International Journal of Image Processing (IJIP)

International Journal of Image Processing (IJIP)

Volume 4, Issue 6, 2011

Editor in Chief Professor Hu, Yu-Chen

International Journal of Image Processing (IJIP)

CSC Publishers

Editorial Preface

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

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

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

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

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

Editorial Board Members
International Journal of Image Processing (IJIP)

Table of Content

Volume 4, Issue 6, December 2011

Pages

Omnidirectional Thermal Imaging Surveillance System Featuring Trespasser and Faint Detection

Wai-Kit Wong   wkwong@mmu.edu.my
Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.

Zeh-Yang Chew   chew5011@msn.com
Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.

Hong-Liang Lim   fly_dragon85@hotmail.com
Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.

Chu-Kiong Loo   ckloo@mmu.edu.my
Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.

Way-Soong Lim   wslim@mmu.edu.my
Faculty of Engineering and Technology,
Multimedia University,
75450 JLN Ayer Keroh Lama,
Melaka, Malaysia.

Abstract

This paper proposed an efficient omnidirectional thermal imaging surveillance system featuring trespasser and faint detection. In this thermal imaging system, the omnidirectional scenes in a monitored site such as old folks home, nursing home, hospital etc. are first captured using a thermal camera attached to a custom made hyperbolic IR (infrared radiation) reflected mirror. The captured scenes to be monitored with trespasser or faint detection are then fed into a laptop computer for image processing and alarm purposes. Log-polar mapping is proposed to map the captured omnidirectional thermal image into panoramic image, hence providing the observer or image processing tools a complete wide angle of view. Two effective human behavioral detection algorithms namely: Human head detection algorithm and home alone faint detection algorithm are also designed for monitored the trespasser or fainted people detection. The observed significances of this new proposed omnidirectional thermal imaging system include: it can cover a wide angle of view (360º omnidirectional), using minimum hardware, low cost and the output thermal images are with higher data compression. Experimental results show that the proposed thermal imaging

surveillance system achieves high accuracy in detecting trespasser and monitoring faint detection for health care purpose.

Keywords: Monitoring and Surveillance, Thermal Imaging System, Trespasser Detection, Faint Detection, Omnidirectional System, Image Processing and Understanding.

1. INTRODUCTION
Conventional surveillance system that employs digital security cameras are great to keep residences safe from thief, vandalism, unwanted intruders and at the meantime can work as a health care facility for nursing purposes. However, such surveillance system normally employs human observers to analyze the surveillance video. Sometime this is more prone to error due to lapses in attention of the human observer [1]. There is a fact that a human's visual attention drops below acceptable levels when assigned to visual monitoring and this fact holds true even for a trained personnel [2],[3]. The weakness in conventional surveillance system has raised the need for a smart surveillance system where it employs computer and pattern recognition techniques to analyze information from situated sensors [4].

Another problem encountered in most conventional surveillance systems is the change in ambient light, especially in an outdoor environment where the lighting condition is varies naturally. This makes the conventional digital color images analysis task very difficult. One common approach to alleviate this problem is to train the system to compensate for any change in the illumination [5]. However, this is generally not enough for human detection in dark. In recent time, thermal camera has been used for imaging objects in the dark. The camera uses infrared (IR) sensors that capture IR radiation coming from different objects in the surrounding and forms IR image [6]. Since IR radiation from an object is due to the thermal radiation, and not the light reflected from the object, such camera can be conveniently used for trespasser or faint detection in night vision too.

Thermal imaging trespasser detection system is a type of smart surveillance system which is used to detect human objects (trespasser) even in poor lighting condition. The system can be employed to secure a place when and where human should not exist. A simple trespasser detection algorithm was proposed in [7]. The algorithm is regional based whereby an IR object that occupy more than certain number of partitions in a thermal image is considered as a human being and vice-versa. However, the algorithm is having two major concerns. First concern is distance, in which if a human being that is far away from the imaging system will not identified as a human. The second concern is if an animal (such as cat, dog etc.) is moving too close to the system (which occupy more than the threshold partitioned), it will be miss-considered as a human being too. Therefore, in this paper, a more effective trespasser detection algorithm with human head detection capability is proposed.

The second approach for the proposed omnidirectional thermal imaging surveillance system in this paper is with faint detection feature. Faint is one of the major problems that happen amongst the elderly people, patients or pregnant women which may cause them suffering physical injuries or even mental problems. Fainting normally occurs when the person falls and his or her head hits on the floor or on hard items. An emergency medical treatment for fainting mainly depends on the response and rescue time. Therefore, detection of faint incidents is very important in order to have the immediate treatment for this population. Various solutions have been developed to recognize faint motion. One of the common ways is using the wearable press button. This allows the person who had fall down to press the button to call for help. However, this system does not help if the person faint instantly. Furthermore, wearable faint motion sensor which makes use of the acceleration sensor and inclination sensor to recognize the motion automatically subsequently signals the alarm [8]. But, this system will not help if the person forgets to wear it. Thus, a possible solution is the use of automatic surveillance vision based systems.

Recently, thermal camera has been used for moving human detection [9], but it is not in omnidirectional view. If a single thermal camera is to monitor a single location, then for more

locations in different angle of view, there required more thermal cameras. Hence, it will cost more, beside complicated the surveillance network. In this paper, an effective surveillance system is proposed which includes three main features:

1) 360 degree viewing using a single thermal camera, surrounding location can be monitored. This achieving wide area coverage using minimum hardware.
2) Effective trespasser detection system able to detect trespassers, even in a poor lighting condition.
3) Effective automatic health care monitoring system that will raise alerts/alarm whenever any human fainted case arises.

Experimental results show that the proposed thermal imaging surveillance system achieves high accuracy in detecting trespasser and monitoring faint detection for health care purpose. The paper is organized as follows: Section 2 briefly comments on the omnidirectional thermal imaging surveillance system, section 3 summarizes log-polar image geometry and the mapping techniques for unwarping the captured omnidirectional thermal image into panoramic form. Then in section 4, it presents the proposed human head detection algorithm for trespasser detection and section 5 shows the home alone faint detection algorithm for health care surveillance purpose. Section 6 contains the experimental results. Finally in section 7 will draw conclusion and envison of future developments.

2. OMNIDIRECTIONAL THERMAL IMAGING SURVEILLANCE SYSTEM MODEL

The proposed omnidirectional thermal imaging surveillance system model in this paper is shown in Fig. 1. This system requires a custom made IR reflected hyperbolic mirror, a camera mirror holder, a fine resolution thermal camera and a laptop or personal computer installed with Matlab programming (version R2007b or later) and an alarm signaling system. The alarm signaling system can be as simple as a computer's speaker.
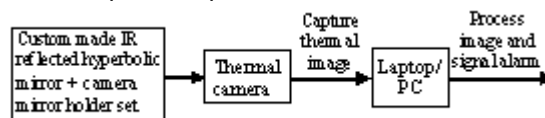
FIGURE 1: Omnidirectional Thermal Imaging Surveillance System Model

A. Custom made IR reflected hyperbolic mirror
The best shape of practical use omnidirectional mirror is hyperbolic. As derived by Chahl and Srinivasan in [10], all the polynomial mirror shapes (conical, spherical, parabolic, etc) do not provide a central perspective projection, except for the hyperbolic one. They also shown that the hyperbolic mirror guarantee a linear mapping between the angle of elevation θ and the radial distance from the center of the image plane ρ. Another advantage of hyperbolic mirror is when using it with a camera/imager of homogenous pixel density, the resolution in the omnidirectional image captured is also increasing with growing eccentricity and hence it will guarantee a uniform resolution for the panoramic image after unwarping.

The research group of OMNIVIEWS project from Czech Technical University further developed MATLAB software for designing omnidirectional mirror [11]. From the MATLAB software, omnidirectional hyperbolic mirror can be designed by inputting some parameters specify the mirror dimension. The first parameter is the focal length of the camera *f*, in which for the thermal camera in use is 12.5 mm and the distance *d* (ρz-plane) from the origin is set to 2 m. The image plane height *h* is set to 20 cm. the radius of the mirror rim is chosen t_1=3.6cm as modified from Svoboda work in [12], with radius for fovea region 0.6 cm and retina region 3.0 cm. Fovea angle is set in between 0º to 45º, whereas retina angle is from 45º to 135º. The coordinates as well as the plot of the mirror shape is generated using MATLAB and shown in Fig. 2. The coordinates as well as mechanical drawing using Autocad are provided to precision engineering company to fabricate/custom made the hyperbolic mirror. The hyperbolic mirror is milling by using aluminum bar and then chrome plating with a chemical element named chromium. Chromium is regarded with great interest because of its lustrous (good in IR reflection), high corrosion resistance, high melting point and hardness. The fabricated mirror is shown in Fig. 3.
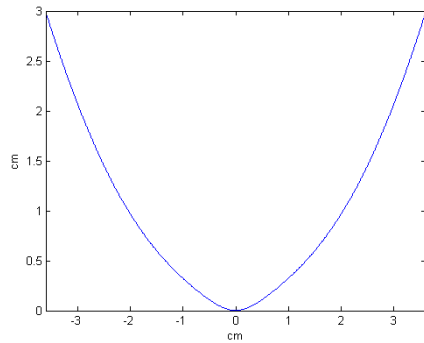
FIGURE 2: Mirror coordinates plot in MATLAB
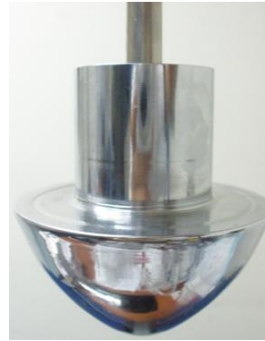
FIGURE 3: Fabricated mirror

B. Camera-mirror holder
The camera mirror holder is self designed and custom made with aluminum material as shown in Fig. 4.

C. Thermal camera
The thermal camera used in this paper is an affordable and accurate temperature measurement mode: ThermoVision A-20M is manufactured by FLIR SYSTEM [13]. The thermal camera has a temperature sensitivity of 0.10 of range from -20ºC to 350ºC. However, for human detection, the temperature range is set to 30ºC to 40ºC. The thermal camera can capture thermal image with fine resolution up to 320 X 240 pixels offering more than 76,000 individual measurement points per image at a refresh rate of 50/60 Hz. The A-20M features a choice of connectivity options. For fast image and data transfer of real-time fully radiometric 16-bit images, an IEEE-1394 FireWire digital output can be chose. For network and/or multiple camera installations, Ethernet connectivity is also available. Each A-20M can be equipped with its own unique URL allowing it to be addressed independently via its Ethernet connection and it can be linked together with router to form a network. Therefore, it is best outfitted for human behavioral based surveillance system, such as trespasser and faint detection.

A problem encountered in thermal camera selection is the existence of the haloing effect in uncalibrated ferroelectric barium strontium titanate (BST) sensors. Haloing effect is the presence of halos around objects having a high thermal contrast with the background [14]. A-20M is chosen because it uses the uncooled microbolometer FPA detector technology which does not produce the haloing effect.

D. Laptop/PC
A laptop or PC can be used for image processor, either place on site or in a monitoring room. Matlab ver R2007b programming is chosen to be used because it has user friendly software for performing log-polar mapping technique to unwrap the omnidirectional thermal image into panoramic form and it can partitioned the panoramic thermal images easily according to each single location to be monitored, process them smoothly with the trespasser or faint detection algorithm user programmed in, and alarm will be triggered once a human being is detected in a tested image for trespasser detection mode or faint motion is detected for faint detection mode. The overall fabricated system model is shown Fig. 4.

FIGURE 4: Overall fabricated omnidirectional thermal imaging surveillance system model

3. LOG-POLAR MAPPING

In this section, log-polar mapping is proposed for unwarping the captured omnidirectional thermal images into panoramic form providing the observer or image processing tools a complete wide angle of view. Log-polar geometry or log-polar transform in short, is an example of foveated or space-variant image representation used in the active vision systems motivated by human visual system [15]. It is a spatially-variant image representation in which pixel separation increases linearly with distance from a central point [16]. It provides a way of concentrating computational resources on regions of interest, whilst retaining low-resolution information from a wider field of view. One advantage of this kind of sampling is data reduction. Foveal image representations like this are most useful in the context of active vision system where the densely sampled central region can be directed to pick up the most salient information. Human eyes are very roughly organized in this way.

In robotics, there has been a trend to design and use true retina-like sensors [17], [18] or simulate the log-polar images by software conversion [19], [20]. In the software conversion of log-polar images, practitioners in pattern recognition usually named it as log-polar mapping. The advantages of log-polar mapping is that it can unwarp an omnidirectional image into panoramic image, hence providing the observer and image processing tools a complete wide angle of view for the surveillance area's surroundings and preserving fine output image quality in a higher data compression manner. The spatially-variant grid that represents log-polar mapping is formed by i number of concentric circles with N samples over each concentric circle [15]. An example of a spatially-variant sampling grid is shown in Fig. 5.

The log-polar mapping use in this paper can be summarized as following: Initially, omnidirectional thermal image is captured using a thermal camera and a custom made IR reflected hyperbolic mirror. The geometry of the captured omnidirectional thermal image is in Cartesian form (x_1, y_1). Next, the Cartesian omnidirectional thermal image is sampled by the spatially-variant grid into a log-polar form (ρ, θ) omnidirectional thermal image. After that, the log-polar omnidirectional thermal image is unwarped into a panoramic thermal image (x_2, y_2), another Cartesian form. Since the panoramic thermal image is in Cartesian form, subsequent image processing task will become much easier.

The center of pixel for log-polar sampling is described by [15]:

$$\rho(x_1, y_1) = \ln_b \left(\frac{R}{\rho_o} \right) \qquad (1)$$

$$\theta(x_1, y_1) = \left(\frac{N}{2\pi} \right) \tan^{-1} \left(\frac{y_1}{x_1} \right) \qquad (2)$$

The center of pixel for log-polar mapping is described as:

$$x_2(\rho, \theta) = \rho \, \cos \left(\frac{2\pi\theta(x_1, y_1)}{N} \right) \qquad (3)$$

$$y_2(\rho, \theta) = \rho \, \sin \left(\frac{2\pi\theta(x_1, y_1)}{N} \right) \qquad (4)$$

where R is the distance between given point and the center of mapping $= \sqrt{x_1^2 + y_1^2}$,

ρ_o is the scaling factor which will define the size of the circle at $\rho(x_1, y_1) = 0$,
b is the base of the algorithm [15],

$$b = \frac{N + \pi}{N - \pi} \qquad (5)$$

N is the number of angular samples over each concentric circle.
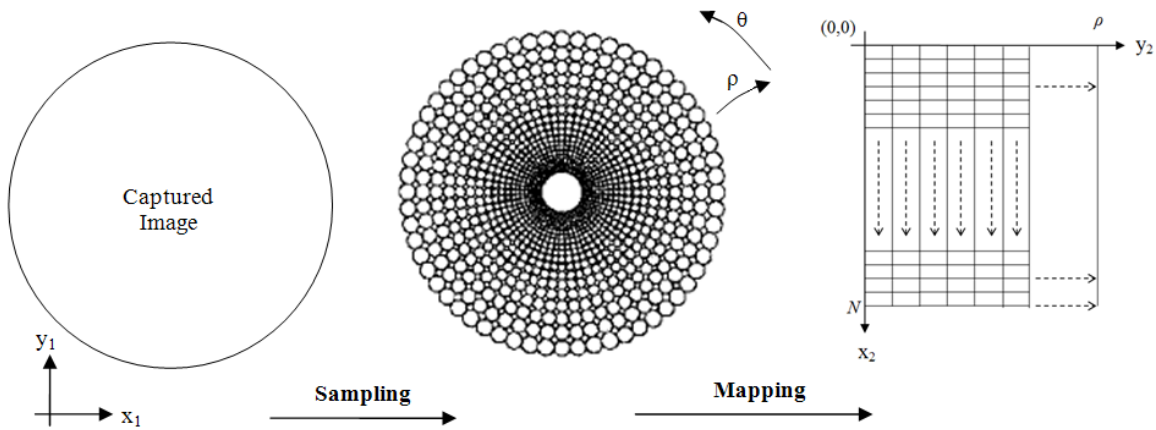
FIGURE 7: A graphical view of log-polar mapping.

A graphical view illustrating the log-polar mapping is shown in Fig. 5 [15]. To sample the Cartesian pixels (x_1, y_1) into log-polar pixel (ρ, θ), at each center point calculated using (1) and (2), the corresponding log-polar pixel (ρ_n, θ_n) covers a region of Cartesian pixels with radius:

$$r_n = br_{n-1} \qquad (6)$$

where n = 1,2,3,.....,N-1. Fig. 6 shows the circle sampling method of log-polar mapping [15], [21], where A, A', B and B' points are the centre of pixel for log-polar sampling.

Wai-Kit Wong, Zeh-Yang Chew, Hong-Liang Lim, Chu-Kiong Loo & Way-Soong Lim
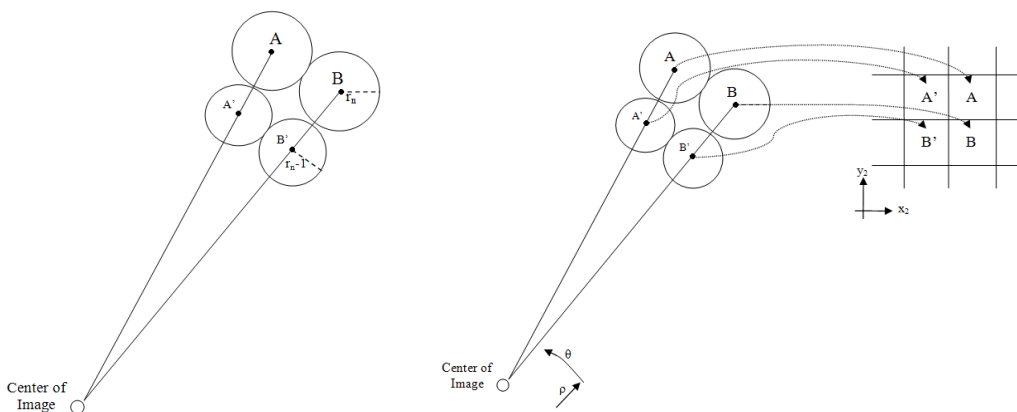
FIGURE 8: Circular Sampling Structure in Log-Polar Mapping.   Fig. 7. Unwarping Process.

The intensity value in each individual log-polar pixel equals the mean intensity value of all pixels inside the sampling circle on the original Cartesian image (x_1,y_1):

$$mean = \frac{total\ (\,x_1,y_1\,)\ pixel\ value}{total\ number\ of\ (\,x_1,y_1\,)\ pixel} \qquad (7)$$

The region of Cartesian pixels on the panoramic image (x_2,y_2) is covered by an individual log-polar pixel on the log-polar (ρ,θ) omnidirectional image. Therefore, the pixels in that specific region on the panoramic image (x_2,y_2) will have the same intensity with respect to the corresponding individual log-polar pixel. Fig. 7 shows the mapping or unwarping of log-polar pixel onto its corresponding Cartesian pixel (x_2,y_2), as described by (3) and (4).

4. HUMAN HEAD DETECTION ALGORITHM FOR TRESPASSER DETECTION

In this section, an effective trespasser detection algorithm modified from L.H.Lee's work in [22] is proposed. The proposed algorithm is discussed in detail as below.

A. Algorithm for Trespasser Detection

Step 1: Acquire thermal image through hyperbolic reflector. Please refer to Fig. 8 for the example of image captured.
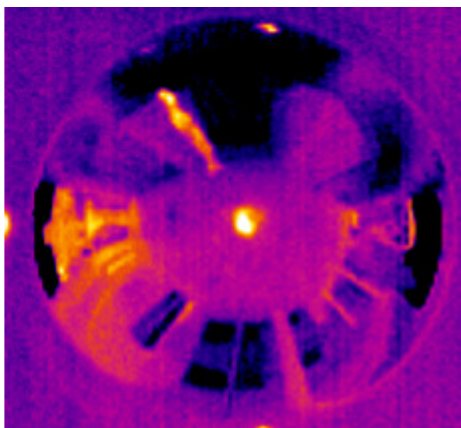
FIGURE 8: Thermal image captured through hyperbolic reflector

Step 2: Image unwarping: Unwarp the acquired thermal image into panoramic image. Please refer to Fig. 9 for the example on resulting panoramic image.
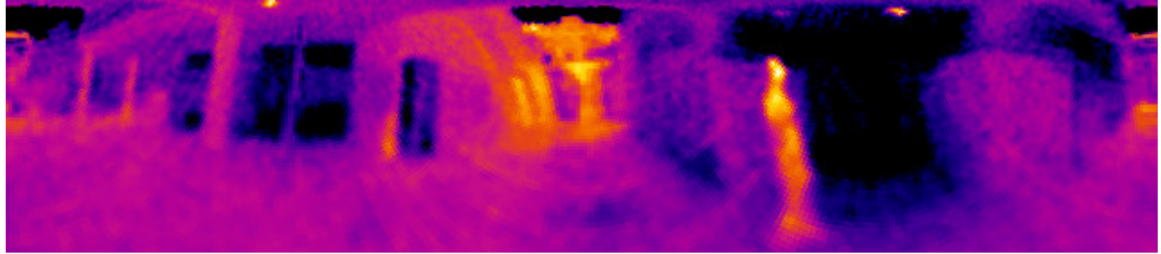
FIGURE 9: Panaromic view of the inspected scence after the thermal image is unwarpped

Step 3: Image cropping: Crop the image to obtain the thermal image of the interested area only. Please refer to Fig. 10 for the example on the resulting cropped image.
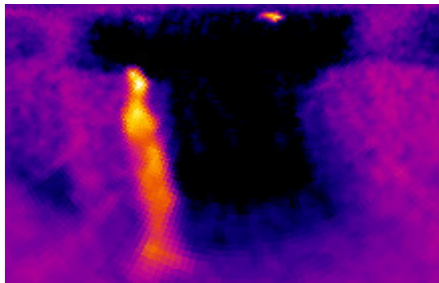
FIGURE 10: Thermal image after cropping process

Step 4: Binary image conversion: Convert Thermal image into purely black & white image (BW) using:

$$BW(i, j) = \begin{cases} 1, T_l \leq Temp(i, j) \leq T_h \\ 0, otherwise \end{cases} \quad (8)$$

Temp (i,j) = Temperature at point (i,j) of thermal image.
T_l and T_h are the minimum and maximum possible human body temperature.
i,j are the pixel's row and column coordinate respectively.

Step 5: Objects identification: Identify objects inside the BW where a group of discontinuous white pixels are considered as a single object.

Step 6: Boundary extraction: Find boundary line of each identified object inside BW and record it into an array of coordinates.

Step 7: Head detection: For each object's boundary, perform the Head detection algorithm as explained in section B.

Step 8: If human being is detected, trigger alarm.

B. Algorithm for Head Detection

Step 1: Starting point identification: Calculate the starting point for head top detection. The starting point shall be the highest point (smallest y-coordinate value) among the intersection points of the boundary with the horizontal middle line where horizontal middle line is given by x = x_m and x_m is the mean of all x-coordinates in the boundary of the object. Please refer Fig. 11(a) for better understanding.

Step 2: Peak points detection: From the starting point, follow the boundary in clockwise direction and search for the first peak point encountered (P_{Peak1}). Again, from the starting point, follow the boundary in anticlockwise direction and search for the first peak point encountered (P_{Peak2}). Peak point is defined as the point in which it has the smallest Y-coordinate value in compare to all the D proceeding points. D is the number of next-of-point to be tested. Refer to Fig. 11(b) for better understanding.
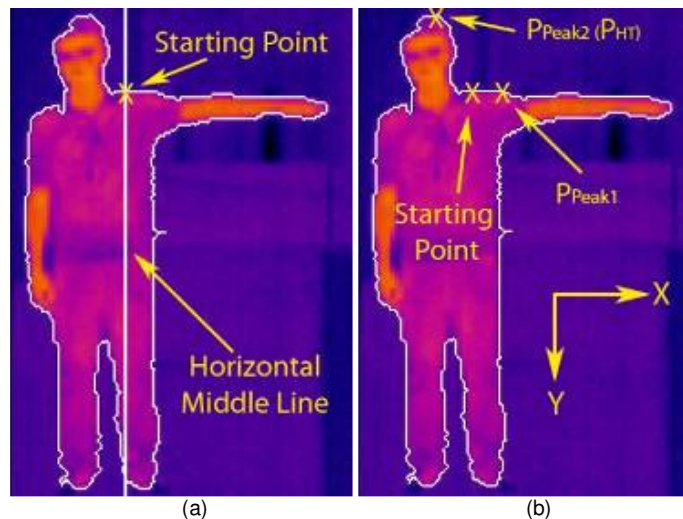
FIGURE 11: (a)Horizontal middle line and the starting point as in step1. (b) Detection of (c/w from Starting point) and (couter c/w from starting point)

Step 3: Head top point detection: Compare P_{Peak1} and P_{Peak2} obtained in step 2. Record the highest point (with smaller Y-coordinate) as the head top point, P_{HT}. For example, in Fig. 11(b), P_{Peak2} is higher thanP_{Peak1}. Thus $P_{HT} = P_{Peak2}$.

Step 4: Boundary line splitting: Split the boundary into left boundary (Bl) and right boundary (Br) from the head top point towards bottom. Take only one point for each y-coordinate to filter out unwanted information such as raised hands. (Refer Fig. 12 for better understanding).
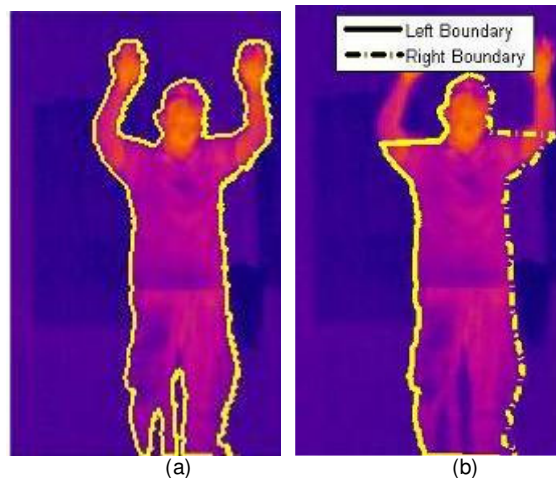
FIGURE 12: a) Original object boundary. (b) Left and Right object boundary after the splitting process (step 4).

Bl=[xl$_i$,yl$_i$] , Br=[xr$_i$,yr$_i$]   (9)

Where xl_i, yl_i, xr_i, yr_i are the pixels' y-coordinates and x-coordinates for left boundary and right boundary respectively.

I=1…N is the index number.

N=size of the boundary matrix (N is the number of pixels for left and right boundary)

Step 5: Left significant points detection: Search downwards along left boundary (Bl) from P_{HT} for the first leftmost point encountered (Pl$_p$). Next, search for the rightmost point right after Pl$_p$ which is Pl$_d$. Refer Fig. 13 for better understanding.

Pl$_p$=(xl$_{lp}$, yl$_{lp}$), Pl$_d$=(xl$_{ld}$, yl$_{ld}$)   (10)

Where subscript lp = index number of the first leftmost point

Subscript ld = index number of the rightmost point right after Pl_p

Step 6: Right significant points detection: Search downwards along right boundary (Br) from P_{HT} for the first rightmost point encountered (Pr_p). Next, detect the leftmost point right after Pr_p which is Pr_d. Refer Fig. 6 for better understanding.

$Pr_p = (xr_{rp}, yr_{rp})$, $Pr_d = (xr_{rd}, yr_{rd})$ (11)
Where subscript rp = index number of the first rightmost point
subscript rd = index number of the first leftmost point right after Pr_p
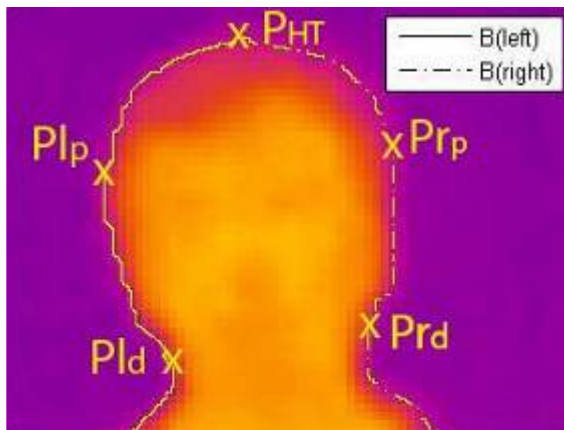
FIGURE 13: Example of points found in step 3, step 5 and step 6.

Step 7: Head symmetric test:
Define h_l = Vertical distance between P_{HT} and Pl_d
h_r = Vertical distance between P_{HT} and Pr_d
Test the ratio between h_l and h_r. If h_l/h_r or $h_r/h_l > 2$, then this object is not considered as a human being and we can skip the next subsequent steps in this algorithm and proceed with the next object. Else, the object has the possibility to be considered as a human being. Continue step 8 for further detection.

Step 8: Neck-body position test: Calculate Δx, which is the distance between x_c and x_m where x_c = horizontal center between Pl_d and Pr_d and x_m is obtained in step1.
Define w_n =horizontal distance between Pl_d and Pr_d.
If $\Delta x \geq 2w_n$, then this object is not considered as a human being and we can skip the next subsequent steps in this algorithm and proceed with next object. Else if $\Delta x < 2w_n$, then the object has the possibility to be considered as a human being. Continue step 9 for further detection.

Step 9: Curve tests:

I)   TOP CURVE TEST.

Define $s_t = floor(min(lp,rp)/(F/2))$ as the step size for top curve test.
Calculate:

$$C_{tl} = \begin{cases} 1, yl_{1+s_t*k} \leq yl_{1+s_t*(k+1)} \\ 0, otherwise \end{cases}$$

$$C_{tr} = \begin{cases} 1, yr_{1+s_t*k} \leq yr_{1+s_t*(k+1)} \\ 0, otherwise \end{cases} \quad (12)$$

$$C_t = \sum C_{tl} + \sum C_{tr}$$

where k=0, …,F/2-1
F is the step size partition variable. F is even integer and F≥2.

For example, if F=6,s_t=8, then the y-coordinates tested is as shown in Fig. 14. The same concept goes for left curve test and right curve test.
Note: the symbol '*' means multiply.
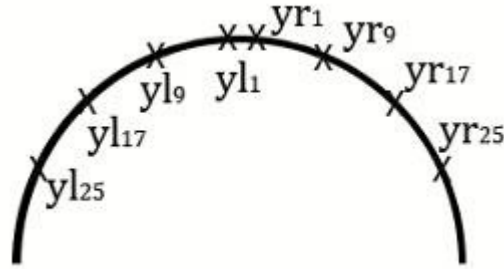
FIGURE 14: Example on top curve test.

II)   <u>LEFT CURVE TEST</u>

Define s_l = floor(min(lp,ld-lp)/(F/2)) as the step size for left curve test.
Calculate:

$$C_{l1} = \begin{cases} 1, xl_{lp+s_l*k} \leq xl_{lp+s_l*(k+1)} \\ \quad 0, \text{otherwise} \end{cases}$$

$$C_{l2} = \begin{cases} 1, xl_{lp-s_l*k} \leq xl_{lp-s_l*(k+1)} \\ \quad 0, \text{otherwise} \end{cases} \quad\quad (13)$$

$C_l = \sum C_{l1} + \sum C_{l2}$
where k=0, …,F/2-1

III)   <u>RIGHT CURVE TEST</u>

Define s_r = floor(min(rp,rd-rp)/(F/2)) as the step size for right curve test.
Calculate:

$$C_{r1} = \begin{cases} 1, xr_{rp+s_r*k} \geq xr_{rp+s_r*(k+1)} \\ \quad 0, \text{otherwise} \end{cases}$$

$$C_{r2} = \begin{cases} 1, xr_{rp-s_r*k} \geq xr_{rp-s_r*(k+1)} \\ \quad 0, \text{otherwise} \end{cases} \quad\quad (14)$$

$C_r = \sum C_{r1} + \sum C_{r2}$
Where k=0, …,F/2-1

Step 10: Human identification:
Define curve test condition:
$C_t \geq F-1$ - (15)
$C_l \geq F-1$ - (16)
$C_r \geq F-1$ - (17)
Check condition (15),(16) and (17). If any two or more conditions are true, the object is verified as a human being. Else, the object is not considered as a human being.

5. HOME ALONE FAINT DETECTION ALGORITHM

In this section, an effective faint detection algorithm for monitoring home alone personal is proposed for the omnidirectional thermal imaging surveillance system.

Step 1: Global parameters definition: Define global parameters to be used in the designed algorithm such as H_p = human's height in previous image, H_c = human's height in current image, W_p = human's width in previous image and W_c = human's width in current image, f = accumulator for human lying image, F = total frames (time) to decide whether a human being is fainted, S = smallest possible human being's size (in terms of number of pixels in group) in an image, H_{max} = maximum height of human of image being capture and initialize it to 0.

Step 2: Image acquisition: Acquire image of background image from thermal camera, unwarp it and store as RGB image, B. Acquire image from thermal camera, unwarp it and store as RGB image, I. An example of B and I is shown in Fig 15.

(a)

(b)
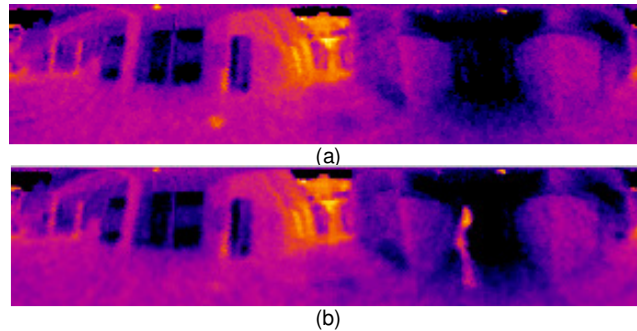
FIGURE 15: Thermal image capture on site corresponding (a) Example of *B* (b) Example of *I*

Step 3: Background subtraction: Subtract image *B* from image *I* and store it as image, *Im.*
Step 4: *Binary image conversion:*
 a) Convert *Im* to gray scale image, *G*. An example of G shown in Fig 16
 b) Convert *G* to binary image, *B1* using minimum possible human body temperature threshold, *T*. An example of *B1* shown in Fig 17.
Step 5: *Noise Filtering:*
 a) Removes from *B1* all connected components (objects) that have fewer than *S* pixels, producing another binary image, *B2*. An example of *B2* is shown in Fig. 18.
 b) Creates a flat, disk-shaped structuring element, *SE* with radius, *R*. An example of structuring element is shown in Fig. 19.
 c) Performs morphological closing on the *B2*, returning a closed image, *B3*. Morphological definition for this operation is dilates an image and then erodes the dilated image using the same *SE* for both operations. An example of *B3* shown in Fig. 20.
 d) Perform Hole Filling in *B3* and returning a filled image *A*: A hole may be defined as a background region surrounded by a connected border of foreground pixels. An example of *A* shown in Fig. 21.
Step 6: Human's height statistical search: Summing each pixels contents in every single column of *A* to form a statistical model as shown in Fig 22(a). From the statistical model, search for H_c (the highest value in the plot). An example of searching H_c is shown in Fig 23.
Step 7: Human's width statistical search: Summing each pixels contents in every single rows of *A* to form a statistical model as shown in Fig 22(b). From the statistical model, search for W_c (the highest value in the plot). An example of searching W_c is shown in Fig 23.
Step 8: If $H_c > H_{max}$, set $H_c = H_{max}$.
Step 9: *Faint detection:*
 IF $H_c \geq W_c$ and $H_c \geq$ (95% of H_{max})
 THEN the human is considered standing or walking. Continue on step 10.
 IF $H_c \geq W_c$ and $H_c <$ (95% of H_{max})
 THEN the human is considered fall down. Continue on step 10.
 IF $H_c < W_c$ and $W_c \neq W_p$ and/or $H_c \neq H_p$ and $f < F$,
 THEN the human is fall down. Continue on step 10.
 IF $H_c < W_c$ and $W_c = W_p$ and $H_c = H_p$ and $f < F$,
 THEN the human is considered possible faint. Continue on step 11.
 IF $H_c < W_c$ and $W_c = W_p$ and $H_c = H_p$ and $f \geq F$,
 THEN the human is considered faint. Continue on step 12.
Step 10: Reset fainted frame counter and start new cycle: Set $f = 0$, update $H_p = H_c$ and $W_p = W_c$. Repeat step 2.
Step 11: Increase frame counter by 1 and start new cycle: Set $f = f + 1$, update $H_p = H_c$ and $W_p = W_c$. Repeat step 2.
Step 12: *Faint case:* Signals alarm until operator noted and performs rescue action and resets the system manually.

FIGURE 16: Example of image *G*

FIGURE 17: Example of image *B1*
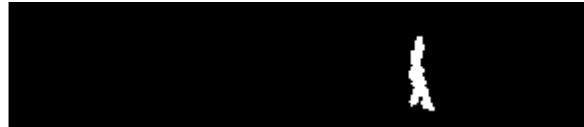
FIGURE 18: Example of image *B2*
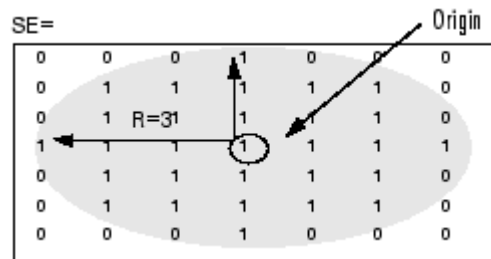
FIGURE 19: Disk shaped structuring elements

FIGURE 20: Example of image *B3*
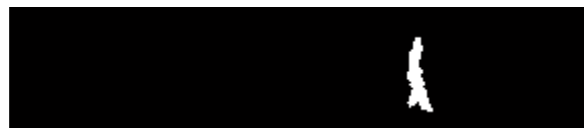
FIGURE 21: Example of image *A*
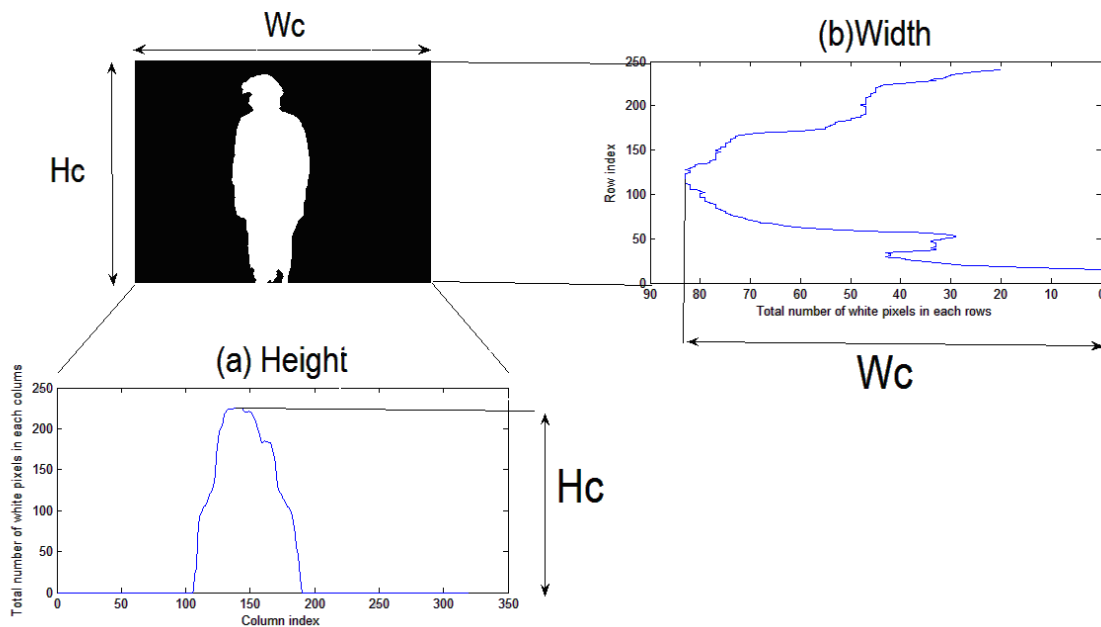
FIGURE 22: Statistical models: (a) human height statistical search and (b) human width statistical search.
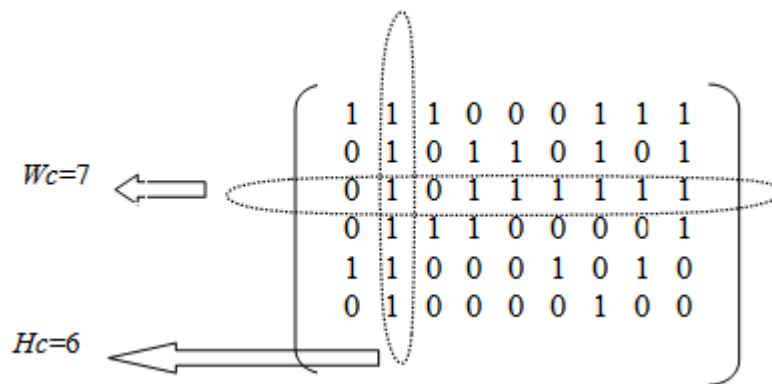
FIGURE 23: Examples of searching *Hc* and *Wc*

6. EXPERIMENTAL RESULTS

In this section, the application of the proposed omnidirectional thermal imaging system for trespasser and faint detection are briefly illustrated. The omnidirectional thermal images for a wide area test site are collected at the main office entrance in Faculty of Engineering and Technology (FET), Multimedia University. An omnidirectional image captured by using digital camera on the site is shown in Fig. 24. An omnidirectional thermal image is also captured by using thermal camera on the site is shown in Fig. 25. The unwarped form of Fig. 24 (digital color panoramic form) is shown in Fig. 26 whereas the unwarp form of Fig. 25 (thermal image panoramic form) is shown in Fig. 27 respectively. In Fig. 27, the log-polar mapping process is by 4:1 reduction mapping scale, which means that 320 X 240 omnidirectional thermal image's Cartesian pixels are mapped to one fourth of the thermal image Cartesian pixels (320 X 60) in panoramic view, with four fold data compression compare to original omnidirectional thermal image as in Fig. 25. The captured thermal images are tested for trespasser and faint detection algorithm proposed in section 4 and 5 in the subsection below.

A.) Experimental Test on Human Head Detection Algorithm for Trespasser Detection

A series of experiment is done to determine the optimum parameters stated in the human head detection algorithm and to examine the performance of the algorithm in its optimum setting. The

FIGURE 24: Case studies of trespasser and faint detection captured at main office entrance in FET (digital color form).
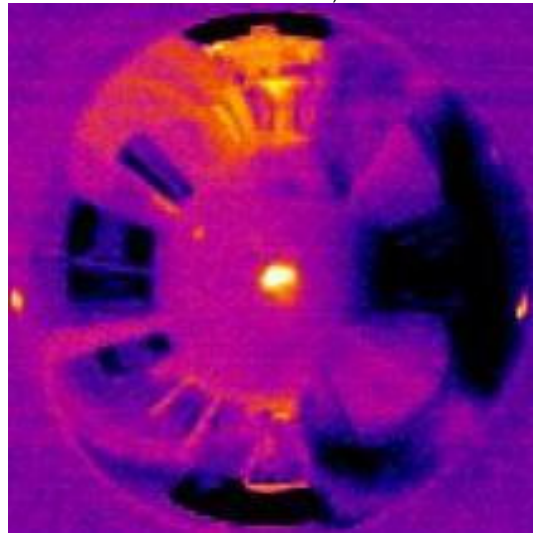
FIGURE 25: Case studies of trespasser and faint detection captured at main office entrance in FET (thermal image).
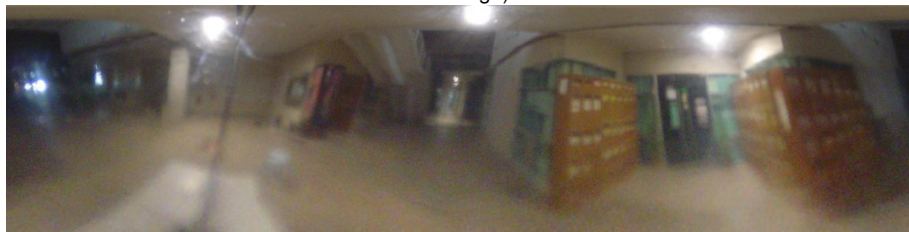
FIGURE 26: Unwarp form of Fig. 24 (digital color panoramic form)
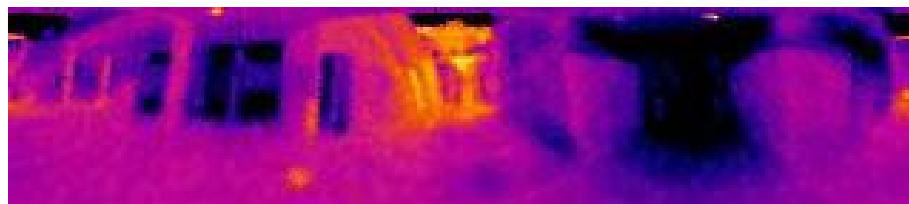
FIGURE 27: Unwarp form of Fig. 25 (thermal image panoramic form)

testing site is the main office entrance of Faculty of Engineering and Technology, Multimedia University, which simulates the door of a nursing home, for tracking Alzheimer patients from leaving nursing home without permission and theft during night time. The omnidirectional thermal imaging tool is setup with the height of 1.5m. This setting is chosen because it gives the best view of the testing site. A total of 10,000 thermal images with test subjects (human being or animal) roaming randomly in the area visible to the proposed system are taken.

To determine optimum value for parameter T_l, a random sample image is chosen and converted into B&W image using step 2 of the Algorithm for Trespasser Detection with value of T_l ranging from 0 to 510 (sum of R and G component in RGB image. B component is excluded because it is not proportional to the change of temperature). Perform the binary image conversion repeatedly with increasing step size of 10 for T_l and search for the optimum T_l where the noise can be minimized and the human is not distorted in the resulting image. The optimum setting found for T_l is 150 in this experiment. For example, if T_l =130 is used, excessive noise will be introduced. If T_l =170 is used, there will be too much distortion to human being in the resulting image. Refer Fig. 28 for better understanding.
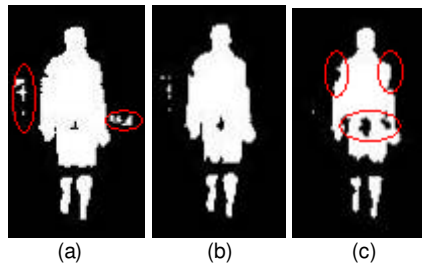
FIGURE 28: (a) T_l =130 (b) T_l =150 (c) T_l =170.

To determine the optimum value for parameter T_h, perform the binary image conversion repeatedly with decreasing step size of 10 for T_h and search for the minimum value of T_h which does not influence the appearance of the human object. The optimum value found for T_h is 430 in this experiment. For example, if T_h =400 is used, the image of human being is distorted. If T_h =460 is used, there will be no improvement for the image. Lower T_h value is preferred because it will filter out more high temperature noise component. Refer Fig. 29 for better understanding.
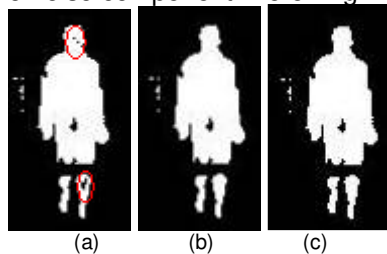
FIGURE 29: (a) T_h =400 (b) T_h =430 (c) T_h =460.

The accuracy of the proposed algorithm is then evaluated using 'operator perceived activity' (OPA) in which the proposed algorithm is evaluated with respect with the results interpreted by a human observer [7],[23]. Firstly, the thermal images are tested using the proposed algorithm. Then, the result is compared with the result of the human observer. The accuracy of the proposed algorithm is the percentage of interpretation (trespasser or not) agreed by both the human observer and the proposed algorithm.
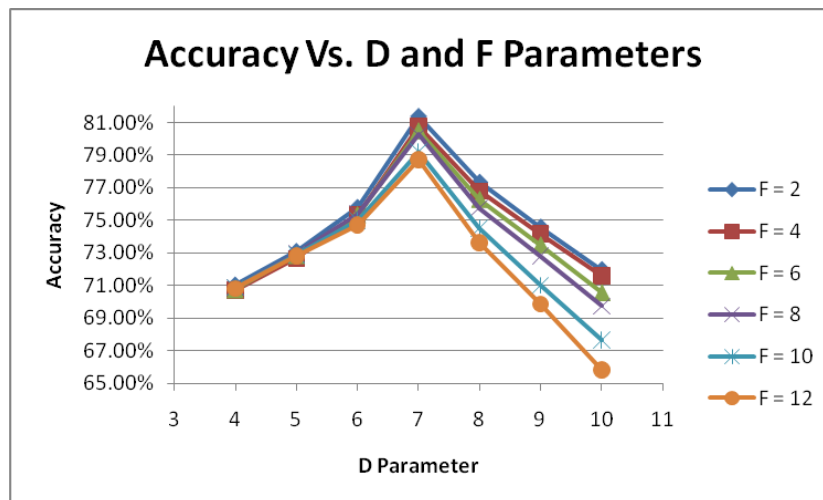
FIGURE 30: Accuracy of proposed algorithm for different combination of D and F.

To determine parameter D and F, all of the 10,000 images in group 1 are tested with different combinations of D and F. As shown in the graph in Fig. 30, the optimum value for parameter D and F are 7 and 2 which contribute to accuracy of 81.38%. From the observation, the smallest the step size partition variable (F), the accurate the trespasser detection is. D is optimum at 7 proceeding points, more than 7 proceeding points or lesser will degrade the overall performance of the trespasser detection accuracy. The proposed algorithm is able to operate in a very fast manner whereby the routine time required to capture a thermal image, unwarped into panoramic form, detect the existence of a trespasser and trigger the alarm is only 2.27 seconds.

