformed such as DCT (Discrete Cosine Transformation). The DCT coefficients of the transformed cover image will be quantized, and then modified according to the secret data [2].

2.1 Least Significant Bit (LSB) Insertion Technique

The simplest form of spatial domain image steganography is implemented by inserting the secret data into the least significant bits. Different algorithms would insert the binary form of the secret data in 1, 2, 3 or 4 – LSBs of the cover image. So, it is simple to implement for RGB, Gray Scale or Binary Images and less susceptible to detection by Human Vision System (HVS).[6]

In LSB steganography, the least significant bits of the cover media's digital data are used to conceal the message. The simplest of the LSB steganography techniques is LSB replacement. LSB replacement steganography flips the last bit of each of the data values to reflect the message that needs to be hidden.

Consider an 8-bit bitmap image where each pixel is stored as a byte representing a grayscale value. Suppose the first eight pixels of the original image have the following grayscale values: [3]

(11010010 01001010 10010111 10001100 00010101 01010111 00100110 01000011)

To hide the letter C whose binary value is 10000011, we would replace the LSBs of these pixels to have the following new grayscale values:

(11010011 01001010 10010110 10001100 00010100 01010110 00100111 01000011)

Note that, on average, only half the LSBs need to change. The difference between the cover (i.e. original) image and the stego image will be hardly noticeable to the human eye [3].

3 BRAILLE METHOD

3.1 Braille Definition

Braille is a tactile method of reading and writing for blind people developed by Louis Braille (1809–1852), a blind Frenchman. The Braille system uses six raised dots in a systematic arrangement with two columns of three dots, known as a Braille cell [8], a 6-dot matrix which is the basis of Braille [7]. By convention, the dots in the left column are numbered 1, 2 and 3 from top to bottom and the dots in the right column are numbered 4, 5 and 6 from top to bottom [8].

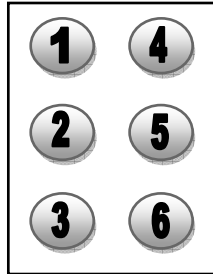


FIGURE 1: 6 – Dots Matrix (Braille Cell).

The six dots of the Braille cell are configured in 64 possible combinations (including the space which has no dots present). The 63 Braille characters with dots are grouped in a table of seven lines. This table is used to establish "Braille order" for listing Braille signs [8].

Line 1:	⠠	⠡	⠢	⠣	⠤	⠥	⠦	⠧	⠨	⠩
Line 2:	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠
Line 3:	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠
Line 4:	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠
Line 5:	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠
Line 6:	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠
Line 7:	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠	⠠

TABLE 1: 63 Braille Characters With Dots.

Line 1 is formed with characters in the upper part of the cell, using dots 1, 2, 4 and 5. Line 2 adds dot 3 to each of the characters in Line 1. Line 3 adds dots 3 and 6 to each of the characters in Line 1. Line 4 adds dot 6 to each of the characters in Line 1. Line 5 repeats the dot configurations of Line 1 in the lower part of the cell, using dots 2, 3, 5 and 6. Line 6 is formed with characters using dots 3, 4, 5 and 6. Line 7 is formed with characters in the right column of the cell, using dots 4, 5 and 6 [8].

3.2 Braille Characters and Signs [9]

3.2.1 Letters of the Alphabet

	a		d		g		j		m		p		s		v		Y
	b		e		h		k		n		q		t		w		z
	c		f		i		l		o		r		u		x		

TABLE 2: Braille's Letters of the Alphabet.

3.2.2 Numerals

	1		3		5		7		9
	2		4		6		8		0

TABLE 3: Braille's Numerals.

3.2.3 Punctuations

	,	comma		.	Full stop		'	Single quote
	;	semicolon		?	Question mark		-	Hyphen
	:	colon		!	Exclamation mark		"	Double quote

TABLE 4: Braille's Punctuations.

3.2.4 Other Symbols

	&	Ampersand		+	Plus		>	Greater than
	[Left bracket		=	Equals		/	Oblique stroke
	@	At		<	Less than)	Right bracket
]	Right bracket		*	Asterisk		_	Underline
	(Left bracket		\$	Dollar		%	Percent

TABLE 5: Braille's Other Symbols.

Finally, the total number of characters listed above is (26 Letters + 10 Numbers + 9 Punctuations + 15 Symbols) so, the total will be 60 characters. The rest of characters to reach 64 characters are (space) and the following table:-

	Text continuation		Lower Case Indicator		Upper Case Indicator
--	-------------------	--	----------------------	--	----------------------

TABLE 6: Braille's Rest of Characters.

At this point of the paper we discussed the LSB image steganography technique and the Braille method of reading and writing for blind people. So, how we will use the Braille method for hiding

the secret message in the cover image?, before explaining our new method we will discuss some previous image steganography methods that based on LSB embedding technique.

4 PREVIOUS WORK

In [10], the author proposed a new LSB method that Instead of using the LSB-1 of the cover for embedding the message, LSB-3 which is the 3rd Least significant bit of the pixel has been used and LSB-1,2 may be modified according to the bit of the message, to minimize the difference between the cover and the stego-cover. After the experiments have been done the tables showed that the LSB – 1 has results better than LSB - 3 in which the embedding capacity is the same but the PSNR of LSB – 3 is lower than LSB – 1. After modifications have been done to LSB – 1 and LSB – 2 the PSNR values were enhanced but still lower than the original LSB – 1 method.

In [11] the authors proposed a method depends on LSB and edge detection where edge pixels were selected in order to hide the data by using the “Least Significant Bit Insertion”. The algorithm randomly selects the edge-pixels of the image to hide the message because It is not noticeable when a single pixel is modified when its surrounding pixels are least like it. But from our point of view by using LSB in embedding in any region in the image the modifications won't be noticeable because the LSB modifies only the last bit which means the pixel will be increased or decreased by one.

In [12] the authors proposed a method that hide the pixel information of the source image in the destination video frames such that each row of pixel (consisting of 8 bits) is hidid in first rows of multiple frames of the target. This makes the hiding so complex and it becomes very difficult to analyze. But what is new in the proposed method? The authors used the original LSB – 1 but by hiding data in multiple frames of the cover video without any modification.

In [1], the authors proposed a method that take the cover image, secret message and secret key, then transfers the secret message into text file, then convert the text file into a zip text file (Compressed File) and convert zip text file to binary codes. Finally the message is embedded by using 2 LSBs. The author here used the zip file for securing the secret message, also it has been compressed. Here they used the original LSB but with last 2 LSBs not LSB – 1 only.

In the next section we will explain the proposed LSBraile method in details and its 6 – bits representation table will be constructed.

5 PROPOSED LSBraile Method

Our proposed method is using the Braille method representations of the characters as each character can be represented by only 6 dots using the 6 – dots matrix which called (Braille Cell). The method will start by representing these characters (dots) as binary digits each of which consists of 6 bits only, not eight bits as in original LSB embedding method which uses the binary representation from the ASCII table. So, by using this representation we can save 2 pixels from each secret byte embedding process or more than one-fourth of the maximum hiding capacity for each cover image, which will increase the maximum hiding capacity (MHC) and enhance the PSNR of the LSB embedding technique.

$$\text{Original LSB MHC (Bytes)} = M \times N / 8 \quad (1)$$

$$\text{LSBraile MHC (Bytes)} = M \times N / 6 \quad (2)$$

As shown in previous equations M is width, N is height of the cover image. As an example if we have 512 x 512 cover image by using the original LSB with 8bit representation of the text, the cover image will hold up to $(512 \times 512 / 8) = 32,768$ byte, but by using our new LSBraile method, the same image can hold $(512 \times 512 / 6) = 43,690$ bytes, which is larger than the original LSB MHC.

Now, we will list Braille method characters and their corresponding 6 – bits representation in a table, in which each black dot will be represented by 1 and each white (empty) dot will be represented by 0.