B.) Experimental Test on Home Alone Faint Detection

In this section, the proposed algorithm for faint detection will be briefly discussed. The testing site is at the lobby around the main entrance of the Faculty of Engineering and Technology, Multimedia University, which simulates the activities area of a nursing home/hospital. For experimental purpose, 17,878 sample images include background image had been captured to test the accuracy of the proposed faint detection algorithm. At the test site, omnidirectional thermal images had been captured using the omnidirectional thermal imaging system, unwarped into panoramic form, performed image filtering, image processing for faint detection and signal alarm or not. These images included different poses of human such as standing or walking, falling, and fainting. The system routine time including the time for capture in omnidirectional thermal image, unwarping into panoramic form, image filtering, faint detection and signal alarm or not is 1 frame per second. Thus *F* as an example is set to 7 so that if a person lies on the same place with same size for 7 second, the system will classify the person is fainting.

The thermal image temperature range has been set from 25°C to 38°C. Human shape's parameters *S* and *R* are approximate from one of the testing image at a distance of 5 meters from the imaging system. From that image, the human's total pixels are only 30. Hence, *S* is set to 30 and *R* for *SE* is set to 2 respectively. Then, different *T* are tried in order to get the approximate human shape. *T* found to be falls within the range of 100 to 126. The *T* value depends on the temperature of the environment. At night time, *T* value is about 120 to 126, whereas in the afternoon, the *T* value is about 100 to 120.

Some example images of different human poses are shown in Fig. 31-33 (each for standing, falling (bending body) and fainting respectively). To differentiate between human standing motion and human bending body motion, a parameter, *p* has been set according to the factor of *Hmax*. The *p* value are found achieved highest accuracy when *p* = 95% of *Hmax* from Fig 6.14. Hence *p* value was set to 95% of *Hmax*. The proposed algorithm had been tested at night time with poor lighting condition at outdoor. The table I show the accuracies of the proposed faint detection algorithm on site.
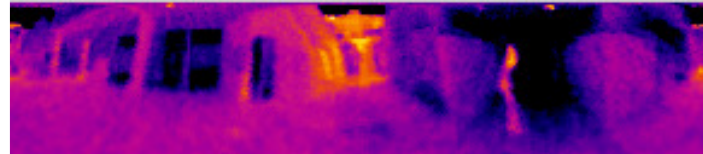
(a)

(b)

FIGURE 31: Standing motion (a) Thermal image (b) Black and white image
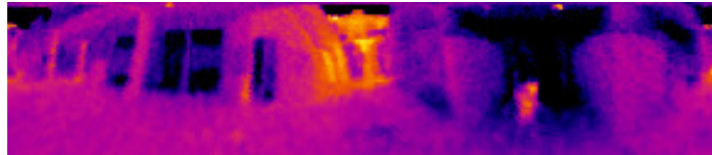
(a)

(b)

FIGURE 32: Falling/bending body motion (a) Thermal image (b) Black and white image
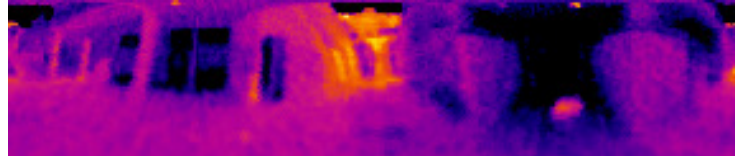
(a)
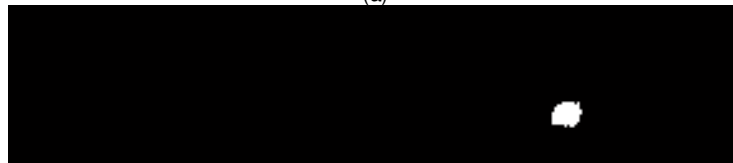
(b)

FIGURE 33: Fainting motion (a) Thermal image (b) Black and white image
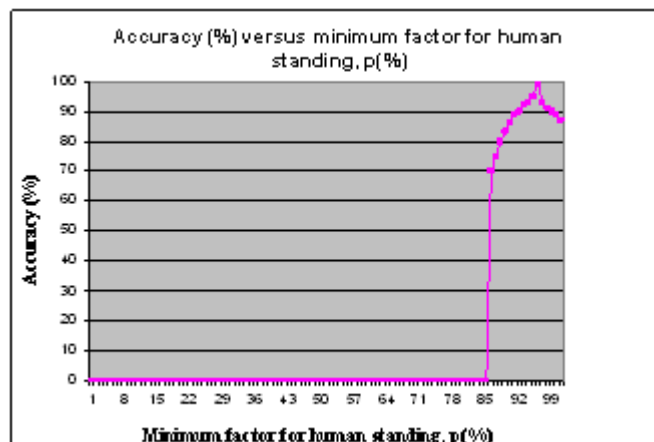
Fig. 34: Graph of accuracy versus *p*

Motion	Test frames	Success	Fail	Accuracy (%)
Standing , walking	6550	6498	52	99.20
Falling	4578	3983	595	86.78
Fainting	6750	6043	707	89.50
Total	17878	16524	1354	92.42

TABLE I.   OUTDOOR MOTION DETECTION AT NIGHT WITH POOR AMBIENT LIGHT CONDITION

The results in table I show that standing or walking motion has highest accuracy compare to falling and fainting motion with 99.20%. The reason that falling motion has the lowest accuracy (with 86.78%) is because some of the falling motion is considered as standing motion in the proposed system. Besides, 89.5% of the fainting motion was detected correctly among the total tested images. Lastly, the proposed algorithm was well performed on poor lighting conditions at night with an average accuracy of 92.42%. Overall, the system can classify human behavioral best in normal standing or walking motion and good in detecting fainting motion as well no matter in good or poor lighting condition. The home alone faint detection surveillance system also functions in a fast way whereby the routine time required to capture in a thermal image, unwarp into panoramic form, detect whether there is a faint suspected case until the signal alarm or not is only 1.85 seconds. When there is a fainting motion detected, the alarm will be trigger until the operator performs the rescue action and resets the system manually.

7. CONCLUSION

In this paper, the usage of hyperbolic IR reflected mirror is proposed in a thermal imaging system for omnidirectional thermal imaging smart surveillance purposes In the proposed system, both human detection and faint detection algorithm has been implemented in order to facilitate smart surveillance functionality which is able to greatly increase the effectiveness of the surveillance system. The trespasser detection algorithm is used mainly to detect unauthorized movement to and fro the building while the faint detection algorithm is used to detect the event where human being faint inside the premise.

For human head detection algorithm, the thermal images are first converted into binary (Black and White) images. The presence of human being is then analyzed based on the shape of the objects (head detection) in the binary image. From the experimental results, it is shown that the proposed trespasser detection algorithm is able to achieve an accuracy of 81.38% with a routine time of 2.27 seconds. For home alone faint detection algorithm, it enables the faint event to be detected even in poor lighting condition. The experimental results show that the proposed faint detection algorithm in the omnidirectional surveillance system achieved high accuracy in monitoring the faint event in the poor lighting condition with the accuracy of 89.50% with a routine time of 1.85 seconds.

There are some distortion on the captured thermal image whereby the temperature surrounding will affect the accuracy for both the proposed trespasser and faint detection algorithm. For example, the temperature surrounding which is nearly equal to the human body will make the human body undistinguishable from the temperature surrounding. Besides, the clothes of the human with lower temperature (human sweating) will causes part of the actual human body temperature changed.

In future, a plan is proposed to combine visible/infrared image fusion model into the currently proposed model to improve the performance of the trespasser and faint detection. In additional to that, it is also planning to employ microprocessor modules such as FPGA (field programmable gate array) and ARM (Advanced RISC Machine) for the image processing and analyzing tasks instead of a computer to effectively reduce the costs and power consumption of the proposed system. These topics will be addressed in future works.

8. REFERENCES

[1] A. Hampapur, L. Brown, J. Connell, S. Pankanti, A. Senior et al., "Smart Surveillance: Applications, Technologies and Implications", Information, Communications and Signal Processing, Vol. 2, p.p. 1133-1138.

[2] A. Hampapur, L. Brown, J. Connell, A. Ekin, N. Haas, M. Lu et al., "Smart Video Surveillance", IEEE Signal Processing Mag., March 2005, p.p. 39-51.

[3] M.W. Green, "The appropriate and effective use of security technologies in U.S. schools, A guide for schools and law enforcement agencies", Sandia National Laboratories, Albuquerque, NM, NCJ 178265, Sep 1999.

[4] C. Shu, A. Hampapur, M. Lu, L. Brown, J. Connell et al., "IBM Smart Surveillance System (S3): A Open and Extensible Framework for Event Based Surveillance", in Advanced Video and Signal Based Surveillance (AVSS 2005), 2005, p.p. 318-323.

[5] C. Lu, and M. S. Drew, "Automatic Compensation for Camera Settings for ImagesTaken Under Different Illuminants", Technical paper, School of Computer Science, Simon Fraser University, Vancouver, British Columbia, Canada, 2007, p.p. 1-5.

[6] Thermographic camera. Retrieved from Wikipedia, the free encyclopedia Web Site: http://en.wikipedia.org/wiki/Thermal_camera.

[7] W.K. Wong, P.N. Tan, C.K. Loo and W.S. Lim, "An Effective Surveillance System Using Thermal Camera", 2009 International Conference on Signal Acquisition and Processing (ICSAP2009) 3-5 Apr, 2009, Kuala Lumpur, Malaysia, p.p. 13-17.

[8] M.R. Narayanan, S.R. Lord, M.M. Budge, B.G. Cellar and N.H. Novell, "Falls Management: Detection and Prevention, using a Waist-mounted Triaxial Accelerometer", 29 th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2007, p.p.4037-4040.

[9] J. Han and B. Bhanu, "Fusion of color and infrared video for moving human detection", ACM Portal, Pattern Recognition, p.p. 1771-1784.

[10] J. Chahl and M. Srinivasan, "Reflective surfaces for panoramic imaging", *Applied Optics*, 36(31), Nov 1997, p.p.8275-85.

[11] S. Gachter, "Mirror Design for an Omnidirectional Camera with a Uniform Cylindrical Projection when Using the SVAVISCA Sensor", Research Reports of CMP, OMNIVIEWS Project, Czech Technical University in Prague, No. 3, 2001. Redirected from: http://cmp.felk.cvut.cz/projects/omniviews/

[12] T. Svoboda, *Central Panoramic Cameras Design, Geometry, Egomotion.* PhD Theses, Center of Machine Perception, Czech Technical University in Prague, 1999.

[13] http://www.flirthemography.com

[14] J.W. Davis and V. Sharma, "Background-Subtraction in Thermal Imagery Using Contour Saliency", International Journal of Computer Vision 71(2), 2007, p.p. 161-181.

[15] H. Araujo, J. M. Dias, "An Introduction To The Log-polar Mapping", *Proceedings of 2nd Workshop on Cybernetic Vision*, 1996, p.p. 139-144.

[16] C. F. R. Weiman and G. Chaikin, "Logarithmic Spiral Grids For Image Processing And Display", *Computer Graphics and Image Processing*, Vol. 11, 1979, p.p. 197-226.

[17] LIRA Lab, Document on specification, *Tech. report*, Espirit Project n. 31951 – SVAVISCA-available at http://www.lira.dist.unige.it.

[18] R. Wodnicki, G. W. Roberts, and M. D. Levine, "A foveated image sensor in standard CMOS technology", *Custom Integrated Circuits Conf*. Santa Clara, May 1995, p.p. 357-360.

[19] F. Jurie, "A new log-polar mapping for space variant imaging: Application to face detection and tracking", *Pattern Recognition,Elsevier Science*, 32:55, 1999, p.p. 865-875.

[20] V. J. Traver, "Motion estimation algorithms in log-polar images and application to monocular active tracking", PhD thesis, Dep. Llenguatges.

[21] R. Wodnicki, G. W. Roberts, and M. D. Levine, "A foveated image sensor in standard CMOS technology", *Custom Integrated Circuits Conf*. Santa Clara, May 1995, p.p. 357-360.

[22] Ling Hooi Lee, "Smart Surveillance Using Image Processing and Computer Vision Techniques.", Bachelor Degree thesis, Multimedia University, Melaka, Malaysia.

[23] J. Owens, A. Hunter and E. Fletcher, "A Fast Model–Free Morphology–Based Object Tracking Algorithm", British Machine Vision Conference, 2002,p.p. 767-776 .

A Novel Cosine Approximation for High-Speed Evaluation of DCT

Geetha K.S   geethakomandur@gmail.com
Assistant Professor, Dept of E&CE
R.V.College of Engineering,
Bangalore-59, India

M.UttaraKumari   dr.uttarakumari@gmail.com
Professor, Dean of P.G.Studies
Dept of E&CE
R.V.College of Engineering,
Bangalore-59, India

Abstract

This article presents a novel cosine approximation for high-speed evaluation of DCT (Discrete Cosine Transform) using Ramanujan Ordered Numbers. The proposed method uses the Ramanujan ordered number to convert the angles of the cosine function to integers. Evaluation of these angles is by using a 4th degree polynomial that approximates the cosine function with error of approximation in the order of 10^{-3}. The evaluation of the cosine function is explained through the computation of the DCT coefficients. High-speed evaluation at the algorithmic level is measured in terms of the computational complexity of the algorithm. The proposed algorithm of cosine approximation increases the overhead on the number of adders by 13.6%. This algorithm avoids floating-point multipliers and requires $N/2\log_2 N$ shifts and $(3N/2 \log2 N)- N + 1$ addition operations to evaluate an N-point DCT coefficients thereby improving the speed of computation of the coefficients

Keywords: Cosine Approximation, High-Speed Evaluation, DCT, Ramanujan Ordered Number.

1. INTRODUCTION

High-speed approximation to the cosine functions are often used in digital signal and image processing or in digital control. With the ever increasing complexity of processing systems and the increasing demands on the data rates and the quality of service, efficient calculation of the cosine function with a high degree of accuracy is vital. Several methods have been proposed to evaluate these functions [1]. When the input/output precision is relatively low(less than 24 bits), table and addition methods are often employed [2, 3]. Efficient methods on small multipliers and tables have been proposed in [4]. Method based on the small look up table and low-degree polynomial approximations with sparse coefficients are discussed in [5, 6].

Recently, there has been increasing interest in approximating a given floating-point transform using only very large scale integration-friendly binary, multiplierless coefficients. Since only binary coefficients are needed, the resulting transform approximation is multiplierless, and the overall complexity of hardware implementation can be measured in terms of the total number of adders and/or shifters required in the implementation. Normally the multiplierless approximation are discussed for implementing the discrete cosine transform (DCT) which is widely used in image/video coding applications. The fast bi-orthogonal Binary DCT (BinDCT) [7] and Integer DCT (IntDCT) [8, 9] belong to a class of multiplierless transforms which compute the coefficients

of the form ωn. They compute the integer to integer mapping of coefficients through the lifting matrices. The performances of these transforms depend upon the lifting schemes used and the round off functions. In general, these algorithms require the approximation of the decomposed DCT transformation matrices by proper diagonalisation. Thus, the complexity is shifted to the techniques used for decomposition.

In this work, we show that using Ramanujan ordered numbers; it is possible to evaluate the cosine functions using only shifts and addition operations. The complete operator avoids the use of floating-point multiplication used for evaluation of DCT coefficients, thus making the algorithm completely multiplierless. Computation of DCT coefficients involves evaluation of cosine angles of multiples of $2\pi/N$. If N is chosen such that it could be represented as $2^{-l} + 2^{-m}$, where l and m are integers, then the trigonometric functions can be evaluated recursively by simple shift and addition operations.

2. RAMANUJAN ORDERED NUMBERS

Ramanujan ordered Numbers are related to π and integers which are powers of 2. Ramanujan ordered Number of degree-1 was used in [10,11] to compute the Discrete Fourier Transform. The accuracy of the transform can be further improved by using the Ramanujan ordered number of degree-2. This is more evident in terms of the errors involved in the approximation.

2.1 Definition : Ramanujan ordered Number of degree-1

Ramanujan ordered Numbers of degree-1 $\Re_1(a)$ are defined as follows:

$$\Re_1(a) = \left[\frac{2\pi}{l_1(a)}\right] \quad where \; l_1(a) = 2^{-a}$$

(1)

a is a non-negative integer and $[\cdot]$ is a round off function. The numbers could be computed by simple binary shifts. Consider the binary expansion of π which is 11.00100100001111… If a is chosen as 2, then $l_1(2) = 2^{-2}$,and $\Re_1(2) = [11001.001000.......] = 11001$. i.e., $\Re_1(2)$ is equal to 25. Likewise $\Re_1(4)$ =101. Thus the right shifts of the decimal point (a+1) time yields $\Re_1(a)$.

Ramanujan used these numbers to approximate the value of π. Let this approximated value be $\hat{\pi}$ and let the relative error of approximation be ε, then

$$\hat{\pi} = \frac{1}{2}\left[\Re_1(a)\iota(a)\right] \quad \hat{\pi} = (1+\in)\pi$$

(2)

These errors could be used to evaluate the degree of accuracy obtained in computation of DCT coefficients.

(a)	$\Re(a)$	$\hat{\pi}$	Upper bound of error
0	6	3.0	4.507×10^{-2}
1	13	3.25	3.4507×10^{-3}
3	50	3.125	5.287×10^{-3}

TABLE 1: Ramanujan ordered Number of Degree-1.

Table 1 clearly shows the numbers which can be represented as Ramanujan ordered -numbers of order-1. Normally, the digital signal processing applications requires the numbers to be power of 2; hence higher degree numbers are required for all practical applications.

2.2 Definition: Ramanujan ordered Number of degree-2

The Ramanujan ordered Number of degree-2 [11] are defined such that $2\pi/N$ is approximated by sum or difference of two numbers which are negative powers of 2. Thus Ramanujan Numbers of degree-2 are,

$$\Re_{2i}(l,m) = \left[\frac{2\pi}{\iota_{2i}(l,m)}\right] \quad for\ i=1,2 \tag{3}$$

$$\iota_{21}(l,m) = 2^{-l} + 2^{-m} \quad for\ m > l \geq 0$$

$$\iota_{22}(l,m) = 2^{-l} - 2^{-m} \quad for\ (m-1) > l \geq 0 \tag{4}$$

Where l and m are integers, Hence

$$\Re_{21}(3,5) = 40 \quad \Re_{21}(1,3) = 10$$

Ramanujan ordered Number of degree-2 and their properties are listed in the table 2 below.

(l,m)	$\Re(l,m)$	$\hat{\pi}$	Upper bound of error
0,2	5	3.125	5.28×10^{-3}
1,2	8	3.0	4.507×10^{-2}
4,5	67	3.140	5.067×10^{-4}

TABLE 2: Ramanujan ordered Number of Degree-2.

The accuracy of the numbers increase with the increase in the degree of the Ramanujan ordered numbers at the expense of additional shifts and additions. State-of-art technologies in Image processing uses the block processing techniques for applications like image compression or image enhancement. The standardized image block size is 8×8, which provides us an opportunity to use Ramanujan ordered numbers to reduce the complexity of the algorithms. Table III shows the higher degree Ramanujan ordered Numbers and their accuracies.

$(l,m,p.....)$	$\Re(l,m,p....)$	$\hat{\pi}$	Upper bound of error	Computational complexity	
				Shifts	Adds
1,2	8	3. 0	4.507×10^{-2}	2	1
1,2,5	8	3.125	5.28×10^{-3}	3	2
1,2,5,8	8	3.1406	3.08×10^{-4}	4	3

TABLE 3: Ramanujan ordered Number of Higher Degree.

Table 3 shows that the error of approximation decreases with the increase in the degree of Ramanujan ordered numbers, but the computational complexity also increases. Hence the choice of Ramanujan ordered number of degree-2 is best validated for the accuracy and the computational overhead in computation of the cosine functions.

3. COSINE APPROXIMATION USING RAMANUJAN ORDERED NUMBER

Method of computing the cosine function is to find a polynomial that allows us to approximate the cosine function locally. A polynomial is a good approximation to the function near some point, $x = a$, if the derivatives of the polynomial at the point are equal to the derivatives of the cosine curve at that point. Thus the higher the degree of the polynomial, the better is the approximation. If p_n denotes the nth polynomial about $x = a$ for a function f, and if we approximate $f(x)$ by $p(x)$ at a point x, then the difference $R_n(x) = f(x) - p(x)$ is called the nth remainder for f about $x = a$. The plot of the 4th order approximation along with the difference plot is as shown in figure 1.
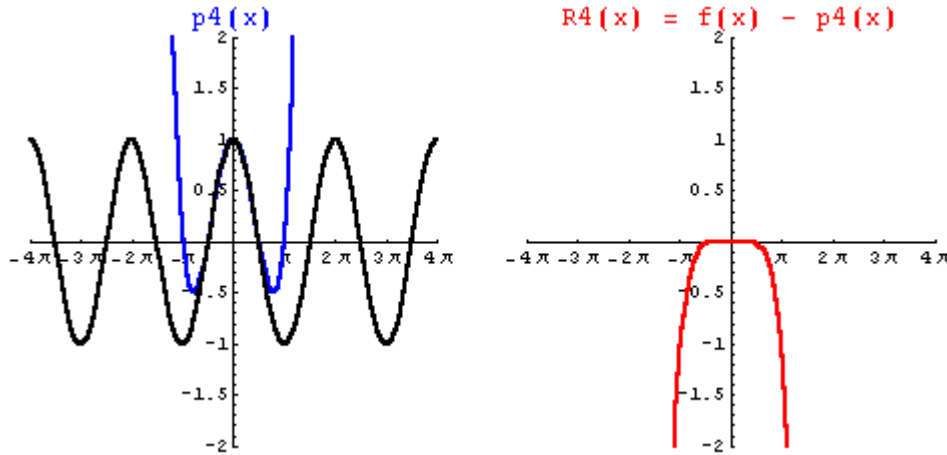
FIGURE 1: Plot of f(x)=cos(x) and the remainder function .

Figure 2 indicates that the cosine approximation at various angles with 4th degree polynomial is almost close with the actual values. The accuracy obtained at various degrees of the polynomial is compared in table 4. Thus we choose the 4th degree polynomial for the cosine approximation.
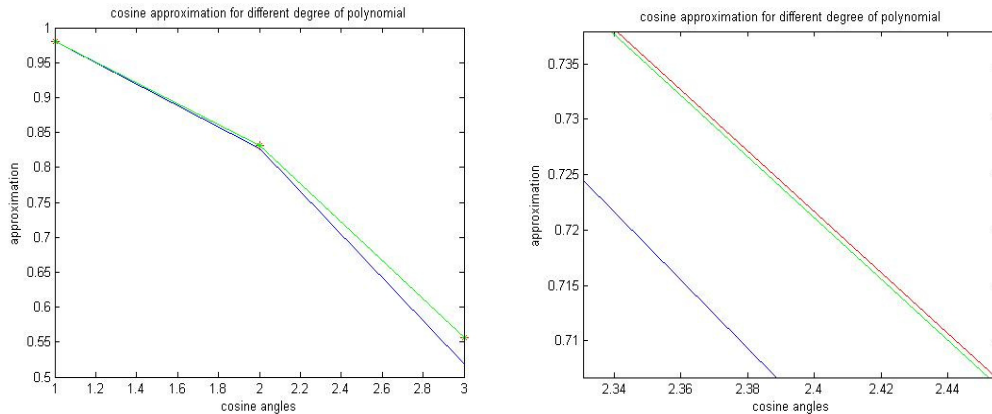
FIGURE 2: Expanded version of cosine approximation.

cos(5π/16)
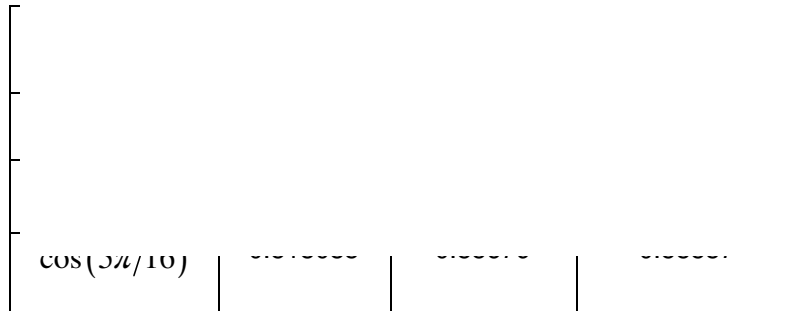
TABLE 4: Cosine approximation in comparison with the Actual value

The evaluation of the cosine function using the polynomial approximation is explained through the application of computing the Discrete Cosine Transform coefficients which uses the cosine as its basis function. Discrete Cosine Transforms (DCT) is widely used in the area of signal processing, particularly for transform coding of images. Computation of DCT coefficients involves evaluation of cosine angles of multiples of 2π/N.

The input sequence $\{x_n\}$ of size N, is transformed as, $\{y_k\}$. The transformation may be either DFT or DCT. The DCT defined as [12]

$$y_k = \frac{2\varepsilon_k}{N} \sum_{n=0}^{N-1} x_n \cos\left(\pi k \frac{(2n+1)}{2N}\right)$$

$$for \ k = 0,1...N-1$$

$$\varepsilon_k = \begin{cases} \frac{1}{\sqrt{2}} & for \ k = 0 \\ 1 & otherwise \end{cases}$$

(5)

Neglecting the scaling factors, the DCT kernel could be simplified as

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

$$for \ 0 \le n \le N-1, \ 0 \le k \le N-1$$

(6)

DCT coefficients are computed by evaluating the sequences of type

$$\{c_n / c_n = p\cos(2\pi n / N), n = 0,1,2...(N-1), p \in \mathbb{R}\}$$

(7)

where $\mathbb{R}$ is the set of real numbers. These computations are done via a Chebyshev-type of recursion. Let us define

$$w(M,p) = \{w_n / w_n = p\cos(2\pi n/M)\}$$

$$n = 0,1,2,....\Psi, p \in \mathbb{R}$$

(8)

$$\Psi = \left(\frac{M}{4} - 1\right), M = \beta N$$

(9)

$$\beta = \begin{cases} 1 & if \ 4 \ divides \ N \\ 2 & if \ 2 \ divides \ N \ but \ not \ 4 \\ 4 & otherwise \end{cases}$$

(10)

The use of β facilitates the computation of W (M, p) by considering cosine values from the first quadrant of the circle. Then the sequence $\{c_n\}$ can be evaluating recursively $\{W_n\}$.

The W (M, p) sequence is estimated as follows. Let us define $\frac{2\pi}{N} 2^{-2}$ as x and hence $c_n = \cos(nx)$. We approximate x by $\hat{x}$ equal to $2^{-l} + 2^{-m}$ with l and m being non-negative integers. Thus, we approximate c_n's by t_n's, where α is equal to $\hat{x}^2/2 - \hat{x}^4/4!$. Since $\hat{x}$ is a Ramanujan ordered Number of degree-2, α is of the form $2^{-c} + 2^{-d}$, where c and d are integers. Then

$$t_0 = p$$
$$t_1 = p(1-\alpha)$$
$$\vdots$$
$$t_{n+1} = 2(1-\alpha)t_n - t_{n-1}$$
$$n = 1, 2, \cdots, (\Psi - 1)$$

(11)

The above equations show that the recursive equations can be computed by simple shift and addition operations.

4. SIMULATION RESULTS

To evaluate the performance of the proposed cosine approximation, the following cosine angles were simulated in matlab. The actual value is the angles evaluated using the inbuilt COS function and the values obtained from the proposed approximation are tabulated in Table 5. The deviation of the approximated results from the actual value is tabulated as error in table 5. The error due to approximation is of the order of 10^{-3} which is acceptable for image coding applications.

Function	Actual value	Proposed Approximation	Error
$\cos(\pi/16)$	0.980785	0.9810	-2.15×10^{-4}
$\cos(3\pi/16)$	0.8314696	0.83312	-1.6504×10^{-3}
$\cos(5\pi/16)$	0.55557	0.5571	-1.53×10^{-3}

TABLE 5: Comparison of the proposed cosine approximation

The cosine approximation being multiplierless using Ramanujan ordered Number proved to be more efficient and simple when being applied to image coding. We tested the efficacy of the proposed algorithm by replacing the existing floating-point DCT/IDCT block from the JPEG encoder/decoder. To prove the efficiency of the proposed algorithm standard test images like

Cameraman and Lena images were considered for JPEG encoder/decoder. The floating-point DCT block evaluates 2-D DCT on the 8x8 block of the image by implementing the direct function as given in equation (5). Most commonly used algorithms [13, 14] use lookup tables to access the cosine values to compute the DCT coefficients and hence require floating-point multipliers in hardware implementation of the algorithm. Table 6 shows the comparison of the algorithms in terms of number of adders and table 7 gives the comparison in terms of shifters/multipliers required for the hardware implementation. The computational complexity is reduced by completely avoiding the floating-point multiplications and is replaced by N/2log₂N shift operations. However the number of addition operations required to compute N DCT coefficients is (3N/2 log2 N)- N + 1 which increases by 13.6% when compared with the number of adders required for the floating-point DCT.

Transform	Floating-point DCT [13]	Proposed Ramanujan DCT
8 x 8	29	36
16 x 16	81	96
32 x 32	209	240
64 x 64	513	576

TABLE 6: Computational Complexity in terms of additions

Transform	Floating-point DCT [13] (Floating-point multiplications)	Proposed Ramanujan DCT (Shifts)
8 x 8	12	12
16 x 16	36	36
32 x 32	80	80
64 x 64	384	384

TABLE 7: Computational Complexity in terms of Multiplications

The performance of the proposed algorithm is compared with the existing floating-point DCT in terms of PSNR and the Compression Ratio achieved and the results are tabulated in Table 8. The PSNR is computed using the MSE as the error metric.

$$MSE = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} \left[I(m,n) - I^{'}(m,n) \right]^2$$

$$PSNR = 20 * \log_{10} \left(255 / \sqrt{MSE} \right)$$

(12)

Compression Ratio =Uncompressed Image size / Compressed image size.
Where N x M is the size of the image, I(m,n) is the original image and I`(m,n) is the reconstructed image.
The results show that the reduction in PSNR for color images like Football is around 10% and reduction in PSNR for smooth transition images like Lena, Cameraman is around 4%, and is improved by 4% for sharp transition images like Circuit board images.

Image	Proposed Ramanujan DCT		Floating-point DCT[13]	
	Compression Ratio	PSNR(dB)	Compression Ratio	PSNR(dB)
Circuit	5.30:1	72.93	5.6:1	69.01
Football	5.62:1	69.24	5.97:1	77.20
Lena	4.20:1	62.04	4.56:1	65.77
Medical Image	5.74:1	62.91	5.74:1	68. 1

TABLE 8: Performance comparison of Ramanujan DCT and Floating-point DCT
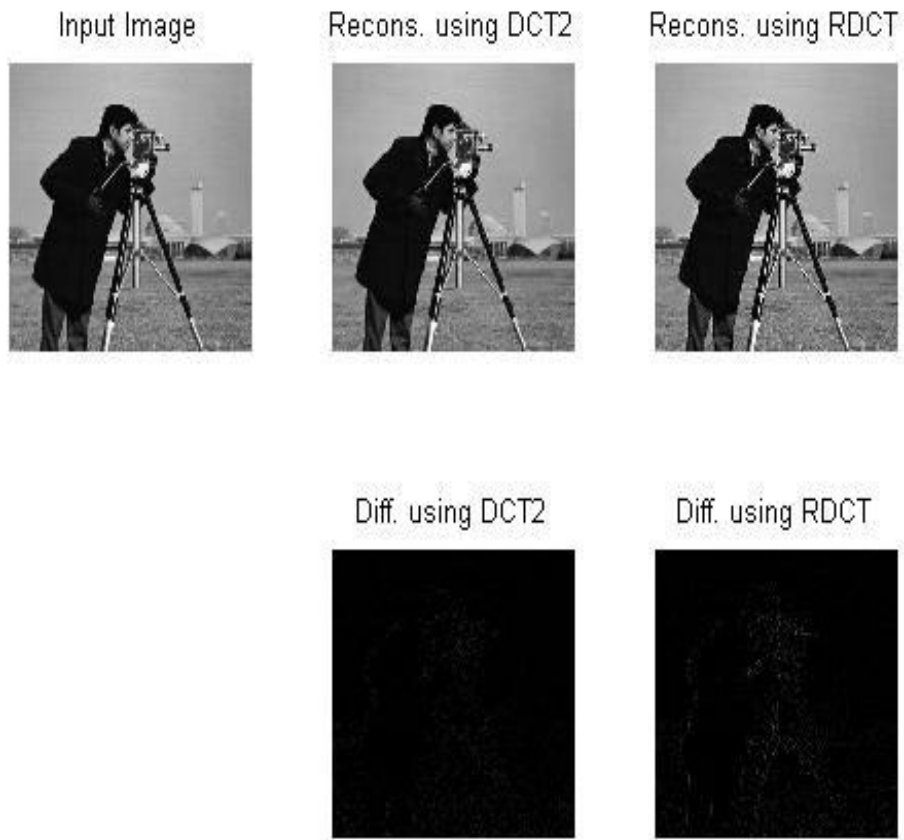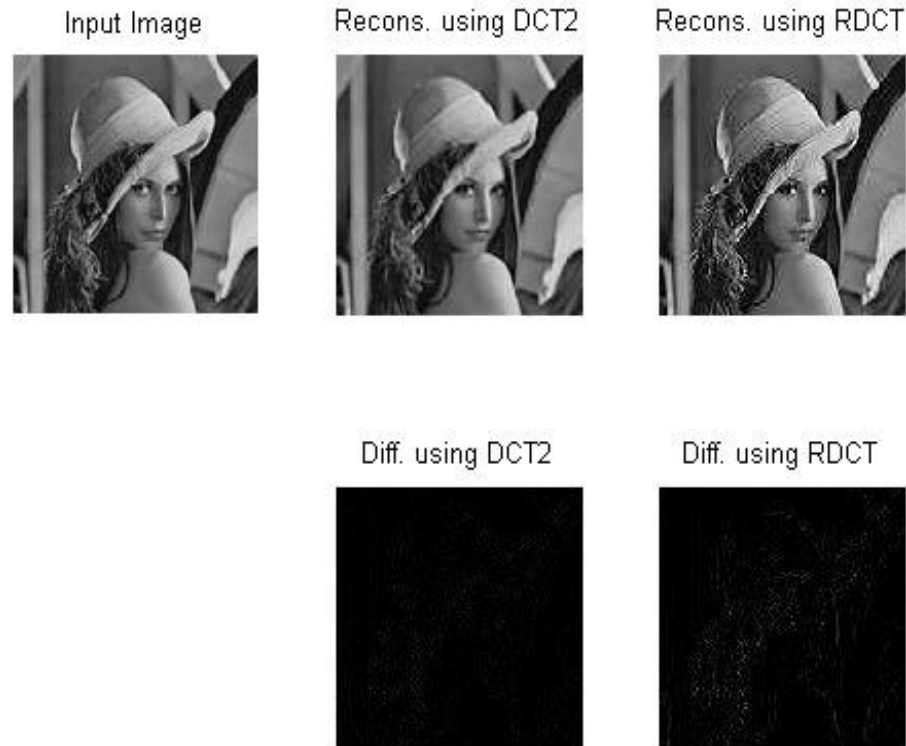
FIGURE3: The Original image and Reconstructed image using floating-point DCT and Ramanujan DCT.

Figure 3 shows the original image and the reconstructed image obtained using both RDCT and DCT2 function applied in JPEG compression standard technique. The difference image obtained between the original and reconstructed image shows that the error in cosine approximation is very negligible.

5. CONCLUSIONS

We have presented a method for approximation of the cosine function using Ramanujan ordered number of degree 2. The cosine function is evaluated using a 4th degree polynomial with an error of approximation in the order of 10^{-3} . This method allows us to evaluate the cosine function using only integers which are powers of 2 thereby replaces the complex floating-point multiplications by shifters & adders. This algorithm takes $N/2 \log_2 N$ shifts and $(3N/2 \log2 N) - N + 1$ addition operations to evaluate an N-point DCT coefficients. The cosine approximation increases the overhead on the number of adders by 13.6%. The proposed algorithm reduces the computational complexity and hence improves the speed of evaluation of the DCT coefficients. The proposed algorithm reduces the complexity in hardware implementation using FPGA. The results show that the reconstructed image is almost same as obtained by evaluating the floating-point DCT.

Geetha K.S & M.UttaraKumari

6. REFERENCES

1. J.-M.Muller, Elementary Functions: Algorithms and Implementation, Birkhauser, Boston, 1997.

2. F.de Dinechin and A.Tisserand, "Multipartite table methods," IEEE Transactions on Computers, vol.54, no.3,pp. 319-330, Mar. 2005.

3. M.Schulte and J. Stine, "Approximating elementary functions with symmetric bipartite tables," IEEE Transactions on Computers, vol. 48, no. 8, pp. 842-847, Aug. 1999.

4. J. Detrey and F.de Dinechin, "Second order function approximation using a single multiplication on FPGAs," in 14th International Conference on Field-Programmable Logic and Applications, Aug.2004, pp.221-230, LNCS 3203.

5. Arnad Tisserand, "Hardware operator for simultaneous sine and cosine evaluation," in ICASSP 2006, vol III, pp.992-995.

6. N.Brisebarre, J.-M.Muller, A.Tisserand and S.Torres, "Hardware operators for function evaluation using sparse-coefficient polynomials," in Electronic Letters 25, 42(2006), pp. 1441-1442.

7. T Tran, "The BinDCT: Fast multiplier less approximation of the DCT". IEEE Signal Proc. Vol. 7, No. 6, pp. 141-44, 2000.

8. Y Zeng, L Cheng, G Bi, and A C Kot, "Integer DCT's and Fast Algorithms". IEEE Trans Signal Processing, Vol. 49, No. 11, Nov. 2001.

9. X Hui and L Cheng, "An Integer Hierarchy: Lapped Biorthogonal Transform via Lifting Steps and Application in Image Coding". Proc. ICSP-02, 2002.

10. Nirdosh Bhatnagar, "On computation of certain discrete Fourier transforms using binary calculus".Signal Processing-Elsevier Vol43,1995.

11. Geetha.K.S, M.Uttarakumari, "Multiplierless Recursive algorithm using Ramanujan ordered Numbers," in IETE Journal of Research, vol. 56, Issue 4, JUL-AUG 2010.

12. K.R.Rao, P.Yip, "Discrete Cosine Transform Algorithms, Advantages Applications". New York: Academic 1990.

13. H.S. Hou, "A Fast Recursive Algorithms for Computing the Discrete Cosine Transform". IEEE Trans. Acoust., Speech, Signal Processing, Vol.35, pp 1455-1461, Oct 1987.

14. N.I.Cho, S.U.Lee, "A Fast 4X4 DCT Algorithm for the Recursive 2-D DCT". IEEE Trans. Signal Processing, Vol.40, pp 2166-2173. Sep 1992.

An Evolutionary Dynamic Clustering Based Colour Image Segmentation

Amiya Halder   amiya_halder@indiatimes.com
Assistant Professor
Dept. of Computer Science & Engineering
St. Thomas' College of Engg. & Technology
Kolkata, West Bengal, India

Nilavra Pathak   nilavrapthk7@gmail.com
Student
Dept. of Computer Science & Engineering University
St. Thomas' College of Engg. & Technology
Kolkata, West Bengal, India

Abstract

We have presented a novel Dynamic Colour Image Segmentation (DCIS) System for colour image. In this paper, we have proposed an efficient colour image segmentation algorithm based on evolutionary approach i.e. dynamic GA based clustering (GADCIS). The proposed technique automatically determines the optimum number of clusters for colour images. The optimal number of clusters is obtained by using cluster validity criterion with the help of Gaussian distribution. The advantage of this method is that no a priori knowledge is required to segment the color image. The proposed algorithm is evaluated on well known natural images and its performance is compared to other clustering techniques. Experimental results show the performance of the proposed algorithm producing comparable segmentation results.

Keywords: Segmentation, Clustering, Genetic Algorithm, Clustering Metric, Validity Index.

1. INTRODUCTION

Image segmentation, i.e., the partitioning of an image into relevant regions, is a fundamental problem in image analysis. Accurate segmentation of objects of interest is often required before further analysis can be performed. Despite years of active research, fully automatic segmentation of arbitrary images is still seen as an unsolved problem.

Colour image segmentation emerges as a new area of research. Colour image segmentation can solve many contemporary problems in medical imaging, mining and mineral imaging, bioinformatics, and material sciences. Naturally, color image segmentation demands well defined borders of different objects in an image. So, there is a fundamental demand of accuracy. The segmented regions or components should not be further away from the true object than one or a few pixels. So, there is a need for an improved image segmentation technique that can segment different components precisely. Image data have some particular characteristics that differentiate them from other form of data. Image data may have corrupted values due to the usual limitations or artifacts of imaging devices. Noisy data, data sparsity, and high dimensionality of data create difficulties in image pixel clustering. As a result, image pixel clustering becomes a harder problem than other form of data. Although there are some existing algorithms for unsupervised color image segmentation, none of them has been found to be robust in determining an accurate number of components or segments. Image segmentation is a very important field in image

analysis, object recognition, image coding and medical imaging. Segmentation is very challenging because of the multiplicity of objects in an image and the large variation between them. Image segmentation is the process of division of the image into regions with similar attributes.

The method for the segmentation of digital images can be broadly classified in: (a) edge and line oriented segmentation, (b) region growing methods, (c) clustering, and (d) region splitting methods. Edge and line oriented segmentation work on either individually analyzing each data band or considering the whole vector space. After edge extraction, a processing should be applied to create objects and segments that represent the elements present in the scene. Region growing and splitting method deals commonly with feature extraction and thresholding. In many object based image segmentation applications, the number of cluster is known a priori, but our proposed scheme is automatically determined the optimal number of clustering. The proposed technique should be able to provide good optimal results whereas the K-means algorithm which may get stuck at values which are not optimal [16]. Some of the several unsupervised clustering algorithms developed include K-means [13, 14], fuzzy K-means, ISODATA [11], self-organizing feature map (SOM) [15], particle swarm optimization (PSO) [10], learning vector quantizers (LVQ) [12], Dynamic GA based Clustering [17] etc.

Colour images are more complex than gray scale images as instead of a single intensity value for a pixel, each pixel is usually denoted by three component values such as Red, Green and Blue. Clustering based methods are ideal to use for gray scale images can be easily extended to cope with higher dimensity, although, the increased dimensionality also leads to more computationally expensive process. Various segmentation techniques have been developed for image segmentation include Unsupervised Colour Textured Image Segmentation Using Cluster Ensembles and MRF Model [18], Determination of Number of Clusters in k-means Clustering and application in Colour Image Segmentation [20], Unsupervised Colour Image Segmentation based on Gaussian Mixture Model [21] etc.

This paper presents automatic image segmentation of colour images using GA-based clustering. One natural view of segmentation is that we are attempting to determine which components of a data set naturally "belong together". Clustering is a process whereby a data set is replaced by clusters, which are collections of data points that "belong together". Thus, it is natural to think of image segmentation as image clustering i.e. the representation of an image in terms of clusters of pixels that "belong together". The specific criterion to be used depends on the application. Pixels may belong together because of the same colour or similarity measure.

Natural colour images are particularly noisy due to the environment they were produced. Therefore, it is hard to develop a robust and faithful unsupervised technique for automatic determination of number of objects in a colour image. Although there are a few existing approaches for unsupervised colour image segmentation, none of them has been found robust in all situation. Initially we tried SNOB [8], a Minimum Message Length (MML) based unsupervised data clustering approach to address this problem [19].

Genetic Algorithm based clustering technique has been used to cluster datasets [1-5]. In this paper we present an efficient image segmentation algorithm using automatic GA based clustering algorithm. The result of this algorithm produced a better result to compare with other techniques.

The rest of the paper is organized in the following manner: Section II introduces Genetic Algorithms. Section III gives an overview of the GA based clustering approach. Proposed algorithm is discussed in section IV. Section V summarizes the experimental results.

2. GENETIC ALGORITHM

Genetic Algorithm (GA) is a population-based stochastic search procedure to find exact or approximate solutions to optimization and search problems. Modeled on the mechanisms of evolution and natural genetics, genetic algorithms provide an alternative to traditional optimization

techniques by using directed random searches to locate optimal solutions in multimodal landscapes [6,7]. Each chromosome in the population is a potential solution to the problem. Genetic Algorithm creates a sequence of populations for each successive generation by using a selection mechanism and uses operators such as crossover and mutation as principal search mechanisms - the aim of the algorithm being to optimize a given objective or fitness function.

An encoding mechanism maps each potential solution to a chromosome. An objective function or fitness function is used to evaluate the ability of each chromosome to provide a satisfactory solution to the problem. The selection procedure, modeled on nature's survival-of-the-fittest mechanism, ensure that the fitter chromosomes have a greater number of offspring in the subsequent generations.

For crossover, two chromosomes are randomly chosen from the population. Assuming the length of the chromosome to be l, this process randomly chooses a point between 1 and l-1 and swaps the content of the two chromosomes beyond the crossover point to obtain the offspring. A crossover between a pair of chromosomes is affected only if they satisfy the crossover probability.

Mutation is the second operator, after crossover, which is used for randomizing the search. Mutation involves altering the content of the chromosomes at a randomly selected position in the chromosome, after determining if the chromosome satisfies the mutation probability. In order to terminate the execution of GA we specify a stopping criterion. Specifying the number of iterations of the generational cycle is one common technique of achieving this end.

3. GENETIC ALGORITHM BASED CLUSTERING

The searching capability of GAs can be used for the purpose of appropriately clustering a set of n unlabeled points in N-dimension into K clusters [1]. In our proposed scheme, the same idea can be applied on image data. We consider a colour image of size mxn and every pixel has Red, Green and Blue components. The basic steps of the GA-clustering algorithm for clustering image data are as follows:

3.1 Encoding

Each chromosome represents a solution which is a sequence of K cluster centers. For an N-dimensional space, each cluster center is mapped to N consecutive genes in the chromosome. For image datasets each gene is an integer representing an intensity value of the three components Red, Green and Blue.

3.2 Population initialization

Each of the P chromosomes is initialized to K randomly chosen points from the dataset. Here P is the population size.

3.3 Fitness computation

The fitness computation is accomplished in two steps.
First, the pixel dataset is clustered according to the centers encoded in the chromosome under consideration, such that each intensity value $x_i(r,g,b)$ of colour image combined with three component red, green and blue (24 bit), i = 1, 2, ..., mxn is assigned to cluster with center $z_j(r,g,b)$, j = 1, 2, ..., K,

$$\text{if } || x_i(r,g,b)-z_j(r,g,b)|| < || x_i(r,g,b)-z_p(r,g,b)||, p = 1,2,...,K, \text{ and } p \neq j. \tag{1}$$

The next step involves adjusting the values of the cluster centers encoded in the chromosome, replacing them by the mean points of the respective clusters. The new center $Z_i(r,g,b)$ for the cluster C_i is gives by

$$Z_i(r, g, b) = \frac{1}{n_i} \sum_{x_j \in C_i} X_j(r, g, b) , i=1,2,.....,K. \tag{2}$$

Subsequently, the clustering metric M is computed as the sum of Euclidean distances of each point from their respective cluster centers given by

$$M = \sum_{i=1}^{K} M_i.$$
(3)

$$M_i = \sum_{x_j \in C_i} \| x_j(r,g,b) - z_i(r,g,b) \|$$
(4)

The fitness function is defined as

f = 1/M.
(5)

A low value of intra-cluster spread is a characteristic of efficient clustering. Hence our objective is to minimize the clustering metric M i.e. maximize f.

3.4 Selection

This paper employs the Roulette Wheel selection - a proportional selection mechanism in which the number of copies of a chromosome, that go into the mating pool for subsequent operations, is proportional to its fitness.

3.5 Crossover

In this paper we use single-point crossover with a fixed crossover probability of μ_c. The procedure followed is the same as that described in section 2.

3.6 Mutation

Each chromosome undergoes mutation with a fixed probability μ_m. A number δ in the range [0, 1] is generated with uniform distribution. If the value at a gene position is v, after mutation it becomes

v ± δ * v,   v ≠ 0

v ± δ ,   v = 0.
(6)

3.7 Termination criterion

We execute the processes of fitness computation, selection, crossover, and mutation for a predetermined number of iterations. In every generational cycle, the fittest chromosome till the last generation is preserved - elitism. Thus on termination, this chromosome gives us the best solution encountered during the search.

4. PROPOSED METHOD

We proposed a new segmentation algorithm that can produce a new result according to the values of the clustering. We consider a colour image f of size mxn. The proposed algorithm is:

1. Repeat steps 2 to 8 for K=2 to K=K_{max}.
2. Initialize the P chromosomes of the population.
3. Compute the fitness function f_i for i=1,…,P, using equation (5).
4. Preserve the best (fittest) chromosome seen till that generation.
5. Apply Roulette Wheel selection on the population.
6. Apply crossover and mutation on the selected population according to sections 3.5 and 3.6 respectively.
7. Repeat steps 3 to 6 till termination condition is reached.
8. Compute the clustering Validity Index for the fittest chromosome for the particular value of K, using equation (7), in order to determine the validity of the clustering on the given dataset.
9. Cluster the dataset using the most appropriate number of clusters determined by comparing the Validity Indices of the proposed clusters for K=2 to K=K_{max}.

4.1 Validity Index

The cluster validity measure used in the paper is the one proposed by Turi [8]. It aims at minimizing the validity index given by the function,

$$V = y \times \frac{intra}{inter} \tag{7}$$

The term intra is the average of all the distances between each pixel x(r,g,b) and its cluster centroid $z_i(r,g,b)$ which is defined as

$$intra = \frac{1}{N} \sum_{i=1}^{K} \sum_{x \in C_i} \| x(r,g,b) - z_i(r,g,b) \|^2 \tag{8}$$

Where ||x(r,g,b) – z_i(r,g,b)|| means the Euclidean distance, which is calculated as

$$\sqrt{(x_{red} - z_{ired})^2 + (x_{green} - z_{igreen})^2 + (x_{blue} - z_{iblue})^2},$$

where N is the total number of pixels, C_i is the cluster number, z_i is the centroids of cluster C_i, K is the total number of clusters. Intra cluster dependency is the sum of square of Euclidean distance of every element from the centroids of the cluster to which it belongs.

On the other hand, inter is the inter cluster dependency which gives the idea about the extent to which each clusters are related. The higher this value the better clustering is obtained. It is represented as

$$inter = \min(\| z_i(r,g,b) - z_j(r,g,b) \|^2), \quad where \tag{9}$$

$$i = 1,2,...K - 1, \ j = i + 1,..., K$$

and Z_i and Z_j are the centroids. Intra cluster dependency is the minimum of the square of Euclidean distances of each centroids from the other.

Lastly, y is given as

$$y = c \times N(2,1) + 1 \tag{10}$$

Where c is a constant (selected value is 25), N(2,1) is a Gaussian distribution function with mean 2 and standard deviation 1, where the variable is the cluster number and is given as

$$N(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(k-\mu)^2}{2\sigma^2}} \tag{11}$$

Where K is the cluster number and μ=2 and σ=1 as per Turi's thesis on clustering. This is done to negate the occurrence of lower number of clusters 2 or 3.

This validity measure serves the dual purpose of
- minimizing the intra-cluster spread, and
- maximizing the inter-cluster distance.

Moreover it overcomes the tendency to select a smaller number of clusters (2 or 3) as optimal, which is an inherent limitation of other validity measures such as the Davies-Bouldin index or Dunne's index.

5. EXPERIMENTAL RESULTS

The proposed algorithm has been applied to well known natural colour images such as Lena, mandrill and peppers etc. All the results have been reported in Table 1. Column 2 of Table 1 represent the optimal range for the number of clusters for the images of Lena, mandrill and peppers etc have also been copied from [8] which is based on visual analysis by a group of ten expert people. These results have been compared to those of SOM and DCPSO [9] and to that of snob [8]. The results tabulated here for each image is the mean of 10 simulations. Figure1 are shown the some segmented output images using our proposed clustering method. The result of TABLE 1 is shows better than SOM and snob and always find a solution within the optimal range. The performance of the proposed algorithm with DCPSO produced comparable segmentation results [9].

The assumptions used for the implementation of the proposed algorithm are given as follows. The value of the parameter, c, for the validity index referred to from [8], is set to 25. The size of the population, P, is taken as 30, crossover rate, μ_c, as 0.9 and mutation rate, μ_m, as 0.01 [6]. The algorithm uses number of iterations as the terminating condition and it is set to 20. The value of K_{max} is empirically set for the individual images.

Table 1: Experiment of some natural images

Image	Optimal Range	Proposed Method	DCPSO using V	SOM	Snob
Lena	5 to 10	5	6.85	20	31
Peppers	6 to 10	6.7	6.25	20	42
Mandrill	5 to 10	6.5	6	20	39
Jet	5 to 7	6	5.3	14	22
Duck	-	7.2	-	-	-
Baboon	-	5.9	-	-	-
F16	-	6.6	-	-	-
Parrot	-	7.1	-	-	-

Original image   Segmented image by proposed
                                                  method
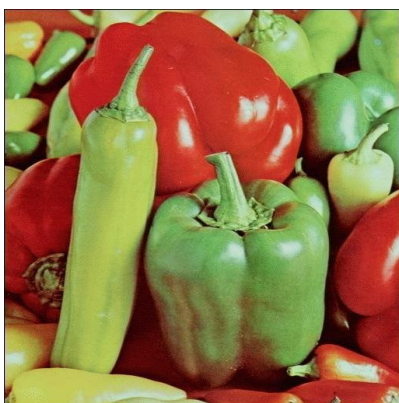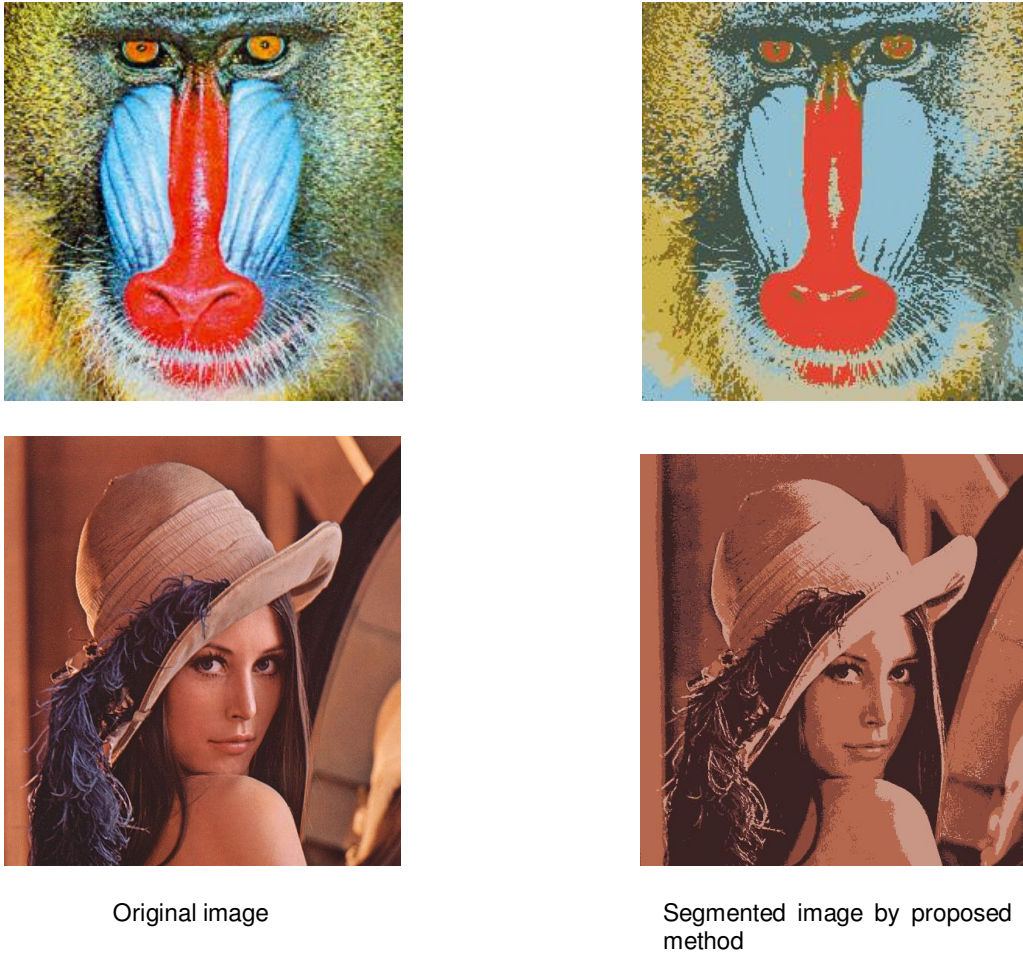
FIGURE 1: Experimental results of some natural images

6. CONSLUSIONS

This paper presented a GA based dynamic colour image segmentation algorithm based on clustering technique which determines the optimal clustering of an image dataset, with minimum user intervention. This implies that the user does not need to predict the optimal number of clusters, required to partition the dataset, in advance. Comparison of the experimental results with that of other unsupervised clustering methods, show that the technique gives satisfactory results when applied on well known natural images. Moreover results of its use on images from other fields (MRI, Satellite Images) demonstrate its wide applicability. Hence, we can conclude that the proposed GA based dynamic clustering approach using a cluster validity measure [8], can be used as an efficient unsupervised image segmentation mechanism.

7. REFERENCES

1. Ujjwal Maulik, Sanghamitra Bandyopadhyay, "Genetic algorithm-based clustering technique", Elsevier Science Ltd., 1999.

2. Qin Ding and Jim Gasvoda, "A Genetic Algorithm for Clustering on Image Data" in International Journal of Computational Intelligence Vol-1 No-1, 2004.

3. Hwei-Jen Lin, Fu-Wen Yang and Yang-Ta Kao, "An Efficient   GA-based Clustering Technique", in Tamkang Journal of Science and Engineering Vol-8 No-2, 2005.

4. Rafael C. Gonzalez, Richard E. Woods, Digital Image Processing, Pearson Education, 2002.

5. R. C. Dubes, A. K. Jain, "Clustering techniques: the user's dilemma", Pattern Recognition, 1976.

6. M. Srinivas, Lalit M. Patnaik, "Genetic Algorithms: A Survey".

7. D. E. Goldberg, Genetic Algorithms in Search, Optimization and Machine Learning, Addison-Wesley, 1989.

8. R. H. Turi, "Clustering-Based Color Image Segmentation", PhD Thesis, Monash University, Australia, 2001.

9. Mahamed G. H. Omran, Andries P Engelbrecht and Ayed Salman, "Dynamic Clustering using Particle Swarm Optimization with Application in Unsupervised Image Classification", PWASET Volume 9, 2005.

10. DW van der Merwe, AP Engelbrecht, "Data Clustering using Particle Swarm Optimization".

11. G Ball, D Hall, "A Clustering Technique for Summarizing Multivariate Data", Behavioral Science, Vol. 12, 1967.

12. LV Fausett, "Fundamentals of Neural Networks", Prentice Hall, 1994.

13. E Forgy, "Cluster Analysis of Multivariate Data: Efficiency versus Interpretability of Classification", Biometrics, Vol. 21, 1965.

14. JA Hartigan, Clustering Algorithms, John Wiley & Sons, New York, 1975.

15. T Kohonen, "Self-Organizing Maps", Springer Series in Information Sciences, Vol 30, Springer-Verlag, 1995.

16. S.Z. Selim, M.A. Ismail, K-means type algorithms: a generalized convergence theorem and characterization of local optimality, IEEE Trans. Pattern Anal. Mach. Intell.6 (1984) 81-87.

17. Dipak Kumar Kole and Amiya Halder ," An Efficient Image Segmentation Algorithm using Dynamic GA based Clustering", International Journal of Logistics and Supply Chain Management , Vol. 2, No. 1,pp.17-20, 2010.

18. Mofakharul Islam, John Yearwood and Peter Vamplew "Unsupervised Color Textured Image Segmentation Using Cluster Ensembles and MRF Model", Advances in Computer and Information Sciences and Engineering, 323–328. © Springer Science+Business Media B.V. 2008.

19. C.S. Wallace, and D..L. Dow, "MML clustering of multi-state, poisson, von mises circular and gaussian distribution", Statistics and Computing,Vol.10(1), Jan. 2000, pp.73-83.

20. R. Siddheswar and R.H. Turi, "Determination of Number of Clusters in k-means Clustering and application in Color Image Segmentation", In Proceedings of the 4th Intl. Conf. on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99), vol. Calcutta, India, 1999 pages: 137-143.

21. Wu Yiming, Yang Xiangyu, and Chan Kap Luk, "Unsupervised Color Image Segmentation based on Gaussian Mixture Model", In Proceedings of the 2003 Joint Conf. of the 4th Intl. Conf. on Information, Communications and Signal Processing, Vol. 1(15-18 Dec. 2003), pages: 541-544.

Phei-Chin Lim, Narayanan Kulathuramaiyer & Dayang NurFatimah Awg. Iskandar

Towards Semantic Clustering – A Brief Overview

Phei-Chin Lim   pclim@fit.unimas.my
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan, Sarawak, Malaysia

Narayanan Kulathuramaiyer   nara@fit.unimas.my
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan, Sarawak, Malaysia

Dayang NurFatimah Awg. Iskandar   dnfaiz@fit.unimas.my
Faculty of Computer Science and Information Technology
Universiti Malaysia Sarawak
94300 Kota Samarahan, Sarawak, Malaysia

Abstract

Image clustering is an important technology which helps users to get hold of the large amount of online visual information, especially after the rapid growth of the Web. This paper focuses on image clustering methods and their application in image collection or online image repository. Current progress of image clustering related to image retrieval and image annotation are summarized and some open problems are discussed. Related works are summarized based on the problems addressed, which are image segmentation, compact representation of image set, search space reduction, and semantic gap. Issues are also identified in current progress and semantic clustering is conjectured to be the potential trend. Our framework of semantic clustering as well as the main abstraction levels involved is briefly discussed.

Keywords: Image Clustering, Semantic Gap, Semantic Clustering, Concept Description, Symbolic Description.

1. INTRODUCTION

Advancement in digital imaging devices, technology and cost-decrease in storage devices contributed to the creation of large-scale digital images in various domains. The question now is how to effectively extract semantically meaningful information (knowledge) from these image collections. One of the fundamental of understanding and learning is to discover the natural groupings of images based on similarity of multiple characteristics or latent aspects of meaning. In this paper, the former referred to image clustering while the latter leads to semantic clustering.

One of the reasons for writing this paper is that we hardly find any work in reviewing image clustering methods by highlighting the challenges and provide association between image clustering and semantic clustering. Semantic clustering has attracted many research efforts after the year 2000 in terms of papers published. A simple experiment is conducted where we searched for publications containing the phrases *image clustering* and *semantic clustering* using

the digital libraries of ACM [1], IEEE [2], ScienceDirect [3] and SpringerLink [4] for the 2000 to 2010. Searches included the specific search phrase in publication title only and in publication title or abstract without restriction on research field. As observed in Figure 1 and 2, publication on *semantic clustering* is lower because papers found should be subset of the papers found using *image clustering*. The number of publication for both search terms in Figure 2 are higher as expected, as the search also include abstract field in addition to title field only as recorded in Figure 1. This difference indicates that the research in *semantic clustering* is still in its infancy stage where many publications have not included the term semantic clustering in the title of the paper but have started to appear in the abstract of these publications. Another observation is that the actual growth of publications on *semantic clustering* started in year 2005 and increased gradually throughout the years, which is in accordance with augmented research focus on the issue of semantic gap. Note should be taken that the number of publications shown in this paper is for reference purposes and not to be taken as quantitative proof of the actual publication count.

For the purpose of completeness and better readability for the uninitiated, we have introduced key applications of image clustering and their challenges in Section 2. Discussion on the open issues and summarization of the reviewed works is presented in Section 3. General organization of our solution is illustrated in Section 4 before we conclude in Section 5.
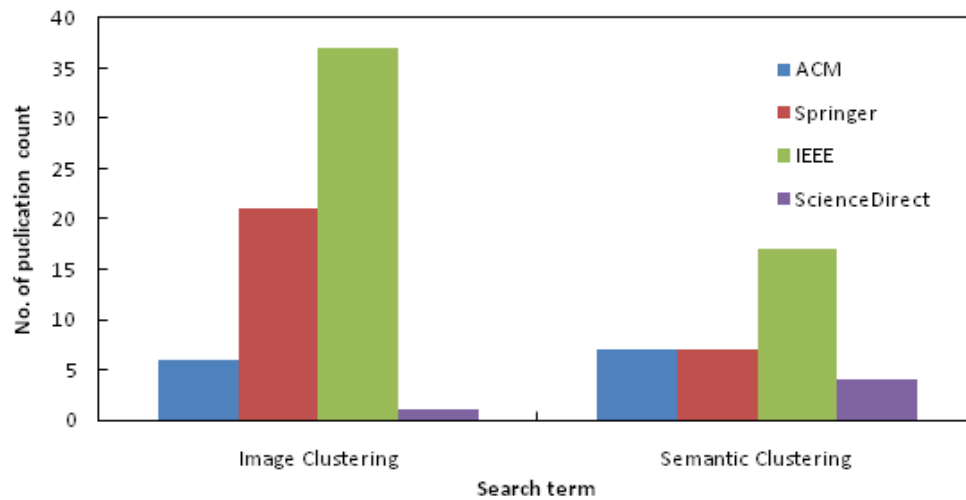
FIGURE 1: Publication count on papers with terms *image clustering* and *semantic clustering* in title only.
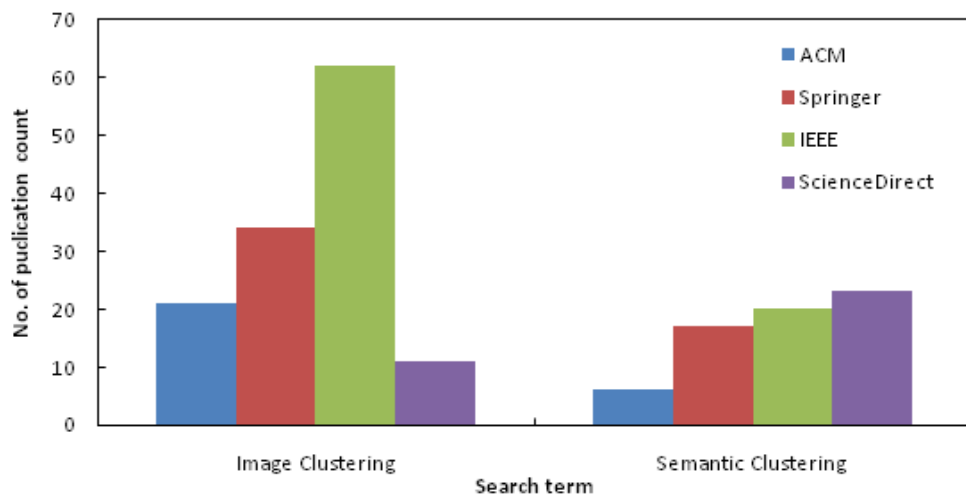
FIGURE 2: Publication count on papers with terms *image clustering* and *semantic clustering* in title or abstract.

2. IMAGE CLUSTERING

There is no universally agreed upon definition for clustering [5], hence, image clustering is generally accepted as the application of cluster analysis in order to support image understanding. Consequently, it inherited the challenges faced by cluster analysis. Interested readers may refer to the vast literature on cluster analysis [5, 6, 7] which will not be cover in this paper. In the following section, reviews on application of image clustering are organized based on the problem addressed and focus is given on employing an unsupervised method to solve these problems.

2.1 Image Segmentation

Initially, unsupervised clustering approach is mostly engaged in Geographic Information System for identifying and segmenting images into desired regions prior to being transformed into medical image analysis. Most work addressing the segmentation-classification problem requires certain level of a priori knowledge. In [8], domain knowledge about real brain CT images is quantified before incorporation into clustering algorithm to cluster region of interest. Images are then classified as normal or abnormal based on the similarity of region of interest. Counterpart of Principal Component Analysis, non-negative matrix factorization is applied on brain CT images to extract visual features and histogram-based semantic features for identifying normal and tumor CT images in [9] while a density function based image clustering analysis is used for the segmentation and identification of abdomen CT images for diagnosis purposes in [10]. In [11], a semantically supervised clustering approach is used to classify multispectral information into geo-images. Again, a priori knowledge is incorporated in the clustering process as selection criteria of the training data. Other work on detecting interesting regions by mean of clustering can be found in [12, 13].

2.2 Compact Representation

Image collections are usually heterogeneous, which makes extraction of representation a hard task. Recently, there have been growing interests in employing unsupervised methods to improve the way of representing images sets. Commonly used cluster representation schemes includes representing cluster by their centroid or by a set of distant points, nodes in a classification tree and conjunctive logical expressions [6] or newly emerging graphical representation of heat map [14] which is a multi-feature representation.

Authors in [15] addressed the problem of image representation and clustering based on learning a lower dimensional representation of the image manifold embedded in the ambient tensor space. Listed contributions including a more natural representation by treating images as matrices, computationally efficient compared to traditional dimensionality reduction algorithms which only applicable to vector data. An unsupervised neural network learning and self-organizing tree algorithm that automatically constructs a top-to-bottom hierarchy is proposed in [16]. Objects are firstly cluster according to similarity between objects in term of colour, shape and texture features before clustering images. Comparison with traditional hierarchical agglomerative clustering showed promising accuracy. Other work can be found in [17].

2.3 Search Space Reduction

Assumption and representation of each image as a whole do not really fit what a user is focusing on in an image, which is a part of an image or region in image with semantic meaning. Therefore, representing each image as a set of regions or objects is the desirable setting. Yet, this leads to expanded search space which makes retrieval efficiency a critical issue. In order to tackle the issue, we summarize works into two categories based on the instant clustering algorithm is applied.

First category comprises of works where image region clustering is done offline in the pre-processing stage and will be re-activated only when growth of new images reaches a limit. Reduction of search space is achieved by performing clustering before image retrieval. A massive 74% search space reduction is achieved in [18]. The strength of the algorithm is that the number of clusters is learned from user query log before being refined using outlier detection method.

Authors in [19] proposed an interactive approach where user's feedback on initial search results is required for analyzing user interest. Both works make use of Corel image sets.

Clustering is performed online in the post-processing stage in the second category. In [20], search results from initial query are clustered into semantic groups together with learned labels which provide overview of the content of the results. Users are able to browse through each cluster easily. Similar work can be found in [21, 22]. Humans have higher tendency to use high-level abstract concepts (image semantics) during querying or browsing of image collection rather than using low-level visual features. Consequently, huge amount of irrelevant search results are returned due to the semantic gap described in the previous section. For this reason, cluster search results are converted into symbolically similar clusters in order to filter out the relevant/irrelevant images in [23].

2.4 Semantic Gap

Content-based Image Retrieval (CBIR) is the technology that organizes digital images by their visual content which was haunted by the critical challenge of the semantic gap being defined as "the lack of coincidence between the information that one can extract from the visual data and the interpretation that the same data have for a user in a given situation" [24]. Several taxonomies of methods in addressing the problem of semantic gap were suggested in [25] and [26]. Clustering is adapted in CBIR to solve the semantic gap issue starting mid 2000's onwards.

In a typical CBIR system, query results are a set of images sorted by feature similarities but there is no guarantee that the returned images will be similar in semantic as shown in Figure 3 (a). Figure 3 (b) showed the result of retrieved images that are similar in semantic and content [27], where butterfly wing images were firstly clustered into semantic shape categories before performing similarity search to ensure only images in the same shape category are considered in finding similar images to the query. CLUE [28] approach the same issue differently by grouping semantically similar images with respect to query and return image clusters instead of a set of ordered images. Similar work was done in [20, 21, 29].

(a)

(b)
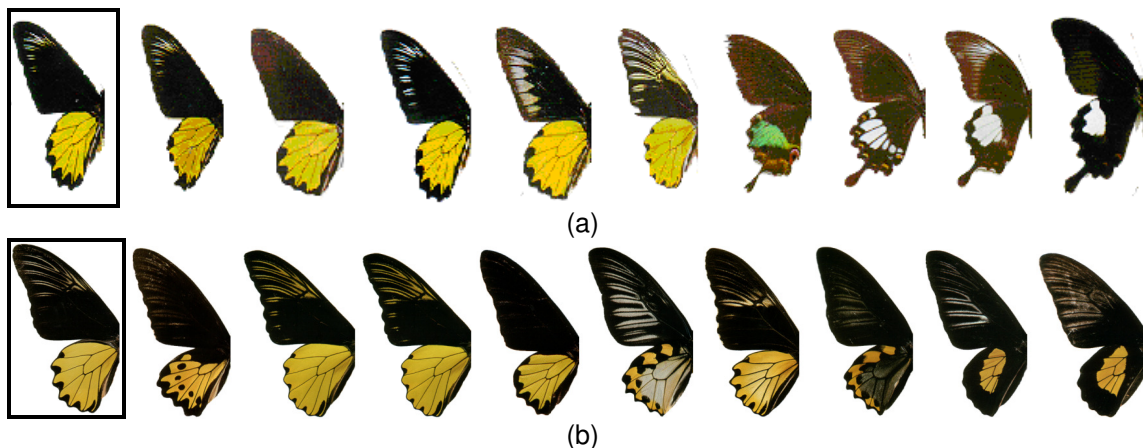
FIGURE 3: Ten most similar images retrieved (query image in highlighted), (a) conventional similarity-based approach, (b) semantic shape category from [27].

3. OPEN ISSUES

In brief, image clustering is the organization of images into different clusters (groups, classes or categories) based on similarity such that images in the same cluster share some common characteristics. A crucial issue of image clustering on large-scale image repository is compact representation for faster image indexing and searching. The derived clusters provide summarization of the image content that can be utilized for tasks such as image search and browsing. Refer to Table 1 for summarization of selected works that employed clustering algorithms.

Phei-Chin Lim, Narayanan Kulathuramaiyer & Dayang NurFatimah Awg. Iskandar

Problem		Authors	Feature	Approach	Cluster Description
Image Segmentation		Pan et.al [8]	Object's gray level, size, location, elongation, direction	Object clustering followed by image clustering	Normal/Abnormal
		Liu et.al [9]	Visual features; histogram-based semantic features	Non-negative Matrix Factorization	Normal/Tumor
		Torres et.al [11]	Spatial semantics (geometrical & topological properties)	Maximum Likelihood	Categories label (water, urban, crop, landslide, vegetation etc)
		Han et.al [13]	Texture homogeneity	Self-organizing feature map, agglomerative clustering	Nil
Search Space Reduction	Before retrieval	Liu et.al [18]	3 texture features, 2 shape features, 27 colour features	Affinity matrix clustering, network flow - based outlier detection (refine clustering result)	Nil
		Zhang & Chen [19]	13 colour features, 6 texture features	Genetic algorithm	Categories label (horse, eagle, flower, fish, glass etc)
	After retrieval	Wang et.al [22]	6 color moments, 44 auto-correlogram, 14 colour-texture moments; keywords (from image titles & description)	K-means; Image Frequency-Inverse Keyword Frequency ; Random Walk with Restarts	Annotations
		Tahayna et.al [23]	11 colour words, 11 texture words, 8 shape words	Three-layer fuzzy partitioning	Nil
Compact Representation		He et.al [15]	32x32 dimensional matrix	K-means	Nil
		Wang & Khan [16]	Object's shape, colour & texture features	Dynamic Growing Self-Organizing Tree	Nil
Semantic Gap		Lim et.al [27]	Shape features	k-Means	Visual concept & its characterization
		Chen et.al [28]	3 average colours, 3 texture features	K-means; Graph-theoretic clustering	Nil
		Wang et.al [20]	Textual features (phrases)	Salient phrase ranking, Image Search Result Clustering	10-50 final cluster names (dependent on query word) after merge & prune, other cluster
		Gao et.al [29]	Colour histogram, texture histogram	Kernel-based clustering	Nil

TABLE 1: Summarization of selected works on image clustering.

In real life, humans tend to query images using high-level abstract concepts rather than visual features. Correlation between image content and the associated text is assumed to be strong but this may not be the case. Hence, most keyword-based image search tends to return a large amount of irrelevant images to a given query. Another spectrum of the issue is due to the computation of image similarities using visual content alone, by minimizing intra-cluster difference or maximizing inter-cluster differences in the feature space. As a result, two semantically similar images may lie far from each other in the feature space while two completely different images may stay close to each other in the same feature space (refer to example in Figure 3).

Other researchers tried to incorporate textual features to solve the mentioned issue [30, 31, 32] with a certain level of success. These textual features are extracted from external sources such as ontologies, Web pages [33] and camera metadata [34]. Different sources face different challenges. Ontologies are scarce and domain dependent where creating one from scratch is time-consuming and laborious. Information provided by Web pages is unstructured in nature where reliability is an open issue. Researchers seem to be gaining interest in utilizing the rich embedded information in camera metadata, yet thorough investigation is required to prove that any value in camera metadata is correlated with image semantic.

Another observation is that the reviewed techniques build clusters solely on the basis of numerical measure. Works that provides explanation through concepts or semantic properties is scarce but is the more desirable future trend. These concepts or semantic properties explain why a set of objects conform into a cluster.

4. TOWARDS SEMANTIC CLUSTERING

Semantic clustering originated from information system field to solve text classification problem. In this paper, we refer semantic clustering as the concept of unsupervised learning to group unstructured images based on latent aspect of meaning.

Even though research for solving the semantic gap is moving towards semantic clustering, most of the work is focusing on direct mapping of visual features to semantic concepts. Little attention is given on employing a symbolic-level abstraction for the mapping, which confronts the human behavior of using high-level concepts in finding images of interest. Figure 4(a) the general direct mapping approaches while Figure 4(b) illustrated our proposed solution respectively.
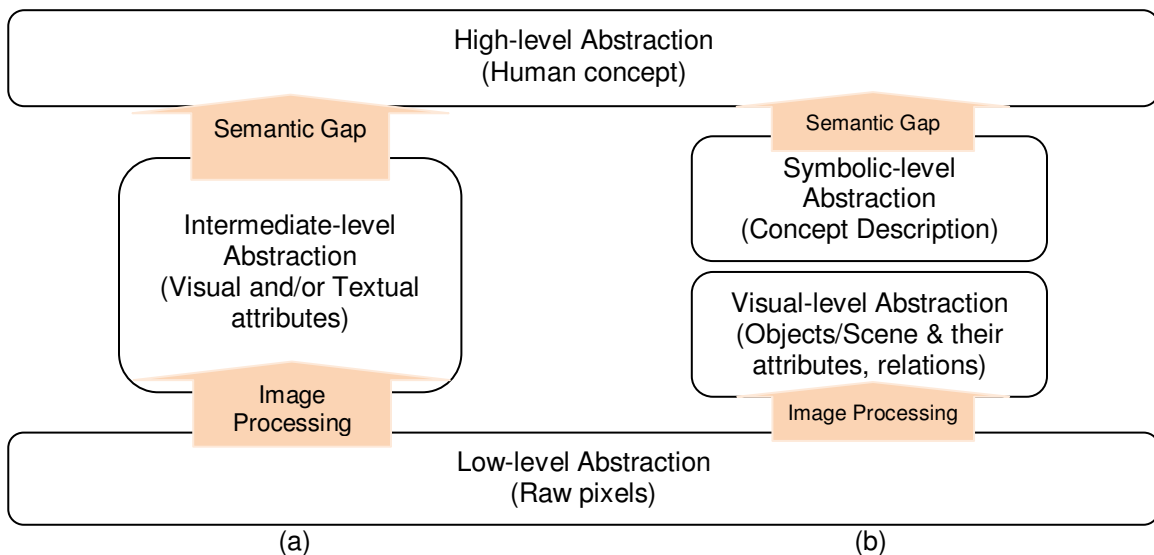
FIGURE 4: Evolution of approaches in bridging the semantic gap: (a) Visual-based; (b) proposed approach.

Our solution is inspired from how author in [23] bring together visual semantics and visual features for automatic image classification. A symbolic-level is inserted on top of visual-level

abstraction aiming to map visual features to semantic concepts in order to form concept description. Concept descriptions are logical statements describing each cluster's member, which aim at providing human understandable interpretation and user-friendly searching method. Comparison of selected works which complied with the abstraction levels in Figure 4 is listed in Table 2 and described accordingly.

Abstraction level of Images	Selected work			
	Awang Iskandar [35] & [36]	Park & Ramamohanarao [37]	Lim [26]	Shin, Kim & Kim [38]
High-level	Object/Character {Bob} Spatial relation {Bob RIGHT OF panel}	Right Apical Edge {Sharply defined} Right Diaphragm {Not sharply defined} … Class {Left Ventricle failure influencing dense ribs OLD inflammatory change at the right apex, bullous change at the left apex Ischemic heart disease and CAL}	Relative Hindwing size {very large} Relative Forewing-Hindwing ratio {Forewing very long} … Genera {*Troides*}	Abstract Emotion {Cute} Object {Bubble}
Symbolic-level	Region tagging→semantic tags/keywords (e.g., Character {Bob}: Region 1 = Body part {head}, Colour {grey}; Region 2 = Body part {eye}, Colour {black})	Symbolic description (e.g. a height of 2cm described as "normal"; height of 2.3cm described as "slightly high")	Visual feature → domain concept (e.g. $\dfrac{L_f}{L_h} \to$ wing ratio) Domain concept → linguistic value (e.g. wing ratio {Forewing very long,…etc})	Visual keywords (e.g. pink, violet, blue, green & circles)
Visual-level	Shape & colour (region area, region mean grey value, compactness & shape boundary), spatial location	Ratio, presence, position, width, size, angle & density	Shape ($\dfrac{L_f}{L_h}$, $\dfrac{4*\pi*area}{perimeter^2}$, etc)	130 Kobayashi colour & pattern (Wavelet transform)
Low-level				

TABLE 2: Examples from selected works that complied with the abstraction levels in Figure 4.

Although these selected works are successfully mapped with all the abstraction levels, richness of high-level concept is different. For example, the high-level concept [35] is object label ("comic character") where the usage is to query relevant comic strips, while [38] tried to map colour and pattern to emotional concepts ("Cute", "Classic") in textile domain. Emotional concepts are abstract attributes to describe images and the most difficult indexing level because it is highly subjective and assessments among different users vary greatly. For example, sample textile image in Table 2 may represent "Cute" to one person but "Romantic" to another. Both works do not provide concept description as compared to [26] and [37].

In [37], image features are firstly extracted and analyzed before converted to symbolic descriptions by fuzzy functions (refer to example given in Table 2). Hence, an expert (radiologist) can work with linguistic terms instead of numerical description of object features. Same in [26], visual features are extracted from salient regions before being converted to domain concepts and then further characterization. As final result, each butterfly genera is represented by their corresponding descriptions which is human understandable.

5. CONCLUDING REMARKS

This paper provided an overview of works on image clustering and discussed the challenges faced which led to the shift to semantic clustering. Recent works mostly formed clusters using low-level visual features and directly mapped clusters to high-level human concepts, which is limited in semantic richness. Even though there are attempts to incorporate textual information to enrich image understanding, there is still lack in providing cluster (concept) description. Therefore, a system that is capable to describe clusters symbolically which highlight the interpretability of clusters is highly desired.

6. REFERENCES

1. ACM – Association for Computing Machinery, http://www.acm.org. Accessed November 2010.

2. IEEE – Institute of Electronic and Electrical Engineers, http://ieeexplore.ieee.org. Accessed November 2010.

3. ScienceDirect, http://www.sciencedirect.com. Accessed November 2010.

4. SpringerLink, http://www.springerlink.com. Accessed November 2010.

5. B.S. Everitt. "*Cluster Analysis*" (3rd Edition), Edward Arnold, Ltd., London, UK, 1993.

6. A.K. Jain, M.N. Murty and P.J. Flynn. "*Data Clustering: A Review*". ACM Computing Survey, 31 (3): 264-323, 1999.

7. R. Xu and D. Wunsch. "*Survey of Clustering Algorithms*". IEEE Transactions on Neural Networks, 16 (3): 645-678, 2005.

8. H. Pan, J. Li and W. Zhang. "*Medical Image Clustering with Domain Knowledge Constraint*". W. Fan, Z. Wu, and J. Yang (Eds.): WAIM 2005, LNCS 3739*:* 719 – 724, 2005.

9. W. Liu, F. Peng, S. Feng, J. You, Z. Chen, J. Wu, K. Yuan and D. Ye. "*Semantic Feature Extraction for Brain CT Image Clustering using Nonnegative Matrix Factorization*". D. Zhang (Ed.): ICMB 2008, LNCS 4901*:* 41–48, 2007.

10. Y. Song, C. Xie, Y. Zhu, C. Li and J. Chen. "*Density Function Based Medical Image Clustering Analysis and Research*". K. Elleithy et al. (eds.): Advances in Computer, Information, and Systems Sciences, and Engineering, 149–155, 2006*.*

11. M. Torres, G. Guzman, R. Quintero, M. Moreno and S. Levachkine. "*Semantic Decomposition of LandSat TM Image*". B. Gabrys, R.J. Howlett, and L.C. Jain (Eds.): KES 2006, Part I, LNAI 4251*:* 550 – 558, 2006.

12. M.S. Hossain, K.A. Rahman, M. Hasanuzzaman and V.V. Phoha. "*A Simple and Efficient Video Image Clustering Algorithm for Person Specific Query and Image Retrieval*". ICIMCS'09. Kunming, Yunnan, China, November 23-25, 2009.

13. E. Han, J. Yang, H. Yang and K. Jung. "*Automatic Mobile Content Conversion using Semantic Image Analysis*". J. Jacko (Ed.): Human-Computer Interaction, Part III, HCII 2007, LNCS 4552: 298–307, 2007.

14. L. Wilkinson and M. Friendly. "*The History of the Cluster Heat Map*". The American Statistician, 63(2):179-184, 2009.

15. X. He, D. Cai, H. Liu and J. Han. "*Image Clustering with Tensor Representation*". MM'05. Singapore, 132-140, November 6-11, 2005.

16. L. Wang and L. Khan. "*A New Hierarchical Approach for Image Clustering*". V. Petrushin and L. Khan (Eds.), Multimedia Data Mining and Knowledge Discovery, 41-57, Springer, London, 2006.

17. S. Xia and E.R. Hancock. "*Clustering using Class Specific Hyper Graphs*". N. da Vitora Lobo et al. (Eds.): SSPR&SPR 2008, LNCS *5342:* 318–328, 2008.

18. Y. Liu, X. Chen, C. Zhang and A. Sprague. "*Semantic Clustering for Region-based Image Retrieval*". Journal of Visual Communication and Image Representation, 20 (2): 157-166, February 2009.

19. C. Zhang and X. Chen. "OCRS*: An Interactive Object-based Image Clustering and Retrieval System*". Multimedia Tools Application, 35: 71-89, 2007.

20. S. Wang, F. Jing, J. He, Q. Du and L. Zhang. "*IGroup: Presenting Web Image Search Results in Semantic Clusters*". CHI'07. San Jose, California, USA, April 28-May 3, 2007.

21. R.H. Leuken, L. Garcia, X. Olivares and R. Zwol. "*Visual Diversification of Image Search Results*". WWW 2009. Madrid, Spain, April 20-24, 2009.

22. C. Wang, F. Jing, L. Zhang and H. Zhang. "*Scalable Search-based Image Annotation*". Multimedia Systems, 14: 205-220, 2008.

23. B. Tahayna, M. Belkhatir and Y. Wang. "*Clustering of Retrieved Images by Integrating Perceptual Signal Features within Keyword-based Image Search Engines*". P. Muneesawang et al. (Eds.): PCM 2009, LNCS 5879*:* 956–961, 2009.

24. A.W.M. Smeulders, M. Worring, S. Santini A. Gupta and R. Jain. "*Content-based image retrieval at the end of the early years*". IEEE Transactions on Pattern Analysis and Machine Intelligent, 22(12): 1349–1380, 2000.

25. Y. Liu, D. Zhang, G. Lu and W. Ma. "*Region-based image retrieval with high-level semantic color names*". 11th International Multimedia Modelling Conference. Melbourne, Australia, 180-187, 2005.

26. P.C. Lim, "*A Generalized Framework for Mapping Low-level Visual Features to High-level Semantic Features*", Master's thesis, Universiti Malaysia Sarawak, 2008.

27. P.C. Lim, N. Kulathuramaiyer, F. Abang & Y.C. Wang, "*Classification of Butterfly Wing Images*". International Conference on Intelligent Systems. Kuala Lumpur, Malaysia, December 1-3, 2005.

28. Y. Chen, J.Z. Wang & R. Krovetz, "*CLUE: Cluster-Based Image Retrieval of Images by Unsupervised Learning*". IEEE Transactions on Image Processing, 14(8): 1187-1201, 2005.

29. Y. Gao, J. Fan, H. Luo and S. Satoh, "*A Novel Approach for Filtering Junk Images from Google Search Results*". S. Satoh, F. Nack, and M. Etoh (Eds.): MMM 2008, LNCS 4903*:* 1–12, 2008.

30. P.A. Moellic, J.E. Haugeard and G. Pitel, "*Image Clustering based on a Shared Nearest Neighbours Approach for Tagged Collections*". CIVR'08. Niagara Falls, Ontario, Canada, July 7-9, 2008.

31. L. Cao, J. Luo & T.S. Huang, "*Annotating Photo Collections by Label Propagation according to Multiple Similarity Cues*". MM'08. Vancouver, British Columbia, Canada, October 26-31, 2008.

32. M. Ferecatu, N. Boujemaa & M. Crucianu, "*Semantic Interactive Image Retrieval Combining Visual and Conceptual Content Description*". Multimedia Systems, 13: 309-322, 2007.

33. W. Lu, R. Xue, H. Li & J. Wang, "*A Strategy of Semantic Information Extraction for Web Image*", S. Wang, L. Yu, F. Wen, S. He, Y. Fang & K.K. Lai (Eds.), Business Intelligence: Artificial Intelligence in Business, Industry & Engineering, 2nd International Conference on Business Intelligence and Financial Engineering. Beijing, China, July 24-26, 2009.

34. S. Yang, S. Kim and Y.M. Ro. "*Semantic Home Photo Categorization*", IEEE Transactions on Circuits and Systems for Video Technology, 17(3): 324-335, 2007.

35. D.N.F. Awang Iskandar. "*Image Retrieval using Automatic Region Tagging*", PhD dissemination, School of Computer Science and Information Technology, Royal Melbourne Institute of Technology University, March 2008.

36. D.N.F. Awang Iskandar, J.A. Thom and S.M.M. Tahaghoghi, "*Content-based Image Retrieval Using Image Regions as Query Examples*". A. Fekete and X. Lin (Eds.): 19th Australasian Database Conference (ADC2008). CRPIT, 75: 39-75. Wollongong, NSW, Australia, 2008.

37. M. Park and K. Ramamohanarao. "*Automatic extraction of semantic concepts in medical images*". IEEE International Conference on Image Processing. Singapore, October 24-27, 2004.

38. Y. Shin, Y. Kim and E.Y. Kim. "*Automatic textile image annotation by predicting emotional concepts from visual features*". Image and Vision Computing, 28(3): 526-537, 2010.

Usman Zakir, Iffat Zafar & Eran A. Edirisinghe

Road Sign Detection and Recognition by Using Local Energy based Shape Histogram (LESH)

Usman Zakir   u.zakir@lboro.ac.uk
Department of Computer Science,
Loughborough University
Loughborough, LE113TU, UK

Iffat Zafar   i.zafar@lboro.ac.uk
Department of Computer Science,
Loughborough University
Loughborough, LE113TU, UK

Eran A. Edirisinghe   e.a.edirisinghe@lboro.ac.uk
Department of Computer Science,
Loughborough University
Loughborough, LE113TU, UK

Abstract

This paper describes an efficient approach towards road sign detection and recognition. The proposed system is divided into three sections namely; Colour Segmentation of the road traffic signs using the HSV colour space considering varying lighting conditions, Shape Classification using the Contourlet Transform considering occlusion and rotation of the candidate signs and the Recognition of the road traffic signs using features of a Local Energy based Shape Histogram (LESH). We have provided experimental results and a detailed analysis to justify that the algorithm described in this paper is robust enough to detect and recognize road signs under varying weather, occlusion, rotation and scaling conditions.

Keywords: Road Signs, HSV, Contourlet Transform, LESH, Autonomous Vehicles

1. INTRODUCTION

Road sign detection and recognition has drawn considerable research attention in recent years due to its challenging nature as a computer vision problem. Road signs have a direct impact on ones daily life be it as a driver, cyclist or as a pedestrian as possible life threats can be easily formed due to their ignorance. In recent years a number of Driver Assistance Systems have been implemented including vision based algorithms claiming to be efficient towards road sign detection and recognition. Generally road signs consist of three properties; firstly they are represented by colours such as Red, Green, Blue, and Brown etc. Secondly they consist of a particular shape such as Circular, Triangular, Octagonal, Square etc. The inner contents of the road signs represent the third property, which may vary depending on the application of the road sign. In this paper we have highlighted the importance of using these properties separately by considering different problems including various lighting conditions, scaling, angular rotation and occlusion.

The perceptual colour of a road sign appears to be different due to varying lighting conditions. This makes it difficult to extract the accurate colour information of a sign. The weather defined by

Usman Zakir, Iffat Zafar & Eran A. Edirisinghe

rain, fog, snow etc., and time of the day defined by day, dusk, night etc., play an important role in creating the above mentioned variations of illumination.

The size of a road sign as appearing in the scene has an impact on its detection and identification accuracies. Signs that appear small will not be even detected as picking up the colour or shape will be a challenging issue for even the best computer vision algorithms. Therefore it is important to include a system functionality, which keeps track of the candidate sign from the point it first become visible in the scene until a reasonable size that will enable its recognition.

Further the detection and recognition in the presence of likely angular rotation of a road sign is also a further challenging computer vision problem which needs to be addressed and resolved. The detection and recognition of a road sign can also be affected by occlusion, i.e., due to the presence of objects in the field of view. In the proposed approach the verification of the road sign is completed at multiple stages. This helps to overcome partial occlusion at a high success rate.

To accommodate the recognition ability of all categories of road signs it is important to distinguish road signs not only by their content but also by their colour and shape information. Table 1 below shows different categorical divisions of road signs according to their colour and shape information.

Triangle	Red Circle	Hexagon	Blue Circle	Square

TABLE 1: Different road signs categorized according to colour and shape information

For clarity of presentation the paper is organized into the following sections: Section 2 focuses on a literature review in the area of road sign detection and recognition. Section 3 presents the proposed road sign detection and recognition system. It is further categorized into three important sections A: Colour based segmentation of a road sign using the HSV colour space, B: Shape Classification using Contourlet Transform (CT) and C: Recognition using features from the Local Energy based Shape Histogram (LESH) and Support Vector Machine (SVM) based classification. Section 4 illustrates the experimental setup and results. The conclusion and future work are presented in Section 5.

2. STATE OF THE ART

Generally road sign detection and recognition literature can be divided into two groups that concentrate on 1: Road Sign Detection and 2: Road Sign Recognition. The Road Sign Detection consists of the procedure of extracting the candidate road sign from a scene. Majority of the work in detection is initiated using colour information of the road sign. The segmentation of the candidate sign from the scene is carried out by employing a colour space. The most popular colour spaces used for this purpose are HSI, RGB, CIElab, YCbCr, CYMK and HSV. In the colour based segmentation approaches images are first converted to a designated colour space and then a segmentation algorithm e.g. thresholding, is applied [2]. In [1], [3], [4], colour based segmentation of road signs have been achieved by first transforming the original image to the HSI colour space and subsequently marking the desired colour pixels (such as Red, Green, Blue etc.) by a white pixel. Pixels that are outside the threshold values are treated as background or noise. Thus a binary image is formed in which white pixels represents the desired coloured area and black pixels represent noise or background. In [5], [6], segmentation has been achieved using RGB (Red, Green, Blue) colour space and the desired pixels are extracted by using threshold values for each colour. These threshold values are obtained on the basis of changing illumination conditions during different times of the day [3]. In [2], [8], [9], segmentation is performed by

transforming the images to HSV (Hue, Saturation and Value) colour space to obtain thresholds using Otsu's algorithm [10]. In [11], though the segmentation is performed in the HSV colour space, pixels of interest are obtained by employing a set of fuzzy rules. In [12], CIElab colour space is employed with a Gaussian model to target the colour information and in [13] chromatic channels are transformed to binary images using Otsu's algorithm. YCbCr colour space is employed in [14] and adaptive thresholding has been performed to obtain the pixel of interest. In [15] CIECAM97 colour space is employed to segment out the road signs and results are compared with HSI, RGB and CIELUV colour space segmentation. Recently a number of attempts have been made to use combined colour spaces and combined colour space models in road sign colour detection. In [16] a joint colour space segmentation approach has been adopted. The Hue of the HSV colour space and image chrominance (U, V) values from YUV space have been jointly used. The results of the two colour spaces are combined by a logical AND operation. 256 RGB and HSL transforms are used in [17] to construct colour distinctions of the image by following simple thresholding. RGB and HSI colour spaces are jointly used to threshold the image in [18]. HSI threshold values for Blue and Red colours are tabulated. Joint colour space has been employed [19] in which HSI is used to extract the chromatic information and RGB is used to extract the achromatic information. Adaptive chrominance and luminance thresholding has been achieved in [20] by employing joint colour space i.e. CIElab and HSI. In [21] a four colour space based colour classifier has been introduced to segment road signs captured under various weathers and lighting conditions.

In conjunction with colour based approaches, shape based approaches are also seldom used in the segmentation of road signs. In this case the colour information is either used as a pre processor or never used at all. Two shape measures are used for circular and triangular road signs by using a fuzzy shape recognizer in [22]. Distance to Border (DtB) vectors are obtained of the segmented road sign from four different views and shape has been classified by using a linear Support Vector Machine (SVM) [4].For shape determination Hough Transform [18] and median filter are applied to detect the triangle and circular road sign shapes [23].The circular shapes are identified through Fast Radial symmetry detection method (FRS) and the triangular and square shapes are identified by using Harris corner detection algorithm [10].Difference of candidate background histogram and template histogram is used to obtain the shape information of the candidate sign [24].Fast Fourier Transform (FFT) is employed to retrieve shape signatures by comparing it with shape references in [25] and [26]. The 2D correlation coefficient is derived to represent the shape similarity by correlating the candidate ROI with 2D binary template in [27]. A coarse-to-fine scheme is proposed for candidate shape verification in [28]. Hopfield neural networks are used to determine triangular and rectangular shapes by obtaining information related to angle, gradient and distance between edges [29]. The RANdom Sampling and Consensus (RANSAC) algorithm is used to examine and determine the circular shapes in [30]. The shape context and its invariance to rotation and scale are determined by employing corners as salient points in [31]. Geometric properties of different road sign shapes are computed by using the Affine Moment Invariants in [11].