Char	Black Dots Index in Cell	Binary Representation (6 bits)						Char	Black Dots Index in Cell	Binary Representation (6 bits)					
		1 st	2 nd	3 rd	4 th	5 th	6 th			1 st	2 nd	3 rd	4 th	5 th	6 th
a	1	1	0	0	0	0	0	6	1, 2, 4, 6	1	1	0	1	0	1
b	1, 2	1	1	0	0	0	0	7	1, 2, 4, 5, 6	1	1	0	1	1	1
c	1, 4	1	0	0	1	0	0	8	1, 2, 5, 6	1	1	0	0	1	1
d	1, 4, 5	1	0	0	1	1	0	9	2, 4, 6	0	1	0	1	0	1
e	1, 5	1	0	0	0	1	0	0	1, 2, 3, 4, 5, 6	1	1	1	1	1	1
f	1, 2, 4	1	1	0	1	0	0	,	2	0	1	0	0	0	0
g	1, 2, 4, 5	1	1	0	1	1	0	;	2, 3	0	1	1	0	0	0
h	1, 2, 5	1	1	0	0	1	0	:	2, 5	0	1	0	0	1	0
i	2, 4	0	1	0	1	0	0	.	2, 5, 6	0	1	0	0	1	1
j	2, 4, 5	0	1	0	1	1	0	?	2, 6	0	1	0	0	0	1
k	1, 3	1	0	1	0	0	0	!	3, 4, 5, 6	0	0	1	1	1	1
l	1, 2, 3	1	1	1	0	0	0	'	3	0	0	1	0	0	0
m	1, 3, 4	1	0	1	1	0	0	-	3, 6	0	0	1	0	0	1
n	1, 3, 4, 5	1	0	1	1	1	0	"	4	0	0	0	1	0	0
o	1, 3, 5	1	0	1	0	1	0	&	1, 2, 3, 4, 6	1	1	1	1	0	1
p	1, 2, 3, 4	1	1	1	1	0	0	[1, 2, 3, 5, 6	1	1	1	0	1	1
q	1, 2, 3, 4, 5	1	1	1	1	1	0	@	2, 3, 4, 6	0	1	1	1	0	1
r	1, 2, 3, 5	1	1	1	0	1	0]	2, 3, 4, 5, 6	0	1	1	1	1	1
s	2, 3, 4	0	1	1	1	0	0	+	2, 3, 5	0	1	1	0	1	0
t	2, 3, 4, 5	0	1	1	1	1	0	=	2, 3, 5, 6	0	1	1	0	1	1
u	1, 3, 6	1	0	1	0	0	1	<	2, 3, 6	0	1	1	0	0	1
v	1, 2, 3, 6	1	1	1	0	0	1	*	3, 5	0	0	1	0	1	0
w	2, 4, 5, 6	0	1	0	1	1	1	>	3, 5, 6	0	0	1	0	1	1
x	1, 3, 4, 6	1	0	1	1	0	1	/	3, 4	0	0	1	1	0	0
y	1, 3, 4, 5, 6	1	0	1	1	1	1)	3, 4, 5	0	0	1	1	1	0
z	1, 3, 5, 6	1	0	1	0	1	1	_	3, 4, 6	0	0	1	1	0	1
1	1, 6	1	0	0	0	0	1	(4, 5	0	0	0	1	1	0
2	1, 2, 6	1	1	0	0	0	1	\$	4, 5, 6	0	0	0	1	1	1
3	1, 4, 6	1	0	0	1	0	1	%	4, 6	0	0	0	1	0	1
4	1, 4, 5, 6	1	0	0	1	1	1	Space	Empty Cell	0	0	0	0	0	0
5	1, 5, 6	1	0	0	0	1	1								

TABLE 7: Braille's Characters 6 – Bits Representation.

5.1 LSBraile Embedding Algorithm: Message Embedding Using LSBraile Method

Input : Cover Image C ; Secret Message M.

Output : StegoImage S.

Steps:

- 1) Split C into 3 channels Red (R), Green (G), Blue (B).
- 2) Split M into characters; $M = \{ m_1, m_2, m_3 \dots, m_n \}$.
- 3) Take m_i from M
- 4) Convert m_i into Braille 6 – bits representation.
- 5) Take 6 pixels from B.
- 6) Apply LSB on m_i 's 6 – bits and the 6 pixels of B.
- 7) Repeat steps from 3 to 6 until the whole M has been embedded in C.
- 8) Merge the 3 channels R, G, B again to construct the StegoImage S.

In the next section we will discuss in details the results that obtained from the proposed LSBraile method including Maximum Hiding Capacity (MHC) and Peak Signal to Noise Ratio (PSNR).

6 EXPERIMENTAL RESULTS

In this section, we will evaluate the performance of our new proposed method by comparing between the MHC of original LSB and LSBraile methods and comparing the PSNR values by taking the same messages and the same cover images then apply the original LSB method, 2 other LSB methods and our new LSBraile method. The results obtained from these comparisons are recorded and can be summarized in the following tables:-

Cover Image	Maximum Hiding Capacity (Byte)	
	Original LSB	LSBraile
128 x 128	2,048	2,730
256 x 256	8,192	10,922
512 x 512	32,768	43,690
1024 x 1024	131,072	174,762

TABLE 8: Comparison Between MHC of Original LSB and Our LSBraile.

As shown in table 8, the proposed LSBraile method has a higher embedding capacity than the original LSB method. So, our LSBraile method succeeded in obtaining more capacity for the same cover image than original LSB method or our LSBraile method has higher MHC than original LSB method.

Before discussing PSNR comparison tables, as we said before our LSBraile represents each character by 6 bits only not 8 bits. So, the PSNR will be higher than the original LSB because we saved 2 pixels from being modified for each secret message character, but we decided to prove our LSBraile efficiency by applying it and comparing it with some LSB methods.

Cover Image (256 x 256)	Message Capacity (Bytes)	PSNR	
		Original LSB	LSBraille
Boat	8,192	55.91603	57.19327
Bird	8,192	55.92812	57.18989
Flinstone	8,192	55.88576	57.16223

TABLE 9: Comparison Between PSNR of Original LSB and Our LSBraille.

In table 9 we compared our proposed LSBraille method with original LSB method by using 8,192 characters (bytes) secret message and 265 x 265 cover images (boat, bird, flinstone) and we found that our LSBraille method has more PSNR values than Original LSB method which means the stego image quality of our method will be higher than the stego image quality of the original LSB method.

Cover Image (256 x 256)	Message Capacity (Bytes)	PSNR	
		LSB – 3	LSBraille
Boat	8,192	39.1132	57.19327
Bird	8,192	39.0955	57.18989
Flinstone	8,192	39.1188	57.16223

TABLE 10: Comparison Between PSNR of LSB – 3 and Our LSBraille.

Also, in table 10 we compared our proposed LSBraille method with LSB – 3 method by using 8,192 characters (bytes) secret message and 265 x 265 cover images (boat, bird, flinstone) and we found that our LSBraille method has more PSNR values than LSB – 3 method which means that the stego image quality of our proposed method will be higher.

Cover Image (256 x 256)	Message Capacity (Bytes)	PSNR	
		Modified LSB – 3	LSBraille
Boat	8,192	42.4163	57.19327
Bird	8,192	42.4062	57.18989
Flinstone	8,192	42.2932	57.16223

TABLE 11: Comparison Between PSNR of Modified LSB – 3 and Our LSBraille.

Moreover, in table 11 we compared our proposed LSBraille method with Modified LSB – 3 method by using 8,192 characters (bytes) secret message and 265 x 265 cover images (boat, bird, flinstone) and we found that our LSBraille method has more PSNR values than Modified LSB – 3 method which means that the stego image quality of our proposed method will be also higher.

Finally, as shown in tables 9, 10 and 11, after the comparisons have been done among our proposed LSBraille method and original LSB, LSB – 3 [10], Modified LSB – 3 [10] methods by using secret message consists of 8,192 characters and 3 different 256 x 256 cover images (boat, bird, flinstone) we found that our LSBraille method has more PSNR values than other LSB methods which means the stego image quality of our method will be higher than the quality of other LSB methods.

7 CONCLUSION

As shown in comparison tables, after doing the same experiments using the LSBraille, original LSB and 2 other LSB methods, the PSNR values of the proposed LSBraille method were higher than other LSB methods. This means we are succeeded to improve the LSB method's PSNR

(stego image quality). Also we are succeeded to improve the maximum hiding capacity (MHC) as shown in table 8.

Finally, we can say that, our proposed LSBBraille method proved its efficiency in image steganography field by enhancing the maximum hiding capacity (MHC) and the stego image quality (PSNR).