Both the recognition and classification of road signs has been carried out by employing series of approaches which include Multi Layer Perceptrons (MLP), Neural Networks [32] with feed forward topology in [33]. To recognize road signs a template matching technique was proposed in [22]. Different types of Neural Networks are employed in the recognition and classification of road signs such as LWN++ in [34], 3-layered [35], Hopfield [36], back propagation [37] and ART1 [38]. Scale Invariant Feature Transform (SIFT) is used in the recognition of road signs [39] and similarity measures among features are obtained through objective functions [40]. SVM Gaussian kernel is used for content recognition in [41]. A two class SVM classifier is used in [42], one versus all SVM classifier in conjunction with RBF kernel is proposed in [43] and SVM polynomial kernel is used with regards to AdaBoost naive Bayes in [44]. The classification of road signs has been performed by Normalized Correlations in [45] and Normalized Cross Correlation is introduced in [46]. Other recognition methods such as, Fuzzy Shape recognizers [11], Adaptive Hausdorff distance based on similarity weights [47] and Gabour Wavelets [18] filters have also been used. The next section aims to address the above research gaps in designing a robust road

Usman Zakir, Iffat Zafar & Eran A. Edirisinghe

sign detection and recognition system capable of performing under wide variations of illumination and environmental conditions.

3. SYSTEM OVERVIEW

The proposed road sign detection and recognition system comprises of three stages, which are detailed in this section. Figure 1 gives an overview of the framework.
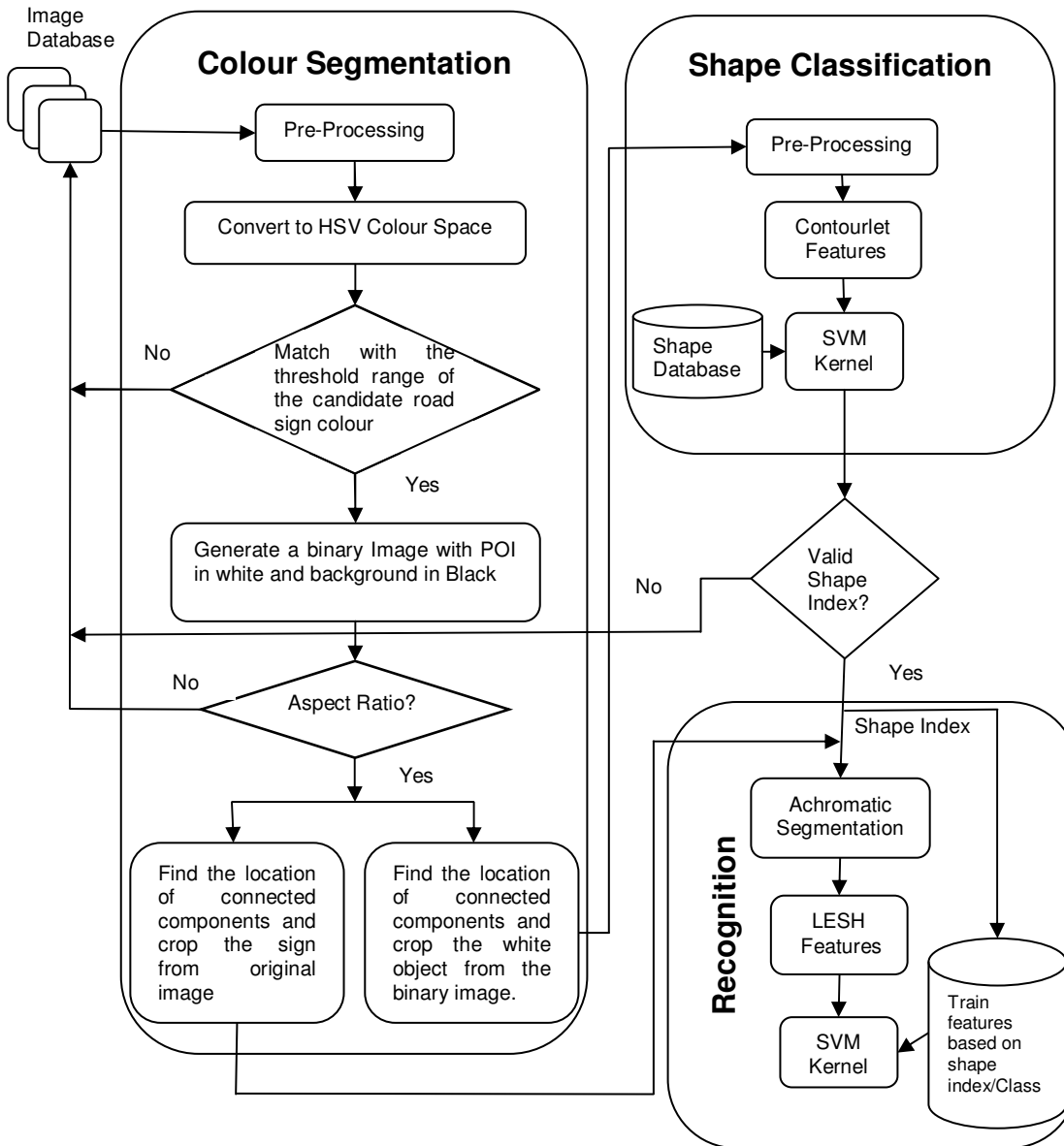
FIGURE 1: Overview of proposed road sign Detection and Recognition framework

A. *Colour Segmentation*: The road sign detection and recognition system is initiated by the colour segmentation process. This process involve the use of a colour constancy algorithm followed by a HSV colour space based segmentation of road signs using threshold values described in [2].

B. *Shape Classification*: The segmented Region of Interest (ROI) is further processed to remove the false detections and to identify those geometric features which truly represent

road signs. Contourlet Transforms with Support Vector Machine (SVM) kernel are applied to measure the shape properties at this stage.

C. *Recognition*: The road signs have different meanings according to their contents. The recognition stage comprises of the use of Local Energy based Shape Histogram features with SVM kernel to recognize various contents of road signs at this stage.

3.1 Colour Segmentation

As mentioned in section 2 the segmentation process of the road sign is initiated by employing a colour space. The possible challenges such as variation in illumination and exposure to rain/sun light may affect the colour definition of the road signs. We have employed HSV colour space to segment the colour of the road sign due to its ability to enhance segmentation results [2] as compared to other available colour spaces, specifically in the outdoor environments. RGB images taken from the camera are first passed through a pre-processing stage which helps to maintain the colour constancy in different illumination conditions [49]. It is then transformed to HSV colour space using the following equation [48].

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

$$S = 1 - \left(\frac{3}{R+G+B}\right)\min(R,G,B) \qquad (2)$$

$$V = \max(R,G,B) \qquad (3)$$

Where R,G and B are representing the three channels of a RGB image. Pixels of interest (POI) can be segmented by using all three of the above components as shown in equation 1-3 i.e. H, S and V. The Hue (H) component contains the colour information while Saturation(S) generates different shades of a particular colour and Value (V) indicates the brightness or darkness of the colour components respectively. Every colour in H component of HSV colour gamut has an angular value which varies according to change appears in S and V components. If S and V components remain constant i.e. S = 100% and I = 100% then Red, Green and Blue colours can be found at $0^o, 120^o$ and 240° anti clock wise respectively. These values are determined by taking the histogram of H, S and V components. In the segmentation process a desired pixel which belongs to Red, Blue or Green colours are represented by 1 or white colour and rest of the pixels are treated as background or 0 as shown in the Figures 2(b) and 3(b). At this stage we have to utilize a rule based approach to keep the segmentation process as fine as possible. At first instance median filtering and thinning are utilized for improving the segmentation consistency. Further to this, objects are selected as road sign candidate and discarded as noise according to their aspect ratio. A priority definition of the colours is embedded with the aspect ratio of the candidate sign to help in selection and rejection process which is shown in Table 2. Red colour has high priority and it can appear to be Triangle, Hexagon or Circular shape in the scene. Blue colour can appear as rectangle and circle where as Green coloured signs can appear as Rectangular shape in the image. The corresponding bounding box of the segmented object is analyzed according to its centre and corner points as illustrated in the Figure 4. The segmented objects within the aspect ratio should meet this sign criteria definition. The Figure 2(c) and 3(c) are the results after removing objects with unsuitable aspect ratio and sign criteria.

(a)   (b)   (c)
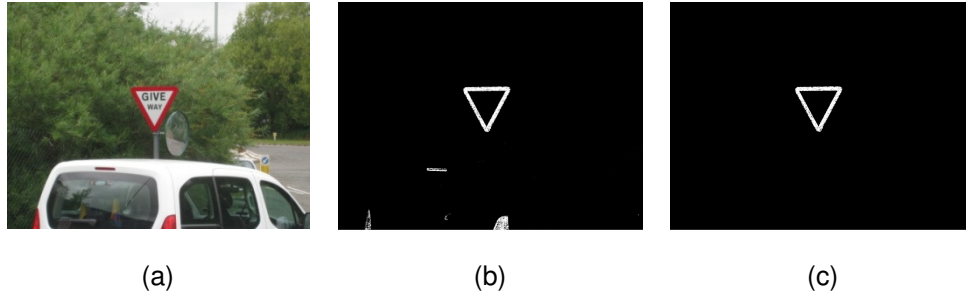
FIGURE 2: Red Colour Segmentation

(a)   (b)   (c)
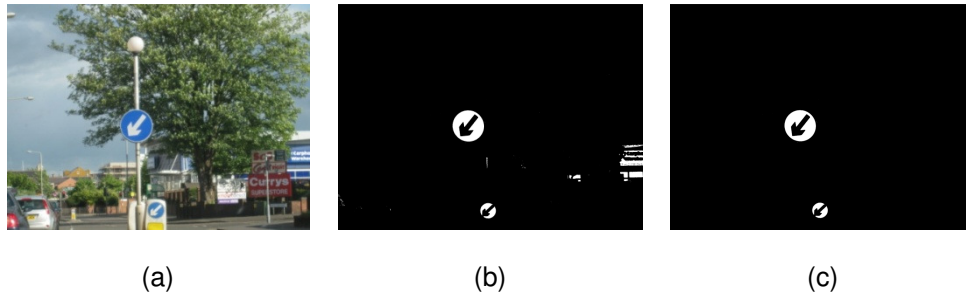
FIGURE 3: Blue Colour Segmentation

Colour	Priority	Shape
Red	1	Triangle, Hexagon, Circle
Blue	2	Rectangle, Circle
Green	3	Rectangle

TABLE 2: Colours with their shapes and priority
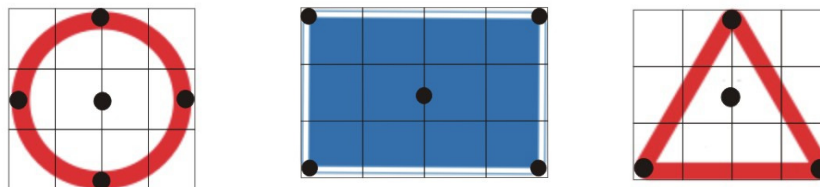
FIGURE 4: The Sign Criteria

3.2 Shape Classification

Road Signs are represented by basic geometric shapes i.e. Triangular, Circular, Hexagonal and Square/Rectangle. This analysis of the road sign, consisting of one of the shapes mentioned earlier is carried out after Colour Segmentation of the road signs. It not only helps to remove non-sign objects but also lessen the burden at recognition stage.

The segmented objects resulting from Colour Segmentation stage are pre-processed at the first stage of the Shape Classification. The pre-processing involves edge sharpening and filling the holes inside the objects. Figure 5 shows the procedure of pre-processing stage in the Shape Classification where as Figure 5(b) shows processed objects which is the output of the pre-processing stage. The Contourlet Transform [50] that is successfully used in image compression and enhancement domains has been employed here to extract the shape features of the pre-processed objects. The shape features comprises of contour edges of an object along 2 dimensional (2D) spaces. These contours are stored for training or testing purposes in the later

Usman Zakir, Iffat Zafar & Eran A. Edirisinghe

Shape Classification stage. The 3 level frequency decomposition is used due to its capability of producing diagonal direction edges. It is a very helpful technique for distinguishing circular shapes with hexagonal shapes. The comparison of 3-level frequency decomposition with 2-level frequency decomposition is presented in Section 4. The proposed frequency decomposition contains 'Haar' attributes of Laplacian pyramid filter and a Directional filter bank, as illustrated in Figure 6.
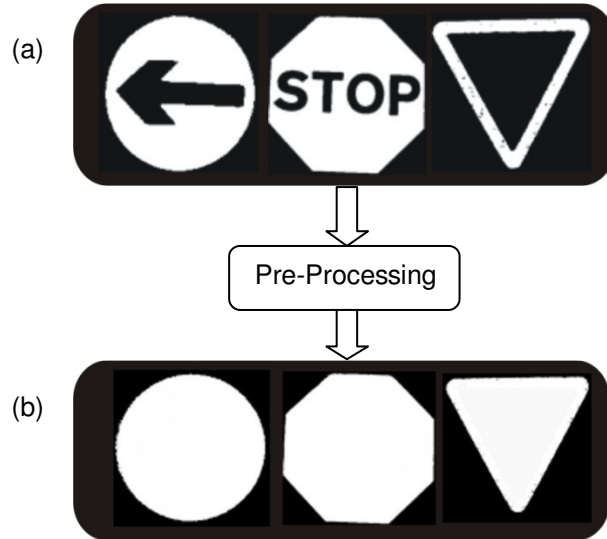
FIGURE 5: Shape Classification Pre-processing procedure (a) Binary Objects (b) Processed Objects
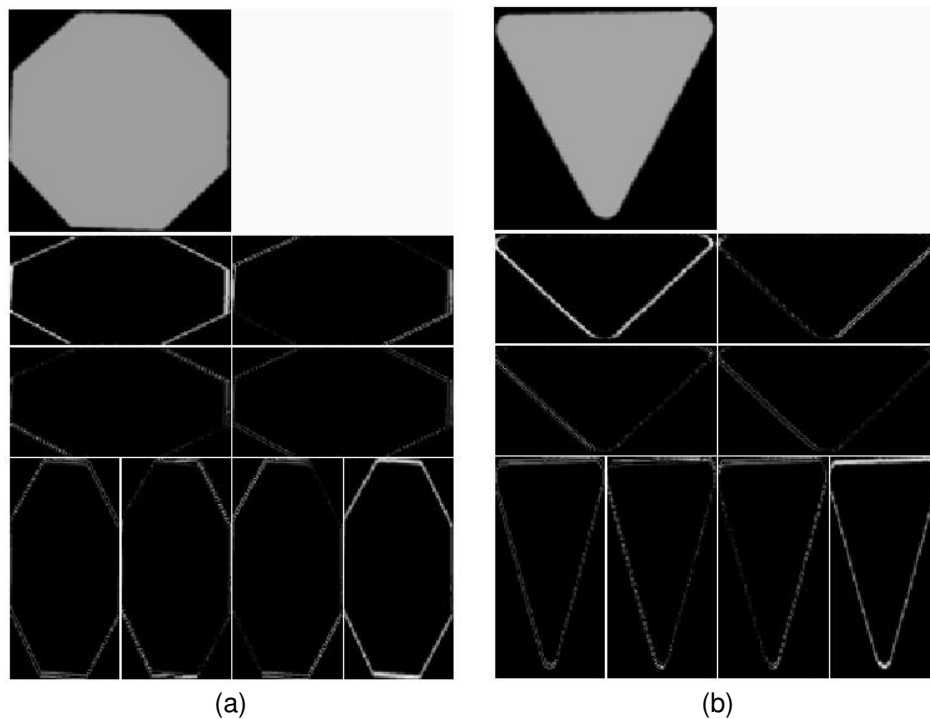
FIGURE 6: Contourlet Transform 3-level frequency decomposition (a) Hexagonal Shape (b) Triangular Shape

The extracted Contourlet features of different shapes are fed in to Support Vector Machine (SVM) which helps to classify the exact shape of the candidate sign. The SVM classifier first proposed in [51] is an effective and powerful method for general purpose pattern recognition and classification. SVM based kernels are used to map the input data to the high dimensional feature space and process it in linear form. In our experiments we have tested few popular kernels functions, $k(u,v)$ defined as follows:

$$k(u,v) = \langle u,v \rangle \tag{4}$$

$$k(u,v) = (u.v+1)^d \tag{5}$$

$$k(u,v) = e^{\frac{|u-v|^2}{\sigma^2}} \tag{6}$$

In most applications it has been generally concluded that, a low degree polynomial kernel shown in equation (5) or a Radial Basis Function (RBF) as kernel shown in equation (6) works quite well. In our case, a polynomial kernel with degree 1 and 2 provides the best results for shape classification. The method described here for shape classification is invariant to scale, rotation and partial occlusion problems. The method has the capability to classify any geometric shape due to its distinguishing multi channels feature analysis.

3.3 Recognition

Once the candidate shape is classified, it initiates the process of Recognition and classification of the road sign contents. The Recognition process comprises of the LESH features extraction of the road sign contents and training/testing of these features is carried out by employing SVM kernel classifier. The candidate road signs which are extracted through Colour Segmentation and hence validated as the road sign shape by Shape Classification are further processed to obtain the valid road sign contents for feature extraction.

The road sign contents are normally represented with black and white colours. In this context achromatic segmentation is introduced to extract the black and white areas of the candidate road sign from rest of the image. The objects which are meeting the criteria of aspect ratio are considered as valid road sign contents and vice versa. The black and white objects after achromatic segmentation and in accordance to aspect ratio are shown in Figure 7.

(a)   (b)   (c)   (d)   (e)
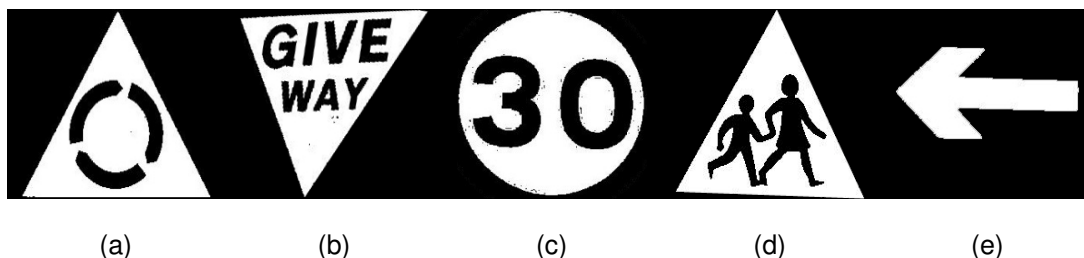
FIGURE 7: Road signs after achromatic segmentation

The Local Energy Model was first introduced in [52] proving that features can be extracted at those points from an image where local frequency components represent maximum uniformity. The extended framework of local energy model is given in Equation (8) and is normalized by the summation of noise cancellation factor T, Sine of phase deviation and factor W, which is the weighting of the frequency spread. Further details of this extended framework can be found in [53].

$$E = \frac{\sum_n W(x) \lfloor A_n(x)(\cos(\phi_n(x) - \bar{\phi}(x)) - |\sin(\phi_n(x) - \bar{\phi}(x))|) - T \rfloor}{\sum_n A_n(x) + \varepsilon} \tag{8}$$

The Local Energy gives reliable information to extract the interest points from an image in an invariant manner to illumination and noise. This raw energy indicates the corners, contours or edges of underlying shape in an image. LESH [54] features are obtained firstly by dividing the candidate image into 16 sub-regions and then the local energy information is calculated for each sub-region along 8 different orientations with the help of Gabor Wavelets kernels [55].

$$G_{u,v}(z) = I(z) * \Psi_{u,v}(z) \tag{9}$$

Where $z = (x, y)$ represents the image position, the symbol "$*$" is convolution operator and $\Psi_{u,v}(z)$ can be calculated as equation (10).

$$\Psi_{u,v}(z) = \frac{\|k_{u,v}\|^2}{\sigma^2} e^{\left(\frac{-\|k_{u,v}\|^2 \|z\|^2}{2\sigma^2}\right)} \left[e^{ik_{u,v}z} - e^{-\sigma^2/2} \right] \tag{10}$$

The orientation label map is produced representing labels of orientation of pixels containing largest energy across all scales in an image. The local histogram h is calculated in equation (12) as follows:

$$w_r = \frac{1}{\sqrt{2\pi}\sigma} e^{\left[(x-r_{xo})^2 + (y-r_{yo})^2\right]/\sigma^2} \tag{11}$$

$$h_{r,b} = \sum w_r \times E \times \delta_{Lb} \tag{12}$$

Where w is the Gaussian weighting function of region r calculated as in equation (11), E represents the local energy computed as equation (8), δ_{Lb} represents Kronecker's delta, L orientation label map and b current bin. From the above description it can be seen that the LESH descriptor of a shape is 8x16=128 dimensional feature vector.

Each object shown in Figure 7(a-e) is representing a separate class of objects. These objects are normalized into fixed size dimensions and converted to gray level images. The next stage is to extract the LESH features of these normalized images. LESH features are obtained by computing the local energy along each filter orientation of image sub-region and its histogram is generated. The overall histogram represents the concatenated histograms computed along each sub-region of the image. Figure 8 shows the LESH features representation of 30 speed limit and round about signs respectively. These features from different classes are classified with the help of multiclass SVM polynomial kernel which is explained earlier in section 3.2.

4. EXPERIMENTAL SETUP AND RESULTS

Using a standard photographic camera, mounted of a car, we have obtained 1200 images of different road signs. These signs were captured during various ambient and lit-up levels of illumination and weather conditions. The road signs collected were limited to ones with basic colours i.e. Red, Green and Blue. Figure 9 illustrates some samples of images taken from a Canon IXUS80IS digital camera at a resolution of 2592 x 1944. The images are pre-processed before converting them to the HSV colour space. This involves the stabilization process of colour
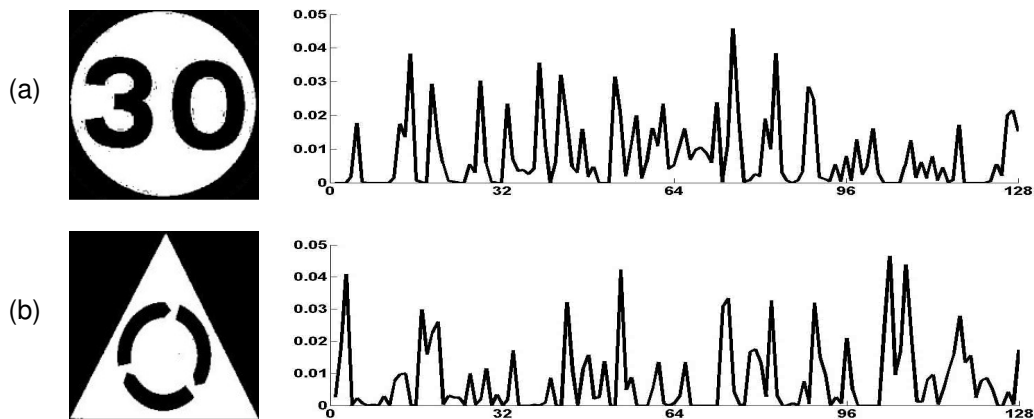
FIGURE 8: LESH feature representation for (a) 30 speed limit sign (b) round about road sign

FIGURE 9: Original coloured images contain red, blue and green colour road signs captured under different environmental conditions

constancy at a particular illumination level. As shown in Figure 1 from section 3.1 the processed images are then transform to HSV colour space and POI are obtained using the threshold values [2] for Red, Blue and Green coloured road signs. The output image is then later converted to binary image in which the white pixels represents the POI and vice versa. The bounding boxes of the white region are analyzed according to their aspect ratios and the sign criteria.

The objects which meet the criteria of being a candidate sign are cropped from both the original image and the binary image. The location information of these cropped signs is also noted for later usage. Figure 10 shows some results after Colour Segmentation. The objects segmented as candidate sign are passed through the Shape Classification module where experiments are carried out on 5 different shape classes. The inputs of the Shape Classification module are the cropped objects from the binary image as shown in the Figure 10(c). These objects are pre processed at first instance in this module which consists of normalization of the object in terms of size and removal of the contents.

The objects are resized to 100x100 fixed size dimensions then algorithms for edge sharpening and to fill the necessary holes are applied as shown in Figure 5. The shape feature selection is done by using Contourlet Transform, which represents the object shape with respect to its

(a)   (b)   (c)

FIGURE 10: Colour Segmentation using HSV colour space (a) Road Signs in the natural environment (b) Cropped Road Signs from the original Image (c) Cropped Road Signs from the Binary Image

Index	Shape	Level*	Pyramidal and Directional Filters	Trained Instances
1		3	Haar	50
2		3	Haar	50
3		3	Haar	50
4		3	Haar	50
5		3	Haar	50

TABLE 3: Shapes considered for obtaining Contourlet features, Level* = Sub band Decomposition

Shape Index	Classified Shapes with Level			
	Level	Success	Level	Success
1	2	75%	3	100%
2	2	0%	3	100%
3	2	75%	3	100%
4	2	100%	3	100%
5	2	100%	3	100%

TABLE 4: Success percentage at different decomposition Levels

Usman Zakir, Iffat Zafar & Eran A. Edirisinghe

relevant class. Table-3 shows the details about the shape classes we have considered for our training and testing purposes. Table-4 shows the success percentage while using different decomposition levels on each class. It is observed that as Hexagonal shapes require more sides to represent the shape, when using the Level 2 decomposition which only uses the horizontal and vertical edges of the shape, the shape classifier gives the wrong interpretation of Hexagons, i.e. they are classified as of having circular shape. The Level 3 decomposition gives directional edges at angle 2π by considering all sides of hexagon shape. The classification of shape involves the comparison between testing object instances having unknown class with the offline trained object instances named by defined classes. We have performed offline training of 50 instances as per shape class which was proven experimentally to classify any geometrical shape at significantly high accuracy. Figure 11 and Table-4 are showing the results of Shape Classification in which it can be observed that by using the Level 3 decomposition of Contourlet Transform the classifier produces 100% success rate.
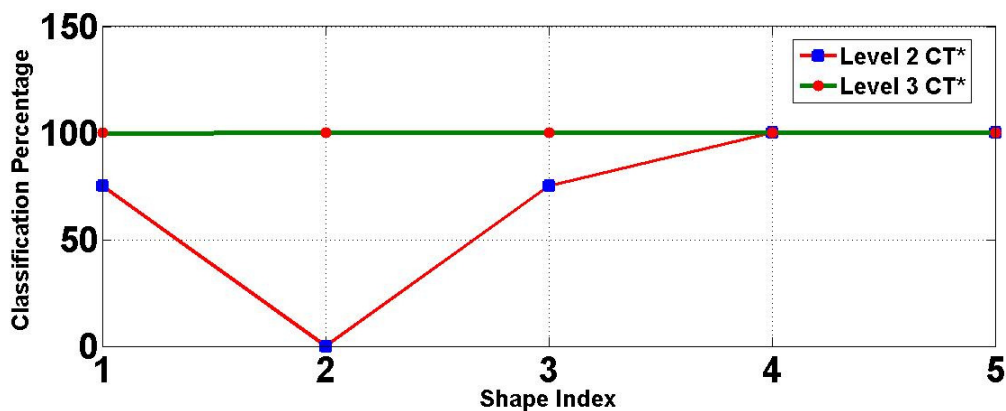
FIGURE 11: SVM classification results, CT = Contourlet Transform

The recognition module takes the input which is the combination of shape index and the cropped road sign from the original image as shown in figure 1. At first stage of recognition the contents of the road sign are captured by applying achromatic segmentation. This segmentation uses the V component of the HSV colour space to obtain the intensities of the pixels in the image. The image is then converted to binary image by using suitable threshold parameters and in accordance to its pixel intensities results. The images used in our experiments are resized to 128x128 fixed sized dimensions and later converted as gray level image. Figure 12 and 13 show LESH feature representation of '*Give Way*' and '*15 Speed Limit*' road sign contents respectively. Figure 12(a-b) and 13(a-b) are representing the original images with colour information and their corresponding pre processed images respectively. The extracted LESH features of these images are presented in Figure 12(c) and 13(c) respectively. The SVM classifier is trained offline on the LESH features of 25 classes containing 25 instances of each class.

The shape index information is fed in to the trained feature database of 25 classes. This helps to retrieve only the specific road sign content which matches with the shape index. The SVM kernel testing set is compared with the retrieved query results from feature database. This makes the functionally of the recognition task more robust, accurate and efficient.

Table 5 shows the recognition accuracy results during various illumination conditions. The results show that the best accuracy figures are indicated during daylight conditions and the worst during fogy situations. Rainy conditions do not appear to have affected the performance of the algorithm significantly.
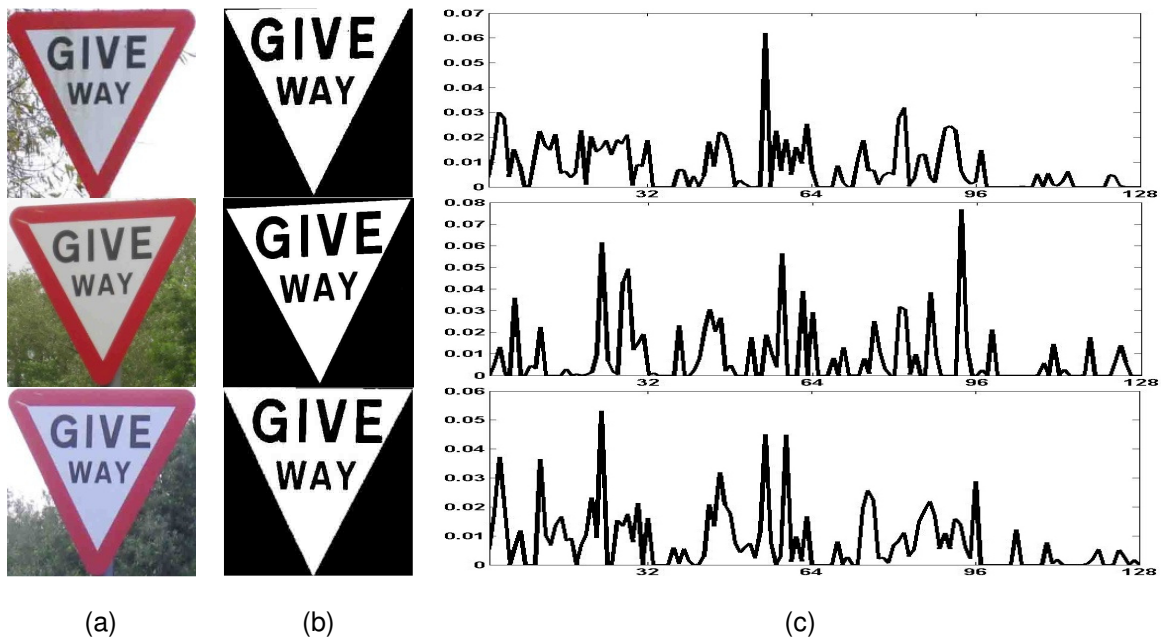
FIGURE 12: LESH feature representation for '*Give Way'* road sign
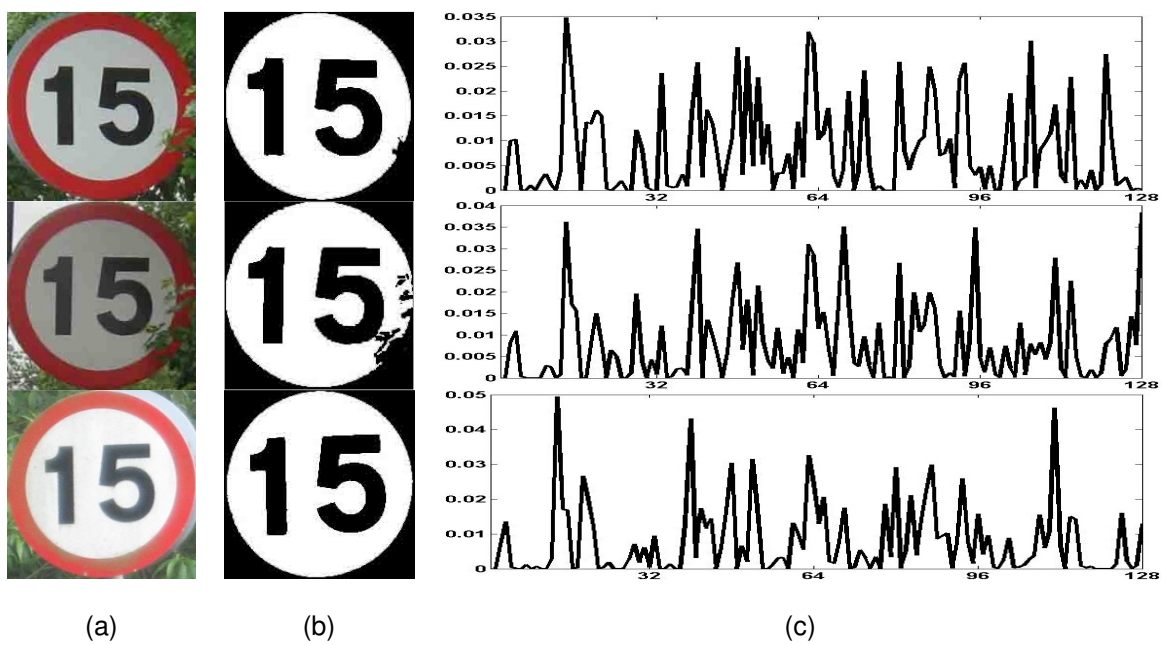
FIGURE 13: LESH feature representation for '15 speed limit' road sign

	Ambient Illumination			
	Day light	Raining	Fog	Night
Traffic Signs	560	430	70	110
Correct Recognition	552	426	61	103
False Recognition	8	4	9	7

TABLE 5: Recognition results during various levels of ambient illumination

5. CONSLUSION & FUTURE WORK

In this paper we have presented a novel approach towards road sign detection and recognition. The system utilizes a robust method of *Colour Segmentation* by employing the HSV colour space and using empirically determined threshold values suitable for various illumination conditions. A novel *Shape Classification* methodology is proposed in which all possible road sign shapes are classified by introducing the use of Contourlet Transform with SVM classifier. The *Recognition* stage introduces the SVM classifier with the Local Energy based Shape Histogram (LESH) features. We have provided experimental results to prove the effectiveness of this approach under varying levels of illumination and environmental conditions. Overall accuracy figures of 96-98% have been reported. We are currently working on real time application of the algorithm within an in-car navigation system.

6. REFERENCES

1. S. M. Bascón, S. L. Arroyo, P. G. Jiménez, H. G. Moreno and F. L. Ferreras. *"Road-Sign Detection and Recognition Based on Support Vector Machines".* IEEE Transactions on Intelligent Transportation Systems, 8(2):264-278, 2007

2. U. Zakir, A. N. J. Leonce and E. A. Edirisinghe. *"Road sign segmentation based on colour spaces: A Comparative Study".*In Proceedings of the 11th Iasted International Conference on Computer Graphics and Imgaing, Innsbruck, Austria, 2010

3. Y.Y. Nguwi and S.Y. Cho. *"Two-tier Self-Organizing Visual Model for Road Sign Recognition".* International Joint Conference on Neural Networks, IJCNN, 2008

4. Y.Y. Nguwi and A.Z. Kouzani. *"Detection and classification of road signs in natural environments".* Neural Computing & Applications, 265–289, 2008

5. V.Andrey and K.H.Jo. *"Automatic Detection and Recognition of Traffic Signs using Geometric Structure Analysis".* SICE-ICASE International Joint Conference, Korea, 2006

6. J. Wu and Y. Tsai. *"Enhanced Roadway Inventory Using a 2-D Sign Video Image Recognition Algorithm".* Computer-Aided Civil and Infrastructure Engineering, 369–382, 2006

7. M. Benallal and J. Meunier. *"Real-time colour segmentation of road signs".* Canadian Conference on Electrical and Computer Engineering, 2003.

8. P. Medici, C. Caraffi, E. Cardarelli and P.P. Porta. *"Real Time Road Signs Classification".* Proceedings of the IEEE International Conference on Vehicular Electronics and Safety Columbus, USA., 2008

9. W.Y. Wu, T.C. Hsieh and C.S. Lai. *"Extracting Road Signs using the Color Information",* World Academy of Science, Engineering and Technology, 32, 2007.

10. C.F. Paulo and P.L. Correia. *"Automatic Detection and Classification of Traffic Signs".* Eighth International Workshop on Image Analysis for Multimedia Int, Services, Greece, 2007

11. H. Fleyeh. *"Traffic Sign Recognition by Fuzzy Sets".* IEEE Intelligent Vehicles Symposium Eindhoven University of Technology Eindhoven, Netherlands, 2008

12. L. D. Lopez and O. Fuentes. *"Colour-Based Road Sign Detection and Tracking".* Springer, Lecture Notes in Computer Science, 2007

13. G.K. Siogkas and E.S. Dermatas, Detection. *"Tracking and Classification of Road Signs in Adverse Conditions".* IEEE Mediterranean Electrotechnical Conference, MELECON, 2006

14. M.L. Eichner and T.P. Breckon. *"Integrated Speed Limit Detection and Recognition from Real-Time Video".* In Proceedings of IEEE Intelligent Vehicles Symposium ,2008

15. X. Gao, K. Hong, P. Passmore, L. Podladchikova and D.Shaposhnikov. *"Colour Vision Model-Based Approach for Segmentation of Traffic Signs".* EURASIP Journal on Image and Video Processing, 2008.

16. W. G. Shadeed, D. I. Abu-Al-Nadi, and M. J. Mismar. *"Road traffic sign detection in color images".* In Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems, 2003

17. D. Matsuura, H. Yamauchi and H. Takahashi. *"Extracting Circular Road Signs Using Specific Color Distinction and Region Limitation".* Systems and Computers in Japan, 2007

18. H. Huang, C. Chen, Y. Jia and S. Tang. *"Automatic Detection and Recognition of Circular Road Sign".* IEEE/ASME International Conference on Mechtronic and Embedded Systems and Applications, MESA, 2008

19. N. Bose, M. Shirvaikar and R. Pieper. *"A Real Time Automatic Sign Interpretation System for Operator Assistance".* Proceedings of the 38th Southeastern Symp on System Theory Tennessee Technological Univ. Cookeville, TN, USA, 2006

20. A. V. Reina, R. Sastre, S. Arroyo and P. Jimenez. *"Adaptive traffic road sign panels text extraction".* Proceedings of the 5th WSEAS International conference on Signal Processing, Robotics and Automation, Madrid, Spain, 2006

21. U. Zakir, I. Zafar, E. A. Edirisinghe. *"A Novel Approach to In-Car Road Sign Recognition".* The 14th International Conference on Machine Design and Production, Cyprus 2010

22. R. Malik,J. Khurshid and S.N. Ahmad. *"Road Sign Detection and Recognition using colour segmentation, Shape analysis and Template Matching".* Proceedings of the Sixth International Conference on Machine Learning and Cybernetics, Hong Kong, 2007

23. W.J. Kuo and C.C. Lin. *"Two- Stage Road Sign Detection and Recognition".* IEEE International Conference on Multimedia and Expo, 2007

24. V.Andrey and K.H.Jo. *"Automatic Detection and Recognition of Traffic Signs using Geometric Structure Analysis".* SICE-ICASE International Joint Conference, Korea, 2006

25. A. V. Reina, R. J. L. Sastre, S. L. Arroyo and P.G. Jimenez. *"Adaptive traffic road sign panels text extraction".* Proceedings of the 5th WSEAS International Conference on Signal Processing, Robotics and Automation, Madrid, Spain, 2006

26. P. Gil-Jim´enez, S. Lafuente-Arroyo, H. G´omez-Moreno, F. L´opez-Ferreras and S. Maldonado-Basc´on. *"Traffic sign shape classification evaluation II: FFT applied to the signature of Blobs".* 607- 612 ,2005

27. J. Wu and Y. Tsai. *"Enhanced Roadway Inventory Using a 2-D Sign Video Image Recognition Algorithm".* Computer-Aided Civil and Infrastructure Engineering, 21: 2006, 369–382

28. L.W. Tsai, J.W. Hsieh, C.H. Chuang, Y.J. Tseng, K.C. Fan, C.C. Lee. *"Road sign detection using eigen colour".* IET Computer Vision 2(3): 164–177, 2008

29. G.H. Kim, H.G. Sohn and Y.S. Song. *"Road Infrastructure Data Acquisition Using a Vehicle-Based Mobile Mapping System".* Computer-Aided Civil and Infrastructure Engineering, 2(1): 346–356, 2006

30. M.L. Eichner and T.P. Breckon. *"Integrated Speed Limit Detection and Recognition from Real-Time Video".* In Proceedings of IEEE Intelligent Vehicles Symposium, 2008

31. P. Silapachote and A. H. R. Weiss. *"A Hierarchical Approach to Sign Recognition".* Proceedings of the Seventh IEEE Workshop on Applications of Computer Vision,2005

32. Y. Y. Nguwi and A. Z. Kouzani. *"Detection and classification of road signs in natural environments".* Neural Computing & Applications, 17: 265–289, 2008

33. P. Medici, C. Caraffi., E. Cardarelli and P. P. Porta. *"Real Time Road Signs Classification".* Proceedings of the IEEE International Conference on Vehicular Electronics and Safety Columbus, USA, 2008

34. A. Broggi, P. Cerri, P. Medici, P. P. Porta. *"Real Time Road Signs Recognition".* Proceedings of the IEEE Intelligent Vehicles Symposium Istanbul, Turkey, 2007

35. H. Ohara, I. Nishikawa, S. Miki, N. Yabuki. *"Detection and Recognition of Road Signs using simple layered Neural Networks".* Proceedings of the 9th International Conference on Neural Information Processing, 2002

36. G. H. Kim, H. G. Sohn and Y. S. Song. *"Road Infrastructure Data Acquisition Using a Vehicle-Based Mobile Mapping System".* Computer-Aided Civil and Infrastructure Engineering, 21: 346–356, 2006

37. N. Kehtarnavaz, A. Ahmad. *"Traffic Sign Recognition in Noisy Outdoor Scenes".* IEEE, Intelligent Vehicles. 1995

38. A. D. L. Escalera., J. M. Armingol and M. Mata. *"Traffic sign recognition and analysis for intelligent vehicles".* Image and Vision Computing, 21: 247–258, 2003

39. A. A. Farag and A. E. Abdel-Hakim. *"Detection, Categorization and Recognition of Road Signs for Autonomous Navigation".* Proceedings of Acivs (Advanced Concepts for Intelligent Vision Systems), Brussels, Belgium, 2004

40. W. G. Shadeed, D. I. Abu-Al-Nadi, and M. J. Mismar. *"Road traffic sign detection in color images".* In Proceedings of the 10th IEEE International Conference on Electronics, Circuits and Systems, 2: 890–893,2003

41. S. Lafuente-Arroyo, S. Maldonado-Bascon, P. Gil-Jimenez, J. Acevedo- Rodriguez, and R. Lopez-Sastre. *"A tracking system for automated inventory of road signs".* IEEE Intelligent Vehicles Symposium,166-171, 2007

Usman Zakir, Iffat Zafar & Eran A. Edirisinghe

42. L. Simon, P. J. Tarel and R. Bremond. *"Alerting the Drivers about Road Signs with Poor Visual Saliency".* Proceedings of the IEEE Intelligent Vehicle Symposium, China, 2009

43. C. Kiran, L. Prabhu, R. Abdu, and K. Rajeev. *"Traffic sign detection and pattern recognition using support vector machine".* Seventh International Conference on Advances in Pattern Recognition,2009

44. A. Z. Kouzani. *"Road-Sign Identification Using Ensemble Learning".* Proceedings of the IEEE Intelligent Vehicles Symposium Istanbul, Turkey,2007

45. A. D. L. Escalera, J. M. Armingol, J. M. Pastor and F. J. Rodríguez. *"Visual sign information extraction and identification by deformable models for intelligent vehicles".* IEEE Transactions on Intelligent Transportation Systems, 5(2):,2004

46. G. K. Siogkas and E. S. Dermatas. *"Detection, Tracking and Classification of Road Signs in Adverse Conditions".* IEEE Mediterranean Electrotechnical Conference, MELECON, 2006

47. Y. Y. Nguwi and S. Y. Cho. *"Two-tier Self-Organizing Visual Model for Road Sign Recognition".* International Joint Conference on Neural Networks, IJCNN, 2008

48. M. Kalčič and J.F.Tasič. *"Colour Spaces – perceptual, historical and applicational background".* In Eurocon Computer as a Tool Conference Proceedings, 2003

49. J. van de Weijer, T. Gevers, A. Gijsenij. *"Edge-Based Color Constancy".* IEEE Transactions on Image Processing, 16(9):2207 – 2214, 2007

50. M. N. Do and M. Vetterli. *"The Contourlet Transform: An efficient directional multiresolution image representation".* 14(12): 2091 – 2106, 2005

51. C. Cortes and V. Vapnik. *"Support-Vector Networks".* Machine Learning, 20(3):273-297,1995

52. M. C. Morrone and R. A. Owens. *"Feature Detection from Local Energy".* PR Letters 6, 303–313, 1987

53. P. D. Kovesi. *"Phase Congruency: A Low-Level Image Invariant".* Psychological Research 64, 136–148, 2000

54. M. S. Sarfraz and O. Hellwich.*" An Efficient Front-end Facial Pose Estimation System for Face Recognition".* In International Journal of Pattern Recognition and Image Analysis, distributed by Springer, 18(3):434–441,2008

55. S. Shiguang, C. Xilin C, and G. Wen. *"Histogram of Gabor Phase Patterns (HGPP): A Novel Object Representation Approach for Face Recognition".* IEEE Transactions on Image Processing 16 (1): 57–68, 2007

Jack Collins, Christopher Kurcz, Curtis Lisle, Yanling Liu & Enrique Zudaire

Assessment of Vascular Network Segmentation

Jack Collins   collinja@mail.nih.gov
ABCC-ISPSAIC-Frederick, Inc., NCI-Frederick
Frederick, 21702, USA

Christopher Kurcz   kurczce@mail.nih.gov
ABCC-ISPSAIC-Frederick, Inc., NCI-Frederick
Frederick, 21702, USA

Curtis Lisle   curtislisle@knowledgevis.com
KnowledgeVis LLC
Maitland, 32751, USA

Yanling Liu   liuy5@mail.nih.gov
ABCC-ISPSAIC-Frederick, Inc., NCI-Frederick
Frederick, 21702, USA

Enrique Zudaire   zudairee@mail.nih.gov
Angiogenesis Core Facility, NCI
NIHGaithersburg, 20877, USA

Abstract

We present an analysis framework to assess the quality and accuracy of vessel segmentation algorithms for three dimensional images. We generate synthetic (in silico) vessel models which act as ground truth and are constructed to embody varying morphological features. These models are transformed into images constructed under different levels of contrast, noise, and intensity. To demonstrate the use of our framework, we implemented two segmentation algorithms and compare the results to the ground truth model using several measures to quantify the accuracy and quality of segmentation. Furthermore, we collect metrics which describe the characteristics of the vessels it fails to segment. Our approach is illustrated with several examples. Funded by NCI Contract No.HHSN261200800001E.

Keywords: Vessel Segmentation, Network Comparison, Quantitative Analysis, Segmentation Quality, Segmentation Accuracy.

1. INTRODUCTION

Extracting vascular networks from three dimensional images remains a difficult and exciting challenge. Many segmentation algorithms have been designed to target different types of vasculature imaged using various modalities and are applied to a wide range of problems in disease diagnosis, flow dynamics, surgical planning, etc... (see e.g. [2][1][3]). Despite the sophistication of these methods, it is often difficult to choose a particular segmentation algorithm given a collection of similar images since a given method may make implicit assumptions about image or vessel characteristics. Furthermore, it is not clear how to predict or quantify the quality or accuracy of a given method onthose images due to a lack of ground truth data.

Jack Collins, Christopher Kurcz, Curtis Lisle, Yanling Liu & Enrique Zudaire

In this paper, we provide a methodology to quantify the accuracy and effectiveness of a vessel segmentation algorithm. We provide tools for constructing a model vessel tree which defines the vessel centerline locations and vessel radius and serves as ground truth. This model is transformed into an image with specified contrast, noise and intensity profile and serves as input to the segmentation algorithm under scrutiny. At a minimum we require the segmentation algorithm to produce a binary image depicting vessel location. Additional data including vessel radius, curvature, and local coordinate system may also be specified. This data is then transformed into a graph (network) representation and aligned with the model tree to determine the quality and accuracy of segmentation. We illustrate these operations in Figure 1.

We emphasize that our approach quantifies segmentation quality and accuracy by comparing attribute graphs and not images. By using the model vessel tree as a starting point and simulating the imaging process we can assess the performance of a segmentation algorithm under a variety of image characteristics. Furthermore, we do not require a trained observer to extract the centerlines from a gold standard data set since the model vessel tree contains centerline location as an attribute. Morphological features (e.g. vessel radius, curvature, tortuosity, etc...) are maintained as attributes of the model vessel tree and we can assess the segmentation performance as a function of these features.

We provide metrics describing the correctly segmented vessels, missed vessels, false positives and changes in network topology to assess segmentation quality. Beyond the quality assessment, we determine the accuracy of segmentation by computing the deviation of the centerline and radius between the ground truth model and the segmentation estimation. Additional assessment includes collecting features where that algorithm fails (e.g. did the algorithm fail to detect thin or highly curvedvessels?). Thus, we provide a framework to determine how well a particular segmentation algorithm performs as a function of the image quality defined by contrast, noise, and intensity profile. This paper is organized as follows. In Section 2, we describe the model tree creation, transformation of the model tree to an image, segmentation of the image, and comparison of the segmented result to the model. Next, in Section 3, we perform a quality and accuracy assessment for several models. Finally, in Section 4, we summarize the results for the different models and describe possible extensions and improvements to our approach.
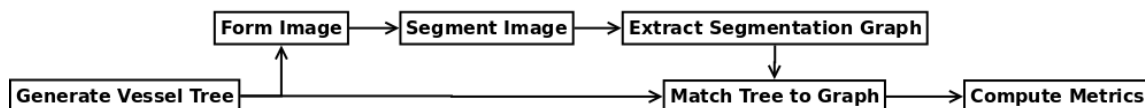
FIGURE 1: Flow Chart of Major Operations in Vessel Segmentation Assessment Framework

2. MATERIALS AND METHODS

In this section we describe the process of model vessel tree generation, model tree to image formation, segmentation of vessels from a noise corrupted image, and a method to compare the model to the segmentation result.