8 REFERENCES

- [1] R. Ibrahim and T. S. Kuan. "Steganography Imaging System (SIS): Hiding Secret Message Inside An Image". *Proceedings of the World Congress on Engineering and Computer Science, WCECS 2010, San Francisco, USA, Vol. (I), Oct 20-22, 2010*.
- [2] A. A. Radwan, A. swilem and A. H. Seddik. "A High Capacity SLDIP (Substitute Last Digit In Pixel)". *Fifth International Conference on Intelligent Computing and Information Systems ICICIS 2011, Cairo, Egypt, 30 Jun - 3 Jul 2011*.
- [3] V. K. SHARMA and V. SHRIVASTAVA. "A Steganography Algorithm For Hiding Image In Image By Improved LSB Substitution By Minimize Detection". *Journal of Theoretical and Applied Information Technology*. 36(1). 15th Feb, 2012.
- [4] A. M. Al – Shatnawi. "A New Method In Image Steganography With Improved Image Quality". *Applied Mathematical Sciences*, 6(79), pp. 3907 – 3915, 2012.
- [5] A. H. Seddik. "Enhancing The (MSLDIP) Image Steganographic Method (ESLDIP Method)". *International Conference on Graphic and Image Processing, Proc. of SPIE 8285(3I)*. © SPIE 2011.
- [6] S. Chutani and H. Goyal. "LSB Embedding In Spatial Domain - A Review Of Improved Techniques". *International Journal of Computers & Technology*, ISSN: 2277-3061, 3(1), Aug. 2012.
- [7] UNESCO. "World Braille Usage". *National library service for the blind and handicapped, library of congress, Washington, D.C, USA, 1990*.
- [8] C. Simpson. "The Rules of Unified English Braille". *Round Table on Information Access for People with Print Disabilities Inc, Australia, Version I: June 2010*.
- [9] Braille Authority of the United Kingdom. "British Braille". *Royal National Institute of the Blind. Bakewell Road. Orton Southgate, Peterborough, Cambridgeshire, 2004*.
- [10] A. I. Abdul-Sada. "Hiding Data Using LSB-3". *J. Basrah Researches (Sciences)*, 33(4). pp. 81-88, DEC. 2007.
- [11] N. Jain, S. Meshram and S. Dubey. "Image Steganography Using LSB and Edge – Detection Technique". *International Journal of Soft Computing and Engineering (IJSCE)* ISSN: 2231-2307, 2(3), Jul. 2012.
- [12] S. Singh and G. Agarwal. "Hiding Image To Video: A New Approach Of LSB Replacement". *International Journal of Engineering Science and Technology*, 2(12), pp 6999-7003, 2010.

Establishment of an Efficient Color Model from Existing Models for Better Gamma Encoding In Image Processing

T. M. Shahriar Sazzad

*Department of Computer Science
University of St Andrews
St Andrews, UK*

tss5standrews@gmail.com

Sabrin Islam

*Department of Computer Science
American International University Bangladesh
Dhaka, Bangladesh*

islamsabrin@yahoo.com

Mohammad Mahbubur Rahman Khan Mamun

*EEE, BUET
Dhaka, Bangladesh*

siha24@gmail.com

Md. Zahid Hasan

*Lecturer, Dept. of CSE
Green University
Dhaka, Bangladesh*

hasan.ice@gmail.com

Abstract

Human vision is an important factor in the areas of image processing. Research has been done for years to make automatic image processing but still human intervention can not be denied and thus better human intervention is necessary. Two most important points are required to improve human vision which are light and color. Gamma encoder is the one which helps to improve the properties of human vision and thus to maintain visual quality gamma encoding is necessary.

It is to mention that all through the computer graphics RGB (Red, Green, and Blue) color space is vastly used. Moreover, for computer graphics RGB color space is called the most established choice to acquire desired color. RGB color space has a great effort on simplifying the design and architecture of a system. However, RGB struggles to deal efficiently for the images those belong to the real-world.

Images are captured using cameras, videos and other devices using different magnifications. In most cases during processing, in compare to the original outlook the images appear either dark or bright in contrast. Human vision affects and thus poor quality image analysis may occur. Consequently this poor manual image analysis may have huge difference from the computational image analysis outcome. Question may arise here why we will use gamma encoding when histogram equalization or histogram normalization can enhance images. Enhancing images does not improve human visualization quality all the time because sometimes it brightens the image quality when it is needed to darken and vice-versa. Human vision reflects under universal illumination environment (not pitch black or blindingly bright) thus follows an approximate gamma or power function. Hence, this is not a good idea to brighten images all the time when better human visualization can be obtained while darkening the images. Better human visualization is important for manual image processing which leads to compare the outcome with the semi-automated or automated one. Considering the importance of gamma encoding in image processing we propose an efficient color model which will help to improve visual quality for manual processing as well as will lead analyzers to analyze images automatically for comparison and testing purpose.

Keywords: Gamma, Human Vision, RGB, HSI, HSB, Light.

1. INTRODUCTION

A color space can be defined as the mathematical illustration of a set of colors. In the areas of image processing there are different color models available of which RGB (mainly used for computer graphics), YUV, YIQ, or YCbCr (used for video systems) and CMYK (used for color printing) are most popular. However, it is to mention that, for instinctive ideas of hue, saturation and brightness; the above three color models are not directly related at all. For this perspective, HSI, HSV or HSB are suitable color models for programming simplicity, end user manipulation and processing purposes although all of these color models is derived from the RGB information supplied by devices such as cameras and scanners [1,2,3,4].

Color Model	Classifications
Munsell	Device dependent
RGB, CMY(K)	Device dependent
YIQ, YUV, YCbCr	Device dependent
HSI, HSV, HSL	User oriented-Device dependent
CIE XYZ, CIE L*U*V*, CIE L*a*b*	Device independent, color Metric

Table 1: Color Models Classifications.

Color Model	Application Area
Munsell	Human visual system
RGB	Computer graphics, Image processing, Analysis, Storage
CMY(K)	Printing
YIQ, YUV	TV broadcasting, Video system
YCbCr	Digital video
HSI, HSV, HSL	Human visual perception, Computer graphics, processing, Computer Vision, Image Analysis, Design image, Human vision, Image editing software, Video editor
CIE XYZ, CIE L*U*V*, CIE L*a*b*	Evaluation of color difference, Color matching system, advertising, graphic arts, digitized or animated paintings, multimedia products

Table 2: Application Areas of Color Models.

It is to mention that all through the computer graphics RGB (Red, Green, and Blue) color space is vastly used. Moreover, for computer graphics RGB color space is called the most established choice to acquire desired color. RGB color space has a great effort on simplifying the design and architecture of a system. However, RGB struggles to deal efficiently for the images those belong

to the real-world. Moreover, processing images with the help of RGB color model is not an efficient method either.

Various types of color model have been established already. One main color model is RGB color model where 3 different colors are added together in different ways to produce a wide range of colors. As for example for a 24 bit RGB color image, a total number of colors can be $(2^8)^3 = 16,777,216$.

RGB color model is used to represent and display images in electronic systems. It is to mention that RGB color model is device dependent as Red, Green and Blue levels are different from manufacturers to manufacturers. Sometimes these colors vary even in same devices over a period of time and hence without a color management RGB color value does not acts as same in devices.

To display RGB colors in hardware a display card named cathode ray tube (CRT) is used to handle the numeric RGB color values and in most CRT displays do have a power-law transfer characteristic with a gamma of about 2.5. In most occasions it has been observed that gamma remains out of consideration. Under these circumstances, an accurate reproduction of the original scene results in an image that human viewers judge as "flat" and lacking in contrast.

To improve the quality of visual perception for color images, the term image enhancement is an important factor. Image enhancement is needed in many areas such as photography, scanning, image analysis etc. Image enhancement approaches fall into two broad categories such as spatial domain and frequency domain methods. The term spatial domain refers to the image plane itself, and approaches in this category are based on direct manipulation of pixels in an image whereas frequency domain processing techniques are based on modifying the Fourier transform of an image.

Color image enhancement is considered the most frequently used method these days using adaptive neighborhood histogram equalization technique [14]. 3D histogram equalization has been proposed using RGB cube [15]. A new approach considering enhancement problem has been established [13, 20]

There are some more techniques available for wavelength based image enhancement which helps to enhances the image edges [19]. It is generally unwise to histogram equalize the components of a color image independently because it causes erroneous color. A more logical approach is histogram normalization while spreading the color intensities uniformly, leaving the color themselves (eg. Hue) enhanced.

Images can be gray-level images or color images. Comparing with color images gray-level images have got only one value for each pixel as images are made with pixel representation. There are many existing algorithm available which helps to enhance the image contrast for gray-level images considering piecewise-linear transformation function named contrast stretching with normalization, stretching with histogram techniques. Most of these available algorithm are not suitable for color images although they are used widely having poor quality and distorted effects [5].

Gray level transformation is proved to be better approach than any other transformation and hence most proposed methods are based on spatial domain approach. Image enhancement using spatial domain works with gray-level transformation or power law transformation. Power law equation is referred to as gamma.

$s = cr^\gamma$; where c and r are positive constants. Value of c= 1 and the value of gamma can vary to set the desired result and the process used to correct power-law transformation phenomena is called gamma correction or gamma encoding.

However, it is to mention that, only enhancing the image does not improve the image quality for better visual perception. Sometimes it is needed to darken the bright images to obtain a better visualization [6]. Gamma is one of the main factor which helps to brighten or darken an image.

The above mentioned techniques are widely used in the areas of image enhancement without much considering the color shifting issues. A color image enhancement technique should not change a pixel value from red to yellow as an example although in some cases color shifting may be necessary while controlling them before it can be applied. Hue is one of the main properties of a color and hence it is not easy to control hue in color enhancement especially in RGB color

model. The color shifting issue has been considered in some research by Gupta et al, Naik et al where it has been suggested that hue should be preserved while applying image enhancement method [16, 17, 18]. These methods keeps hue preserved and avoids color shifting but still there are problems. However, enhancement does not resolve human visualization perfectly because sometimes images need to make dark instead of enhancement. In that case enhancement does not help at all.