2.1 Model Tree Generation

We construct model vessel trees which serve as ground truth for subsequent analysis. Lindenmayer systems [17][18] (*L*-systems) provide a natural framework for defining such objects. This formalism grows a tree from a root location by following a set of rules which specify the direction of growth, branch radius, branch length, angle between branches and number of branches. In Figure 2 we illustrate this procedure with a simple example. We note that *L*-systems have been used in a variety of applications in computational botany, computer graphics, and medicine (see e.g. [10][24][20][21]). Although our model is formally a tree, there is no fundamental limitation for its extension to a more complicated topology.

A model tree contains a collection of nodes (vertices), connections between nodes (edges), and node attributes. We characterize these trees by node location (x,y,z coordinates), radius, node spacing and node type i.e. end, middle, or junction. Furthermore, we generate a local coordinate system at each node given by the tangent vector ($T(s)$), normal vector ($N(s)$), and binormal vector ($B(s)$) where s is arc length. These vectors are used to compute the curvature($\langle dT/ds, N \rangle$)andtorsion ($\langle dB/ds, N \rangle$) at each node location.
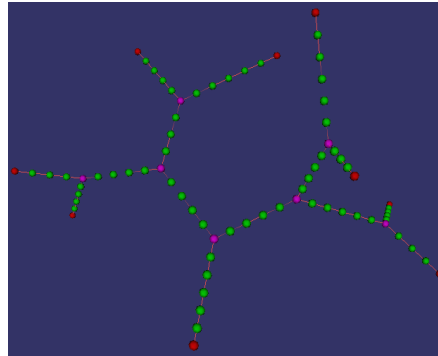
FIGURE 2: Simple Example of Model Vessel Tree. Nodes in Red, Green and Magenta Correspond to End, Middle and Junctions (Respectively).

2.2 Image Generation From Model Tree

We transform a model vessel tree to a volumetric image by simulating the image formation process. This mapping requires specification of spatial resolution, intensity range of the vessel, background intensity, noise level, and intensity profile of the vessel cross-section.

To generate the image, we first convert the vessel model to a binary image by rounding each node location in the model to the nearest voxel index. Then to create vessel thickness, we label all voxels inside the sphere given by each nodes radius. Next, this label map is convolved with a Gaussian (characterized by its standard deviation η) which blurs the boundary of the vessel producing a smooth cross-section intensity profile. At this point we have an image of real values in $[0,1]$ where the values close to one correspond to the interior of large vessels and values close to zero correspond to either the exterior of vessels or the centerline of thin vessels. We use a linear mapping to a specified vessel intensity maximum and minimum ($[0,1] \rightarrow [I_{min}, I_{max}]$) and add a background intensity (I_{back}) to the image. Finally, we complete the process by corrupting the image with additive Gaussian noise (characterized by its standard deviation σ). The image formation steps are illustrated in Figure 3.

This image formation process characterizes images by four simple qualities (i) the relationship between voxel intensities and vessel radius, (ii) contrast between vessel centerline and background, (iii) amount of noise relative to contrast and (iv) the ratio of node spacing to image resolution. Thus, a wide range of images can be created simulating different image acquisition environments.

Jack Collins, Christopher Kurcz, Curtis Lisle, Yanling Liu & Enrique Zudaire

(a)

(b)   (c)   (d)
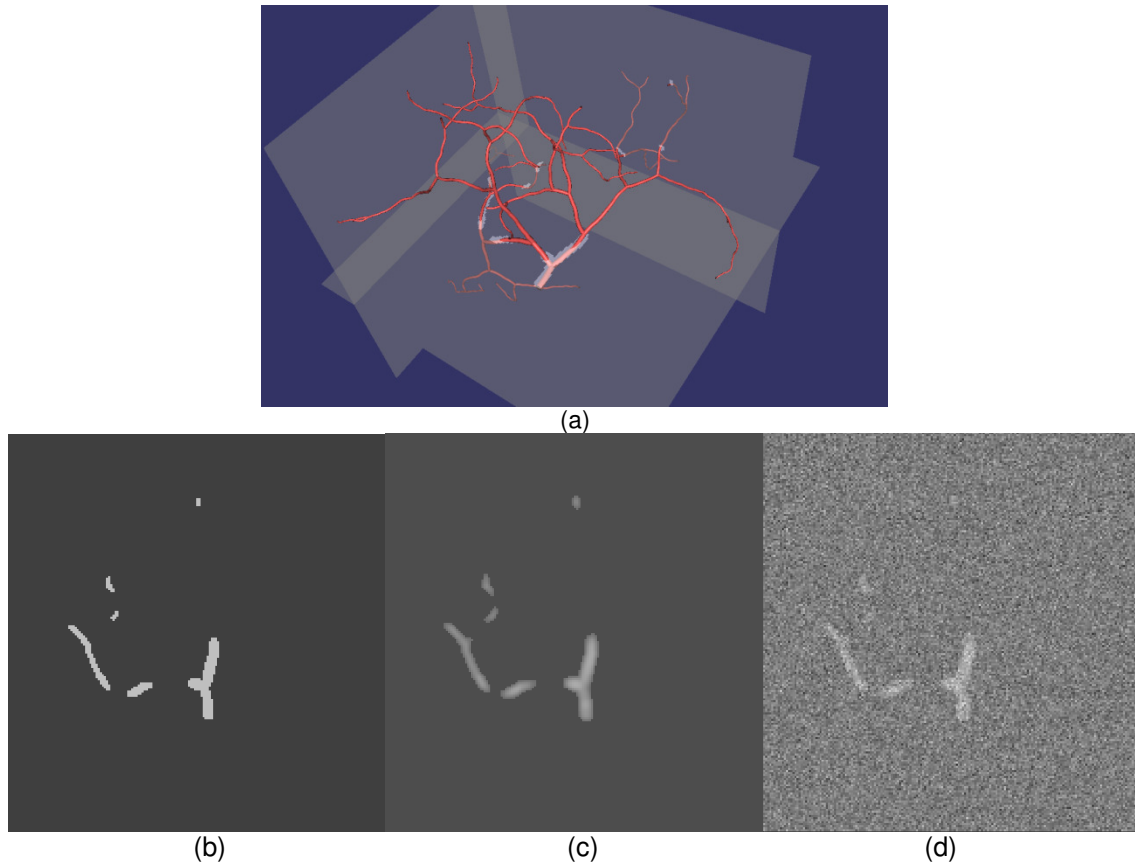
FIGURE 3: Image Formation Process. (a) Generated x, y, z Slices of Label Map Overlayed on Model Tree. For Visualization Purposes The Model Has Been Made Artificially Thinner. (b) A Slice From The Label Map. (c) Slice From (b) After Gaussian Smoothing $(\eta = 2)$, Intensity Rescaling $(I_{min} = 200, I_{max} = 400)$ and Background Added $(I_{back} = 100)$. (d) Slice From (c) After Addition of Gaussian Noise $(\sigma = 100)$.

2.3   Vessel Segmentation

There is a vast amount of literature and wide range of methods for vessel segmentation (see e.g.[13][19] for a review). We emphasize that the main point of this paper is not the development of vessel segmentation algorithms but the assessment of their performance. Toward this end, we implemented two vessel segmentation algorithms for demonstration purposes. In what follows, we denote $I(x)$ the image intensity at position $x \in \mathbb{R}^3$.

2.3.1   Vesselness

The first segmentation method assumes a cylindrical model for vessels and uses the eigenvalues of the Hessian matrix to enhance vessel voxels (see e.g. [7][25]). We begin by considering the voxels above a specified threshold $(\{x | I(x) > \epsilon_1\})$. For each specified scale, τ, we form the Hessian matrix

$$H_\tau(x) = \tau^2 \begin{pmatrix} \partial_{xx} G_\tau * I & \partial_{xy} G_\tau * I & \partial_{xz} G_\tau * I \\ \partial_{yx} G_\tau * I & \partial_{yy} G_\tau * I & \partial_{yz} G_\tau * I \\ \partial_{zx} G_\tau * I & \partial_{zy} G_\tau * I & \partial_{zz} G_\tau * I \end{pmatrix} \tag{2.1}$$

where $G_\tau(x) = 1/(2\pi\tau^2)^{3/2} e^{-\|x\|^2/2\tau^2}$, * denotes the convolution operator, and ∂_x denotes the partial derivative with respect to x. The eigenvalues of (2.1) are computed and ordered so that $|\lambda_{1,\tau}| \leq |\lambda_{2,\tau}| \leq |\lambda_{3,\tau}|$. For each scale, we compute

$$V_\tau(x) = \begin{cases} 0, & \lambda_{2,\tau} > 0 \text{ or } \lambda_{3,\tau} > 0 \\ \left(1 - e^{-\frac{a_\tau^2}{2\alpha^2}}\right) e^{\frac{b_\tau}{2\beta^2}} \left(1 - e^{-\frac{c_\tau^2}{2\gamma^2}}\right), & otherwise \end{cases} \tag{2.2}$$

where $c_\tau = \sqrt{\lambda_{1,\tau}^2 + \lambda_{2,\tau}^2 + \lambda_{3,\tau}^2}$, $b_\tau = |\lambda_{1,\tau}|/\sqrt{\lambda_{2,\tau}\lambda_{3,\tau}}$ and $a_\tau = |\lambda_{2,\tau}|/|\lambda_{3,\tau}|$. Finally, we determine the vesselness, $V(x) = max_\tau V_\tau(x)$, and construct a label map, $L_V = \{x \,|\, V(x) \geq \epsilon_V\}$, by considering the voxels above a specified vesselness threshold, ϵ_V.

2.3.2 Fast Marching
The second segmentation method uses the Fast Marching method (see e.g. [27][12, §9.3.1]). This method propagates an outward moving front satisfying

$$|\nabla T| F(x) = 1 \tag{2.3}$$

from an initial surface S given a speed $F(x)$. The solution, T, to (2.3) represents the arrival time at location x. We use a sigmoidal speed function,

$$F(x) = \frac{1}{1 + e^{-\frac{A(x) - \kappa}{\beta}}}$$

where $A(x)$ is the average intensity over a $3 \times 3 \times 3$ neighborhood. To construct the label map, $L_F = \{x \,|\, T(x) \leq \epsilon_F\}$, we consider all voxels whose arrival time is less than a specified threshold, ϵ_F.

2.3.3 Post Processing
To aid subsequent analysis we apply the hole filling algorithm [12, §6.6.4] to the label map constructed from either segmentation algorithm (L_V or L_F). We determine the vessel centerline using the skeletonization algorithm [9] and the vessel radius at each centerline location using the distance transform [31].

From the labeled centerline image we generate a data structure which contains the locations of the vessel centerline at each point and refer to this structure as the *segmentation graph* throughout the rest of this paper. We note that its topology may deviate from a tree due to segmentation errors. The segmentation graph is constructed by converting each centerline voxel to a node in the segmentation graph where the node location is given by the voxels physical coordinates. The remaining attributes of the graph are then populated. Since the model has been discretized during the image formation process, the node locations in the segmentation graph may not align with those from the model. In Figure 4 we provide an example, where the spacing between points in the model is 1 but, the spacing between points in the segmentation graph is either 1, $\sqrt{2}$ or $\sqrt{3}$.
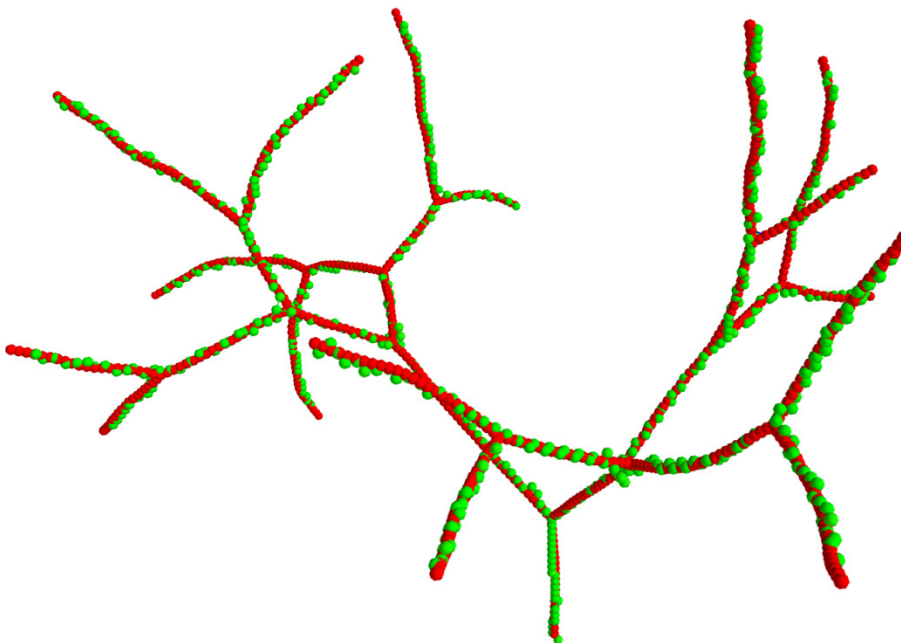
FIGURE4: Model Tree in Red and Segmentation Graph in Green

2.4   Graph Matching

Given a model tree and a segmentation graph we determine the nodes in the segmentation graph that correspond to these in the model tree. This mapping allows us to compare results of segmentation to ground truth. Since the spacing between nodes in the model tree and vessel graph may differ, we use a many-to-one mapping between nodes in the segmentation graph to nodes in the model tree.

This association problem is often referred to as inexact graph matching (see e.g. [4][6][30][32]) which is often posed as a global alignment problem or largest subgraph isomorphism problem. However, the problem presented here is slightly different because we have node attribute information which can be used to assist in the association problem. Thus, we seek a method which optimizes the global alignment while minimizing the pairwise distance between nodes in the model tree and segmentation graph.

For this purpose, we use the IsoRank algorithm [28] which constructs a set of constraints for all pairs of nodes in the model tree and segmentation graph based on neighborhood information. Using these constraints and minimizing the pairwise distances, the IsoRank algorithm establishes a ranking between pairs of nodes in the model tree and segmentation graph. We note that the rankings are computed by solving for the largest eigenvalue of a sparse matrix and the entries of the eigenvector correspond to the pairwise ranks (as in the PageRank algorithm see e.g. [16]). The ranking matrix in combination with a maximum allowable deviation between node positions and maximum number of multiple matches is uses to determine the many-to-one mapping between nodes in the segmentation graph to nodes in the model tree. An example matching is illustrated in Figure 5.

Given the association between nodes in the model tree and segmentation graph we classify nodes in the segmentation graph as either matched, missed or false positives. A matched node in the segmentation graph is associated with one node in the model tree, a missed node in the model tree has no association with a node in the segmentation graph and a false positive is a node in the segmentation graph with no association to a node in the model tree. Notice in Figure 5 thatin a limited number of cases the algorithm can fail to match nodes that appear to correspond to a node in the model tree. We note that this occurs due to our enforcement of a many-to-one mapping. Adopting a many-to-many matching policy would reduce the number of false positives at the cost of matching nodes that may not correspond. Finally, a one-to-one policy would inflate the number of missed nodes since the spacing between nodes in the model tree is less than or equal to the spacing in the segmentation graph.

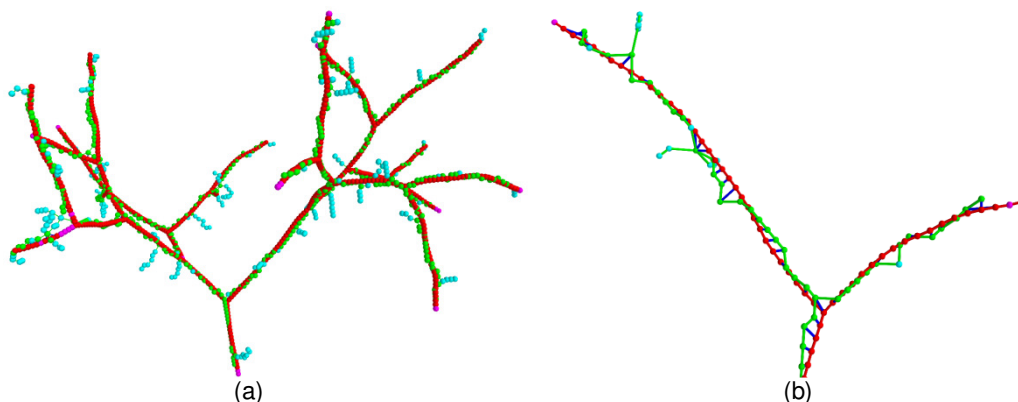**FIGURE5:** Model Tree and Segmentation Graph Matching. (a) Model Tree Nodes in Red, Segmentation Graph Nodes in Green, Association Between Matched Nodes as Dark Blue Cylinders, Missed Nodes in Magenta, False Positives in Light Blue. (b) Close up of (a).

3. RESULTS

In this section we construct two distinct model trees, generate images under a variety of conditions, and compare the segmentation graph to the model tree for the purpose of quantifying segmentation quality and accuracy for the vesselness and fast marching algorithms. The segmentation quality metrics include the (i) the percentage of correctly matched nodes, (ii) identification of the missed node type (i.e. junction, middle or end), (iii) percentage of false positives and (iv) the amount of topological change between model tree and segmentation graph. For the matched nodes, we compute the position and radius accuracy and, furthermore, record the curvature and radius of missed nodes in the segmentation graph. Therefore, this analysis measures how well each segmentation algorithm performs and quantifies the characteristics of vessels each algorithm fails to segment.

We present the quality measures as bar charts displaying the percentage of matched and missed model nodes (see Figures 7(a,d), 11(a,d)). In a similar fashion, we display the percentage of false positives i.e. the number of unmatched segmentation nodes over the total number of model nodes (see Figures 7(b,e), 11(b,e)). We note that the matching algorithm systematically over estimates the number of false positives due to the many-to-one constraint described in Section 2.4. Unfortunately, we do not have a rigorous bound for the error induced but can get a rough estimate by examining the number of false positives under no noise. To measure the quality with respect to changes in the topology we use a bar chart to display the percentage of nodes in each connected component of the segmentation graph (see Figures 7(c,f), 11(c,f)). Recall that a connected component of a graph is a subgraph which contains a path between every pair of nodes. For example, in Figure 7(c) with $\sigma = 0$ there are 4 connected components with the largest component containing approximately 95 percent of the segmentation graphs nodes.

The position accuracy of matched nodes is computed using the max-norm $(\|x\|_\infty = max_i |x_i|)$. Since a vector with unit max-norm is bounded by a box of side length one, it facilitates a straightforward comparison to voxel coordinates. The position accuracy includes error from the quantization of the nodes position during the image formation process and error due to segmentation inaccuracy. To compute the position accuracy, we calculate the distance between every pair of matched nodes and note that due to the many-to-one mapping of segmentation nodes to model nodes we systematically over estimate the position error. In practice, however, we expect the number of many-to-one mappings to be small (compared to the total number of nodes in the model) and thus the effect is minimal. We display the position accuracy as a histogram where the x-axis denotes the error between model and segmentation node position and on the y-axis we display the number of nodes for each bin. We overlay the results from several levels of noise using different colors (see Figures 8(a,c), 12(a,c)). A similar plot is provided for the relative radial error (see Figures 8(b,d), 12(b,d)).

We also characterize the missed nodes by radius and curvature with a histogram plot. In this plotthe x-axis corresponds to the value of radius (curvature) and y-axis corresponds to the percentageof nodes within a radius (curvature) bin that were missed by the segmentation algorithm. We overlaythe results from several levels of noise using different colors (see Figures 9(a,c)(b,d), 13(a,c)(b,d)).

The model trees shown in Figures 6 and 10 share similar topological features but have differentmorphologies. In particular, the models in Figure 6 and Figure 10 differ significantly only inthe curvature of the branches. In fact, the image generation process and segmentation algorithmused identical parameters.

In what follows, we generate images of the model trees (see Section 2.2) with vessel intensity range of $[I_{min} = 200, I_{max} = 400]$, background intensity of $I_{back} = 100$, smoothing factor $\eta = 2$, noise levels $\sigma = 0, 20, 40, 60, 80, 100, 120$, and spacing between nodes and image resolution of 1. We use the same segmentation parameters for each model. For the vesselness algorithm (seeSection 2.3.1), we use an intensity threshold $\epsilon_I = 150$, six logarithmically spaced

scales (τ) from 1 to5, vesselness parameters $\alpha = 0.25, \beta = 0.25$ and $\gamma = 30$, and a vesselness threshold $\epsilon_V = 0.75$.As a post processing step we removed all components with five nodes or less that contain no junctionpoints. For the fast marching algorithm (see Section 2.3.2) we use speed function parameters $\mu = 1$and $\kappa = 180$, a time threshold $\epsilon_F = 500$, and one seed point at the root location of the tree as theinitial contour. To generate the hole filled label map (see Section 2.3.3) we use a majority threshold0 and a maximum of 20 iterations. For the IsoRank algorithm we set $\alpha = 0.5$ (see [28, equation 5]),a maximum of 3 matches of a segmentation node to a model node and a maximum distance of 3 fora match.

3.1 Model Tree 1

This model has 726 nodes with radii between 0.6 and 4. Both algorithms correctly segment an overwhelming majority of nodes with the fast marching algorithm having a slightly greater percentage of matched nodes. The vesselness algorithm induces a larger percentage of false positives and fragments the model tree into several components with the number of components increasing as the noise level increases (see Figure 7). The number of matched nodes generated by the vesselness algorithm tends to decrease as the noise level increases whereas the number of matched nodes generated by the fast marching algorithm is nearly invariant to noise level.The position and radius accuracy are not significantly affected by the amount of noise and the algorithmsperform equally well with regard to the position accuracy. The vesselness algorithm performsslightly better with regards to radius accuracy (see Figure 8). Almost all matched nodes are accurateto within one voxel and approximately 50 percent are inside a box of side length $1/2$ (seeFigure 8(a,c)). The fast marching algorithm generally over estimates the vessel radius since weuse the average intensity over a $3 \times 3 \times 3$ window when evaluating the speed function, and, thusartificially extend the boundary of the vessel.
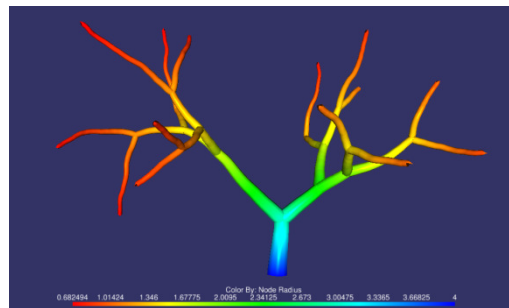
FIGURE 6: Model Tree 1 With 726 Nodes And Radii From 0.6 (red) to 4 (blue). Image of Model Tree Has A Maximum Intensity of 400, A Minimum Intensity of 200 And A Background Intensity of 100.

The fast marching algorithm generates fewer false positives then the vesselness algorithm. Furthermore,the vesselness algorithm tends to produce several components while the fast marchingalgorithm produces only one. The behavior is expected as the fast marching algorithm cannot splitthe contour into multiple closed contours and the vesselness algorithm does not enforce connectivityconstraints. Thus, false positives generated by the fast marching algorithm correspond to additionalshort branches and for the vesselness algorithm correspond to additional components.

The majority of missed nodes have small radii and their percentage increases with the level of noise(see Figure 9(a,c)). In some conditions, we observe that the missed nodes correspond to thosenear the root (see Figure 9(c)). An expected result is that nodes with larger curvature are morelikely to be missed and the percentage increases with the level of noise (see Figure 9(b,d)). Bothalgorithms tend to miss nodes with similar properties and in similar percentages.

Jack Collins, Christopher Kurcz, Curtis Lisle, Yanling Liu & Enrique Zudaire

FIGURE 7: Quality Results Associated With The Model in Figure 6. Top Row Corresponds to Results From The Vesselness Algorithm And The Bottom Row From Fast Marching Algorithm. (a,d) Percentage of Correctly Segmented Nodes (red), Missed Junction Nodes (cyan), Missed End Nodes (green), And Missed Middle Nodes (blue) As A Function of Noise Level ($\sigma = 0, 20, 40, 60, 80, 100, 120$). (b,e) Percentage of False Positives As A Function of Noise. (c,f) Percentage of Nodes Within Each Component of The Segmentation Graph. Each Color Corresponds to A Component.
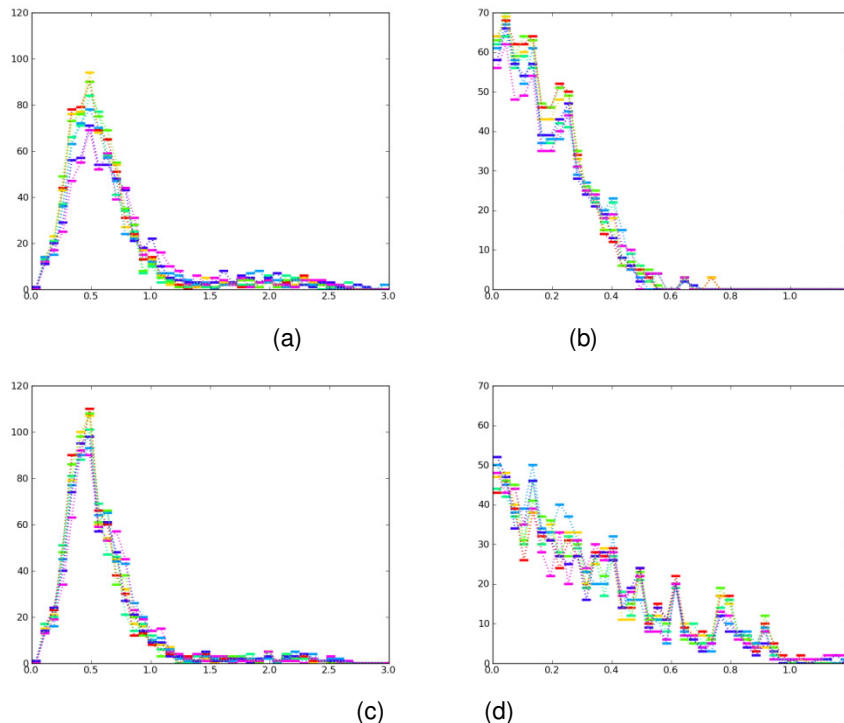
FIGURE 8: Accuracy Results Associated With The Model in Figure 6. Top Row Corresponds to Results From The Vesselness Algorithm And The Bottom Row From Fast Marching Algorithm. Each Histogram Corresponds to The Noise Level 0-red, 20-yellow, 40-green, 80-light blue, 100-dark blue, 120-magenta. (a,c) Absolute Position Error (Using max-norm) Between Every Pair of Matched Nodes. (b,d) Relative Radius Error Between Every Pair of Matched Nodes.

3.2 Model Tree 2

This model has 769 nodes with radii between 0.8 and 3 and is similar to themodel described in Section 3.1 except the vessels are more curved. We compare the quality andaccuracy results to those in Section 3.1 for the purpose of assessing the sensitive of each vesselsegmentation algorithm to curvature.

(a)   (b)

(c)   (d)
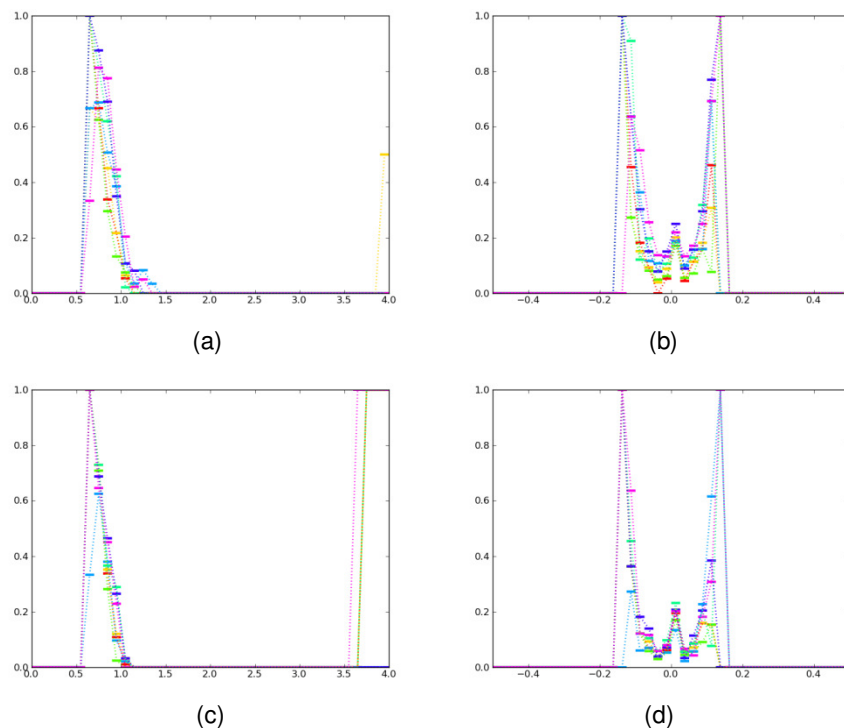
FIGURE 9: Properties of Missed Nodes Associated With The Model in Figure 6. Top Row Corresponds to Results From The Vesselness Algorithm And The Bottom Row From Fast Marching Algorithm. Each Histogram Corresponds to The Noise Level 0-red, 20-yellow, 40-green, 80-light blue, 100-dark blue, 120-magenta. (a,c) Percentage of Missed Nodes Within Radius Interval. (b,d) Percentage of Missed Nodes Within Curvature Interval.

Each algorithms number of matched nodes, false positives and components is essentially the sameas in Section 3.1 (see Figure 11(a,b,c)). The vesselness algorithm produces two large componentswhen $\sigma = 60, 100$ (see Figure 11(c)). In this case, these components are created because of high levels of noise located at a middle branch yielding low vesselness responses. Furthermore, since the eigenvalues of (2.1) are sensitive to changes incurvature (see [15]) we expect the segmentation graph to miss more nodes in curved region andthus create more components in the graph.

The position and radius accuracy follow similar trends as those in Section 3.1 (see Figure 12). Highlycurved nodes are missed as the noise level increases (see Figures 13(b,d)), however, it is difficult tomake a direct comparison because a percentage of missed nodes is given (not an absolute number)and range of curvatures in the two models varies significantly.

Jack Collins, Christopher Kurcz, Curtis Lisle, Yanling Liu & Enrique Zudaire
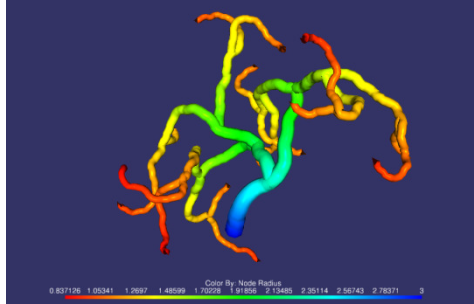
FIGURE 10: Model Tree 2 With 769 Nodes And Radii From 0.8 (red) to 3 (blue). Image of Model Tree Has A Maximum Intensity of 400, A Minimum Intensity of 200 And A Background Intensity of 100.
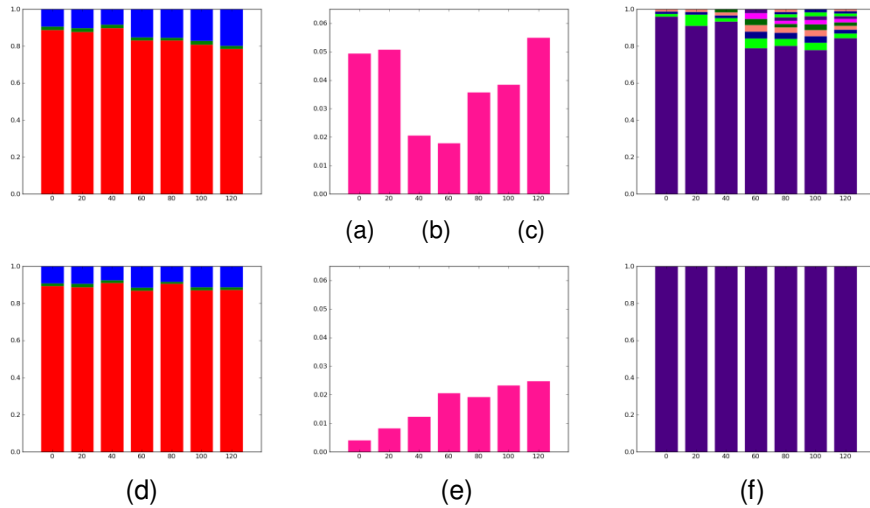

(a)　　(b)　　　(c)

(d)　　　　　　(e)　　　　　　(f)

FIGURE 11: Quality Results Associated With The Model in Figure 10. Top Row Corresponds to Results From The Vesselness Algorithm And The Bottom Row From Fast Marching Algorithm. (a,d) Percentage of Correctly Segmented Nodes (red), Missed Junction Nodes (cyan), Missed End Nodes (green), And Missed Middle Nodes (blue) As A Function of Noise Level ($\sigma = 0,20,40,60,80,100,120$). (b,e) Percentage of False Positives As A Function of Noise. (c,f) Percentage of Nodes Within Each Component of The Segmentation Graph. Each Color Corresponds to A Component.

4. DISCUSSION

We have developed model vessel trees and a quantitative analysis framework to assess the accuracyand quality of vessel segmentation algorithms for three dimensional images. We applied thisframework to two models under varying image qualities using two segmentation algorithms of fundamentallydifferent classes. The results of this assessment can be briefly summarized by (i) bothalgorithms have roughly the same quality with respect to the number of correctly segmented nodes(ii) the vesselness algorithm introduces a significantly larger number of false positives (iii) the vesselnessalgorithm fragments the model tree into several components while the fast marching algorithmretains the tree structure and (iv) the segmentation quality degrades as the vessels become more thin and/or curved. This type of quantitative result is not surprising as the vesselness algorithm implicitlyassumes a cylindrical model and the fast marching algorithm can only move the contour inan outward normal direction and cannot break the closed contour. More importantly, this analysisprovides qualitative information on how and where the algorithm fails.

We would like to point out similar work in [26] where the authors developed a framework for characterizingand comparing coronary artery centerline extraction algorithms. In this work the authorsassembled a publicly available database of Computed Tomography Angiography (CTA) images andderived the ground truth centerline from three trained observers. The images were acquired using astandard protocol and algorithms where ranked based on overlap and accuracy metrics. The majordifference with our approach is that we start with model generation which

defines ground truth ratherthan extracting the vessel tree from the image. This luxury allows us to easily identify the vesselfeatures that the algorithm fails to segment. Furthermore, we do not need to merge the centerlinetraced by each observer and determine the uncertainty this induces.

Generally, we expect this framework to be used as a measure of performance tool for the assessmentof vessel segmentation algorithms under a wide range of scenarios. Depending on the applicationthe vessel segmentation algorithm may apply different principles. For example, techniques to segmentthe coronary artery may significantly vary from algorithms used to extract capillary beds (see e.g. [1][34][23][29]). With the generation of appropriate models, algorithms designed for either casecan be assessed using our framework. Thus, as one designs a segmentation algorithm it is possibleto quickly assess the impact of differing implementations or parameters. Another use of thisframework is to quantify the limits of detectability for a particular algorithm. As shown in Section 3,analysis of missed node distributions versus vessel radius is straightforward and could be extendedto incorporate other image, topological, or morphological properties (e.g. tortuosity, segment length,degree distribution, etc...). This type of analysis is particularly interesting when the imaging deviceis well characterized. In this case, the image generation process can be modified to better modela particular imaging modality. Thus, a candidate segmentation algorithm can be fully characterizedagainst a particular class of data. Although beyond the scope of this paper, we note that our approach may be used for other network structures e.g. neural networks. In this case the metrics of interest may be more closely related to network topology than morphological properties. Finally, given a vessel network, a segmentation algorithm and adesired quality result, one may determine the required image characteristics. This may assist in thedevelopment of imaging protocols with the intention of subsequent quantitative analysis and not justvisual clarity. For example, one may anticipate the need to accurately segment vessels of a certainradius (for a particular problem a biological argument may be made) and may want to ascertain theimage quality required so that a particular segmentation algorithm performs at a given quality level.
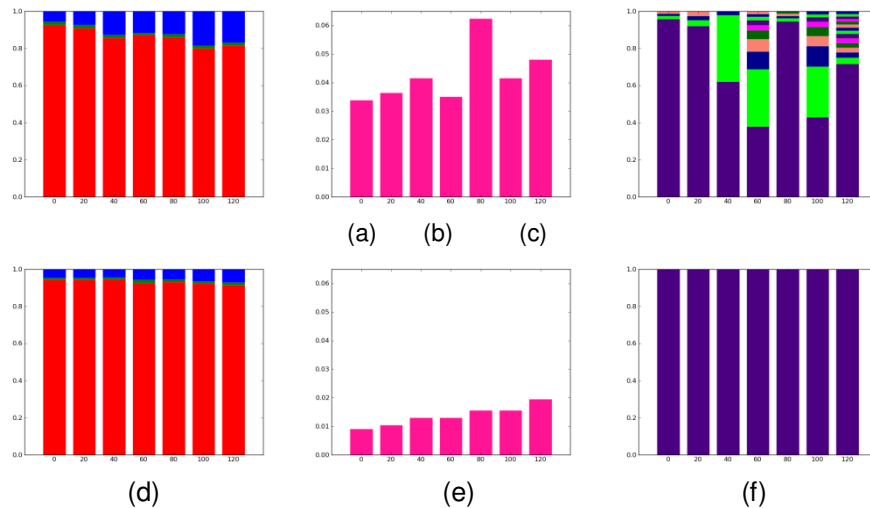

(a)

(b)

(c)

(d)

FIGURE 12: Accuracy Results Associated With The Model in Figure 10. Top Row Corresponds to Results From The Vesselness Algorithm And The Bottom Row From Fast Marching Algorithm. Each Histogram Corresponds to The Noise Level 0-red, 20-yellow, 40-green, 80-light blue, 100-dark blue, 120-magenta. (a,c) Absolute Position Error (Using max-norm) Between Every Pair of Matched Nodes. (b,d) Relative Radius Error Between Every Pair of Matched Nodes.

Jack Collins, Christopher Kurcz, Curtis Lisle, Yanling Liu & Enrique Zudaire

We note that several improvements can be made to our framework and categorize them into modelimprovements and analysis improvements. Depending on the application, a more detailed vascularmodel may be desired than the simple *L*-system approach employed here (see e.g. [14]). For example,a study in segmentation algorithms pertaining to angiogenesis could benefit from a vesselmodel which more accurately reflects the vascular morphology in or near the tumor environment.Such methods typically use a system of reaction diffusion partial differential equations (see e.g. [5])to model the growth of a vessel network. If the imaging modality is well characterized, a better suited noise model (rather than additive Gaussian noise) can be implemented. As pointed out in [33] aGaussian smoothed vessel cross section may be insufficient. Since generating the image from themodel is explicit in our framework, changing the type of convolution or averaging procedure to producea particular type of vessel cross section profile is straightforward. Our model currently neglectsbackground objects and does not explicitly incorporate vessel proximity to one another. A naturalextension to our assessment is multivariate analysis of several properties. For example, we couldexamine if missed nodes have a correlation between radius, curvature or contrast.

The software developed in this paper was written using the following free and open source libraries:Insight Toolkit [12] for image processing tasks, NetworkX [8] for the graph representation, Matplotlib[11] for 2D plotting, and Mayavi [22] for 3D plotting.

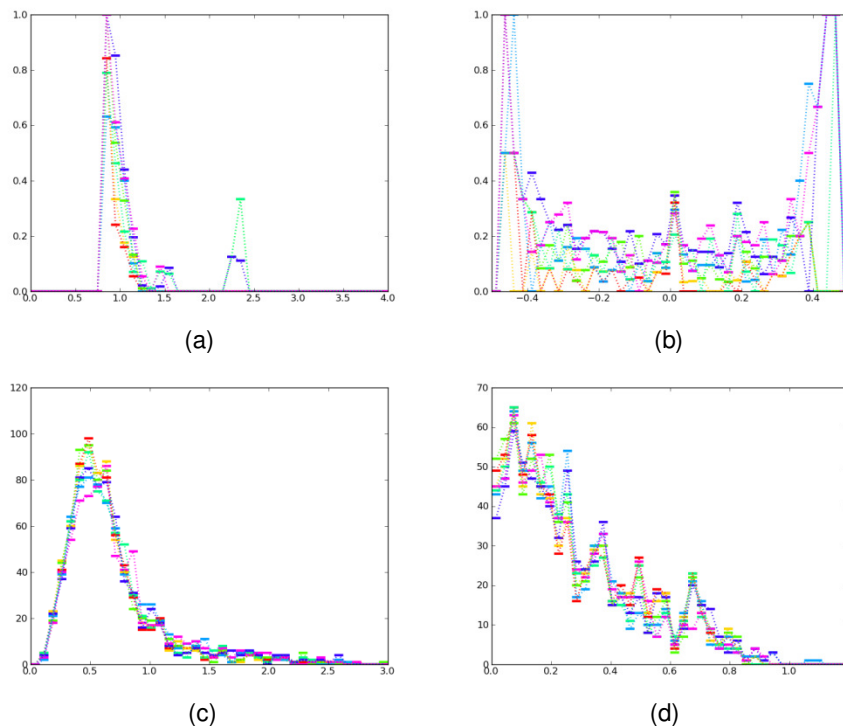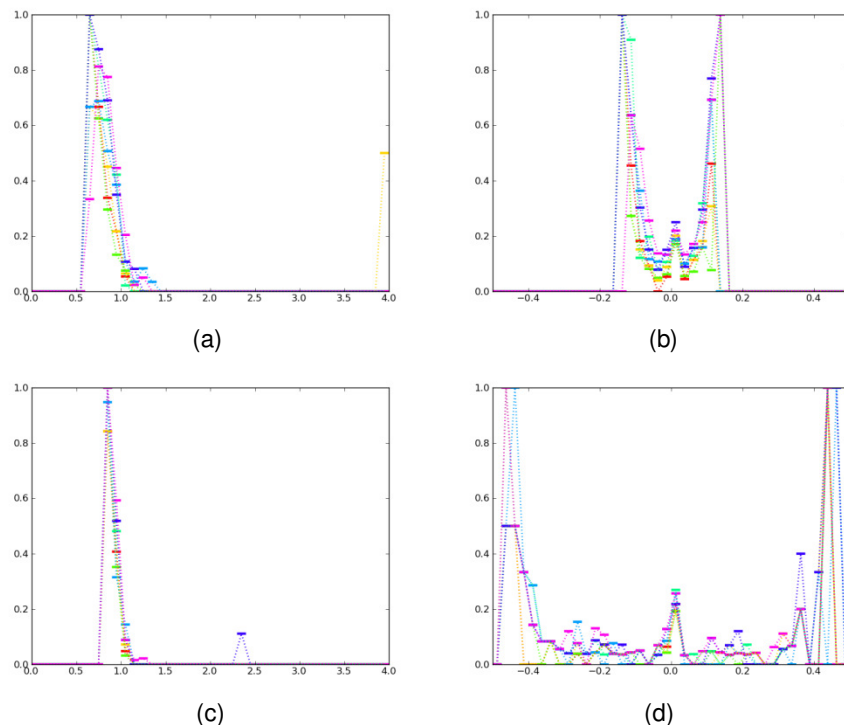**FIGURE 13:** Properties of Missed Nodes Associated With The Model in Figure 10. Top Row Corresponds to Results From The Vesselness Algorithm And The Bottom Row From Fast Marching Algorithm. Each Histogram Corresponds to The Noise Level 0-red, 20-yellow, 40-green, 80-light blue, 100-dark blue, 120-magenta. (a,c) Percentage of Missed Nodes Within Radius Interval. (b,d) Percentage of Missed Nodes Within Curvature Interval.

ACKNOWLEDGMENT

Jack Collins, Christopher Kurcz, Curtis Lisle, Yanling Liu & Enrique Zudaire

Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

5. REFERENCES

1. L. Antiga, M. Piccinelli, L. Botti, B. Ene-Iordache, A. Remuzzi, and D. A. Steinman. *"An image-based modeling framework for patient-specific computational hemodynamics"*. Medical & Biological Engineering & Computing, 46:1097-1112, November 2008.

2. E. Bullitt, M. Ewend, J. Vredenburgh, A. Friedman, W. Lin, K. Wilber, D. Zeng, S. R. Aylward, and D. Reardon.*"Computerized assessment of vessel morphological changes during treatment of glioblastoma multiforme: report ofa case imaged serially by MRA over four years"*. Neuroimage, 47(Supplement 2):T143-151, August 2009.

3. F. Cassot, F. Lauwers, C. Fouard, S. Prohaska, and V. Lauwers-Cances. *"A novel three-dimensional computer-assistedmethod for a quantitative study of microvascular networks of the human cerebral cortex"*. Microcirculation,13:1-18, January 2006.

4. R. M. Cesar, Jr., E. Bengoetxea, I. Bloch, and P. Larrañaga. *"Inexact graph matching for modelbasedrecognition: Evaluation and comparison of optimization algorithms"*. Pattern Recognition, 38(11):2099-2113,2005.

5. M. A. Chaplain, S. R. McDougall, and A. R. Anderson. *"Mathematical modeling of tumor-induced angiogenesis"*.Annual Review of Biomedical Engineering, 8:233-257, 2006.

6. D. Conte, P. Foggia, C. Sansone, and M. Vento. *"Thirty years of graph matching in patternrecognition"*. International Journal of Pattern Recognition and Artificial Intelligence, 18(3):265-298.

7. A. F. Frangi, R. F. Frangi, W. J. Niessen, K. L. Vincken, and M. A. Viergever. *"Multiscale vesselenhancement filtering"*. Medical Image Computing and Computer-Assisted Interventation, 1998.

8. A. A. Hagberg, D. A. Schult, and P. J. Swart. *"Exploring network structure, dynamics, and function usingNetworkX"*. In Proceedings of the 7th Python in Science Conference, Pasadena, CA, USA, August 2008.

9. H. Homann. *"Implementation of a 3D thinning algorithm"*. Insight Journal, 2007. http://hdl.handle.net/1926/1292.

10. H. Honda. *"Description of the form of trees by the parameters of the tree-like body: Effects of the branching angleand the branch length on the shape of the tree-like body"*. Journal of Theoretical Biology, 31(2):331-338, 1971.

11. J. D. Hunter. *"Matplotlib: A 2d graphics environment"*. Computing in Science and Engineering, 9(3):90-95, 2007.

12. L. Ibanez, W. Schroeder, L. Ng, and J. Cates. *"The ITK Software Guide"*. Kitware, Inc., 2005.

13. C. Kirbas and F. Quek. *"A review of vessel extraction techniques and algorithms"*. ACM Computing Surveys,36(2):81-121, 2004.

14. G. Kokai, Z. Toth, and R. Vanyi. *"Modelling blood vessels of the eye with parametric L-systems usingevolutionary algorithms"*. In Proceedings of the Joint European Conference

onArtificial Intelligence in Medicine and Medical Decision Making, Aalborg, Denmark, June 1999.

15. K. Krissian, G. Malandain, N. Ayache, R. Vaillant, and Y. Trousset. *"Model-based detection oftubular structures in 3d images"*. Computer Vision and Image Understanding, 80(2):130-171, 2000.

16. A. N. Langville and C. D. Meyer. *"Google's PageRank and Beyond: The Science of Search Engine Rankings"*.Princeton University Press, Princeton, NJ, USA, 2006.

17. A. Lindenmayer. "*Mathematical models for cellular interactions in development i. filaments with one-sided inputs"*.Journal of Theoretical Biology, 18(3):280-299, 1968.

18. A. Lindenmayer. *"Mathematical models for cellular interactions in development ii. simple and branching filamentswith two-sided inputs"*. Journal of Theoretical Biology, 18(3):300-315, 1968.

19. S. Luo and Y. Zhong. *"Extraction of brain vessels from magnetic resonance angiographic images: concise literaturereview, challenges, and proposals"*. In Proceedings of IEEE Engineering in Medicine and Biology Society, 2:1422-1425, 2005.

20. F. Nekka, S. Kyriacos, C. Kerrigan, and L. Cartilier. *"A model of growing vascular structures"*. Bulletin of Mathematical Biology, 58:409-424, May 1996.

21. W. Palubicki, K. Horel, S. Longay, A. Runions, B. Lane, R. Měch, and P. Prusinkiewicz. *"Self-organizing tree models for image synthesis"*. In Proceedings of ACM SIGGRAPH 2009, New York, NY, USA, 2009.

22. P. Ramachandran. *"MayaVi: A free tool for CFD data visualization"*. 4th Annual CFD Symposium, AeronauticalSociety of India, 2001.

23. X. Qian, M. P. Brennan, D. P. Dione, W. L. Dobrucki, M. P. Jackowski, C. K. Breuer, A. J. Sinusas, and X. Papademetris.*"A non-parametric vessel detection method for complex vascular structures"*. Medical Image Analysis, 13(1):49–61, February 2009.

24. G. Rozenberg and A. Salomaa, editors. *"Lindenmayer Systems: Impacts on Theoretical Computer Science,Computer Graphics, and Developmental Biology"*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2001.