To resolve the above mentioned for human visualization considering two issues 1) color shifting and 2) human visualization we have come up with an idea that gamma encoding is necessary while decomposing the luminance (is an objective term and it is a measure of the amount of light coming off from a source, or reflected from an object) or brightness (perception of how much light is coming from a source or an object, and depends upon the context as well as the luminance) and for saturation instead of histogram equalization, histogram normalization can be applied.

This research aspires to establish an efficient color model for better gamma encoding in image processing from all the existing color models available at this moment.

2. METHODOLOGY

Our proposed gamma encoding technique is based on spatial domain instead of frequency domain approach.

In RGB color model, there are three primary colors considered named Red, Green and Blue where RGB is defined as additive or subtractive model and hence different colors can be preformed using the combination of these primary colors. But for HIS (hue, saturation, intensity) and HSV (hue, saturation, value) or HSB (hue, saturation, brightness) color spaces were developed to distinguish and understand color by human. Hue is the main attribute of a color and thus decides which color the pixel has obtained. However, hue should not be changed at any point because changing the hue changes the color as well as distortion occurs in the image. Moreover, comparing with color space like CIE LUV and CIE Lab, in HSB it is easy to control hue and color shifting. Our main approach is to preserve the hue and apply better human visualization using saturation and brightness and hence we have chosen HSB color space instead of other color space [21, 22, 23].

It is to mention that for traditional image processing such as histograms, equalization HSI color space is one of the best model [7]. However, HSB color space is one of the best for manipulating hue and saturation (to shift colors or adjust the amount of color) and thus it capitulates a better active range of saturation [8].

3. COLOR MODEL CONVERSION

2.1 RGB to HSB

Below equations describes the conversion from RGB to HSB color space. For easier definition we have used maximum and minimum component values as M and m respectively and R for Red, G for Green and B for Blue and C is the difference between maximum and minimum.

$$M = \max (R , G , B) \quad (1)$$

$$m = \min (R , G , B) \quad (2)$$

$$C = M - m \quad (3)$$

Hue is the proportion of the distance around the edge of the hexagon which passes through the projected point, measured on the range [0,1] or in degree [0,360]. Mathematical expression for hue is

$$H' = \left\{ \begin{array}{ll} \text{Undefined,} & \text{if } C = 0 \\ \frac{G-B}{C} \bmod 6, & \text{if } M = R \\ \frac{B-R}{C} + 2, & \text{if } M = G \\ \frac{R-G}{C} + 4, & \text{if } M = B \end{array} \right\} \quad (4)$$

$$H = 60^\circ \times H' \quad (5)$$

2.2 HSB to RGB

Below equations describes the conversions from HSB to RGB.

$$H' = \frac{H}{60^\circ} \quad (6)$$

$$X = C(1 - |H' \bmod 2 - 1|) \quad (7)$$

$$(R_1, G_1, B_1) = \left\{ \begin{array}{ll} (0,0,0) & \text{if } H \text{ is Undefined} \\ (C, X, 0) & \text{if } 0 \leq H' < 1 \\ (X, C, 0) & \text{if } 1 \leq H' < 2 \\ (0, C, X) & \text{if } 2 \leq H' < 3 \\ (0, X, C) & \text{if } 3 \leq H' < 4 \\ (X, 0, C) & \text{if } 4 \leq H' < 5 \\ (C, 0, X) & \text{if } 5 \leq H' < 6 \end{array} \right\} \quad (8)$$

$$m = Y' - (0.30R_1 + 0.59G_1 + 0.11B_1) \quad (9)$$

$$(R, G, B) = (R_1 + m, G_1 + m, B_1 + m) \quad (10)$$

This is a geometric warping of hexagons into circles where each side of the hexagon is mapped onto a 60 degree arc of the circle.

$$S = 0, \text{ if } C = 0 \quad (11)$$

$$S = 1 - \min / \max, \text{ otherwise} \quad (12)$$

S is denoted for saturation

$$I = \frac{1}{3}(R + G + B) \quad (13)$$

where I is denoted as intensity

$$B = \max \quad (14)$$

where B is denoted in HSB as brightness.

2.3 RGB to HSI

Equation (1) describes the conversion from RGB to HSI color space.

$$I = \frac{1}{3}(R + G + B) \quad (15)$$

$$S = 1 - \frac{3}{(R + G + B)} [\min(R, G, B)] \quad (16)$$

$$H = \cos^{-1} \left\{ \frac{0.5[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right\} \quad (17)$$

$$\text{If } B \text{ is greater than } G, \text{ then } H = 360^\circ - H \quad (18)$$

Where R, G and B are three color component of source RGB image, H, S and I it's components of hardware independent on HSI format

2.4 HSI to RGB

As it can be seen that conversion from RGB to HSI is not easy with regard to computing algorithm complexity because it's regarding minimum from three searching (expression 1, as minimum two operators of condition), long cosine function, square root, square computation, additional operation of condition (expression 4) during one pixel conversion. Moreover, it is difficult to convert from HSI color space to standard RGB, where the process depends on which color sector H lies in. For the RG sector ($0^\circ \leq H \leq 120^\circ$), we have the following equations to convert RGB to HSI format:

$$B = I(1 - S) \quad (19)$$

$$R = I \left[1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (20)$$

$$G = 3I - (R + B) \quad (21)$$

For the GB sector ($120^\circ \leq H \leq 240^\circ$):

$$H = H - 120^\circ \quad (22)$$

$$R = I(1 - S) \quad (23)$$

$$G = I \left[1 + \frac{S \cos H}{\cos(60^\circ - H)} \right] \quad (24)$$

$$B = 3I - (R + G) \quad (25)$$

For the BR sector ($240^{\circ} \leq H \leq 360^{\circ}$):

$$H = H - 240^{\circ} \quad (26)$$

$$G = I(1 - S) \quad (27)$$

$$B = I \left[1 + \frac{S \cos H}{\cos(60^{\circ} - H)} \right] \quad (28)$$

$$R = 3I - (G + B) \quad (29)$$

4. GAMMA ENCODER

It is wise to use luma which represents the brightness in an image and can be denoted as Y. Luma is weighted average of gamma-encoding which can be denoted as Y' for R,G and B and hence denoted as R'G'B'.

The equation becomes,

$$Y = 0.2126R + 0.7152G + 0.0722B \quad \text{for luminance}$$

$$Y' = 0.2126R' + 0.7152G' + 0.0722B' \quad \text{for gamma encoding}$$

5. SATURATION

To make the color image soft and better human acceptance it is necessary to use saturation adjustment. We have applied histogram normalization instead of histogram equalization because normalize models stretches image pixel values to cover the entire pixel value range from (0-255) whereas equalize module attempts to equalize the number of pixels in a given color thus uses a single row of pixels.

6. PROCESSING STEPS

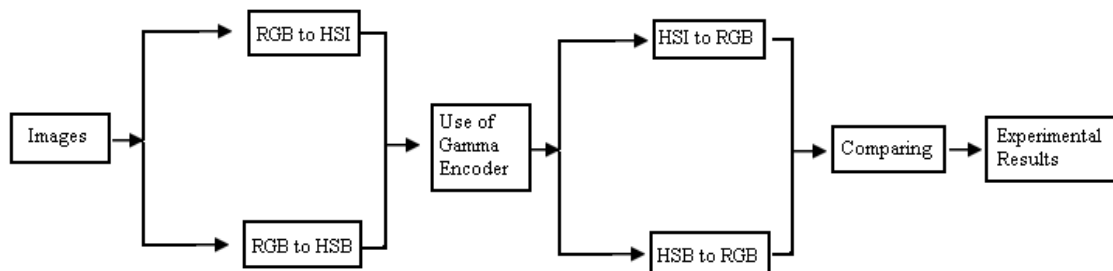


FIGURE 1: Block Diagram of Proposed Work.

7. EXPERIMENTAL RESULTS

To test the performance of our proposed approach we have used three different contrast color images (low contrast or darker from the original outlook, medium contrast or similar to original outlook and high contrast or brighter than original outlook color images). To evaluate the contrast performance we have applied histogram normalization saturation value from 0.4 – 0.6 and gamma correction value ranges from 0.75 – 2.2 in different computers as different computers acts different according to gamma value. It is to mention that gamma value > 1 performs darkening and vice-versa [9, 10, 11, 12].

Figure 2, 3 and 4 images with (a),(b),(c) illustrates that (a) is the original image, (b) is the experimental result obtained using HSI and (c) is the experimental result obtained using HSB.