25. Y. Sato, S. Nakajima, N. Shiraga, H. Atsumi, S. Yoshida, T. Koller, G. Gerig, and R. Kikinis. *"Three-dimensionalmulti-scale line filter for segmentation and visualization of curvilinear structures in medical images"*. Medical Image Analysis, 2(2):143-168, 1998.

26. M. Schaap, C. T. Metz, T. van Walsum, A. G. van der Giessen, A. C. Weustink, N. R. Mollet, C. Bauer, H. Bogunovi,C. Castro, X. Deng, E. Dikici, T. O'Donnell, M. Frenay, O. Friman, M. Hernandez Hoyos, P. H. Kitslaar, K. Krissian,C. Kuhnel, M. A. Luengo-Oroz, M. Orkisz, O. Smedby, M. Styner, A. Szymczak, H. Tek, C. Wang, S. K. Warfield,S. Zambal, Y. Zhang, G. P. Krestin, and W. J. Niessen. *"Standardized evaluation methodology and reference databasefor evaluating coronary artery centerline extraction algorithms"*. Medical Image Analysis, 13(5):701-714, October 2009.

27. J. A. Sethian. *"Level Set Methods and Fast Marching Methods"*. Cambridge University Press, 1999.

28. R. Singh, J. Xu, and B. Berger. *"Global alignment of multiple protein interaction networks with application to functionalorthology detection"*. In Proceedings of the National Academy of Sciences of the United States of America. 105(35):12763-12768, September 2008.

29. M. Sofka and C. V. Stewart. *"Retinal vessel centerline extraction using multiscale matched filters, confidence andedge measures"*. IEEE Transactions on Medical Imaging, 25(12):1531–1546, Dec 2006.

30. S. Thoreson. *"An efficient solution to inexact graph matching with application to computer vision"*. PhD thesis,Norwegian University of Science and Technology, 2007.

31. N.J. Tustison, M.Siqueira, and J.C.Gee.*"N-D Linear Time Exact Signed Euclidean Distance Transform"*. InsightJournal, 2006. http://hdl.handle.net/1926/171.

32. J. R. Ullmann. *"An algorithm for subgraph isomorphism"*. Journal of ACM, 23(1):31-42, January 1976.

33. S. Worz and K. Rohr. *"Segmentation and quantification of human vessels using a 3-D cylindrical intensity model"*.IEEE Transactionson Image Processing, 16(8):1994-2004, Aug 2007.

34. P. J. Yim, P. L. Choyke, and R. M. Summers. *"Gray-scale skeletonization of small vessels in magnetic resonanceangiography"*. IEEE Transactions on Medical Imaging, 19(6):568–576, June 2000.

Colour Face Image Database for Skin Segmentation, Face Detection, Recognition and Tracking of Black Faces Under Real-Life Situations

OJO, John Adedapo   jaojo@lautech.edu.ng
Department of Electronic & Electrical Engineering,
Ladoke Akintola University of Technology (LAUTECH),
Ogbomoso, P.M.B. 4000, Nigeria.

Adeniran, Solomon A   sadenira@oauife.edu.ng
Department of Electronic & Electrical Engineering,
Obafemi Awolowo University (OAU),
Ile Ife, Nigeria.

Abstract

This paper presents a new face database containing only black face images. It is designed to be used for benchmarking algorithms developed for skin segmentation, face detection and recognition under real life situations. The database contains coloured images of black faces in JPEG formats taken over a period of two years (2007-2009). It contains three subfolders, where the images have been grouped for face detection, face recognition and face tracking purposes. There are over 100 images for face detection containing more than 150 face images and 600 images of ten per 60 individuals for face recognition and 10s video sequences of ten black faces. An interactive Graphic User Interface (GUI) was developed in Matlab® (7.4.1a version) for the face database providing basic information about the images as well as creating an interactive and easy to use environment.

Keywords: Skin Segmentation; Face Detection; Face Recognition; Face Tracking

---

1. INTRODUCTION

To benchmark face detection or face recognition algorithms, the use of standard datasets is recommended as this would allow researchers to their compare results. Since there are varieties of large databases, the choice of an appropriate one to select is based on the problem(s) intended to be solved. The common face databases are the AT&T face database [1] and the MIT database (http://www.ai.mit.edu/projects/cbcl). Others that have been mentioned and used are the MATRANORTEL database [2], AR face database [3], and FERET database [4,5].

Image is formed in the eye and in the camera by the amount of illumination reflected by an object. This explains why capturing, detecting and recognition of faces of different skin colour is of high importance. Recent studies have shown that pigmentation (i.e. surface reflectance properties, including albedo, hue, specularity, translucency, and spatial variation in these properties) cues are equally as important as shape information for face recognition [6].

In the face recognition homepage (http://www.face_rec.org/databases/), about 35 face databases were reported to exist as at July 2008. In the handbook of face recognition, 20 face databases for face recognition, 2 for face detection, 5 for non-face and sub-face datasets and 3

for facial expression analysis were reviewed [7]. The summary of some of the face databases is presented in Table 1. They include FERET database [4,5], AT&T (formerly ORL) [1], AR (Purdue) database [3], CMU PIE database [8], Indian face database [9], Chinese Academy of Sciences (CAS) database [10] and VT-AAST image database [11]. In these databases, face images were captured under controlled conditions which are far from what is obtainable in the real life. For example, the cameras used for surveillance have much more lower resolutions compared to the ones reported to have been used in capturing the images in these databases. In addition, surveillance cameras do not have the capacity or added function of providing flash when an image is to be captured, whereas most of the images in the earlier databases used either flashes or were taken under standard lighting conditions.

We venture to bridge the gap in the existing databases especially for the detection and recognition of Black faces. Black surfaces reflect a small fraction of the light rays incident on them. Therefore, under very poor illumination conditions, very little information is made available for processing, detection or recognition. Recent report have countered the notion that color is unimportant for human face recognition and suggested instead that when shape cues in images are compromised (e.g. by reductions in resolution), the brain relies on color cues to pinpoint identity [12,13].

This paper presents a new face database which contains only black face images of different skin tones, middle aged and both male and female gender. It is designed to be used for benchmarking algorithms developed for face detection, recognition and tracking under real-life conditions. Section 2 gives a description of the database, how it was populated and the results of skin segmentation performed on some of the images in the database and section 3 gives the conclusion.

2. BLACK FACE DATABASE (BFDB)

The Black-face Database comprises of two parts, Part A was taken to serve as a pilot to the second Part B and for use in the early part of a research work [14], the Part B is being developed to increase the robustness of the database for benchmarking face detection, recognition and face tracking algorithms, with its own peculiar advantage. The Part A of the database is fully reported in this paper, while work is in progress on Part B of the database. Each Part contains three sub-folders; containing images for face detection, face recognition and video sequence of faces for face tracking.

The face detection folder 'Detect' contains 10 uncompressed pictures containing at least 45 faces. The images were captured at different dates and periods of the day and contain images with different pose, orientation, gender, background and illumination. A summary of the images in the 'Detect' subfolder is shown in Table 2. The face recognition folder contains 123 images of 20 persons taken under different conditions with varying facial expressions, details are shown in Table 3. The pictures were taken between August 2008 and April, 2009. The pictures were taken with Hp 7.3 Mega-pixels camera and they are in JPEG formats. The pictures were later cropped and compressed to 448×336 pixels. Compressing the images did not necessarily affect skin detection but the time of execution was faster than when the images were not compressed [14].

The database contains twenty folders each containing at least five upright and frontal pictures of each subject (person). Each folder comprises of frontal images of each subject taken without the use of flashes. There are three variations under the illumination which are; controlled lighting, labelled as 'bright'; outdoor but dull illumination labelled as 'poor'; and under dark illumination labelled as 'dark'. The variations in facial expressions were 'smile', 'frown' and 'blink'. Shots of subject with their heads slightly rotated to the right and to the left were also taken to achieve variations in pose and were labelled 'right' and 'left' respectively. At least one of the subjects had a scarf on and at least another one was in glasses. Examples of images of two subjects (persons) are shown in Figure 1.

Name or Description of Database and Date	Number of Subjects	Conditions		Image Resolution	Number of Images
FERET (August, 1993 – July, 1996) [4,5]	1199	Facial expression	2	256 × 384 pixels (8 bit grey scale)	14,051
		Illumination	2		
		Pose	9-20		
		Time	2		
AT & T (1992 – 1994) [1]	40	Varying Lighting Facial expression	4	92 × 112 pixels (PGM, grey scale)	400
		Facial details	2		
AR (PURDUE) (1998) [3]	116	Facial expression	4	768 × 578 Pixels (RGB colour images)	3,288
		Illumination	4		
		Occlusion	2		
		Time	**2**		
CMU-PIE (October – December, 2003) [8]	68	Pose	13	640 × 486 pixels	41,368
		Illumination	43		
		Facial expression	3		
Indian Face (February, 2002) [9]	40	Pose	7	640 × 480 pixels (JPEG, 256-grey levels)	440
		Facial expression	4		
CAS-PEAL (August, 2002 – April, 2003) [10]	1040	Pose	21	360 × 480 (Cropped grey scale images)	30,900
	377	Facial expression	6		
	438	Accessory	6		
	233	Illumination	9-15		
	297	Background	2-4		
	296	Distance from camera	12		
	66	Time	2		
VT-AAST (2007) [11]		Pose	3	300 × 225 (JPEG & GIF)	1027
		Orientation	2		
		Race	4		
		Structural Components	3		

TABLE 1: Summary of some existing face databases

FIGURE 1: Examples of images in the face recognition database

Part B of the database contains two folders, 'detect' and 'recognition' folders. The 'detect' folder contains two folders, one contains images captured under good illumination ('Light') while the other contains images captured under poor illumination ("Dark'). The light folder has 50 images which contain at least 72 human faces of varying poses and illumination captured under good lighting condition. Few images were adopted from the internet one or two existing databases. The dark folder has 50 images which contain at least 68 human faces of varying poses and orientations taken under poor illumination. A summary is given in Table 3 and examples of images captured under good and poor illuminations are shown in Figures 2a and 2b respectively.

(a) Images taken under good illumination

(a) Images taken under poor illumination
FIGURE 2: Examples of images in the face detection database

Images	Total Number of faces	Gender		Pose	
		Male	Female	Rotated	Upright
Pic1	1	1		1	
Pic2	5	5			5
Pic3	1	1		1	
Pic4	1		1		1
Pic5	1		1		1
Pic6	1	1			1
Pic7	2		2		2
Pic8	2	2			2
Pic9	1		1	1	
Pic10	7	6	1		7

TABLE 2: Summary of face detection images in Part a of the database

Folders	Gender	Illumination			Rotation		Expression		Blink	Other Effects		
		Bright	Poor	Dark	Right	Left	Smile	Frown	Blink	Scarf	Cap	Glasses
f1	M	√	√	√	√	×	√	×	×	×	×	×
f2	F	√	√	√	√	×	√	√	×	√	×	×
f3	M	√	√	√	√	×	√	√	×	×	×	×
f4	M	√	×	√	√	√	√	√	×	×	×	×
f5	M	√	√	√	√	√	×	√	×	×	×	×
f6	M	√	√	√	√	×	√	√	×	×	×	×
f7	M	√	√	√	√	×	√	√	×	×	×	×
f8	M	√	×	√	√	√	√	√	×	×	×	√
f9	F	√	√	√	√	√	√	√	×	×	×	×
f10	M	√	√	√	√	×	√	×	×	×	×	×
f11	M	√	√	√	√	√	√	√	×	×	×	×
f12	F	√	√	√	√	√	√	√	×	×	×	×
f13	M	√	√	√	√	×	√	√	×	×	×	×
f14	M	√	√	√	√	×	√	√	×	×	×	×
f15	M	√	√	√	√	√	√	√	×	×	×	×
f16	M	√	√	√	×	√	√	√	×	×	×	×
f17	M	√	√	√	√	×	√	√	×	×	×	×
f18	M	√	√	√	√	×	√	√	×	×	×	×
f19	M	√	√	√	√	√	√	√	×	×	×	×
f20	M	√	×	√	√	√	√	×	×	×	×	×

TABLE 3: Summary of face recognition images in Part A of the database

Grouping	Number of Images	Number of Faces
Dark	50	72
Light	50	68
Total	100	140

TABLE 4: Summary of face detection images in Part B of the database

An interactive Graphic User Interface (GUI) was developed in Matlab[TM] 7.4.1a version for the face database to enable an easier use. This is also an improvement over other face databases. Both the database would be copied at a known location on the user's computer and the Matlab file would be copied into the Matlab directory. The user then only needs to type the default filename "face_database in the Matlab command prompt. The interactive GUI would appear and the needed image can be selected. The user would type in the address or location of the database into the query, then select the exact image to be used. The image can then be saved and used for either face detection, training or testing a face recognition algorithm. The default filename for the face detection GUI is Database_detect. The outlooks of the GUIs are in Figures 3 and 4. The database as well as the Matlab® codes can be obtained by contacting the authors.
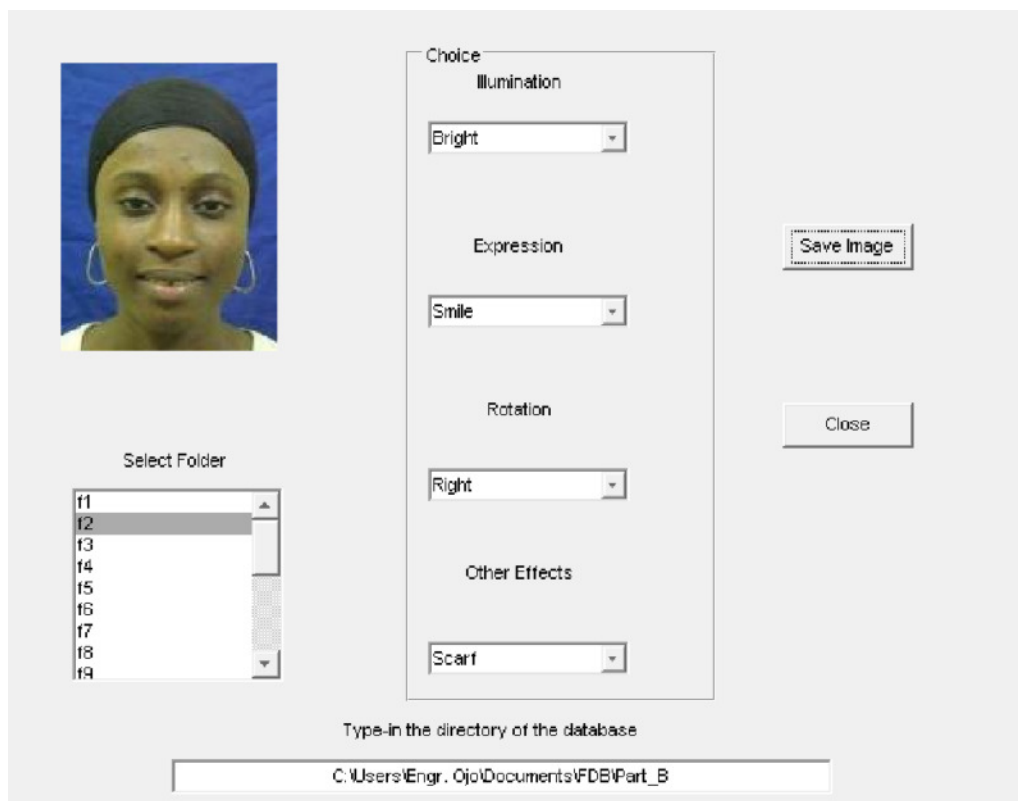
FIGURE 3: Matlab GUI for Face Detection Database

FIGURE 4: Matlab GUI for Face Recognition Database

Images in the database were tested using earlier skin segmentation algorithms [15], and some of the results are shown in Figure 5. Each of these figures shows the original image in the database and the corresponding skin-segmented images. The skin-likelihood areas in the skin segmented images appear white while other areas have deeper shades.
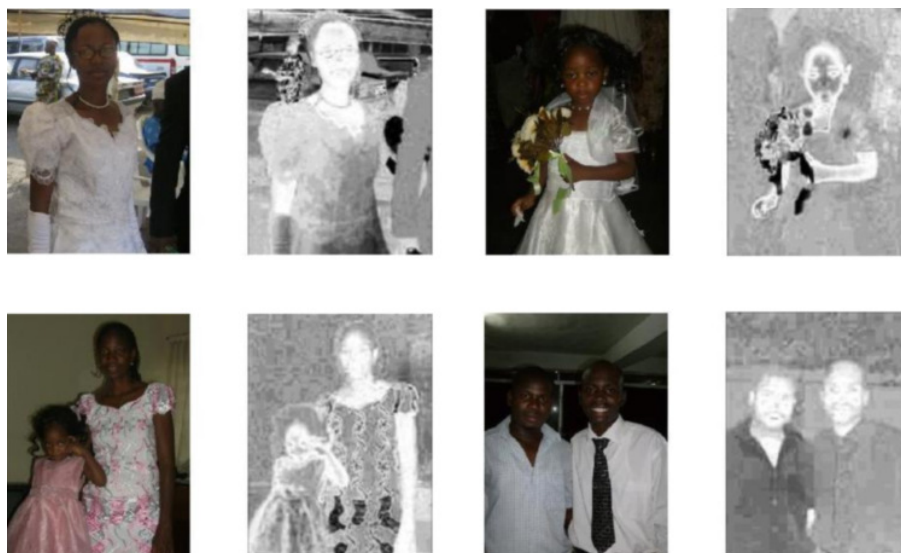
FIGURE 5: Skin segmented images from the database
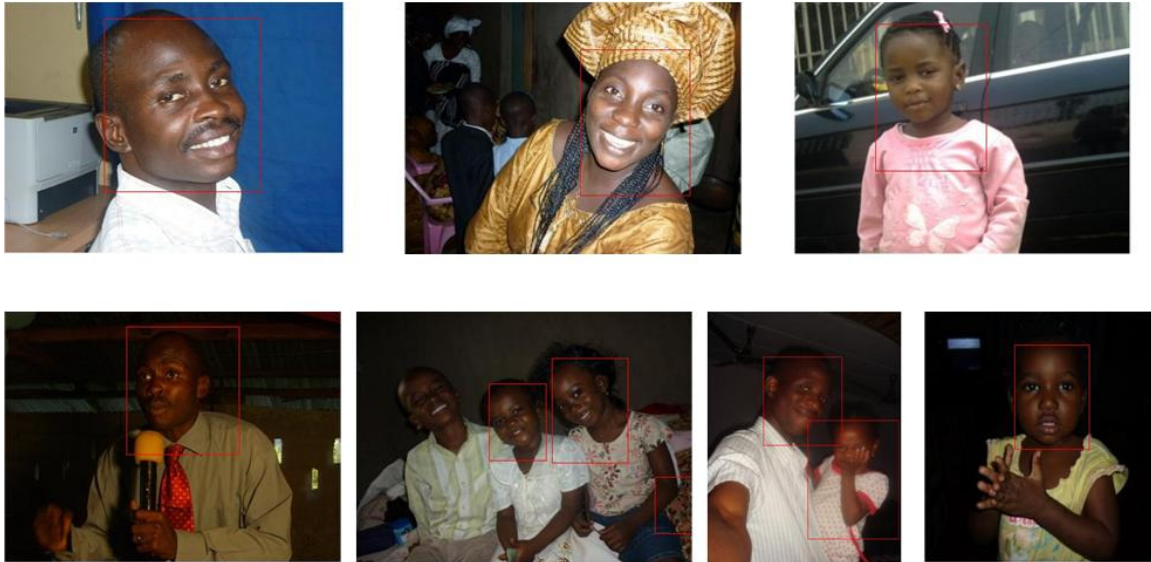
FIGURE 6: Selected images from the database showing detected faces.

Images in the face detection subfolder were tested for presence of faces using a novel face detection algorithm [16]. Some of ths results are shown in Figure 6. Detection rates of 83.8% and 76.4% were achieved for faces captured under good and poor illumination respectively. The same images were tested with earlier algorithms which have been reported to have good performances [17,18], details shown in Table 5. It is observed that although the earlier algorithms performed excellently well when benchmarked on some face detection databases, it was not so with the new database (BFDB). As a result, the new database with its unique feature of large variations under real life situations is expected create new frontiers for researches in face detection and recognition.

	Detection Rates	
Detection Method	**Good illumination**	**Poor illumination**
Hybrid Model [16]	83.8%	76.4%
Viola & Jones [17]	35.29%	31.94%
Pai et al [18]	66.18%	56.9%

TABLE 5: Detection rates for images taken under good Illumination

3. CONSLUSION

This paper has discussed a new face database developed to solve some challenges poised in using appropriate images to benchmark algorithms developed for skin-segmentation, face detection, recognition and face tracking, especially for Black faces. It contains images taken under large variations under real life situations and it gave good results in the study of skin-segmentation.

Ojo, John Adedapo & Adeniran, Solomon A.

4. REFERENCES

1. F. Samaria and A. Harter (1994), Parameterization of a Stochastic Model for Human face Identification. 2nd IEEE Workshop on Applications of Computer Vision, Saratosa FL. December, 1994, pp 138-142 Available at: http://www.uk.research.att.co/faced databased.html [Accessed 20 October 2007]

2. A. Tefas, Y. Menguy, O. Kotropoulos, G. Richard, I. Pitas and P.C. Lockwood. *"Compensating for Variable Recording Conditions in Frontal Face Authentication Algorithms".* IEEE International Conference on Acoustics, Speech and Signal Processing, (6)3561 – 3564, 1999

3. A.R. Martine and R. Benavente (1998), "The AR Face Database," CVC Technical Report 24, Computer Vision Centre (CVC). Technical Report, Barcelona. Available at: http://rvll.ecn.purdue.edu/~alex/alex-face-Db.html [Accessed 12 July 2009]

4. P.J. Phillips, H. Wechsler and P. Rauss *"The FERET Database and Evaluation Procedure for face Recognition Algorithms".* Image and Vision Computing, 16(5):295-306, 1998

5. P.J. Phillips. H. Moon, S.A. Rizvi and P.J. Rauss. *"The FERET Evaluation Methodology for Face-Recognition Algorithms".* IEEE Transaction on Pattern Analysis and Machine Intelligence. 22(10):1090-1104, 2000

6. S.E. Umbaugh. *"Computer Vision and Image Processing: A Practical Approach Using CVIP tools".* Prentice-Hall International, Inc., N.J. USA, pp. 12-24 (1998)

7. R. Gross. *"Face Database".* In S.Z. Li and A.K. Jain (Eds.) Handbook of Face Recognition. Springer-Verlag., pp. 301-328 (2005)

8. T. Sim, S. Baker and M. Bsat. *"The CMU Pose, Illumination and Expression Database".* IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(12):1615-1618, 2003.

9. V. Jain and A. Mukherjee (2002), The Indian Face Database [online], Available at: http://vis-www.is.umass.edu/~vidit/IndianFaceDatabase2002 [Accessed 7 November 2008]

10. W. Gao, B. Cao, S. Shans, D. Zhou, X. Zhang and D. Zhao (2004) CASP-PEAL Large Scale Chinese Face Database and Evaluation Protocols Technical Report. JDL-TR-04FR-001 [online], Joint research Development laboratory. Available at: http://www.jdl.ac.cn/peal/index.html [Accessed 10 June 2008]

11. A.S. Abdallah, M.A. El-Nasr and A.C. Abbott. *"A New Colour Image Database for Benchmarking of Automatic Face Detection and Human Skin Segmentation Techniques".* Proceedings of World Academy of Science, Engineering and Technology, 20:353-357, 2007

12. P. Sinha, B. Balas, O. Yuri, and R. Russell. *"Face Recognition by Humans: Nineteen Results All Computer Vision Researchers Should Know About".* Proceedings of the IEEE, 94(11):1948-1962, November 2006

Ojo, John Adedapo & Adeniran, Solomon A.

13. A. Yip and P. Sinha. *"Role of color in face recognition". Perception*, 31: 995–1003, 2002.

14. J.A. Ojo, S.A., Adeniran, and T.I. Raji. *"Developing Face Database for Detection and Recognition of Black Faces under Real-Life Situations".* In Book of Abstracts, 1st Postgraduate Research Symposium, Faculty of Engineering & Technology, LAUTECH, Ogbomoso, Nigeria, 10 -11th March, 2010

15. J. Cai and A. Goshtasby (1999), *"Detecting Human Faces in Colour-Images".* Image and Vision Computing, (18):63-75, 1999

16. J.A. Ojo, S.A., Adeniran and T.I. Raji (2010). *"Robust Face Detection Algorithm."* Paper accepted for publication in Global Journal of Engineering & Technology, Vol 3, No 3.

17. P. Viola and M. Jones (2004). *"Robust Real-Time Face Detection."* International Journal of Computer Vision, 57(2), 137-154.

18. Y.-T, Pai, S.-J. Raun, M.-C., Shie and Y.-C. Liu (2006). *"A Simple and Accurate Colour Face Detection Algorithm in Complex Background."* IEEE International Conference on Multimedia and Expo. Pp 1545-1548.

Rate Distortion Performance for Joint Source Channel Coding of JPEG Image Over AWGN Channel

Prof. Jigisha N. Patel   jnpatel@eced.svnit.ac.in
Assistant Professor, ECED,
s v national institute of tech.
surat,395007,india

Dr Suprava Patnaik   ssp@eced.svnit.ac.in
Professor, ECED,
s v national institute of tech.
surat,395007,india

Ms.Vaibhavi P. Lineswala   vaibhavilineswala@yahoo.co.in
PG Student, ECED,
s v national institute of tech.
surat,395007,india

Abstract

This paper presents the rate distortion behavior of Joint Source Channel Coding (JSCC) scheme for still image transmission. The focus is on DCT based source coding JPEG, Rate Compatible Punctured Convolution Codes (RCPC) for transmission over Additive White Gaussian Noise (AWGN) channel under the constraint of fixed transmission bandwidth. Information transmission has a tradeoff between compression ratio and received quality of image. The compressed stream is more susceptible to channel errors, thus error control coding techniques are used along with images to minimize the effect of channel errors. But there is a clear tradeoff between channel coding redundancies versus source quality with constant channel bit rate. This paper proposes JSCC scheme based on Unequal Error Protection (UEP) for robust image transmission. With the conventional error control coding schemes that uses Equal Error Protection (EEP), all the information bits are equally protected. The use of the UEP schemes provides a varying amount of error protection according to the importance of the data. The received image quality can be improved using UEP compared to Equal Error Protection (EEP).

Keywords: JPEG, Convolution Code, Puncturing, JSCC, UEP, EEP

1. INTRODUCTION

With rapid growth of data communication infrastructure, there has been an increasing demand for multimedia communication services involving image communication over wireless channels. Two common problems encountered in multimedia communication services are large bandwidth requirement and noisy transmission channels. Communication channels have limited resources such as bandwidth and power, and multimedia sources usually contain significant amount of

redundancy. Therefore data compression (source coding) is necessary [2] [17]. Source coding reduces redundancy and in doing so, it not only introduces distortion in the form of quantization noise, but data dependency is also occurs among the bits from a coded bit stream. This makes the source more sensitive to transmission errors. All the current image coding standards use some form of Variable Length Coding (VLC). To combat the errors introduced by noisy channels, channel coding is often employed to add controlled redundancy.

Error control mechanisms devised for image/video transport can be categorized into four groups: (1) at the source encoder, to make the bit stream more resilient to error (2) at the transport level, including channel coders, packetize/multiplexers (3) Error Concealment at the decoder upon detection of errors, and (4) interaction between the source encoder and decoder, so that the transmitter can adapt its operating based on the loss conditions detected .

According to Shannon's separation theorem [1], source coding and channel coding can be performed separately and sequentially without loss of optimality. However, this does not hold true for practical communication system and one can improve the overall performance by designing the source and channel codes jointly rather than separately, a process called Joint Source-Channel Coding (JSCC). In recent years, extensive research has been carried out in the field of JSCC [3] [4] [10] [15] [20]. It is well known that the theoretical bound for lossless compression is the entropy of the source. In the same way entropy determines the lowest possible rate for lossless compression, Rate Distortion (R-D) theory addresses the same question for lossy compression[18][23].

In 1979 David [3] employs combined source channel approach for 2-D DPCM which has been appropriately matched to the image source. In 1981David and James [4] employs source encoder 2-D block transform coding using the discrete transform (DCT). The approach is an extension of previous work. In 1998 Sherwood and Zegar [6] proposed product channel codes (two dimensional) to protect progressively compressed and packetized images for noisy channels. The main idea is to break the image coder bit stream into packets, encode them with the same Rate compatible punctured convolution code (RCPC) and across packets Reed Solomon (RS) code is used. A nice feature of this particular product code is that decoding the column is unnecessary unless decoding failure. In 2001 Wei Xiang and Steven [5] has presented unequal error protection (UEP) methods to JPEG image transmission using Turbo codes based on importance of data. Simulation results demonstrate the UEP schemes outperforms the equal error protection (EEP) scheme in terms of bit error rate (BER) and peak signal to noise ratio (PSNR). They assume ideal synchronization within the DCT blocks.

In 2005 Yeshan etc. [8] proposed Region of interest (ROI) feature supported by JPEG2000 image compression standard and allows particular region of interest within an image to be compressed at a higher quality than rest of the image. The UEP scheme using Golay code and Hamming code is applied according to importance of data. However ROI feature can useful only specific images. In 2006 Pasteur Poda and Ahmed Tamtaoui [9] proposed UEP scheme using retransmission protocol for JPEG image over Time varying channels. However this proposed solution is not obvious match with real time application. In 2008 Chou Chen etc [10] proposed JPEG image protection using RCPC. To cope up with the synchronization problem, synchronization codeword (Restart marker RM) they periodically inserted after each row into the JPEG image bit stream.

2. SYSTEM OVERVIEW

The standard image transmission model considered for this work is given in Fig 2.1. It consists of source encoder, channel encoder, transmission channel, channel decoder, and source decoder. The source encoder reduces or eliminates any redundancies in the input image, which usually leads to bit savings. The source coded signal is then encoded further using channel encoder to add error protection prior to transmission over a channel and hence increases noise immunity of source encoder's output. At the receiver, channel decoder detects and/or corrects transmission

errors and source decoder decompresses the signal. Most of the practical standards for image compression are lossy, i.e. the volume of data is compressed at the expense of visual quality.
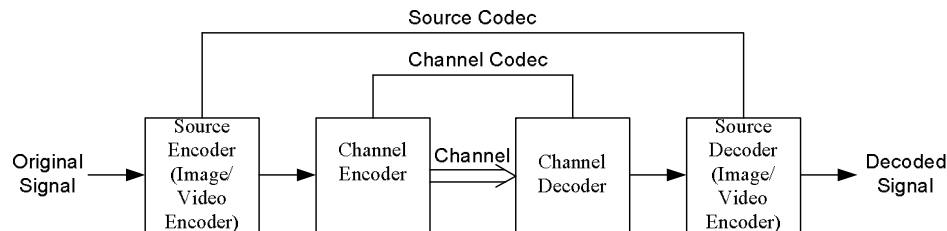
FIGURE 2.1: Image Transmission System

In this paper section III describes design of Source Encoder Decoder used in this simulation. The encoded bit stream is partitioned into two groups, DC and AC coefficients. Section IV describes design of Channel Encoder Decoder. Section V discusses design first for equal error protection using JSCC and Rate Distortion performance to obtain optimum solution. Secondly, joint source-channel coding (JSCC) based on UEP is applied in which RCPC channel encoder applies different channel coding rates to DC and AC coefficients. Highly sensitive DC coefficients are better protected with a lower code rate, while less sensitive AC coefficients higher code rate.

3. JPEG ENCODER DECODER

The Joint Photographic Experts Group (JPEG) standard (1992) is widely used for coding still images (such as photographs). Its main application is storage and transmission of still images in a compressed form, and it is widely used in digital imaging, digital cameras. An overview of image compression standard JPEG is discussed in detail [2] [21]. DCT is widely used in JPEG because of two important properties; high de correlation and energy compaction [25]. Fig. 3.1 shows the basic block diagram of JPEG Encoder.
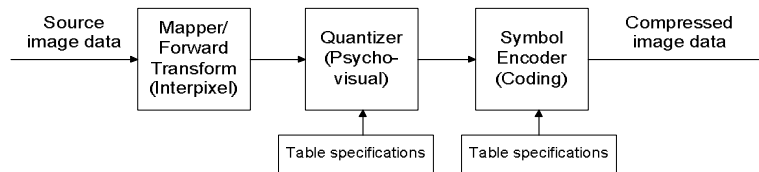
FIGURE 3.1: JPEG Source Codec

JPEG encoder-decoder consists the following steps [17]:
- Converting the base image to 8x8 matrices
- Level shifting by subtracting 128 from each pixel
- DCT transform
- Quantizing and normalizing
- DPCM coding of DC coefficient and Huffman encoding
- Zigzag scanning , Run length encoding and Huffman encoding of AC coefficients
- Denormalization and Dequantization
- Inverse DCT and Level shifting back by adding 128 to each pixel

In our simulation, we have used symbols and specification as given in Table 3.1:

Jigisha N. Patel, Suprava Patnaik & Vaibhavi P. Lineswala

Original File	'cameraman.tif '
Original File Size (S)	256X 256
Bits per pixel of Original file (BPP$_O$)	8 bits/pixel
Total bits after JPEG encoding (Bs)	Depends on Quality factor
Source Encoder Rate (Rs) bits/pixel	Bs/S
Compression Ratio (CR)	(BPP$_O$ X S)/ Bs

TABLE 3.1: Symbols and Specification of JPEG Encoder and Decoder

As Quality Factor (QF) changes the number of nonzero element in each 8X8 block of DCT after quantization varies. This affects finally reconstructed image. In JPEG, stream is partitioned into DC coefficients and AC coefficients. The simulation results for the test image cameramen for different QF are shown in Table 3.2**.**

QF	Bs	Rs =Bs/S Bits/pixel	MSE	PSNR(dB)	CR	Perceptual Quality
1	7758	0.11	1383	16.72	67.58	Not accept
2	9137	0.13	566.82	20.6	57.38	Not accept
5	13235	0.2	252.64	24.1	39.62	Not accept
10	19711	0.3	154.51	26.24	26.59	Not accept
15	25376	0.38	114.98	27.52	20.66	accept
20	30579	0.47	93.19	28.43	17.14	accept
30	39533	0.63	68.79	29.75	13.26	accept
40	47301	0.72	54.5	30.77	11.08	good
50	55134	0.84	44.64	31.63	9.5	good
60	63003	0.96	36.15	32.54	8.32	good
70	74557	1.13	27.17	33.79	7.03	good
90	139454	2.12	6.74	39.84	3.75	good

TABLE 3.2: Evaluation parameters for various QF

Source rate Rs approximately exponentially increases with QF increases as shown in Fig .3.2. The source Rate Distortion (RD) curve for Cameramen image is shown in Fig. 3.3. From the source RD curve it is concluded that as QF increases the source bits rate (R_s bits/pixel) increases, so distortion (MSE) in received image is reduces. Higher compression can be achieved at the cost of visual quality. This curve varies from image to image.
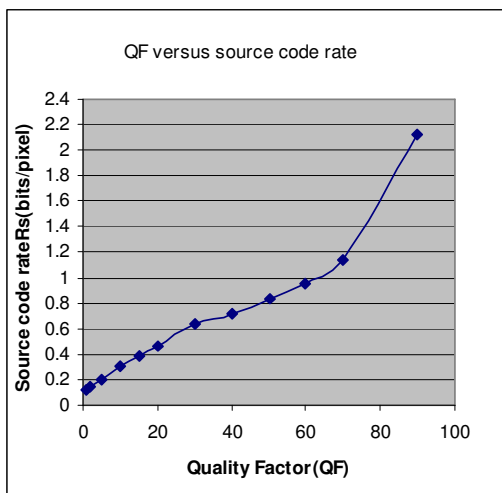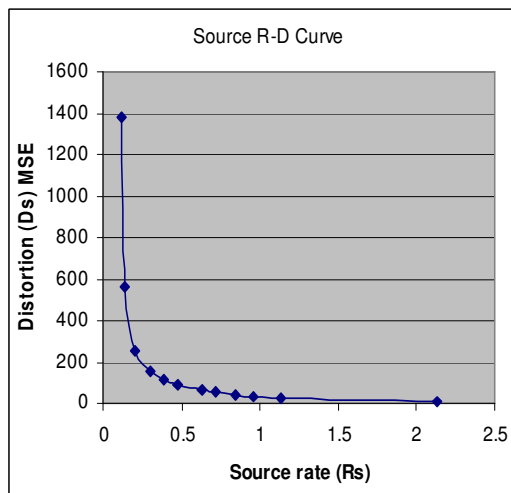

FIGURE 3.2: QF versus Rs

FIGURE 3.3: Source Rate Distortion (RD) curve

Jigisha N. Patel, Suprava Patnaik & Vaibhavi P. Lineswala

4. CHANNEL CODING

The bit stream for compressed image is more susceptible to channel errors. Thus error control coding techniques are used along with compressed image bit stream to minimize the effect of channel errors. Various Error Control Techniques are Automatic Repeat Request (ARQ), Forward Error Correction (EFC), Interleaving, Layered Coding with Unequal Error Protection and Error Concealment. Cyclic Redundancy Check (CRC-16) code is already proposed for error detection and Rate Compatible Punctured Convolution (RCPC) [11] [12] [13] [14] code for error correction. When the same protection is given to all encoded source bits regardless their channel error sensitivity, the method is called Equal Error Protection (EEP). The method of modulating the amount of channel coding based on the required level of protection is known as Unequal Error Protection (UEP). UEP scheme allows us to protect significant bits by allocating a lower channel code rate and less significant bits at a higher channel code rate.

Convolution codes are a powerful class of error correcting codes, providing equal error protection over the information bit stream [18]. Punctured convolution codes were first introduced by Cain, Clast and Geist [11]. Puncturing is the process of deleting (puncturing) some parity bits from the output codeword of lower code rate coder according to a puncturing matrix so that fewer bits are transmitted than in the original coder and hence leading to higher code rate. For a rate 1/N mother code rate encoder, the puncturing pattern can be represented as an N x P matrix, where P is a matrix whose elements are 1's and 0's, with a 1 indicating inclusion and a 0 indicating deletion of bit. In 1988 Hagenauer [13] extended the concept of punctured convolution codes by puncturing a low rate 1/N code periodically with period p to obtain a family of codes with rate $p/(p+l)$ where l can be varied between 1 and (N-1)p.

Fig.4.1 shows convolution code of rate = 1/3 with memory M = 6 and code generator matrix [133 171 145]. The specification for RCPC code is given in Table 4.1.
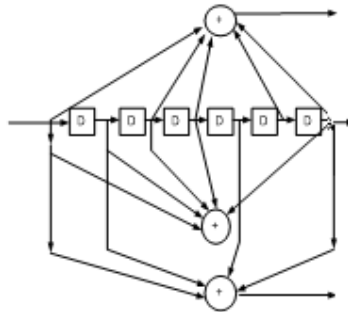
FIGURE 4.1: RCPC Code Generator

Mother code rate (1/N)	1/3
Punctured Code rates, Rc = ($p/p+l$)	8/9, 8/10, 812, 8/14, 8/16, 8/18, 8/20, 8/22, 8/24
Puncture period p	8
Decoder	Soft decision
Memory	6
Code Generator	[133, 171, 145]
Channel type	AWGN
Modulation	BPSK

TABLE 4.1: Specification of RCPC Code

The performance of selected RCPC codes on a Gaussian channel states with soft decision under different values of Es/No are simulated and results are given in figure 4.2. Lower code rate makes lower bit error probabilities, which means better protection for combating the channel errors.
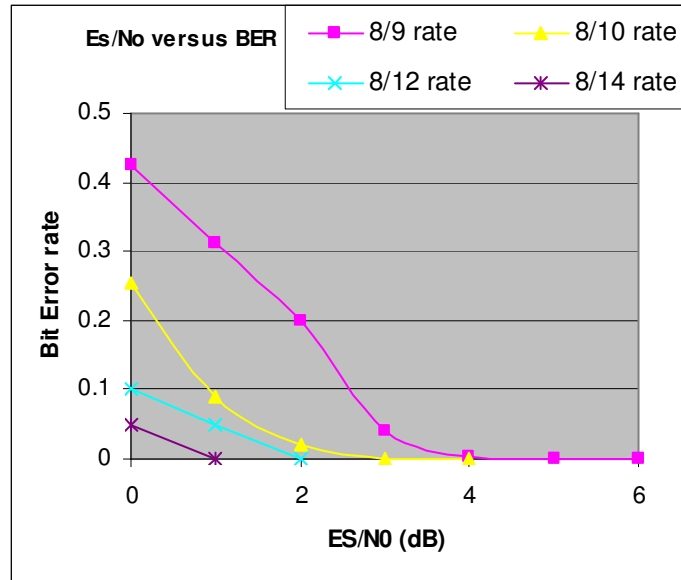
FIGURE 4.2: performance of RCPC Code family

5. SIMULATION RESULT FOR JOINT SOURCE CHANNEL CODING(JSCC)

The goal of JSCC is to distribute the source bits and the channel bits between source coder and channel coder so that the resulting end-to-end distortion is minimized. JSCC has gained significant research attention during the last decade, particularly since the Internet revolution.

The image coding usually involves a rate-distortion trade off. That is, when more bits are spent on coding a picture, less distortion will occurs. Conversely, when fewer bits are spent, more distortion will be observed. The rate-distortion trade off curve is useful in situations when the bit budget is a constrain. Generally the Joint Source Channel Coding (JSCC) schemes achieve the optimal bit allocation between source and channel. In a traditional image coder, the optimization algorithm only considers the distortion caused by quantization of DCT coefficients. However, in a JSCC framework, the rate-distortion tradeoff is extended to include the distortion coming from quantization and channel errors.

(A) Equal Error Protection (EEP) Scheme:

The JPEG encoder output bit stream is partition into DC coefficient and AC coefficient bit stream. These streams are partitioned into consecutive blocks of length B. Then a collection of C no of total CRC bits are derived based only on these B bits (C= 16) are appended with B data bits. Finally M zero bits, where M is the memory size of the convolution coder (M=6), are appended to the end. The purpose of adding M bits is to flush the memory and terminate the trellis to zero state. The resulting block of B + C + M bits is then passed through a Rate Compatible Punctured Convolutional (RCPC) coder. Equal Error Protection (EEP) defined by RCPC code rate is same for both DC and AC Coefficients. The Various parameters analyzed for this system are given below.

(i) If we fixed punctured convolution code rate Rc=8/9 and change the parameter E_S/N_O from 4 dB to 6dB the received image quality can be improved as shown in Fig.5.1.In this simulation the source rate is fixed Rs = 0.47 Bits Per Pixel(QF=20).

(ii) If we fixed QF =20 and E_S/N_O=2dB, variation in channel code rate Rc the required bit budget will change as given in Table 5.1. Here the simulation is done using fixed packet size 256. Total bit budget at input of channel is defined as R_{Total}.

Decompressed image

Decompressed image

QF=20, PSNR=19.3918, CR=17.1454   QF=20, PSNR=28.4367, CR=17.1454
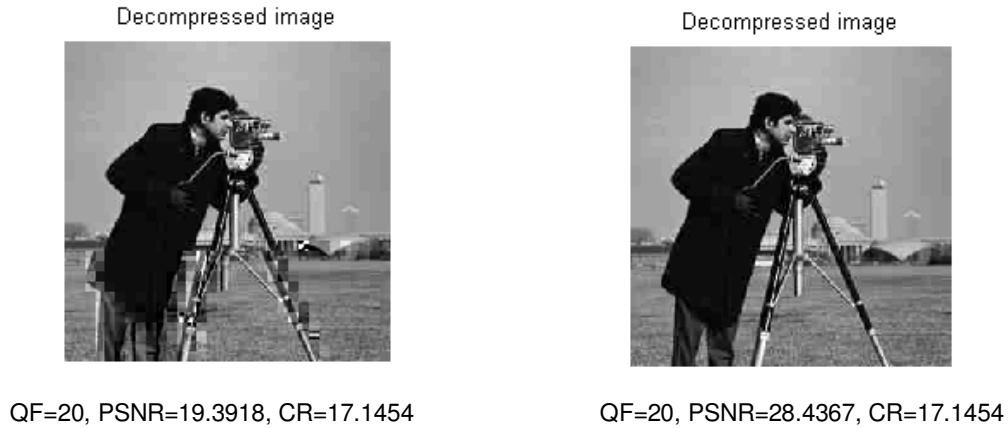
FIGURE 5.1: Rate = 8/9 (a) Es/No = 4dB (b) Es/No = 6dB

Channel Code Rate(Rc)	Packet Length in Bits (B)	Total Packet(P) =30579/B	Total Bits Bc (Px256)	R_{Total} = Bc/S Bits/Pixel	PSNR	Bit Error Rate(BER)	MSE
8/9	205	150	38400	0.58	4.94	0.2358	2.08E4
8/10	183	168	43008	0.65	6.66	0.0248	1.4E4
8/12	149	206	52736	0.8	28.593	0	89.90
8/14	124	247	63232	0.96	28.593	0	89.90
8/16	106	289	73984	1.13	28.593	0	89.90
8/18	92	333	85284	1.3	28.593	0	89.90
8/20	80	383	98048	1.5	28.593	0	89.90
8/22	71	431	110336	1.7	28.593	0	89.90
8/24	64	478	122368	1.87	28.593	0	89.90

TABLE 5.1 QF=20, Es/N0=2dB, packet size 256

As channel code rate reduces, more number of redundancy bits is added. So for fixed QF, bits per pixel of source (Rs) are fixed, but total transmitted number of bits increasing. As code rate reduces BER performance also improves. At channel code rate (8/12) bit error rate (BER) becomes zero and lower than that the entire code rate, PSNR becomes constant for same channel condition (E_S/N_O=2 dB) . This can be considered as optimum channel code rate to generate highest of PSNR.

(B) Optimum JSCC Design for Fixed bit budget (R_{Total}) Using RD performance:

For fixed total bit rate R_{Total} , EEP algorithm searches all possible combinations of source bit rate (R_S) and channel code rate (R_C) to find the best one that minimizes the end-to-end distortion D_{Total}. End to end distortion D_{Total} is measured in terms of Mean Square Error (MSE), it includes both source distortion and channel distortion. With R_{Total} = 1.5 bits/pixel, Es/No = 2 dB, Packet size = 256 bits, CRC size = 16 bits, the simulation results are given in Table 5.2 .The operational RD curve is plotted in Fig. 5.2. Initially as Rs increases, channel code rate is sufficient for correcting channel errors up to Rc =8/12. Up to Rc=8/12 rate channel error can be corrected ,so visual quality improve as Rs increases .But after this point as Rs increases, source distortion

decreases but channel noise immunity also decreases, so total distortion increases. There exist optimal points for which allocation of the available fixed transmission rate bits are optimally allotted between source and channel such that end to end distortion is minimum. From the graph, optimal point (highlighted in Table 5.2 by color) is obtained at R_S = 0.87 Bits/pixel (QF = 54) and R_C channel code rate = 8/12. In other words, to obtain minimum distortion, source should be coded at QF=40 and 8/12 rate RCPC code should be used for channel coding for fixed R_{Total} =1.5 and Es/No=2dB. The simulation results can be repeat for another value of Es/No.

Quality Factor (QF)	Source BPPs (R_S)	Channel Code Rate (Rc)	(R_{Total})	Distortion Es/N0=2dB	Distortion Es/No=2dB (dB)	PSNR
14	0.37	8/24	1.5	115	41.26	27
20	0.47	8/20	1.5	89.9	39.07	28.59
32	0.63	8/16	1.51	63.08	35.99	30.13
40	0.72	8/14	1.49	52.9	34.47	30.89
54	0.87	8/12	1.5	40	32.14	32.05
67	1.07	8/10	1.51	1.09E4	85.61	5.32
73	1.21	8/9	1.5	2.21E4	86.90	4.67

TABLE 5.2: Experimental Results of Optimal Bit Allocation with EEP Scheme
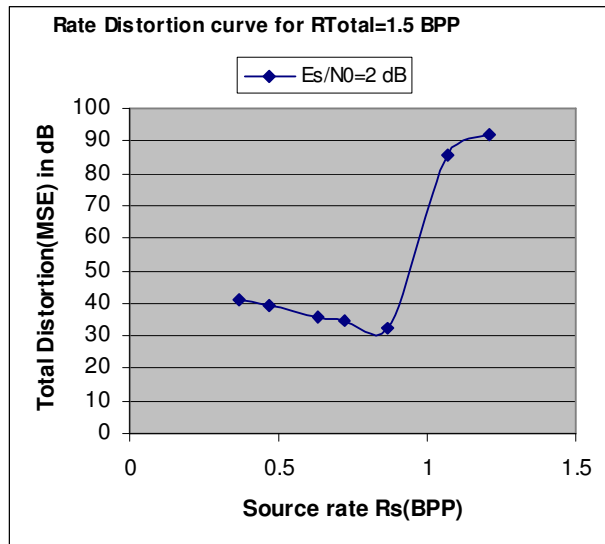

FIGURE 5.2: Operational RD curve

C) Simulation of Unequal Error Protection (UEP) Scheme

In UEP both the DC and AC coefficients have applied different protection channel code rate according to their importance. From Table 5.3.it is concluded that UEP scheme outperforms EEP in terms of end to end distortion for fixed R_{Total}.

Jigisha N. Patel, Suprava Patnaik & Vaibhavi P. Lineswala

R_{Total} (BPP)	CASE	Rs	R_{DC}	R_{AC}	D_{total} (MSE)	PSNR (dB)
1.5	UEP	0.85	8/14	8/12	48.29	31.20
	EEP	0.47	8/20	8/20	89.9	28.59
	EEP	0.63	8/16	8/16	63.03	30.13

TABLE 5.3: EEP and UEP comparison

6. CONCLUSION

Joint source channel coding approach for digital data communications, mainly for information sources like images and video, has registered a great success and is more and more passing to be conventional nowadays. There is a clear tradeoff between channel coding redundancies versus source coding resolution. When few channel redundancy bits carrying quantization information, there is little channel error correction. Though source coding or quantization distortion is small it will cause unacceptable higher distortion due to uncorrected channel errors. On the other hand more redundancy bits at the channel will leave insufficient bit rate to describe the source In this case the channel error correction capability is higher, but the source coding distortion is relatively high, thus again possibly yielding a large total distortion. Between these two extremes there exist optimal choice of a channel code rate and source code rate that minimize the distortion. The optimum point will be shift as channel condition changes. We allocate lower coding rate to higher sensitive DC coefficients bit stream and higher channel coding rate to AC coefficients bit stream for exploiting different sensitivity of source bits.

7. REFERENCES

[1]   E.Shanon, "A mathematical Theory of Communication," The Bell system technical journal, volume 27, pp. 379-423, 1948.

[2]   Gregory K. Wallace, "The JPEG still picture compression standard", Special issue on Digital multimedia systems, Issue 4, vol 34, pp. 30-44, April 1991.

[3]   J W Modestino and D.G Dautt, "Combind source channel coding of images" IEEE trans. Commun. Vol. COM-27, pp.1644-1659, Nov-1979.

[4]   J W Modestino and D.G Dautt, "Combind source channel coding of images using the Block cosine -Transform" IEEE trans. Commun. Vol. COM-29, No-9, pp.1261-1274, Sep-1981.

[5]   W. Xiang, S. Barbulescu, S. Pietrobon, "Unequal error protection applied to JPEG image transmission using Turbo codes", ITW2001, cairns, Sept. 2-7, Australia.

[6]   P. Greg Sherwood and Kenneth Zeger, "Error protection for progressive image Transmission over Memory less and Fading Channels," IEEE Transaction on communications, Vol-46, pp. 12, Dec. 1998.

[7]   Manora Cadera, Hans Jurgen Zepernik, Ian David Holland "Unequal Error protection Schemes for Image Transmission over Fading Channels," Wireless communication systems pp. 203-207, Sept. 2004.

[8]   Yeshan Yatawara, Manora Caldera, Tubagus Maulana Kusuma and Hans Zepernik,

"Unequal Error protection for ROI Coded Images over Fading Channels" IEEE Proceedings of the system Communications, pp. 111-115, Aug. 2005.

[9] Pasteur Poda and Ahmed Tamtaoui, "On the Enhancement of Unequal Error Protection Performance in Images Transmission Over Time Varying Channels," IJCSNS International Journal of Computer Science and Network Security, Vol. 6, No. 9B, Sept. 2006.

[10] Chou-Chen Wang, Tung-Yuen Huang and chung You Yang, "Joint Source Channel Coding for JPEG Compressed Images over Noisy Channel," Congress on Image and Signal Processing, IEEE Computer society pp. 676-680, 2008.

[11] J.B.Cain, G.C Clark, and J M Geist," Punctured convolutional codes of rate (n-1)/n and simplified maximum likelihood decoding" IEEE Tran. Inform. Theory, vol.IT-20, pp.388-389, May 1974.

[12] Y. Yasuda et al., "High rate punctured convolution codes for soft decision viterbi decoding", IEEE Trans. Commu., vol. COM-32, pp. 315-319, Mar.1984.

[13] J. Hagenauer, "Rate compatible punctured convolutional codes (RCPC) and their application" IEEE transaction on communication vol 36 no 4, pp. 389-400, April 1988.

[14] A. J. Viterbi, "Convolutional codes and their performance in communication systems" IEEE Trans Commun. Technol., vol. COM-19, pp.751-772, OCT-1971.

[15] M.J. uf and J.W. Modestimo, "Operational Rate Distortion performance for joint source and channel coding of images" IEEE Trans Image processing, vol. -8, pp.68-84, Mar 1999.

[16] B. Hochwald and K. zegar, "Tradeoff between source and channel coding. 'IEEE Trans Information theory, vol.43, Sept 1997.

[17] Rafael C. Gonzalez & Richard E. Woods, "Digital Image Processing", second edition, Pearson Education publication, 2002.

[18] Sklar, B., "Digital Communications: Fundamentals and Applications", Second Edition, Prentice-Hall, pp 436-460, 2001

[19] K. R. Rao, N. Ahmed, T. Natarajan "Discrete Cosine Transform", IEEE transactions computers, vol. 23, pp. 90-93, Jan 1974.

[20] Zhenyu Wu, "Joint source channel coding for image transmission with JPEG200 over memoryless channels", IEEE transactions on image processing, vol. 14, no. 8, pp 1020-1032, Aug 2005.

[21] Navin Chaddha and Suhas Diggavi, "A Frame-work for Joint Source-Channel Coding of Images over Time-Varying Wireless Channels", IEEE xplore, pp 89-92, 1996.

Improving Performance of Multileveled BTC Based CBIR Using Sundry Color Spaces

Dr. H. B. Kekre   hbkekre@yahoo.com
Senior Professor,
MPSTME, SVKM's NMIMS University,
Mumbai-56, India

Sudeep Thepade   sudeepthepade@gmail.com
Associate Professor and *Ph.D. Research Scholar,*
MPSTME,SVKM's NMIMS University,
Mumbai-56, India

Shrikant P. Sanas   shrikant_sanas@yahoo.co.in
M-tech Student
MPSTME,SVKM's NMIMS University,
Mumbai-56, India

Abstract

The paper presents an extension of content based image retrieval (CBIR) techniques based on multilevel Block Truncation Coding (BTC) using nine sundry color spaces. Block truncation coding based features is one of the CBIR methods proposed using color features of image. The approach basically considers red, green and blue planes of an image to compute feature vector. This BTC based CBIR can be extended as multileveled BTC for performance improvement in image retrieval. The paper extends the multileveled BTC using RGB color space to other nine color spaces. The CBIR techniques like BTC Level-1, BTC Level-2, BTC Level-3 and BTC Level-4 are applied using various color spaces to analyze and compare their performances. The CBIR techniques are tested on generic image database of 1000 images spread across 11 categories. For each CBIR technique, 55 queries (5 per category) are fired on extended Wang generic image database to compute average precision and recall for all queries. The results have shown the performance improvement (ie., higher precision and recall values) with BTC-CBIR methods using luminance-chrominance color spaces (YCgCb, Kekre's LUV, YUV, YIQ, YCbCr) as compared to non-luminance (RGB, HSI, HSV, rgb , XYZ) Color spaces. The performance of multileveled BTC-CBIR increases gradually with increase in level up to certain extent (Level 3) and then increases slightly due to voids being created at higher levels. In all levels of BTC Kekre's LUV color space gives best performance.

Keywords: Content Based Image Retrieval (CBIR), BTC, Color Spaces.

1. INTRODUCTION

From ancient era, images play an important role in human communication. It is basic and common way to express the information. Today with advancement in information and

communication technology most of the information is digitized. Large amount of digital data is generated, transmitted, stored, analyzed and accessed. Mostly, information is in the form of multimedia such as digital images, audio, video, graphics [6]-8. Large numbers of images are generated from various sources on a daily basis. Such images occupy lot of space and are very challenging to search and retrieve from very large image pool. The need for efficient retrieval of images has been recognized by managers and users of large image collections. Efficient indexing techniques for the retrieval of best matching image from a huge database of images are being developed. Content based image retrieval gives efficient solution to these problems [18,19]. Content Based Image Retrieval (CBIR) is used to provide a high percentage of relevant images in response to the query image [12]. The goal of an image retrieval system is to retrieve a set of matching images from an image database [21].

A Content Based Image Retrieval (CBIR) technique takes an image as an input to query and outputs number of matching images to the query image [11]. In CBIR technique, features are used to represent the image content. The features are extracted automatically and there is no manual intervention, and thus eliminating the dependency on humans in the feature extraction stage [10]. The typical CBIR system performs two major tasks. The first one is feature extraction (FE), where a set of features, forming feature vector, is generated to accurately represent the content of each image in the database. A feature vector is much smaller in size than the original image. The second task is similarity measurement (SM), where a distance between the query image and each image in the database using their feature vectors (signatures) is computed so that the top "closest" images retrieved [3], [13], [14], [15].

Many current CBIR system use Euclidean distance [5] on the extracted feature set as a similarity measure. The Direct Euclidian distance between image P and query image Q can be given as equation. where Vpi and Vqi are the feature vectors of image P and query image Q respectively with size 'n'.

$$ED = \sqrt{\sum_{i=1}^{n}\left(Vpi - Vqi\right)^2}$$ (1)

Some of important applications for CBIR technology could be identified as art galleries, museums, archaeology, architecture design, geographic information systems, weather forecast, medical imaging , trademark databases, criminal investigations, image search on the Internet. The thirst of a better and faster image retrieval technique is increasing day by day. The paper discusses the performance improvement in multileveled BTC based CBIR techniques [1] using various color spaces. In all ten different color spaces including RGB color space are considered here with four different levels of BTC for feature extraction resulting into total 40 CBIR methods.

2. BLOCK TRUNCATION CODING (BTC)
Block truncation coding (BTC) is a relatively simple image coding technique developed in the early years of digital imaging more than 29 years ago. Although it is a simple technique, BTC has played an important role in the history of digital image coding in the sense that many advanced coding techniques have been developed based on BTC or inspired by the success of BTC [2]. Block Truncation Coding (BTC) was first developed in 1979 for grayscale image coding [2]. The method first computes the mean pixel value of the whole block and then each pixel in that block is compared to the block mean. If a pixel is greater than or equal to the block mean, the corresponding pixel position of the bitmap will have a value of 1 otherwise it will have a value of 0. Two mean pixel values one for the pixels greater than or equal to the block mean and the other for the pixels smaller than the block mean are also calculated. At decoding stage, the small blocks are decoded one at a time. For each block, the pixel positions where the corresponding bitmap has a value of 1 is replaced by one mean pixel value and those pixel positions where the corresponding bitmap has a value of 0 is replaced by another mean pixel value.

3. CBIR USING MULTILEVELED BLOCK TRUNCATION CODING

The block truncation coding (BTC) technique can be extended to higher levels by considering multiple threshold values to divide the image pixels into higher (upper) and less than or equal to (lower) threshold. The image pixel data is thus divided in to multiple clusters and per cluster the mean value is taken as part of feature vector. At BTC Level 1 only one threshold value is used to divide the pixel data to get two clusters and respective means of these clusters as upper mean and lower mean are computed, resulting in to feature vector of size six (two values per color plane). In next level each cluster can be further divided into two parts with respect to it's mean value resulting into total four clusters per color plane to get feature vector of size twelve (four per plane). Thus BTC can be extended to multiple levels to get BTC Level 2, BTC Level 3, etc. The feature vector extraction for CBIR using multileveled BTC with RGB color space is explained in section A, B and C here.

A)   CBIR using BTC-RGB-Level-1 (BTC-RGB-6) [3,4,9,17]

In original BTC we divide the image into R, B, and G components and compute the mean value of each color component as individual color. Let $I(m,n)=[R(m,n), G(m,n), B(m,n)]$ be the color image image of size mxn. Let the thresholds be MR, MG and MB, which could be computed as per the equations 2, 3 and 4.

$$MR = \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}R(i,j) \tag{2}$$

$$MG = \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}G(i,j) \tag{3}$$

$$MB = \frac{1}{m*n}\sum_{i=1}^{m}\sum_{j=1}^{n}B(i,j) \tag{4}$$

Here three binary bitmaps will be computed as BMr, BMg and BMb. If a pixel in each component (R, G, and B) is greater than or equal to the respective threshold, the corresponding pixel position of the bitmap will have a value of '1' otherwise it will have a value of '0'.

$$BMr(i,j) = \begin{cases} 1, if....R(i,j) > MR \\ 0,...if...R(i,j) <= MR \end{cases} \tag{5}$$

$$BMg(i,j) = \begin{cases} 1, if....G(i,j) > MG \\ 0,....if...G(i,j) <= MG \end{cases} \tag{6}$$

$$BMb(i,j) = \begin{cases} 1, if....B(i,j) > MB \\ 0,....if...B(i,j) <= MB \end{cases} \tag{7}$$

$$UR = \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n}BMr(i,j)} * \sum_{i=1}^{m}\sum_{j=1}^{n}BMr(i,j)*R(i,j) \tag{8}$$

$$UG = \frac{1}{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} BMg(i,j)} * \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} BMg(i,j)*G(i,j) \qquad (9)$$

$$UB = \frac{1}{\sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} BMb(i,j)} * \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} BMb(i,j)*B(i,j) \qquad (10)$$

Two mean colors one for the pixels greater than or equal to the threshold and other for the pixels smaller than the threshold are also calculated [15]. The upper mean color UM(UR, UG, UB) is given as equations 8, 9 and 10.

And the Lower Mean LM= (LR, LG, LB) is computed as following equations 11, 12 and 13.

$$LR = \frac{1}{m*n - \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} BMr(i,j)} * \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}\{1 - BMr(i,j)\}*R(i,j) \qquad (11)$$

$$LG = \frac{1}{m*n - \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} BMg(i,j)} * \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}\{1 - BMg(i,j)\}*G(i,j) \qquad (12)$$

$$LB = \frac{1}{m*n - \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n} BMb(i,j)} * \sum\limits_{i=1}^{m}\sum\limits_{j=1}^{n}\{1 - BMb(i,j)\}*B(i,j) \qquad (13)$$

These Upper Mean and Lower Mean together will form a feature vector or signature of the image. For every image stored in the database these feature vectors are computed and stored in feature vector table.

B) CBIR using BTC- RGB- Level 2 (BTC-RGB-12)[1]

In BTC-RGB-Level 2 the image data is divided into 12 parts using the six means obtained in BTC-RGB-Level 1. Here the bitmap are prepared using upper and lower mean values of individual color components. For red color component the bitmap 'BMUR' and 'BMLR' are generated as given in equations 14 and 15. Similarly for green color component 'BMUG' & 'BMLR' and for blue colour components 'BMUB' & 'BMLB' can be generated.

$$BMUR(i,j) = \begin{cases} 1, if....R(i,j) > UR \\ \\ 0,...if...R(i,j) <= UR \end{cases} \qquad (14)$$

$$BMLR(i,j) = \begin{cases} 1, if....R(i,j) > LR \\ \\ 0,...if...R(i,j) <= LR \end{cases} \qquad (15)$$

Using this bitmap the two mean colour per bitmap one for the pixels greater than or equal to the threshold and other for the pixels smaller than the threshold are calculated [15]. The red plane of the image is divided into two parts as upper red image and lower red image as given by equations 16 and 17.

$$Iur(i,j) = R(i,j) * BMr(i,j),..for.i = 1...m, and.j = 1...n. \qquad (16)$$

$$Ilr(i,j) = R(i,j) *[1 - BMr(i,j)],..for.i = 1...m, and.j = 1...n. \qquad (17)$$

The upper mean color UM (UUR, ULR, UUG, ULG, UUB, ULB) are given as follows.

$$UUR = \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} BMUR(i,j)} * \sum_{i=1}^{m}\sum_{j=1}^{n} BMUR(i,j)*Iur(i,j) \tag{18}$$

$$ULR = \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} BMLR(i,j)} * \sum_{i=1}^{m}\sum_{j=1}^{n} BMLR(i,j)*Ilr(i,j) \tag{19}$$

And the first two components of Lower Mean LM= (LUR, LLR, LUG, LLG, LUB, LLB) are computed using following equations

$$LUR = \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} BMUR(i,j)} * \sum_{i=1}^{m}\sum_{j=1}^{n}\{1-BMUR(i,j)\}*Iur(i,j) \tag{20}$$

$$LLR = \frac{1}{\sum_{i=1}^{m}\sum_{j=1}^{n} BMLR(i,j)} * \sum_{i=1}^{m}\sum_{j=1}^{n}\{1-BMLR(i,j)\}*Ilr(i,j) \tag{21}$$

These Upper Mean and Lower Mean values together will form a feature vector for BTC-12. For every image stored in the database these feature vectors are computed and stored in feature vector table.

C) CBIR using BTC-RGB-Level 3 (BTC-RGB-24)[1] and BTC-RGB-Level 4 (BTC-RGB-48)

Similarly the feature vector for BTC-24 can be found by extending the BTC till level 3. Each plane will give the 8 elements of feature vector. For Red plane we get (UUUR, LUUR, ULUR, LLUR, UULR, LULR, ULLR, LLLR). Also feature vector of size 48 can be generated by extending this process of BTC to next level.

4. COLOR SPACES IN CBIR USING MULTILEVELED BTC

Just as discussed in section 3 for RGB color space, the CBIR using multileveled BTC can be used with other color spaces. Here in all ten color spaces like RGB, HSV[20], XYZ[20], rgb[20], HSI[20], Kekre's LUV [3], YCbCr[17], YUV[9], YIQ[20], Kekre's YCgCb[20] are considered. The ten color spaces along with BTC extended to four levels resul into total 40 CBIR methods. These color spaces can mainly be divided into two categories as luminance-chrominance color spaces (Kekre's LUV, YCbCr, YUV, YIQ, Kekre's YCgCb) and nom-luminance color spaces (RGB, HSI, HSV, XYZ and rgb).

5. IMPLEMENTATION

The implementation of these CBIR techniques is done using MATLAB 7.0. The CBIR techniques are tested on the augmented Wang [15] image database of 1000 variable size images spread across 11 categories of human beings, animals, natural scenery and man-made things. To compare various techniques in various color space, performance is evaluated based on precision and recall. The efficiency of CBIR technique is evaluated based on accuracy, stability and speed. To assess the retrieval effectiveness, we have used the precision and recall as statistical comparison parameters for the BTC-6, BTC-12, BTC-24 and BTC-48 techniques of CBIR on all color spaces. The standard definitions of these two measures are given by following equations.

$$Precesion = \frac{Number\_of\_relevant\_images\_retrieved}{Total\_number\_of\_images\_retrieved} \tag{22}$$

$$Recall = \frac{Number\_of\_relevant\_images\_retrieved}{Total\_number\_of\_relevent\_images\_in\_database}$$

(23)

6. RESULTS AND DISCUSSION

The methods BTC-6, BTC-12, BTC-24 and BTC-48 are applied to the image database of 1000 images using various color spaces. The resulting 40 CBIR methods are tested using 55 random queries (5 per image category). The average precision and recall values of all these queries per CBIR methods are computed and considered for performance comparison.

Figure 1 shows the average precision and recall values plotted against number of retrieved images for RGB-BTC-CBIR at level 1 (RGB-BTC-6), level 2 (RGB-BTC-12), level 3 (RGB-BTC-24) and level 4 (RGB-BTC-48). The conclusion that 'higher the level of BTC used better the performance of CBIR is' can be drawn from the given graph as indicated by higher precision and recall values in BTC 48 and BTC 24 as compared to BTC-12 and BTC-6.

However the distinction in the performance of all these techniques is not very clear. The height of crossover point of precision and recall curves plays very important role in performance comparison of CBIR methods. Ideally this crossover point height should be one. Higher the value of this crossover point better the performance is.

Figure 2a shows the zoomed version of graphs in figure 1 for crossover points of precision and recall curves. Figure 2 b shows the bar graphs indicating the heights of crossover points of precision-recall curves of respective CBIR methods shown in figure 2a. From this figure 2b it can be observed that the best performance (Highest precision-recall crossover point value) is given by BTC-RGB-48. The increase in performance from BTC-RGB-6 to BTC-RGB-24 is gradual in nature while the performances of BTC-RGB-24 and BTC-RGB48 have very slight difference. So the conclusion that multileveled BTC helps to improve the performance of CBIR methods up to BTC-Level 3 (BTC-24) can be drawn from these observations.

Figure 3 gives the comparison of crossover points of the multileveled BTC-CBIR techniques per color space. In all color spaces increasing the BTC level helps in improving the performance of CBIR as indicated by higher crossover point values.

Figure 4 gives the comparison of crossover points for all color spaces per level of BTC-CBIR techniques. In all levels of BTC-CBIR luminance-chrominance color spaces (YCgCb, Kekre's LUV, YUV, YIQ, YCbCr) better than non-luminance (RGB, HSI, HSV, rgb, XYZ) Color spaces and in all Kekre's LUV gives best performance.

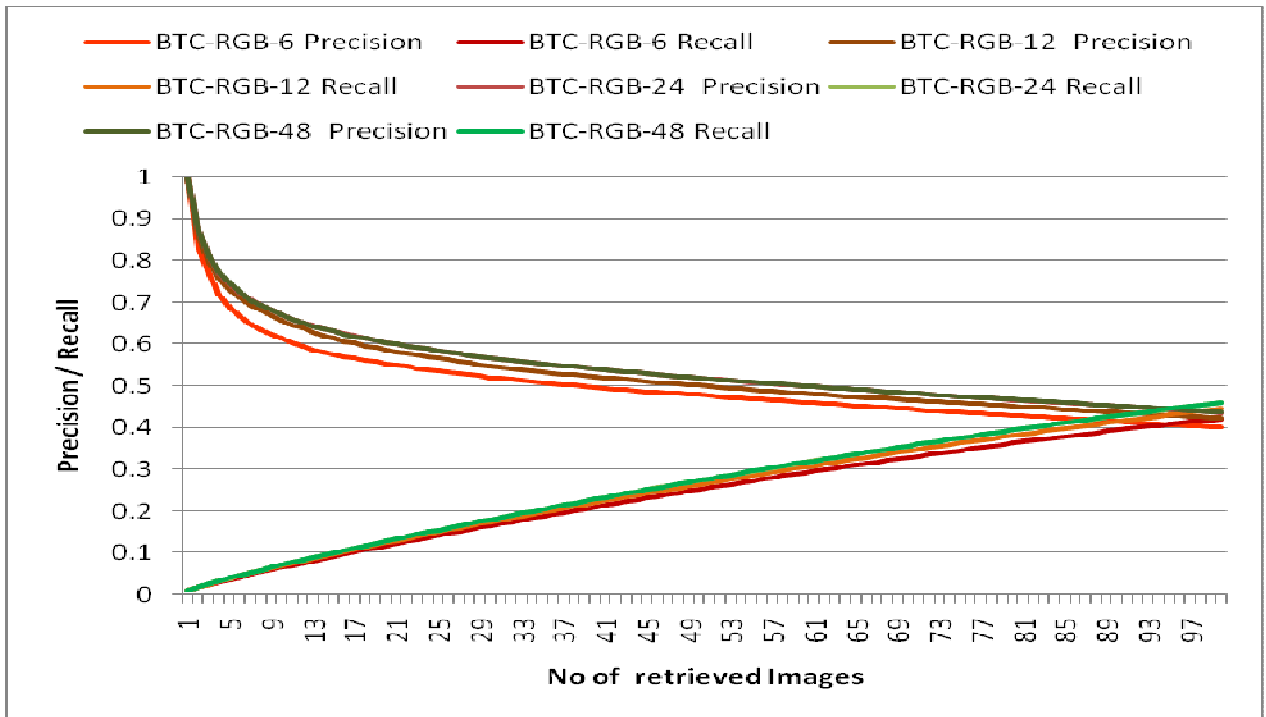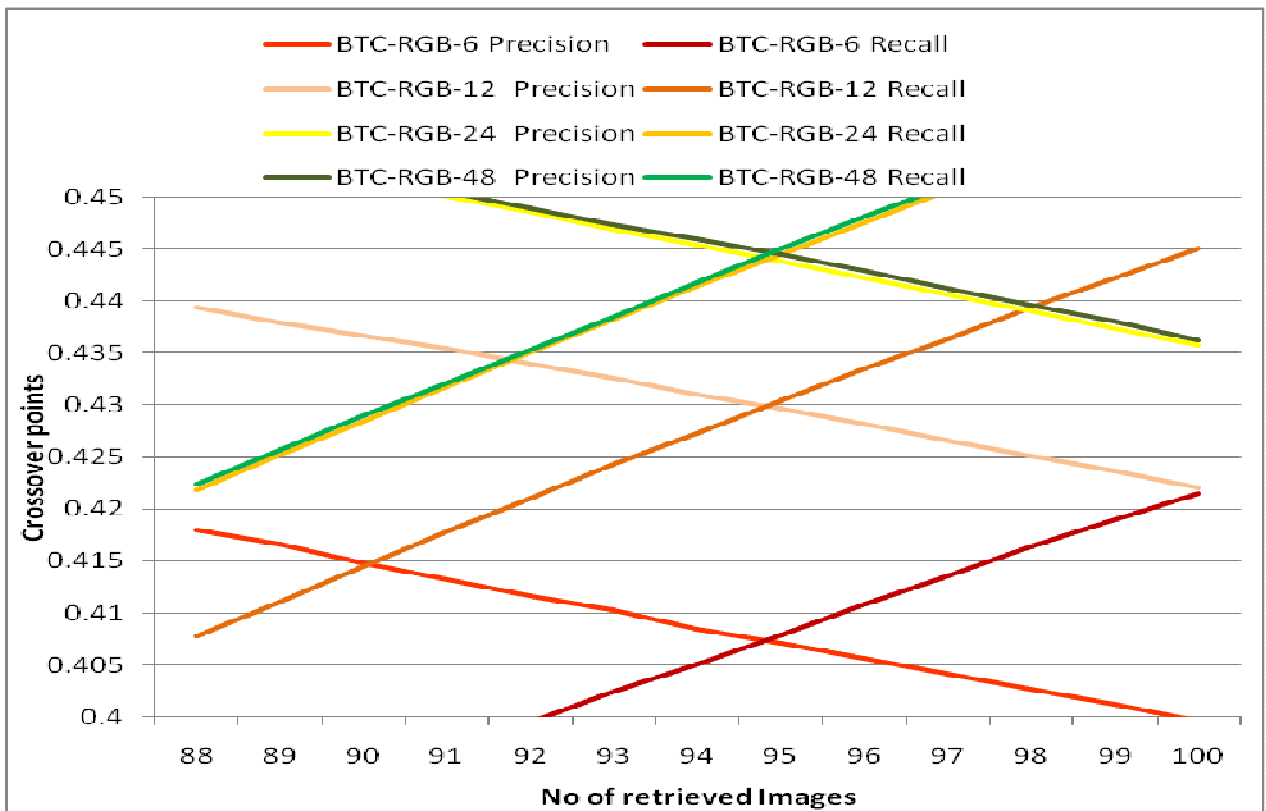**FIGURE 1 :** Precision & Recall plotted against for RGB color space.

FIGURE2.A: Crossover points of Precision-Recall plotted against number of retrieved images.
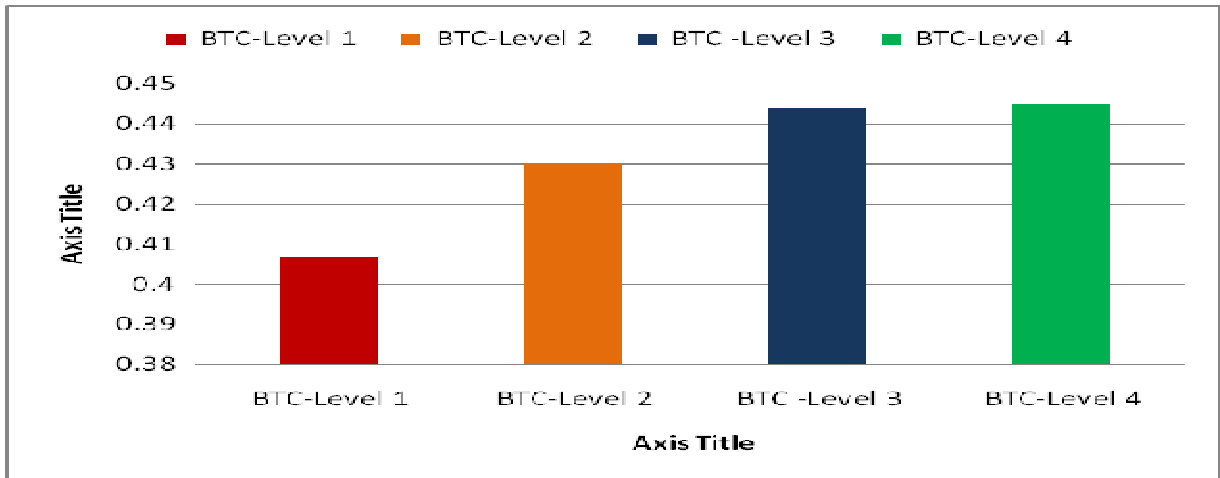
FIGURE 2.B : Performance comparison of discussed CBIR methods using Precision-Recall crossover points for RGB color sapce.
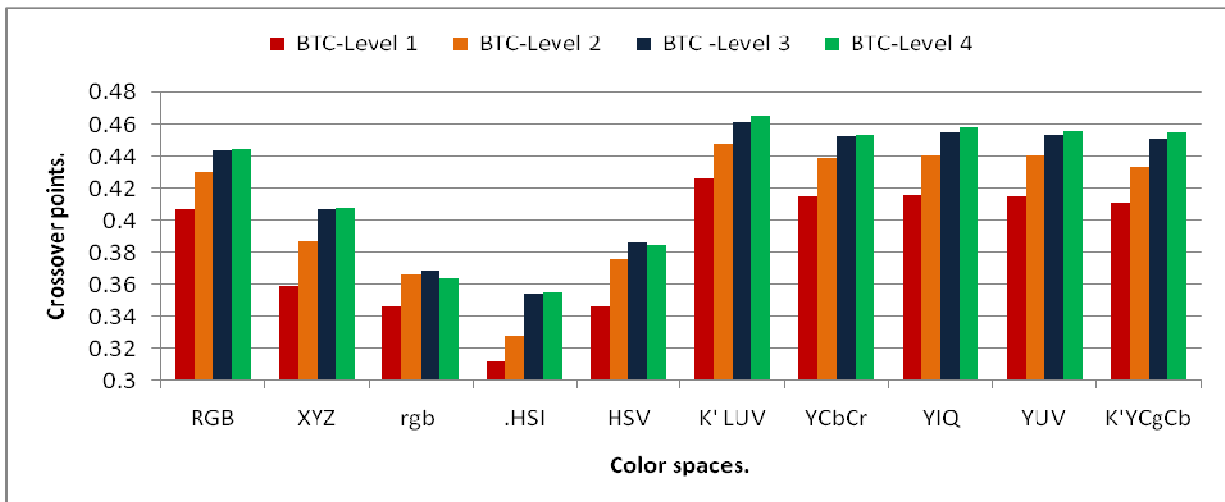

FIGURE 3: Performance Comparison between different levels of BTC for respective color spaces.
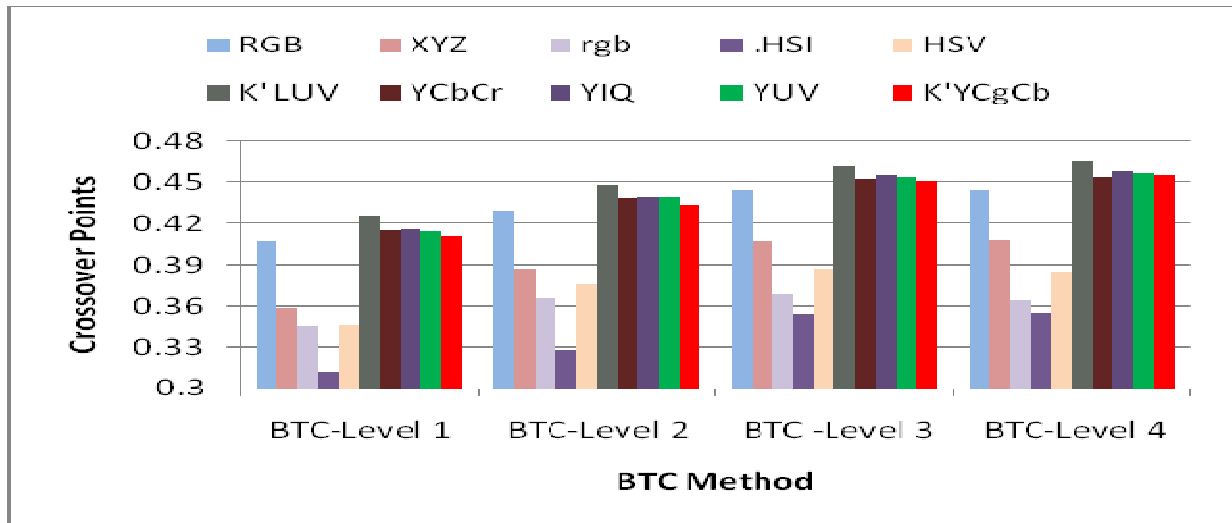
FIGURE 4: Performance Comparison between different color spaces for respective level of BTC.

7. CONCLUSION

The performance of CBIR system depends on the precision and recall. Quite often the crossover point of precision and recall is taken as criteria for judging the performance of CBIR technique on various Color spaces. Kekre's LUV gives better result and the values are 0.4258, 0.448, 0.462, 0.46495 for BTC-6, BTC-12, BTC-24 and BTC-48 respectively. The results have shown that, the performance improvement (i.e. higher precision and recall values) in Luminous & Chrominance Color spaces with BTC-CBIR methods compared to Non-Luminous &Non-Chrominance Color spaces are lower. Performance improves with increased level of BTC. Up to level -3 gradual increases in performance with increasing level is observed while the difference of performance in level-3 and level-4 is negligible due to voids being created at higher levels.

8. REFERENCES

[1] Dr.H.B.Kekre, Sudeep D. Thepade, Shrikant P. Sanas 'Improved CBIR using Multileveled Block Truncation Coding' International Journal of Computer Science and Engineering (IJCSE), Volume 2, Number 7, October 2010 Edition, pp2471-2476. http://www.enggjournals.com/ijcse.

[2] Guoping Qiu, "Color Image Indexing Using BTC," IEEE Transactions on   Image Processing, VOL.12,   NO.1, pp.93-101, January 2003.

[3] Dr.H.B.Kekre, Sudeep D. Thepade, "Boosting Block Truncation Coding using Kekre's LUV Color Space for Image Retrieval", WASET International Journal of Electrical, Computer and System Engineering (IJECSE), Volume 2, Number 3, pp. 172-180, Summer 2008. Available online at http://www.waset.org/ijecse/v2/v2-3-23.pdf

[4] Dr.H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Augmented Block Truncation Coding Techniques", ACM International Conference on Advances in Computing, Communication and Control (ICAC3-2009), pp. 384-390, 23-24 Jan 2009, Fr. Conceicao Rodrigous College of Engg., Mumbai. Is uploaded on online ACM portal.

[5] Dr.H.B.Kekre, Sudeep D. Thepade, "Scaling Invariant Fusion of Image Pieces in Panorama Making and Novel Image Blending Technique", International Journal on Imaging (IJI), www.ceser.res.in/iji.html, Volume 1, No. A08, pp. 31-46, Autumn 2008.

[6] B.G.Prasad, K.K. Biswas, and S. K.Gupta, "Region–based image retrieval using integrated color, shape, and location index", computer vision and image understanding, Oct 2003.

[7] Minh N. Do, Member, IEEE, and Martin Vetterli, Fellow, IEEE," Wavelet-Based Texture Retrieval Using Generalized Gaussian Density and Kullback-Leibler Distance," IEEE Transactions nn Image Processing, Vol.11, No.2, Feb 2002.

[8] Dr.H.B.Kekre, Sudeep D. Thepade, "Rendering Futuristic Image Retrieval System", National Conference on Enhancements in Computer, Communication and Information Technology, EC2IT-2009, 20-21 Mar 2009, K.J.Somaiya College of Engineering, Vidyavihar, Mumbai-77.

[9] Dr.H.B.Kekre, Sudeep D. Thepade, "Using YUV Color Space to Hoist the Performance of Block Truncation Coding for Image Retrieval", IEEE International Advanced Computing Conference 2009 (IACC'09), Thapar University, Patiala, INDIA, 6-7 March 2009.

[10] Dr.H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke,"Energy Compaction and Image Splitting for Image Retrieval using Kekre Transform over Row and Column Feature Vectors", International Journal of Computer Science and Network Security (IJCSNS),Volume:10, Number 1, January 2010, (ISSN: 1738-7906) Available at www.IJCSNS.org.

[11] Dr.H.B.Kekre, Sudeep D. Thepade, Archana Athawale, Anant Shah, Prathmesh Verlekar, Suraj Shirke, "Walsh Transform over Row Mean and Column Mean using Image Fragmentation and Energy Compaction for Image Retrieval", International Journal on Computer Science and Engineering (IJCSE),Volume 2S, Issue1, January 2010, (**ISSN: 0975–3397). Available online at www.enggjournals.com/ijcse.**

[12] Dr.H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features Extracted from Walshlet Pyramid", ICGST International Journal on Graphics, Vision and Image Processing (GVIP), Volume 10, Issue I, Feb.2010, pp.9-18, Available online www.icgst.com/gvip/Volume10/Issue1/P1150938876.html

[13] Khalid Sayood ," Introduction to Data Compression ," University of Nebraska-Lincoln , Second Edition , ISBN:1-55860-558-4, by Academic Press,2000.

[14] Stian Edvardsen,"Classification of Images using color, CBIR Distance Measures and Genetic Programming, "Ph.D. Thesis , Master of science in Informatics, Norwegian university of science and Technology, Department of computer and Information science, June 2006.

[15] Rafael C.Gonzalez, Richard E. Woods," Digital Image Processing," University of Tennessee, Second Edition, ISBN 81-7808-629-8, Pearson Education Pvt. Ltd.,2002.

[16] http://wang.ist.psu.edu/docs/related/Image.orig (Last referred on 23 Sept 2008)

[17] Dr.H.B.Kekre, Sudeep D. Thepade, "Color Based Image Retrieval using Amendment Block Truncation Coding with YCbCr Color Space", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009, pp. 2-14. Available online at **www.ceser.res.in/iji.html (ISSN: 0974-0627).**

[18] Dr.H.B.Kekre, Tanuja Sarode, Sudeep D. Thepade, "Color-Texture Feature based Image Retrieval using DCT applied on Kekre's Median Codebook", International Journal on Imaging (IJI), Volume 2, Number A09, Autumn 2009,pp. 55-65. Available online at **www.ceser.res.in/iji.html (ISSN: 0974-0627).**

[19] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the QBIC system. *IEEE Computer*, 28(9):23–32, 1995.

[20] Dr.H.B.Kekre, Sudeep D. Thepade, A. Athawale, Adib Parkar, "Using Assorted Color Spaces and Pixel Window Sizes for Colorization of Grayscale Images", ACM-International Conference and Workshop on Emerging Trends in Technology (ICWET 2010), Thakur College of Engg. And Tech., Mumbai, 26-27 Feb 2010, uploaded on online ACM Portal.

[21] Dr.H.B.Kekre, Sudeep D. Thepade, Shobhit W., Miti K., Styajit S., Priyadarshini M. "Image Retrieval with Shape Features Extracted using Gradient Operators and Slope Magnitude Technique with BTC", International Journal of Computer Applications (IJCA), Volume 6, Number 8, pp.28-33, September 2010. **Available online at** http://www.ijcaonline.org/volume6/number8/pxc3871430.pdf

A Hybrid Trademark Retrieval System Using Four-Gray-Level Zernike Moments and Image Compactness Indices

Innchyn Her her@mail.nsysu.edu.tw
Professor, Dept. of Electro-mechanical Engineering
National Sun Yat-sen University
Kaohsiung 80424, Taiwan

Kazi Mostafa kazimostafa@gmail.com
Student, Dept. of Electro-mechanical Engineering
National Sun Yat-sen University
Kaohsiung 80424, Taiwan

Huan-Kai Hung hhung23@gmail.com
Student, Dept. of Electro-mechanical Engineering
National Sun Yat-sen University
Kaohsiung 80424, Taiwan

Abstract

There are just too many trademarks out there so that a good automated retrieval system is required to help to protect them from possible infringements. However, it is from people, i.e., the general consumers' viewpoint how similar or confusing two trademarks can be. Thus, in this paper we propose a hybrid system where we patently incorporate human inputs into a computerized trademark retrieval scheme. Various surveys involving general consumers' cognition and responses are conducted, and the results are used as benchmarks in developing the automated part. The core mathematical features used in the scheme are four-gray-level Zernike moments and two new image compactness indices. Experimental results show that this hybrid system, when compared with human-generated results, is able to achieve an average accuracy rate of 95% while that of the closest competing existing method is 65%.

Keywords: Image matching, Information retrieval, Trademarks.

1. INTRODUCTION

Owing to the constant development of commercial activities the number of trademarks used by companies increases dramatically by the year. Consequently, to design a new trademark without infringing others has become a critical and complex issue. With the advent of computers, traditional ways of archiving many documents have been replaced by computerized methods, which feature automated processes and fast and accurate information retrieval. When it comes to trademarks, unfortunately, there is no universally accepted sorting and retrieval scheme so far. For instance, through keywords or *search codes* that are manually assigned to the trademark images, an official retrieval system has been used for some time [1][22]. This procedure, though feasible, contains major drawbacks. It is very difficult for an operator to describe all images evenly objectively and consistently, especially for abstract or complex ones and over a long period of

time. On the other hand, a different approach called Content-Based Image Retrieval (CBIR) proposes an image-based strategy instead of a tag-based one. Each trademark is processed beforehand, to leave and save its pertinent features in the computer. Then for a new trademark application, the features of the new image are compared with the stored data to see if there have been similar entries in the database. Such CBIR trademark retrieval systems are intrinsically less susceptible to human errors. A good CBIR scheme tends to exhibit robustness in terms of image size, position, and orientation variations. In general, these methods are often classified into contour-based schemes and region-based schemes.

A very early effort to digitally define the features of an image was the work by Freeman [2]. He invented the chain code, a typical contour-based approach. Many researchers later refined his method to improve the applicability. For example, Peng and Chen [3] more recently proposed to use the chain code to describe simple and properly segmented trademarks, while encoding them according to the angles. Contour-based methods are usually for simple, binary images. An advantage is that the coded contour is usually invariant with respect to translation and rotation [4]. Yet the newer region-based methods have a broader application range and a larger toolbox to use. For example, there are moments of various kinds, often used as descriptive features of an image. A benefit of using the moments is their insensitivity to noise. As a result, researchers have begun to resort more to region-based techniques in these years. For instance, Yin and Yeh [5] have proposed a method to employ a fuzzy approach to expedite the image classifying and retrieval processes. The image features used were the area, the number of closed objects, the location of the centroid, and symmetry of the image, etc.

The invariant moments, devised by Hu [6] based on geometric moments, have long been valued as useful image descriptive features. But they are comparatively sensitive to noise and the accompanying image reconstruction procedure was difficult. Shih and Chen [7] later suggested to use invariant moments with Fourier transforms and the histogram of image boundary orientations as features in a trademark retrieval scheme. Ciocca and Schettini [8] also demonstrated a way for smartly using the invariant moments. They combined moments with boundary directions and the results from a multi-resolution wavelet transform. Through a relevance feedback scheme they could compute the similarity between trademarks. On the other hand, there were researchers who had chosen to use Zernike moments for trademark retrieval schemes. For example, Kim and Kim [9] constructed a Gamma distribution model with Zernike moments to describe visually salient features of monochrome trademarks. Subsequently, Kim et al. [10] also presented a modified Zernike-moment shape descriptor, by first partitioning a trademark image into an inner and an outer region. More recently, Kamila et al. [11] normalized the regular Zernike moments via geometric moments and used them with binary images.

Regardless of these many research efforts on automated trademark retrieval methodology, the governing institution that supervises the examination and approval of trademarks still, for the time being, relies largely on manual and tag-based methods [1], [12], [13], [14]. From scientific point of view, a trademark is invariably an image, a discrete two-dimensional mathematical function. Using *search codes* to describe and to retrieve an image risks the danger of inaccuracy, inconsistency, and inefficiency. But why do people still stick to them? One blunt answer: Undeniably, trademark is not an engineering or scientific term. It is purely commercial; trademarks are meant to interact with people. They contain subtle feelings difficult to perceive or describe by the computer. For the task of trademark comparison and retrieval, human inputs can never be dismissed.

In light of this we now propose a hybrid retrieval system for trademarks. In the core of our numerical scheme, we use Zernike moments for their resistance to noise, their invariance to rotation, and multi-resolution capabilities. Besides, two new image features are proposed as additional sorting criteria for the computer. However, the term "hybrid" implies that human perception works hand in hand with computer's speed. In this system humans set up a benchmark while the computer emulates humans' cognition and feeds back a list of candidates

for human to make final judgments. We will show that the hybrid system is robust and able to provide an output very usable for a professional trademark examiner.

2. PRELIMINARIES

2.1 Trademarks

A trademark is a mark that identifies one person's goods [13]. In practice, the word "trademark" is used to refer to any class of mark, or a word, symbol, or phrase, used to distinguish a particular manufacturer's product from those of others. The enactment of the trademark law aims to protect the legal use of the trademarks, so that legitimate transactions and fair business morals are guaranteed. It is obvious that there is little room for confusion in the use of trademarks. However, if there is any, in particular deliberate confusion, an act of infringement is very likely evolving somewhere.

Thanks to the collaborated efforts of industrialized nations and the WTO, trademark definition and laws around the world are quite similar now. According to a typical trademark regulation [14], there are four different kinds of trademarks, as depicted in Fig. 1. In this paper, we will deal mostly with graph and sign trademarks for they are more difficult to describe in words.

There are a few rules in determining the similarity of trademarks. The most important aspect is the impression the whole trademark imposes on the consumers. It is by this crude total impression that people decide whether two trademarks are confusingly similar or not. As confirmed by a senior officer of local Patent and Trademark Office [15], people are more impressed by the rough shape of a trademark, rather than by its details or color. Moreover, people tend to neglect the frame or circle outside the trademark. For computers, a framed trademark can be a far cry from an unframed one. For people, strangely, they produce almost identical impressions.
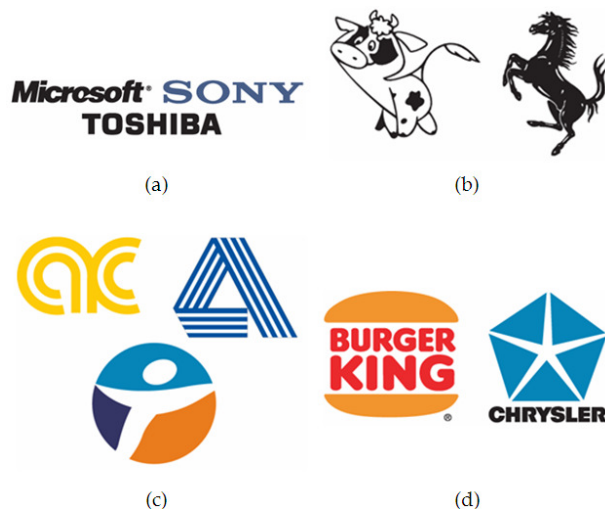
FIGURE 1: Different kinds of trademarks: (a) Word trademarks, (b) Graph trademarks, (c) Sign trademarks, and (d) Combined trademarks.

To verify the above statements, we have conducted a questionnaire survey that involved 100 general consumers [16]. When presented with a single trademark design, 59 consumers said that the color was the most impressing feature, versus 34 for the shape and 5 for its pronunciation. However, when presented with sets of similar trademarks for comparison, a total of 83 consumers confirmed that the rough shapes of the trademarks, rather than their colors or details, were the most discerning feature for trademark identification. In view of this double testimony, we decided to devise two new features solely for the rough shape of the trademark image. They will be discussed in Section 3.2.

2.2 Moments

Moments are handy properties that are repeatedly used in region-based image retrieval systems, while moments in various orders are regarded as descriptive features of an image. In this paper we will use and compare two types of moments, and their formulations are briefly discussed as follows.

The invariant moments of Hu were proposed in 1962 [6]. In this paper author anticipated a theory of two-dimensional moment invariants for planar geometric figures. In order to make the moments invariant to scaling, translation, and rotation, they need to be normalized. Based on the normalized moments, Hu proposed the seven invariants. Hu's invariants can be used as image descriptive features. The computation of image reconstruction from these invariants, however, is no piece of cake. Besides, they can still be affected by the image noise [8], [11], [17].