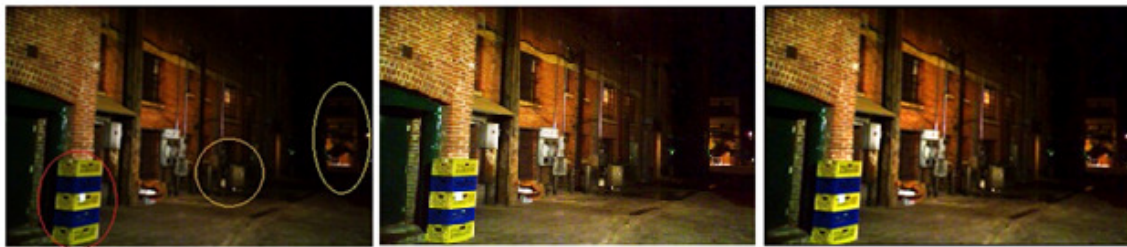


FIGURE 2



FIGURE 3



FIGURE 4

Images used	Using HSI (acceptance rate from users)	Using HSB (acceptance rate from users)	Comparison result
Bright Images (Total 223 images)	83 %	88 %	HSB acceptance rate is high
Dark Images (Total 304 Images)	79 %	89 %	HSB acceptance rate is high

TABLE 3: Detailed comparison between existing approach without gamma and our proposed approach with accuracy. Sample results were collected considering human visual perception.

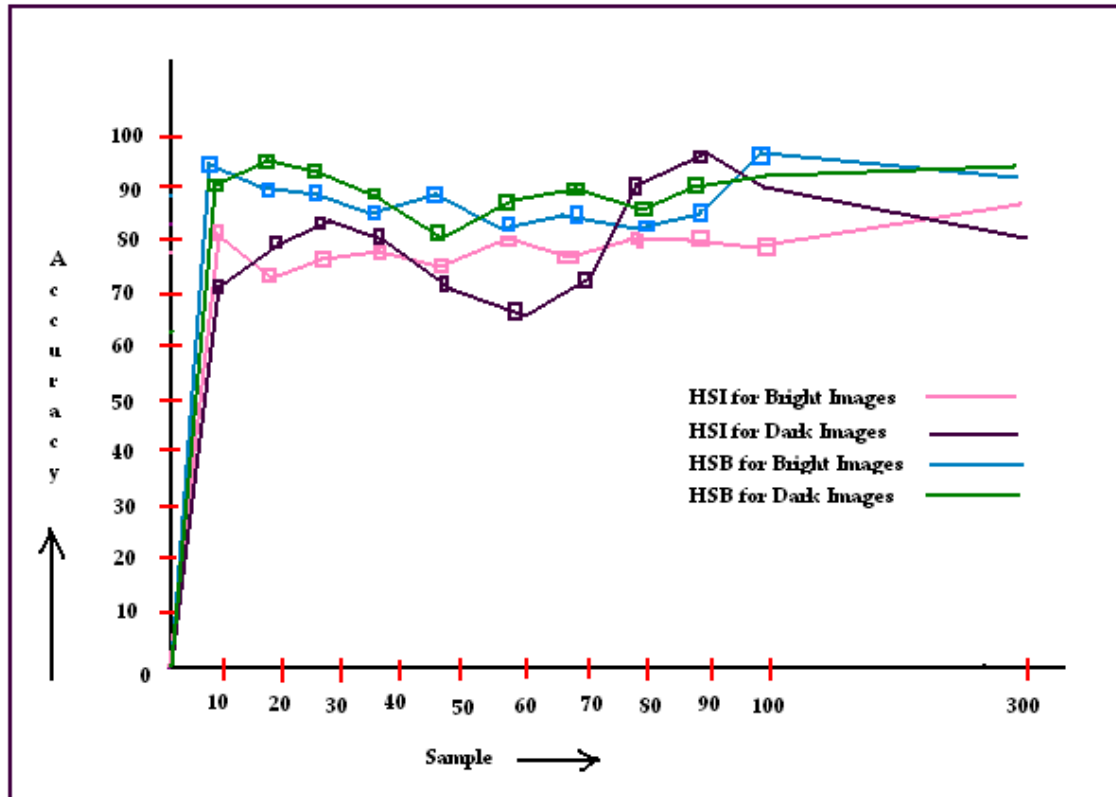


FIGURE 5: (Represents Table 3 in Graphical Form).

From the above Table 3 and Fig: 5; it is clear that HSB works better in compare to HSI for both bright and dark images. Moreover, for dark images using HSI only 79% accuracy is obtained whereas using HSB 89% accuracy has been obtained which proves that especially for dark images use of HSB will be the best approach for image enhancement. For bright images there is accuracy difference of 5% between HSI and HSB and hence it can be said that HSB performs better. However, special care is important when enhancing bright images.

8. CONCLUSION

This paper has proposed an efficient color model for better gamma encoding in image processing from all the existing color models available at this moment. It is difficult to judge an enhanced image result even with a subjective assessment. We claim that HSB color model is more robust than HSI color model or from others because others do produces unrealistic colors and/or over enhanced resultant images. However, there may be still some areas needs to be taken care of as the color enhancement needs to change or shift color using hue although these cases are exceptional and very rare.

9. REFERENCES

- [1] C. Solomon, T. Breckon. 2011. "Fundamentals of Digital Image Processing".
- [2] Nishu, Sunil A. 2012, "Quantifying the defect visibility in digital images by proper color space selection", International journal of engineering research and applications, vol.2, Issue 3, pp. 1764-1767.
- [3] Raunaq M. and Utkarsh U., 2008, "Hue-preserving color image enhancement without gamut problem", Term paper, pp. 1-6.

- [4] Yusuf Abu S., Nijad Al-Najdawi, Sara T., 2011, "Exploiting Hybrid methods for enhancing digital X-Ray Images", International Arab journal of information technology, vol. 8.
- [5] Umesh R., Zhou W., Eero P. S., 2009, "Quantifying color image distortions based on adaptive spatio-chromatic signal decompositions", IEEE international conference on image processing.
- [6] Hana Al-Nuaim, Nouf A., 2011, "A user perceived quality assessment of lossy compressed images", International journal of computer graphics, vol. 2, No. 2, pp. 23-36.
- [7] MD. Zahid Hasan, T. M. Shahriar Sazzad, Md. Hasibur Rahman, November, 2012, "Use of Gamma Encoder for Image Processing considering Human Visualization", International Journal of Computer Applications, volume 58, number 10/9315-3547.
- [8] T. M. Shahriar Sazzad , MD. Zahid Hasan, Fatma Mohammed, December 15, 2012, "Gamma encoding on image processing considering Human visualization, analysis and comparison", International Journal for Computer Science and Engineering.
- [9] R.C. Gonzalez and R.E. woods, 2007, "Digital Image Processing", 3rd Edition, Prentice Hall, Upper Saddle River, NJ.
- [10] Jian-feng Li, Kaun-Quan Wang, David Zhang, 2002, "A New equation of saturation in RGB-TO-HIS conversion for more rapidity of computing", Proceedings of the international conference on machine learning and cybernetics, pp. 1493-1497.
- [11] Papoulis, A., 1968, "Systems and Transforms with Applications in Optics", New York: McGraw-Hill.
- [12] Russ, J.C., 1995, "The Image Processing Handbook". Second ed., Boca Raton, Florida: CRC Press.
- [13] R. E. Blake, 1999, "Partitioning Graph Matching with Constraints", Pattern Recognition, Vol 27, No.3, pp 439-446.
- [14] J. Foley, A. van Dam, S. Feiner and J. Hughes, 1990, "Computer Graphics: Principles and Practice", Second Edition, Addison-Wesley, Reading, MA.
- [15] R. E. Blake and P. Boros, 1995, "The Extraction of Structural Features for Use in computer Vision". Proceedings of the Second Asian Conference on Computer Vision, Singapore.
- [16] V. Bozuloiu, M. Ciuc, R. M. Rangayyan, & C. vertan, 2001, "Adaptive neighborhood histogram equalization of color images", International Journal of Electron Image, Vol 10, No.2, pp 445-459.
- [17] P. E. Trahanias, & A. N. Venetsanopoulos, 1992, "Color image enhancement through 3-D histogram equalization", Proc. 11th IAPR Conf. on Pattern Recognition, The Hague, Netherlands, pp 545-548.
- [18] B. A. Thomas, R. N. Strickland, & J. J. Rodriguez, 1997, "Color image enhancement using spatially adaptive saturation feedback", Proc. 4th IEEE Conf. on Image Processing, Santa Barbara, CA, USA, pp 30-33.
- [19] A. Gupta, & B. Chanda, 1996, "A hue preserving enhancement scheme for a class of color images", Pattern Recognition Letters, Vol 17, No. 2, pp 109-114.

- [20] L. Lucchesse, S. K. Mitra, & J. Mukherjee, 2001, "A new algorithm based on saturation in the xy-chromaticity diagram for enhancement and re-rendition of color images", Proc. 8th IEEE Conf. of Image Processing, Thessaloniki, Greece, pp 1077-1080.
- [21] M. S. Shyu, & J. J. Leou, 1998, "A genetic algorithm approach to color image enhancement", International Journal of Pattern Recognition, Vol 31, No. 7, pp 871-880.
- [22] Y. Kobayashi, & T. Kato, 1999, "A high fidelity contrast improving model based on human vision mechanism", Proc. on Multimedia Computing and Systems, Florence, Italy, pp 578 – 584.
- [23] R. N. Strickland, C. S. Kim, & W. F. McDonnell, 1987, "Digital color image enhancement based on saturation component", International Journal of Optical Engineering, Vol 26, No. 7, pp 609-616.
- [24] Jobson D. J., Rahman Z., & Woodell G. A., "2002, Statistics of visual representation", Proc. SPIE Conf. on Visual Information processing XI, Orlando, FL, USA, pp 25-35.
- [25] L. Lucchesse, S. K. Mitra, & J. Mukherjee, 2001, "A new algorithm based on saturation in the xy-chromaticity diagram for enhancement and re-rendition of color images", Proc. 8th IEEE Conf. of Image Processing, Thessaloniki, Greece, pp 1077-1080.

Selective Median Switching Filter for Noise Suppression in Microstructure Images of Material

P.S.Hiremath

*Professor, Department of Computer Science
Gulbarga University
Gulbarga, 585106, India*

hiremathps53@yahoo.com

Anita Sadashivappa

*Asst. Professor, Department of Computer Science and Engineering
PDA College of Engineering
Gulbarga, 585102, India*

anitaharsoor@gmail.com

Abstract

The image pre-processing is very critical and important task in any digital image analysis system. The eventual success and failure of image analysis depends on the performance of pre-processing techniques applied on the image to be analyzed. In digital images, different noise types are noticed and to attenuate each type of noise, different pre-processing methods have been proposed in literature. The main focus of this paper is on pre-processing the microstructure images. Among many types of noise, impulse noise is the one which is generally noticed in microstructure images. This paper is to present a novel, efficient and suitable pre-processing method for negotiating the impulse noise that is generally present in microstructure images. Through this paper, a new filtering method, selective median switching filter (SMSF) has been proposed. The proposed method is compared with filtering methods those belong to median filter family for their efficiency in negotiating with impulse noise. The efficiency of proposed method is compared with other methods by computing the three image quality assessment methods, namely, mean square error (MSE), peak signal-to-noise ratio (PSNR) and correlation coefficient. The experimental results confirm that the proposed SMSF method is efficient in handling the impulse noise present in microstructure images of material. Also, the proposed SMSF method is quite efficient in preserving the edge information in images.

Keywords: Pre-processing, SMSF, Median Filter, MSE, PSNR, Correlation Coefficient, Microstructure.

1. INTRODUCTION

In general, the images to be processed are not in process-ready condition. Therefore, in any digital image processing application, one or the other pre-processing method is a mandatory first step. In this paper, the microstructure images of cast iron material are considered for study. Some of the sample microstructure images are shown in the Figure 1. There is direct relationship between the microstructure of material and the properties of the material [1]. Therefore, for the analysis of microstructure images, the quality of microstructure image matters. It is generally observed that the microstructure images of material suffer from impulse noise. The amount of noise varies from one material sample to another. Accurate analytical reports can be expected only from good quality microstructure images.

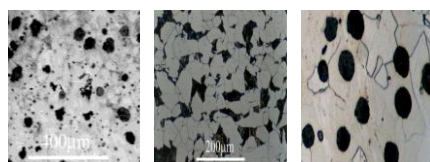


FIGURE 1: Sample Microstructure Images of Cast Iron.

2. IMPULSE NOISE

Impulse noise corruption is very common in digital images. Impulse noise is always independent and uncorrelated to the image pixels and is randomly distributed over the image. Hence unlike the Gaussian noise, for an impulse noise corrupted image all the image pixels are not noisy, a number of image pixels will be noisy and the rest of pixels will be noise free. The 'salt and pepper' and random valued impulse noise are the two different types of impulse noise types. In 'salt and pepper' type of noise, the noisy pixels takes either salt value (gray level 225) or pepper value (grey level 0) and it appears as black and white spots on the images. The presence of impulse noise affects the segmentation results [2,5-8]. False regions are made by considering isolated noisy pixels present in the image and these false contours affect the qualitative and quantitative analysis results.

2.1 De-Noising Methods

The noise usually corrupts images by replacing some of the pixels of the original image with new pixels having luminance values near or equal to the minimum or maximum of the allowable dynamic luminance range. In most of the applications, it is very important to remove impulse noise from image data, since the performances of subsequent image processing tasks are strictly dependent on the success of image noise removal operation. However, this is a difficult problem in any image processing system because the restoration filter must not distort the useful information in the image and preserve image details and texture while removing the noise.

Many methods have been proposed in the literature to deal with impulse noise [3,9,12,18]. Median filtering methods are the general choice to deal with impulse noise. There are many variants of median filtering methods, for experimentation, we have considered most commonly preferred median filtering methods, namely, median filter [9], weighted median filter [11,12,14,20], switching median filter [19]. Each one of the median filter variant has inherent merits and drawbacks. The basic median filter is a simple rank selection filter that attempts to remove impulse noise by changing the luminance value of the center pixel of the filtering window with the median of the luminance values of the pixels contained within the window. Although the median filter is simple and provides a reasonable noise removal performance, it removes thin lines and blurs image details even at low noise densities. The weighted median filter [12,17,19] and the center-weighted median filter (CWM) [10,14] are modified median filters[13,16] that give more weight to the appropriate pixels of the filtering window. These filters have been proposed to avoid the inherent drawbacks of the basic median filter by controlling the tradeoff between the noise suppression and detail preservation. The switching median filter is composed of the median filter with an impulse detector. In this approach, the impulse detector aims to determine whether the center pixel of a given filtering window is corrupted or not. If the center pixel is identified by the detector as a corrupted pixel, then it is replaced with the output of the median filter, otherwise, it is left unchanged. Some extensions of the basic switching median filter including multiple median-based filters in the structure have also been proposed. Adaptive center weighted median (ACWM) [11,15] filter that avoids the drawbacks of the CWM filters and switching median filters and input data will be clustered by scalar quantization (SQ) method, that is resulted in fix threshold for all of images, but modified adaptive center weighted median (MACWM) filter will be used from FCM method, then bound between clusters for any image achieved by information of same image, as a result, clustering of input data to M block would be done better. This study has motivated to test their performance on microstructure images and forced to propose a modified method of median filtering, which is reliable, computationally light and preserve the edge information to a greater extent.

Having known that the microstructure images have the impulse noise, the median filters have been experimented for their optimum performance. The median filters experimented are, basic median filter, weighted median filter, center weighted median filter, switching median filter.

The proposed method includes systematic analysis of performance of each of the noise suppressing filters mentioned in Section 2 in presence of impulse noise of different distributions. Each time after applying the noise suppression filter, means square error (MSE), peak signal-to-

noise ratio (PSNR) and correlation coefficient (CC) are determined. The MSE and PSNR are computed by using the Eqs. (1) and (2), respectively:

$$MSE = \frac{1}{MN} \sum_{y=1}^M \sum_{x=1}^N (I_2(x, y) - I_3(x, y))^2 \tag{1}$$

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

where 255 is the peak gray-level of the grayscale image, I_2 represents the grayscale image, and I_3 represents the filtered image; each of dimension $M \times N$. The summation in Eq.(1) is over all the pixels (x,y) of the images. The PSNR value computed for the filtered image indicates the performance of the filtering method.

The correlation coefficient (CC) is a measure that determines the degree to which two variable's (images) variations are associated. The CC is computed by using the Eq. (3).

$$CC = \frac{\sum_{y=1}^M \sum_{x=1}^N (I_2(x, y) - \bar{I}_2)(I_3(x, y) - \bar{I}_3)}{\sqrt{\sum_{y=1}^M \sum_{x=1}^N (I_2(x, y) - \bar{I}_2)^2} \sqrt{\sum_{y=1}^M \sum_{x=1}^N (I_3(x, y) - \bar{I}_3)^2}} \tag{3}$$

where \bar{I}_2 is the mean intensity of image I_2 and \bar{I}_3 is the mean intensity of image I_3 . The correlation coefficient is a number between 0 and 1. If there is no relationship between the two variables the correlation coefficient is 0 or very low. As the strength of the relationship between the two variables increases so does the correlation coefficient. A perfect fit gives a correlation coefficient of 1.

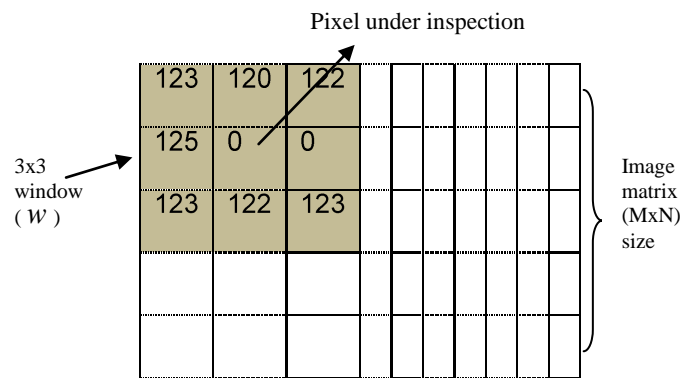


FIGURE 2: Image Matrix With Intensity Values.

2.2 Proposed Selective Median Switching Filter (SMSF)

The SMSF method is a modified method of basic switching median filter. The SMSF is a point-wise linear operation, it operates on a 3×3 neighborhood pixel window (w) around the pixel under inspection (Figure 2). Each pixel undergoes a noise-check process and it is replaced by new intensity value when it is found to be a noise pixel.

The proposed method is given in the following algorithm.

Algorithm: Selective median switching filter

- Step 1: Input RGB microstructure image I_1 and convert to grayscale image I_2 .
- Step 2: Consider 3x3 window W at top-left corner.
- Step 3: Find the median, W_{median} of window elements of W .
- Step 4: Determine the difference (D_i) of each element of window W from the central pixel element of window W .
- Step 5: Determine the mean, D_{mean} , of D_i .
- Step 6: Determine the mean, W_{mean} , of elements of W .
- Step 7: If $D_{mean} \geq W_{mean}$ then replace the central pixel value of window W by W_{median} value; otherwise, leave central pixel value of window W unchanged.
- Step 8: Repeat Steps 2 to 7 by moving the window left to right row-wise over the entire image I_2 to yield the filtered image I_3 .
- Step 9: Output the filtered image I_3 .

3. EXPERIMENTAL RESULTS AND DISCUSSION

For experimentation, we have used microstructure images of various compositions of cast iron drawn from the microstructures library [4]. Each image is added with ‘salt and pepper’ noise in the range 5% to 60% and then subjected to proposed SMSF method for noise suppression. The performance of the proposed filter and other filtering methods is assessed by determining the peak signal-to-noise ratio (PSNR), mean square error (MSE) and correlation coefficients (CC) for the filtered images. The comparison of performance of proposed filter SMSF with other filters is shown in the Figure 3. For better image quality, the PSNR and CC values are higher and MSE values are lower. The value of D_{mean} gives the dispersion of window elements around the central pixel value in eight orientations, while the W_{mean} gives the dispersion around the average pixel values of the window elements.

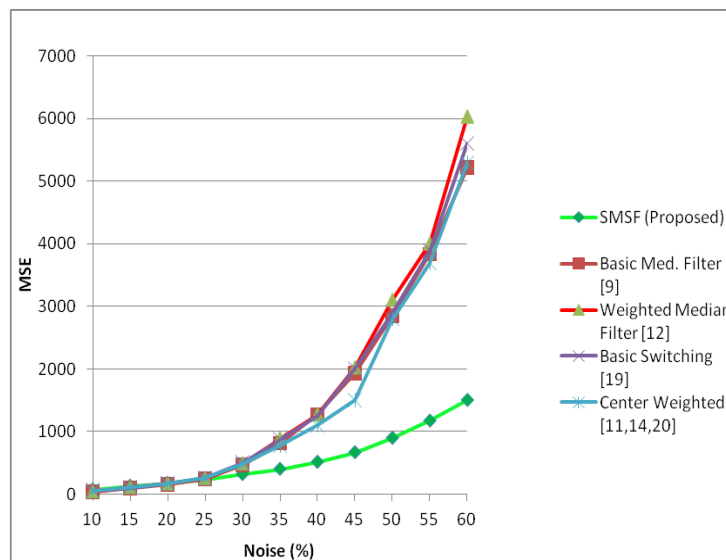


FIGURE 3 (a): Performance Comparison of Proposed Method SMSF And Other Filters In Terms of MSE.

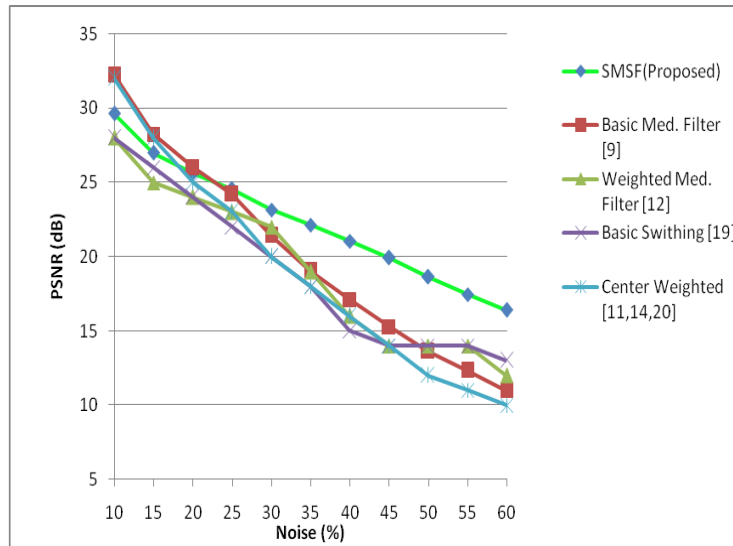


FIGURE 3 (b): Performance Comparison of Proposed Method SMSF And Other Filters In Terms of PSNR.

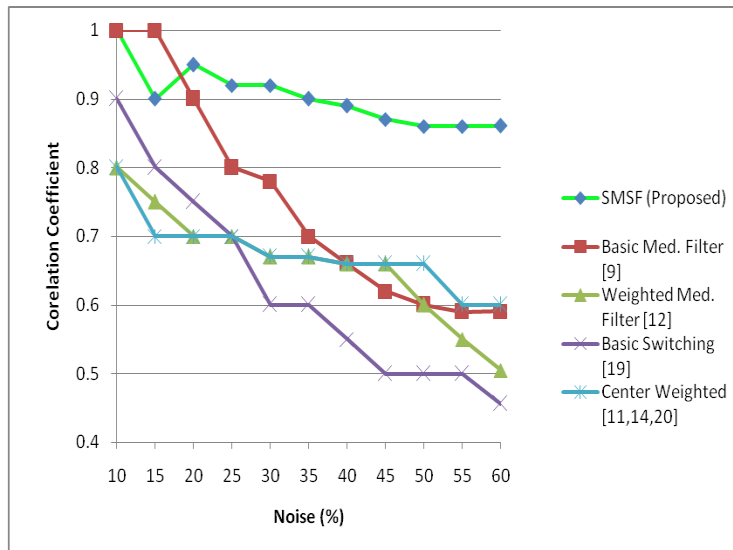


FIGURE 3 (c): Performance Comparison of Proposed Method SMSF And Other Filters In Terms of CC.

It is observed that the proposed method SMSF is capable of preserving the edges of regions of interest (ROI) present in the images to a greater extent as compared to the basic median filter method. The edge profile of regions present in the filtered image is shown in the Figure 4. In the basic median filter method, the central pixel value of the window is replaced unconditionally by median of window elements and hence the edge information is lost. Whereas in the proposed SMSF method, a selective replacement of central pixel value is done based on the condition $D_{mean} > W_{mean}$ in the inspecting window (Step 7 in the Algorithm). It is observed that, in case of uniform regions, the D_{mean} value is less than the W_{mean} , and hence, the centre pixel is not replaced. In case of noisy central pixel value, D_{mean} value is much higher than the W_{mean} and therefore, the central pixel value is replaced by median value. In case of edges, the 3x3 inspection window shows majority of pixel intensity values close to each other, so that

$D_{mean} < W_{mean}$. Hence, the central pixel is not replaced by median value, and therefore, the edges are not blurred.

The proposed method is tested on 50 microstructure images of materials of various compositions and also 20 standard test images used for denoising experiments, e.g. Lena, pout, coin, rice, etc. The sample test images and corresponding resultant filtered images obtained by the proposed method are shown in the Figure5.

Effect of noise on segmentation

The impulse noise affects image segmentation and quantification results. It is observed that the segmentation time with active contours method is directly proportional to the amount of noise present in the image. In an image with 5% of impulse noise, 700 to 800 numbers of iterations are necessary to segment. When, the same image is corrupted by 10% of impulsive noise, the segmentation process requires 1100 to 1300 iterations.

The enhancement of image quality leads to better segmentation results. This is demonstrated by applying active contour method to the SMSF filtered image. The Figure6 shows the segmentation results obtained by active contour method applied to SMSF filtered microstructure image. The segmented images show that there are no false contours or loss of boundary information in the segmented regions. Therefore, one can expect greater accuracy in quantification results in further image analysis procedures.

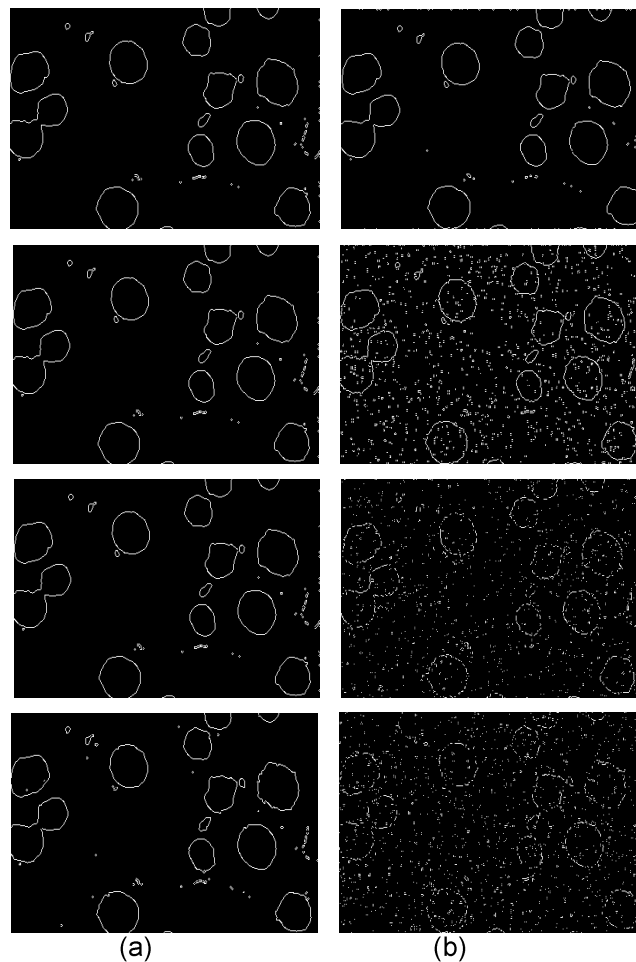


FIGURE 4: Comparison of Edge Preserving Property of Basic Median Filter And Proposed Method At Different Noise Levels, 10%,20%, 30% and 35% (top to bottom). (a) Edge Information After Applying The Proposed Method And (b) Edge Information After Applying The Basic Median Filtering Method.

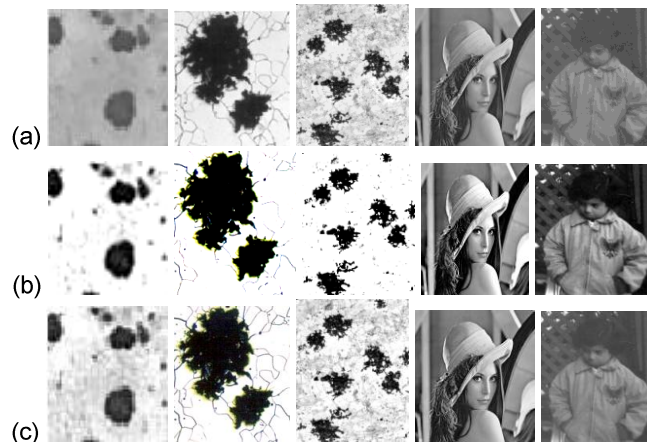


FIGURE 5 (a) Sample Test Images, (b) Resultant Filtered Images Obtained By Proposed SMSF Method And (c) Resultant Filtered Images Obtained by Basic Median Filter Method.

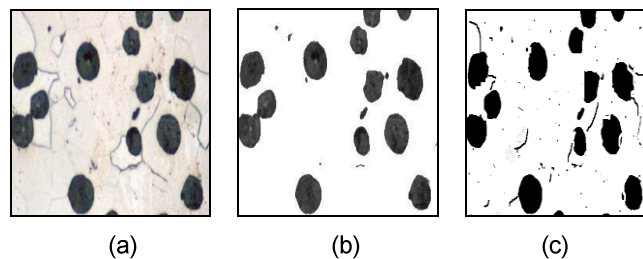


FIGURE 6: (a) Original Microstructure Image, (b) Segmented Image Obtained By Active Contour Method And SMSF Filter and (c) Segmented Image Obtained By Active Contour Method And Basic Median Filter.

4. CONCLUSION

Through this paper, a novel filtering method by name, selective median switching filter for suppressing the impulse noise has been proposed (SMSF). The results of filtering method are quite encouraging and it has proved to be more efficient than the other median filtering methods that are generally preferred to deal with the impulse noise. Unlike other median filtering methods, the proposed method has property of preserving edges while acting on impulse noise quite efficiently. Also, the proposed method plays vital role in quality of the segmentation of images and reduces the amount of time required for segmentation significantly.

5. REFERENCES

- [1] ASM Committee. ASM Handbook: Metallography and Microstructures, Vol.9. USA:ASM International, 2004.
- [2] Gonzalez R.C. and Woods R.E. Digital Image Processing, 3rd Ed. USA:Pearson Publication, 2008.
- [3] H.Sadoghi Yazdi, F.Homayouni. "Impulsive noise suppression of images using adaptive median filter". International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 3, No. 3, pp. 1-12, 2010.
- [4] University of Cambridge. "Micrograph library". Internet:www.doitpoms.ac.uk, Jan 5,2013.
- [5] M. Sonka, V.Hlavac and R. Image Processing, Analysis, and Machine Vision, 3rd Ed. USA: CL Engineering, 2007.

- [6] Pattan Prakash, V.D.Mytri and P.S. Hiremath. "Active Contour Multigrid Model for Segmentation and Automatic Quantification of Material Phases of Cast Iron". International Journal of Computer Applications, Vol. 9. No.4, pp.32-37, 2010.
- [7] Pattan Prakash, V.D.Mytri and P.S. Hiremath. "Performance Analysis of Segmentation Methods for Automatic Quantification of Phases of Cast Iron" in proc. International conference on Computational Vision and Robotics (ICCV), 2010, pp. 174-180.
- [8] J.C. Russ. The Image Processing Cook Book, 2nd Ed., Chapters 2,3,4 and 5,USA:CRC Press,2011.
- [9] S. E. Umbaugh. Computer Vision and Image Processing,USA:Prentice-Hall,1998.
- [10]S.-J. Ko and Y. H. Lee. "Center weighted median filters and their applications to image enhancement". Circuits and Systems (IEEE Transactions), Vol. 38, No. 9, pp.984–993,1991.
- [11]T.C. Lin, P.T. Yu. "A new adaptive center weighted median filter for suppressing impulsive noise in images". International Journal of Information Sciences (Elsevier), No.177, pp.1073–1087, 2007.
- [12]O. Yli-Harja, J. Astola, and Y. Neuvo. "Analysis of the properties of median and weighted median filters using threshold logic and stack filter representation", Signal Processing (IEEE Transactions), Vol. 39, No. 2, pp. 395–410,1991.
- [13]Chen, T. and H.R. Wu. "Application of partition-based median type filters for suppressing noise in images". Image Processing (IEEE Transactions), Vol.6, pp.829-836, 2001.
- [14]Ko, S.J. and Y.H. Lee. "Center weighted median filters and their applications to image enhancement. Circuits and Systems (IEEE Transactions), Vol.38, pp.984-993.1991.
- [15]Lin, T.C. "A new adaptive center weighted median filter for suppressing impulsive noise in images". Journal of Information Sciences (Elsevier), Vol.177, pp. 1073-1087, 2007.
- [16]Lukac, R. "Performance boundaries of optimal weighted median filters". International Journal of Image Graphics, Vol.4,pp.157-182, 2004.
- [17]Smolka, B. and A. Chydzinski. "Fast detection and impulsive noise removal in color images". International Journal of Real-Time Imaging (Science Direct), Vol.11,pp.389-402, 2005.
- [18]Sun, T. and Y. Neuvo. 1994. "Detail-preserving median based filters in image processing", Pattern Recognition Letters, Vol.15, pp. 341-347, 1994.
- [19]Wang, G., D. Li, W. Pan and Z. Zang. "Modified switching median filter for impulse noise removal. Signal Process", Vol.90, pp.3213-3218, 2010.
- [20]Yin, L., R. Yang, M. Gabbouj and Y. Neuvo. "Weighted median filters: A tutorial". Circuits and Systems. II: Analog Digital Signal Processing, Vol. 43: 157-192, 1996.

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. Started with volume 7, 2013, IJIP is appearing with 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/p

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