On the other hand, the Zernike moments, proposed by Teaque in 1980 [18], are constructed on orthogonal Zernike polynomials. These complex number Zernike functions were invented by the Nobel laureate in 1961, defined within a unit circle. A typical Zernike polynomial is expressed as,

$$V_{nm}(x,y) = V_{nm}(\rho,\theta) = R_{nm}(\rho)e^{jm\theta} \tag{1}$$

(x, y) represents an arbitrary point in the unit circle, and within the circle, we have $x^2 + y^2 \leq 1$. The integer n, starting from zero, is the order of the polynomial. m is also an integer that satisfies $|m| \leq n$, and $n - |m|$ must be an even number. ρ is the magnitude of vector (x, y), and θ is the orientation, measuring counterclockwise from the x-axis. j is the unit imaginary number $\sqrt{-1}$. Now the radial component R_{nm} (ρ) of Zernike polynomial is given by:

$$R_{nm}(\rho) = \sum_{s=0}^{\frac{n-|m|}{2}} (-1)^s \frac{(n-s)!}{s!\left(\frac{n+|m|}{2}-s\right)!\left(\frac{n-|m|}{2}-s\right)!} \rho^{n-2s} \tag{2}$$

According to the above expression, we know $R_{n,-m}(\rho) = R_{nm}(\rho)$. Moreover, the complete orthogonal property of Zernike polynomial V(x,y) within the unit circle can be demonstrated by this equation:

$$\iint_{x^2+y^2\leq 1} V_{nm}^*(x,y) V_{pq}(x,y) \, dxdy = \frac{\pi}{n+1}\delta_{np}\delta_{mq}$$

$$\text{and } \delta_{ab} = \begin{cases} 1 & (a=b) \\ 0 & (\text{otherwise}) \end{cases} \tag{3}$$

Where * denotes the conjugate of a complex number.
Zernike moments are the projection of the image function f(x, y) onto the above Zernike orthogonal basis functions [11]. We denote Zernike moment of order n for a continuous function as A_{nm}, and

$$A_{nm} = \frac{n+1}{\pi} \iint_{x^2+y^2\leq 1} f(x,y) V_{nm}^*(\rho,\theta)dxdy \tag{4}$$

For a digital image, this equation is simplified as:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x,y) V_{nm}^*(\rho,\theta), \quad x^2 + y^2 \leq 1 \tag{5}$$

Note that the conjugate of A_{nm} is identical to $A_{n,-m}$, and the magnitude (denoted by Z_{nm}) of Zernike moment also has the following property:

$$Z_{nm} = \|A_{nm}\| = \|A_{n,-m}\| \tag{6}$$

In a typical image analysis, an image is decomposed via above equations into a series of Zernike moments A_{nm} of various orders. Conversely, by properly assembling these moments with

corresponding Zernike polynomials V_{nm}, the original image function $f(x, y)$ can also be reconstructed [16].

$$f(x,y) = f(\rho,\theta) \cong \sum_{n=0}^{N} \sum_{m} A_{nm}(\rho,\theta) V_{nm}(\rho,\theta) \tag{7}$$

Now let us assume that an image $f(\rho, \theta)$, the polar form of $f(x, y)$, has undergone a rotation about the origin by an angle α. The rotated image is then $f^r(\rho,\theta) = f(\rho,\theta-\alpha)$. Rewrite the previous Zernike moment equation in polar form, we have

$$A_{nm} = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho,\theta) R_{nm}(\rho) e^{-jm\theta} \rho \, d\rho \, d\theta \tag{8}$$

But the Zernike moment of the rotated image $f^r(\rho,\theta)$ is

$$A_{nm}^r = \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho,\theta-\alpha) R_{nm}(\rho) e^{-jm\theta} \rho \, d\rho \, d\theta \tag{9}$$

Substituting $\theta = \tilde{\theta} + \alpha$ into above equation yields

$$A_{nm}^r = \left[\frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho,\tilde{\theta}) R_{nm}(\rho) e^{-jm\tilde{\theta}} \rho \, d\rho \, d\tilde{\theta} \right] e^{-jm\alpha} = A_{nm} e^{-jm\alpha} \tag{10}$$

This expression leads to the equivalence of the magnitude:

$$Z_{nm}^r = \|A_{nm}^r\| = \|A_{nm}\| = Z_{nm} \tag{11}$$

Which means invariance of rotation within a unit circle.

3. METHODOLOGY

3.1 Smallest Enclosing Circle

Zernike moments are to be used as a kernel feature in our scheme, and it has an intrinsic property of rotational invariance. However, in order to achieve scaling and translational invariance, additional arrangements are needed. Since Zernike polynomials are defined within the unit circle, a method for conveniently finding the smallest enclosing circle for a trademark is necessary. This will take care of the translation concern. Then, the whole thing will be scaled to a fixed size for use with Zernike moments and other retrieval computations.

There exist many schemes for finding the smallest enclosing circle for different applications. For instance, Berg et al. [19] included in their book a recursive formula as such: For a set $P = \{p_1, p_2, \ldots p_i, \ldots, p_n\}$ of n points in a plane, if D_i is the smallest circle enclosing, $P_i = \{p_1, p_2, \ldots p_i\}$, the following rules hold.
(a) If $p_i \in D_{i-1}$, then $D_i = D_{i-1}$
(b) If $p_i \notin D_{i-1}$, then p_i is on the boundary of D_i. So starting from D_1, eventually we get D_n.

The point-by-point method is simple in formulation (as shown in Figure 2), however computationally intensive in practice, especially for digital images that normally contain hundreds of points (even on the boundary). In this paper, we use a simplified version, based upon the above scheme, for locating the smallest enclosing circle for a trademark image. Firstly, for a given trademark, extract digitally its contour, which should contain all boundary points of the image. Then, search among the boundary points and find the two points that are most distantly apart, say, points a and b in Fig. 2(a).
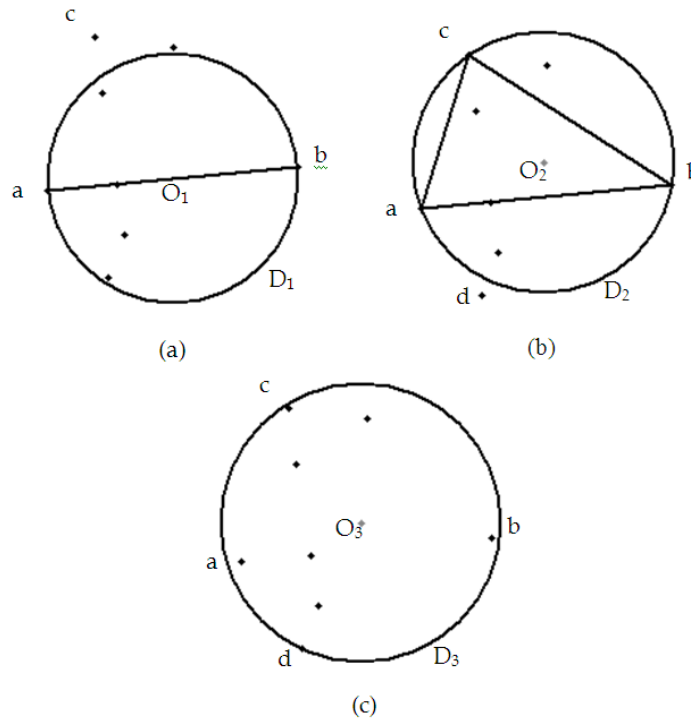
FIGURE 2: A simplified scheme for locating the smallest enclosing circle.

Construct circle D_1 so that line $\overline{ab}$ is a diameter of the circle. Next, if there are some points outside circle D_1, locate among them the point c, which is farthest from the center O_1 of D_1. Construct circle D_2 by points a, b, and c, as shown in Fig. 2(b). Lastly, if there still are some remaining points outside D_2, there is a final modification to make. Find point d that is most distant from O_2 (the center of D_2). Let $\overline{O_2d} + \overline{O_2c}$ be the length of the new diameter, and the new center be adjusted to O_3, which is on line $\overline{O_2d}$. Construct D_3 as illustrated in Fig. 2(c), which is the smallest enclosing circle used in this paper. Within four steps, we swiftly determine a usable smallest enclosing circle for a trademark image. Although the above method was heuristic, the enclosing circle it finds is well defined and suffices for our purpose. In our experience, we have not yet encountered a trademark that has a problem with this simplified scheme. When the enclosing circle is found, a transformation matrix can easily be formulated [20] to translate the trademark to the origin and to scale the image to the prescribed dimension. This dimension was 64×64 in most of our experiments.

3.2   Wrap Contour and Compactness Indices

Besides Zernike moments, which are a region-based method, we also require in this paper some contour-based features that can represent the gross shape of the trademark. We now propose a wrap contour concept, which we think is more describing than the common convex hull concept for two-dimensional objects (Fig. 3). A convex hull, having to maintain a convex shape, often has only loose contact with the object, and therefore poorly represents the contour of the object (Fig. 3b). The wrap contour (Fig. 3c), on the contrary, is allowed to shrink snugly inward the object. Thus, in the end it better represents the approximate shape of the object.

Technically, the smallest enclosing circle of the trademark is first to be located. Then, from the center of the enclosing circle, a radial line is drawn for every possible angle θ (according to image resolution) to intersect the outer contour of the trademark (Fig. 3a). The intersections are recorded as a function r(θ), where r is the distance from the center of the circle. If there is no intersection for some θ, then assign r=0 for that θ. Finally, let the wrap contour r(θ) become a closed and continuous curve in the polar plane, as illustrated in Fig. 3c.
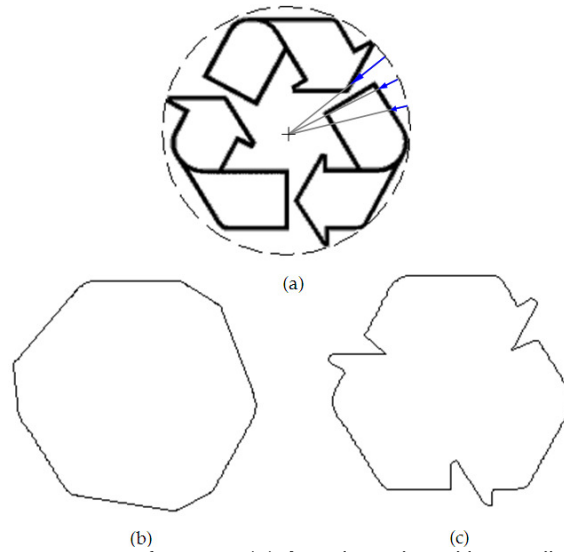
FIGURE 3: Using contours as features. (a) A trademark and its smallest enclosing circle, (b) The convex hull for that trademark, (c) The wrap contour for that trademark.

The wrap contour of a trademark is used in our scheme to help us define two new features called image compactness indices, CI_1 and CI_2. They can be regarded as mixed contour-based and region-based features, and are defined as:

$$\text{Compactness index 1 } (CI_1) = \frac{\text{Area of Wrap Contour}}{\text{Area of Smallest Enclosing Circle}} \tag{11}$$

$$\text{Compactness index 2 } (CI_2) = \frac{\text{Image Area}}{\text{Area of Wrap Contour}} \tag{12}$$

Note that in the above definition of CI_2, the "Image Area" is the foreground area of the trademark. However this area might change when we alter the color depth of the image. We will discuss this in the next section.

3.3 Four-Level Images

Generally, trademark designs are not limited to binary images. In fact, most modern trademarks use vivid colors to be visually conspicuous and attractive to consumers. On the other hand, a trademark normally uses only a scant palette of colors. This is because a trademark is not like a photograph. For trademarks, the simpler the design is, the quicker people tend to remember.

To compute the moments, a color image has to be transformed into a gray-scale image function, and traditionally a scale of 255 gray levels is used for picture-like images. For simple images as trademarks, some researchers suggested using only a two-level (binary) scale [11]. We have found, however, transforming trademarks into binary images sometimes incurs significant losses in the features. Fig. 4 shows examples of such losses in two trademark images, even though the thresholds used were computed by a sophisticated automated scheme [16]. As a compromise, we propose using four gray levels that correspond to levels 0, 85, 170, and 255, in a 0 to 255 grayscale. A trademark image is then transformed, with evenly spaced threshold values, to an image that contains only these four gray levels. By thus using more than two gray values, the loss of image features is effectively reduced. Yet we also take into account the fact that trademarks are less chromatically complex. Our experiments showed that by this scheme the total impression of a trademark image is satisfactorily preserved. Detailed examples of producing Zernike moments and about reconstructing such four-level trademark images from moments can be found in [19].
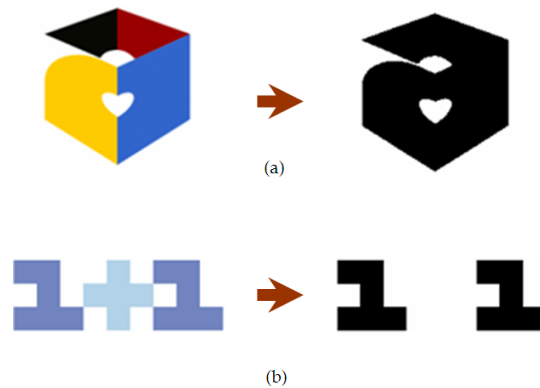
(a)

(b)

FIGURE 4: Losing boundaries from using too few gray levels. (a) Some internal boundaries are missing, (b) The plus sign is gone.

3.4   Exclusive Trademark Features

The feature values of an image are collectively used to identify the image. Thus, these features must be appropriately extracted. For trademarks, there are additional yet unique concerns. First, a trademark can be registered with a principal pattern and that pattern plus a variety of frames, e.g., circular, square, or polygonal shapes. Framed versions are regarded as the same as the original in trademark evaluation. Second, a chromatically altered image of a trademark pattern is also considered equivalent to the original one. For example, a dark-colored trademark on a light background is the same as the inversed, light trademark pattern on a dark background.

In this paper we highlight the quest of emulating the real trademark evaluation process. Hence, in extracting the features of one trademark we may need to consider not only the original design, but also its inversed image, and its core image. The core image refers to the central portion of the original trademark that lies within a smaller concentric circle whose radius is 2/3 of the enclosing circle. This core image is used to exclude the trademark frame and thus reduce the chances of misinterpretation. It is also noted that in real trademark retrieval the most important task is to detect confusingly similar designs, rather than to recover exactly like ones. Care has been taken in using features, so as to make our retrieval system neither under-restraining nor over-restraining. In this paper, we use Zernike moments of orders 0 to 19 as main region-based features for trademarks. The lower-order (0 to 12) moments represent the rough content of the trademark, whereas the higher-order moments delineate the minutia in the trademark image.

3.5   Weighting and Normalization

According to the trademark law, the degree of similarity between trademarks is the seriousness of confusion they cause for consumers. Obviously, this is a subjective judgment, and which may vary among different designs. Thus, we employ for our hybrid system a human-machine interface that allows users to feedback and to adjust the weighting of any feature type. Weighting means to prescribe significance for each feature group, which is among, say, m distinct image feature categories.

Combining the values from different feature categories requires a work called normalization. This process is to ensure that all image features possess influences of similar order. Now let P be an input trademark image, to be compared with the trademark database, which comprises n image entries Q_i, i=1 to n. Regarding the jth feature category, which contains k values, the similarity between images P and Q_i is then defined as a Euclidean distance on the coordinate hyperspace [21], given by

$$\text{dist}\left(P^j, Q_i^j\right) = \left[\sum_{s=1}^{k}\left({}_sp^j - {}_sq_i^j\right)^2\right]^{1/2} \tag{13}$$

where $_sp^j$ and $_sq_i^j$ are corresponding feature values of P and Q_i. Now for the jth feature category, we scan through the entire database for the two extreme values:

$$Min^j = \min_{i \le n}\left\{dist\left(P^j, Q_i^j\right)\right\}$$
$$Max^j = \max_{i \le n}\left\{dist\left(P^j, Q_i^j\right)\right\}$$

(14)

The normalized Euclidean distance, meaning the normalized similarity between images P and Q_i regarding the jth feature category, is then

$$Dist\left(P^j, Q_i^j\right) = \frac{dist\left(P^j, Q_i^j\right) - Min^j}{Max^j - Min^j}$$

(15)

Adding the consideration of the weighting factor w^j for the jth of a total of m feature categories, the gross similarity between images P and Q_i is then:

$$S_G\left(P, Q_i\right) = \sum_{j=1}^{m} w^j \cdot Dist\left(P^j, Q_i^j\right)$$

(16)

4. EXPERIMENTS & RESULTS

In this paper, 2020 trademark images were collected from various sources and compiled to form the experimental database. These images had been of different sizes, colors, and complexity. To begin with, since we wanted to emphasize human perspectives, we recruited ten adult volunteers, aged from 20 to 30, three among them female, and asked them to pick from the sample database some groups of similar images visible to their naked eyes. Listed in Table 1 are the five groups of similar trademarks hand-picked by the volunteers. These human culled images were to be used as benchmarks as we tested our hybrid scheme.
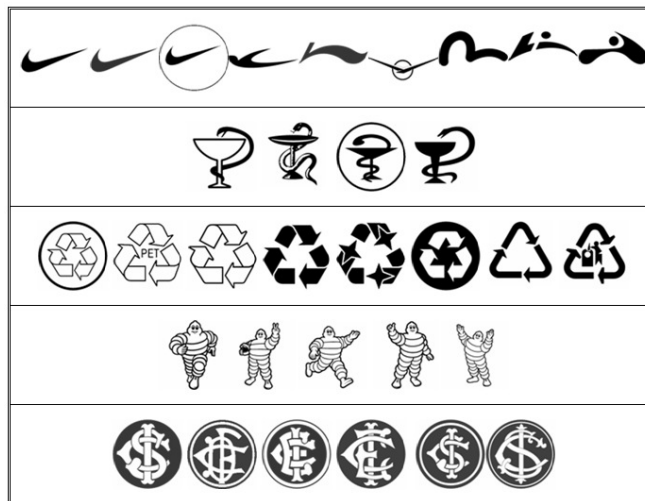
Table 1: Human-Picked Groups of Similar Trademark Patterns

Note that each group in Table 1 actually stands for a certain aspect of trademark identification. The images in the first group, "Nike," all contain an abstract notion of a checkmark. In the second group, "Goblet," either the snake or the goblet can deform its shape or change its color. The third group, "Recycle," is an example of inversed and framed trademark images. The fourth group involves various postures of the "Michelin" Man. These figures are line drawings containing no solid areas. The last group is "Monogram." Monograms are commonly used as trademarks in western countries and are also common victims of infringements.

As for the specific image features used, in each of our experiments we chose either to use only lower-order Zernike moments or to incorporate higher-order moments as well, according to the character of the input image. A collection of eight weighting factors $w^j = \{w^1, w^2, \ldots w^8\}$ was used in the experiments. Their meanings are described as follows. w^1: compare the whole input image with whole images from the database. w^2: compare core images only. w^3: compare core of the input image with the whole images from the database. This is to remove the frame of the input image. w^4: compare the inversed core of the input with the whole database images. w^5: compare CI_1's of whole images. w^6: compare CI_2's of whole images. Lastly, w^7 and w^8: compare CI_1's and CI_2's of the core images, respectively.

We first took a "Monogram" trademark from Table 1 as the input image, and inspected exclusively the effects of using four-gray-level Zernike moments. Shown in Fig. 5 are the results from the conventional binary scheme (Fig. 5a) and from our alternative four-level scheme (Fig. 5b). The input image is shown on the left. Here, all Zernike moments of orders 0 to 19 for the whole images are used as features ($w^j = \{1,0,0,0,0,0,0,0\}$). For each scheme we list, out of 2,020 database images, 60 images closest to the input. The order is from left to right and from top to down. Since our trademark retrieval system was to provide a list of ranked similar entries for human examiners to inspect, a better scheme was the one that produced a better candidate list. We see that Fig. 5b gathers more monogram-like images in its top rows, and it contains fewer samples that are too different to human's eye.

We also tested the effects of the proposed wrap contour concept and the related compactness index feature. We used an outlined "Nike" image as the input. Although to human's eye an outlined trademark gives similar rough impression as the solid one, we found that it literally confounded all the region-based methods. Compared in Fig. 6 are results from Hu's moment invariants (Fig. 6a), conventional binary Zernike moments (Fig. 6b), Kim's improved Zernike moments (Fig. 6c), and our first compactness index CI_1 (Fig. 6d, and $w^j = \{0,0,0,0,1,0,0,0\}$). In this experiment, only the compactness index scheme is capable of yielding meaningful outputs.

(a)

(b)

FIGURE 5: Using different gray-levels with Zernike moments. (a) The top output list from the two-level scheme, (b) The top output list from the four-level scheme.

(a)

(b)

(c)

(d)

FIGURE 6: Search results for the outlined Nike. (a) Hu's moment invariants, (b) Traditional Zernike moments, (c) Kim's modified Zernike, (d) Our compactness index CI_1.

Then we tried using more image features to improve the search efficiency. Again in Fig. 7 we put Hu's method (Fig. 7a), the original Zernike (Fig. 7b), Kim's modified Zernike (Fig. 7c) and our scheme in comparison. A regular "Nike" logo is used as the input. Due to its simple construction, only lower-order moments are required. With the above-mentioned w^j notation, Fig. 7d shows results of the vector {1,4,0,0,0,0,0,0}, denoting a double four-level Zernike search stressing the importance of the core image. Fig. 7e corresponds to the weighting vector {0,0,0,0,1,0,4,1},

showing the effects of combining the two compactness indices, also emphasizing the core. The weightings of Fig. 7f are {1,4,0,0,1,0,4,1}, equivalent to the sum of the weightings of Fig. 7d and Fig. 7e. From these figures we see that while Fig. 7d already produces a better listing than Figs. 7a to 7c, Fig. 7f produces the best result of all. All but one trademark in the "Nike" group in Table 1 are listed in the top 20 (top 1%, out of 2,020) candidates of Fig. 7f.

Similar satisfactory results of the proposed scheme were obtained for a monogram trademark. Again, Figs. 8a to 8c are results from the three existing schemes, and Figs. 8d to 8f are results from three variations of the proposed scheme. The weighting vectors for Figs. 8d to 8f are {5,5,0,0,0,0,0,0}, {0,0,0,0,10,1,0,0}, and {5,5,0,0,10,1,0,0}, respectively. For monogram images we do not need to emphasize the core. Note that different image type may require different weighting factor assortments. In the previous two examples we have not used w^3 and w^4. They are by no means less useful. Many additional examples involving framed or inversed trademarks or even various kinds of noise can be found in [16].

FIGURE 7: Comparison between schemes. (a) Hu's moment invariants, (b) The original Zernike, (c) Kim's modified Zernike, (d) Our double four-level Zernike, (e) Our Compactness indices, (f) Combining double Zernike and Compactness indices.

(d)

(e)

(f)

FIGURE 7: Comparison between schemes. (*Continued*)

(a)

FIGURE 8: Another comparison between schemes. (a) Hu's moment invariants, (b) Original Zernike, (c) Kim's modified Zernike, (d) Our double four-level Zernike, (e) Our Compactness indices, (f) Combining double Zernike and Compactness indices.

FIGURE 8: Another comparison between schemes. (a) Hu's moment invariants, (b) Original Zernike, (c) Kim's modified Zernike, (d) Our double four-level Zernike, (e) Our Compactness indices, (f) Combining double Zernike and Compactness indices. (*Continued*)
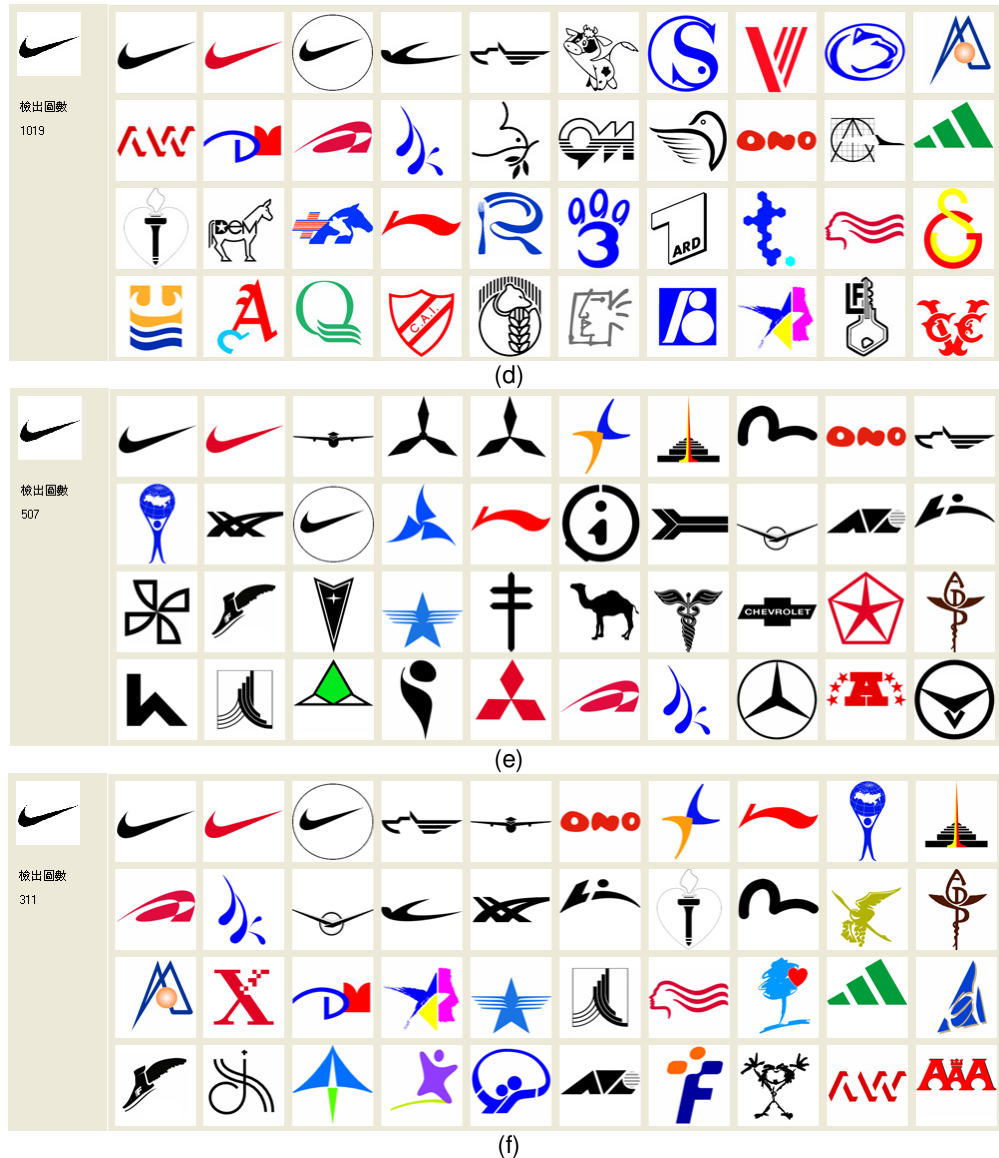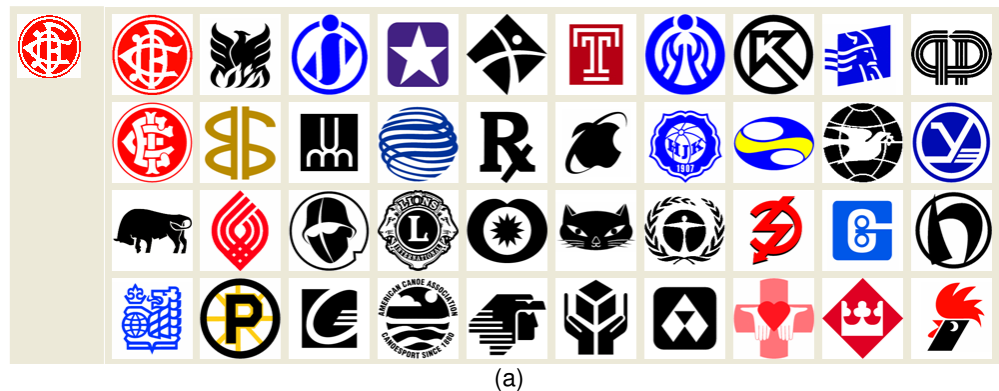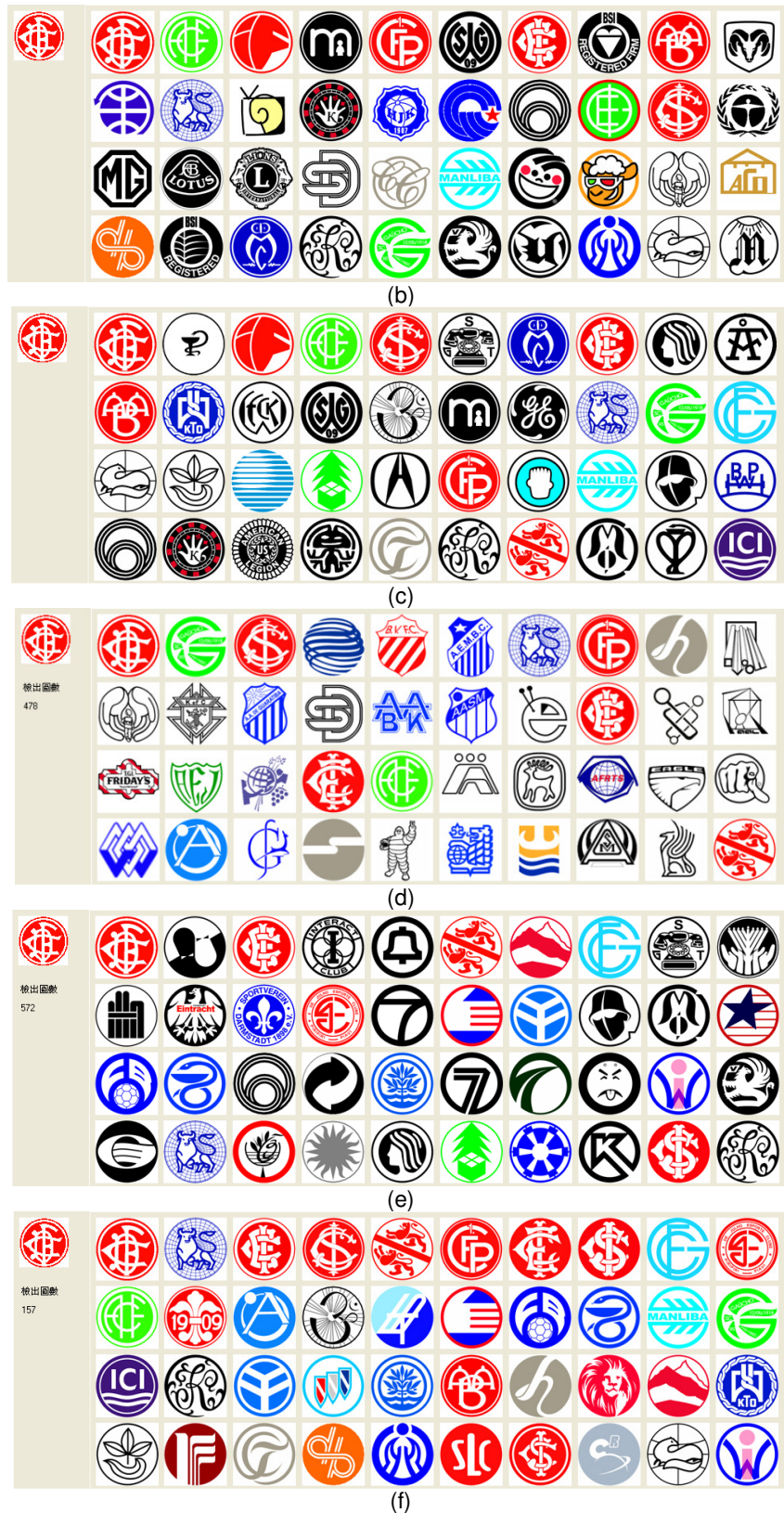
On the whole, for the five groups of trademark images in Table 1, the performance of our method is summarized as follows. In average, both our double Zernike schemes (as Figs. 7d and 8d) and CI-only schemes (as Fig. 7e and 8e) got roughly 60% of human-picked images in their top 40 (top 2%) image listings. This was no better than Kim's modified Zernike (65% in average) and more or less equivalent to traditional Zernike (58% in average). But it is noted that since our double-Zernike and CI's are very different image features, the combination of the two (as Figs. 7f and 8f) yielded very good (95% in average) results. But with what kind of trademark did we show greatest advantage? We will say in dealing with the "Goblet" images, our method outperformed hugely other methods (100% vs. 25%). The next were the "Monogram" and the "Nike" trademarks. For the "Recycle" and the "Michelin" groups, only small improvement was achieved. This was because both traditional Zernike or Kim's modified Zernike already treated them quite well. Hybrid system may not be able to retrieve distributed shape or image like shape. Because we were concentrated only trademark retrieval and this kind of shape is not familiar as a trademark. However, our method was still a competitive choice because we could deal with outlined, framed, and inversed variation of the trademarks.

In all above experiments, a 64×64 image size was used for the input and for all the database images. An input image must go through a series of preprocessing processes such as conversion to four gray-level, obtaining the smallest enclosing circle, and proper geometric transformation. All image features are then extracted, paired with their weighting factors and compared with stored data in the database. The whole process for a 200×200 input image takes about 8.5 seconds on an Intel Celeron 2.66 GHz computer with 512 Mb RAM. Lastly, we think it is interesting to mention that when we doubled the resolution of all images (input and database) to 128×128, only slight change was observed in the results.

5. CONCLUSIONS

We have presented in this paper the framework of a hybrid trademark retrieval system. Surveys of human responses to different trademark aspects were conducted and the results were used as benchmarks for the system. The task of the computer was only to provide, according to the benchmarks, a good ranked list of similar candidates. Both region-based image features (i.e., four-gray-level Zernike moments) and contour-base features (the proposed compactness indices) were used. A simplified method for finding the smallest enclosing circle was also presented. Features of different categories were conjoined via a weighting scheme particularly useful for dealing with framed or inversed trademark variations. Experiments have verified that, when human viewpoints obtained from consumer surveys were used as standards, this hybrid scheme performed considerably better than some existing retrieval methods.

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] United States Patent and Trademark Office, "2007 Design Code Updates", *USPTO Design Search Code Manual*, http://tess2.uspto.gov/tmdb/dscm/dsc_ap.htm, accessed 2010

[2] H. Freeman, "*On the Encoding of Arbitrary Geometric Configurations*". IRE Trans. Electronic Computing, 10: 260-268, 1961

[3] H. L. Peng and S. Y. Chen, "*Trademark Shape Recognition Using Closed Contours*". Pattern Recognition Letters, 18: 791-803, 1997

[4] C. C. Huang and I. Her, "*Homomorphic Graph Matching of Articulated Objects by An Integrated Recognition Scheme*". Expert Systems with Applications, 31: 116-129, 2006

[5]   P. Y. Yin and C. C. Yeh, "*Content-based Retrieval from Trademark Databases*". Pattern Recognition Letters, 23: 113-126, 2002

[6]   M. K. Hu, "*Visual Pattern Recognition by Moment Invariants*". IRE Trans. Information Theory, pp. 179-187, 1962

[7]   J. L. Shih and L. H. Chen, "*A New System for Trademark Segmentation and Retrieval*". Image and Vision Computing, 19: 1011-1018, 2001.

[8]   G. Ciocca and R. Schettini, "*Content-based Similarity Retrieval of Trademarks Using Relevance Feedback*". Pattern Recognition, 34: 1639-1655, 2001

[9]   Y. S. Kim and W. Y. Kim, "*Content-based Trademark Retrieval System Using a Visually Salient Feature*". Image and Vision Computing, 16: 931-939, 1998

[10] H. K. Kim, J. D. Kim, D. G. Sim and D. I. Oh, "*A Modified Zernike Moment Shape Descriptor Invariant to Translation, Rotation and Scale for Similarity-based Image Retrieval*". IEEE International Conference on Multimedia and Expo, 1: 307-310, 2000

[11] N. K. Kamila, S. Mahapatra and S. Nanda, "*Invariance Image Analysis Using Modified Zernike Moments*". Pattern Recognition Letters, 26: 747-753, 2005

[12] United States Patent and Trademark Office, *Trademark Manual of Examining Procedure (TMEP)*, 5th ed., USPTO, 2007

[13] United States Patent and Trademark Office, *U.S. Trademark Law*, USPTO, 2010

[14] Taiwan Intellectual Property Office, *Taiwan Trademark Law*, TIPO, 2003

[15] J. R. Chang, Director, Taiwan Intellectual Property Office, Kaohsiung Branch, Interview by author, 2006

[16] H. K. Hung, "*A Computer-assisted Trademark Retrieval System with Zernike Moment and Image Compactness Indices*". Master thesis, Electro-Mechanical Engineering Department, National Sun Yat-sen University, Taiwan, 2006

[17] S. D. Lin, S. C. Shie, W. S. Chen, B. Y. Shu, X. L. Yang and Y. L. Su, "*Trademark Image Retrieval by Distance-Angle Pair-wise Histogram*". International J. Imaging Systems and Technology, 15: 103-113, 2005

[18] M. Teague, "*Image Analysis via the General Theory of Moments*". J. Optical Society of America, 70: 920-930, 1980

[19] M. de Berg, M. van Kreveld, M. Overmars and O. Schwarzkopf, *Computational Geometry: Algorithms and Applications*, New York : Springer, (2000)

[20] I. Her, "*Geometric Transformations on the Hexagonal Grid*". IEEE Trans. Image Processing, 4: 1213-1222, 1995

[21] I. Her, "*A Symmetrical Coordinate Frame on the Hexagonal Grid for Computer Graphics and Vision*". ASME J. Mechanical Design, 115: 447-449, 1993.

[22] J. Schietse, J. Eakins & R. Veltkamp, "*Practice and Challenges in Trademark Image Retrieval*". ACM International Conference on Image and Video Retrieval, Amsterdam, The Netherlands, 2007

Statistical Models for Face Recognition System With Different Distance Measures

R.Thiyagarajan   thiyagucdm@gmail.com
Department of Electronics and
Instrumentation Engineering
Annamalai University
Annamalai Nagar, India

S.Arulselvi   arulselvi_2k3@yahoo.co.in
Department of Electronics and
Instrumentation Engineering
Annamalai University
Annamalai Nagar, India

G.Sainarayanan   sai.jgk@gmail.com
R&D ICTACT, Chennai
India

Abstract

Face recognition is one of the challenging applications of image processing. Robust face recognition algorithm should posses the ability to recognize identity despite many variations in pose, lighting and appearance. Principle Component Analysis (PCA) method has a wide application in the field of image processing for dimension reduction of the data. But these algorithms have certain limitations like poor discriminatory power and ability to handle large computational load. This paper proposes a face recognition techniques based on PCA with Gabor wavelets in the preprocessing stage and statistical modeling methods like LDA and ICA for feature extraction. The classification for the proposed system is done using various distance measure methods like Euclidean Distance(ED), Cosine Distance (CD), Mahalanobis Distance (MHD) methods and the recognition rate were compared for different distance measures. The proposed method has been successfully tested on ORL face data base with 400 frontal images corresponding to 40 different subjects which are acquired under variable illumination and facial expressions. It is observed from the results that use of PCA with Gabor filters and features extracted through ICA method gives a recognition rate of about 98% when classified using Mahalanobis distance classifier. This recognition rate stands better than the conventional PCA and PCA + LDA methods employing other and classifier techniques.

Keywords: Face Recognition, Statistical Models, Distance measure methods, PCA/LDA/ICA.

1. INTRODUCTION

Human face detection and recognition is an active area of research spanning several disciplines such as image processing, pattern recognition and computer vision with wide range of applications such as personal identity verification, video-surveillance, facial expression extraction,

advanced human and computer interaction. The wide-range variations of human face due to pose, illumination, and expression, result in a highly complex distribution and deteriorate the recognition performance. Hence, there is a need to develop robust face recognition algorithm. Block diagram of a typical face recognition system is shown in Fig. 1. In this block diagram, the preprocessing stage is a filtering process used to reduce noise and dependence on precise registration. Classification is usually one among number of standard methods like minimum distance classifiers, artificial neural networks, etc. Feature extraction is the area that tends to differentiate. This paper addresses the feature extraction method using gabor wavelets and PCA and ICA on face images. A good survey of face recognition system is found in [1]. The methods for face recognition can be divided into two different classes: geometrical feature matching and template matching.
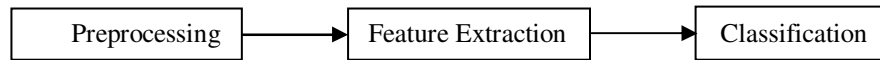
| Preprocessing | → | Feature Extraction | → | Classification |

FIGURE 1. Block diagram of a typical face recognition system.

In the first method, some geometrical measures about distinctive features such as eyes, nose, mouth and chin were extracted. These extracted features were used for the recognition. In the second method, the face image is represented as a two-dimensional array of intensity values and this is compared to a single or several templates representing a whole face. The choice of features for the recognition with the classification method accounts for a good recognition results. These features are first extracted from the raw data using some feature extraction methods. These extracted features should be of lesser dimension and provide good discriminating ability.

In the proposed work, gabor wavelets were used for filtering in the pre processing stage. This results in gabor feature vector representing the face image. Gabor transformed face images exhibit strong characteristics of spatial locality, scale, and orientation selectivity. These images can, thus, produce salient local features that are most suitable for face recognition. The input face image is treated as a two – dimensional pattern of intensity variations. In the face based approach, this two dimensional data helps to define the face image as feature vectors. The well known method for feature extraction is the PCA method. Reconstruction of human faces using PCA was first done by Kirby and Sirovich [2] and recognition of human faces was done by Turk and Pentland [3]. The recognition method, known as eigenface method defines a feature space which reduces the dimensionality of the original data space. This reduced data space is used for recognition.

But the two most common problems with the PCA method are its poor discriminator power for data set from the same class and large computational load. These limitations with the PCA method is overcome by combining Linear Discriminant Analysis (LDA) method which is one of the most important feature selection algorithms in appearance based methods [4]. With direct implementation of LDA method, the dataset selected should have larger number of samples per class for LDA method to extract the discriminating features. With the small sample size problem, the LDA method will results in poor discriminating features. Hence, many LDA based approach first use the PCA to project an image into a lower dimensional space otherwise called as face space. Then perform the LDA to maximize the discriminatory power.

The second problem in PCA-based method is that it will de-correlates the input data using second order statistics, generating a compressed data with minimum mean squared projection error, but it will not account for the higher order dependencies in the input data projected onto them. To overcome this problem ICA is used which minimizes both second order and higher order dependencies in the input and attempts to find the basis along which the data (when projected onto them) are statistically independent[5]. ICA is used to reduce redundancy and represent independent features explicitly. These independent features are most useful for subsequent pattern discrimination and associative recall [6].

In this proposed work, the gabor wavelets are used in the pre processing stage. This gabor filtered face images results in a large gabor feature vector. This larger dataset is reduced to a smaller dimension using the PCA algorithm which extracts the principle features (top eigen vectors). The reduced dataset is further processed using the two feature extraction techniques namely LDA and ICA. This results in the most discriminating and the most independent features of the input data respectively. The classification is done using the minimum distance measure methods like ED, CD and MD methods. The results are compared and tabulated for the resulted recognition rate. In comparison with the conventional use of PCA and LDA alone, the proposed method gives a better recognition rate. The algorithm is tested with the use of ORL face database with 200 training and 200 test samples. The feasibility of the new algorithm has been demonstrated by experimental results. Encouraging results were obtained for the proposed method compared to conventional methods.

This paper is organized as follows: Section 2 deals with the basics of Gabor wavelets and Gabor features. Section 3 deals background of PCA and Eigenfaces. Section 4 deals with LDA. Section 5 details about ICA. Section 6 deals with the various types of distance measure method. The proposed method is reported in section 7. Experimental results are presented in Section 9 and finally, Section 10 gives the conclusions.

2.   GABOR WAVELET

The prime motive behind the feature based approach is the representation of face image in a very compact manner, thus reducing the memory used. Such methods gains importance as the size of the database is increased. This feature based method in a face recognition system aims to find the important local features on a face and represent the corresponding features in an efficient way. The selection of feature points and representation of those values has been a difficult task for a face recognition system. Physiological studies found that simple cells present in human visual cortex can be selectively tuned to orientation as well as spatial frequency. J.G.Daugman[7] has worked on this and confirmed that the response of the simple cell could be approximated by 2D Gabor filters.

Because of their biological relevance and computational properties, Gabor wavelets were introduced into the area of image processing [8,9]. The Gabor wavelets, whose kernels are similar to the 2D receptive field profiles of the mammalian cortical simple cells, exhibit desirable characteristics of spatial locality and orientation selectivity, and are optimally localized in the space and frequency domains. These drawbacks were overcome by Gabor wavelets. The Gabor wavelet is well known for its effectiveness to extract the local features in an image. It has the ability to exhibit the desirable characteristics of capturing salient visual properties such as spatial localization, orientation selectivity, and spatial frequency selectivity. This method is robust to changes in facial expression, illumination and poses. Applying statistical modeling methods like PCA, LDA, ICA for the Gabor filtered feature vector, results in a crisp gabor feature vectors. In order to consider all the gabor kernels, all the features were concatenated to form a single gabor feature vector. This high dimensional gabor vector space is then reduced using the PCA. Further on applying the ICA and LDA on the reduced data, the more independent and discriminating features were obtained.

The Gabor wavelets (kernels, filters) can be defined as follows [7], [8]:

$$\Psi_{\mu,v}(z) = \|k_{\mu,v}\|^2 \exp\left(-\frac{\|k_{\mu,v}\|^2 \|z\|^2}{2\sigma^2}\right)\left(\exp(ik_{\mu,v}z) - \exp\left(-\frac{\sigma^2}{2}\right)\right) \quad (1)$$

Where μ and v in Eqn. 1 define the orientation and scale of the Gabor kernels, $z = (x,y)$ and $\|.\|$ denotes the norm operator, and the wave vector $k_{\mu,v}$ is defined as follows:

$$k_{\mu,v} = k_v \exp(i\Phi_\mu) \quad (2)$$

where $k_v = k_{max} / f_v$ and $\varphi_\mu = \pi\mu/8$, k_{max} is the maximum frequency, and f is the spacing factor between kernels in the frequency domain [9]. The Gabor kernels in Eq. 1 are all self-similar since they can be generated from one filter, the mother wavelet, by scaling and rotation via the wave vector kμ,v. Each kernel is a product of a gaussian envelope and a complex plane wave, while the first term in Eq. 2 determines the oscillatory part of the kernel and the second term compensates for the DC value. The effect of the DC term becomes negligible when the parameter σ, which determines the ratio of the gaussian window width to wavelength, has sufficiently large values. In most cases the use of gabor wavelets of five different scales, $v \in$ {0,….., 4} and eight orientations, $\mu \in$ {0,…..,7}is used for representational bases. Fig. 2 shows the Gabor kernels at five scales and eight orientations, with the following parameters: σ = 2π, $k_{max} = \pi/2$ and $f = \sqrt{2}$.

2.1 Gabor Wavelet Representation

The Gabor wavelet representation of an image is the convolution of the image with a family of Gabor kernels as defined by Eq. 3. Let I(x,y) be the gray level distribution of an image, the convolution of image I and a Gabor kernel ψμ,v is defined as follows.

$$O_{\mu,v}(z) = I(z) * \Psi_{\mu,v}(z)$$

(3)

where z = (x,y) and * denotes the convolution operator, and $O_{\mu,v}$ (z) is the convolution result corresponding to the gabor kernel at orientation μ and scale v. Therefore, the set S = {$O_{\mu,v}$ (z) , for $\mu \in$ {0,…..7} and $v \in$ {0,…..4}} forms the gabor wavelet representation of the image I(z). Applying the convolution theorem, we can derive each $O_{\mu,v}$ (z) from Eq. 3 via the Fast Fourier Transform (FFT):

$$\mathsf{F}\{ O_{\mu,v}(z) \} = \mathsf{F}\{I(z)\} \mathsf{F}\{ \psi_{\mu,v}(z)\}$$ (4)
$$O_{\mu,v}(z) = \mathsf{F}^{-1}\{ \mathsf{F}\{I(z)\} \mathsf{F}\{ \psi_{\mu,v}(z)\}\}$$ (5)

where F and F^{-1} denote the Fourier and inverse Fourier transform, respectively. Fig.2 and 3 show the Gabor wavelet representation (the real part of gabor kernels with five scales and eight orientations and the magnitude of gabor kernels at five different scales). The input frontal face image as shown in Fig 4 is preprocessed using these kernels and the resultant convolution output of the image and the kernels are as shown in Fig. 5 (Real part of the convolution output of a sample image) and Fig. 6 (Magnitude of the convolution output of a sample image). To encompass different spatial frequencies, spatial localities, and orientation selectivities into a single augmented feature vector, concatenate all these gabor representations.
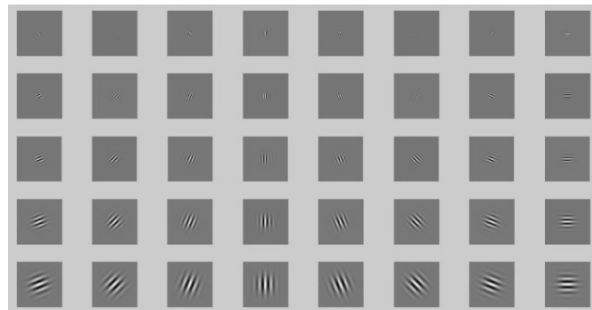

FIGURE 2. Gabor Kernels (Real part with five scales and eight orientations)

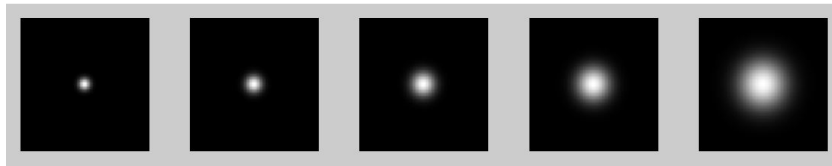**FIGURE 3.** Magnitude of Gabor Kernel at five different scales.
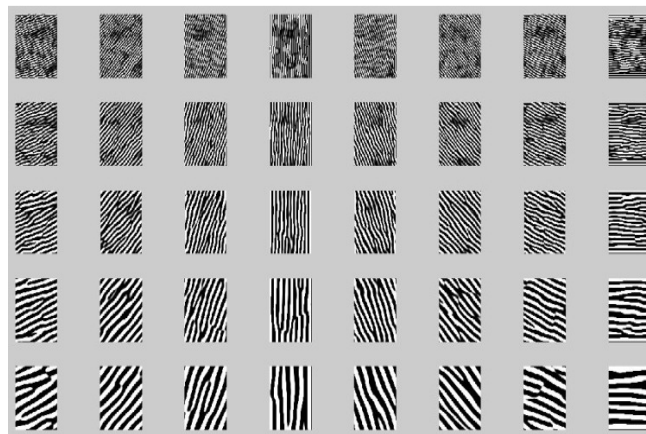

FIGURE 4. Sample frontal image

FIGURE 5. Real part of the convolution output of the sample image

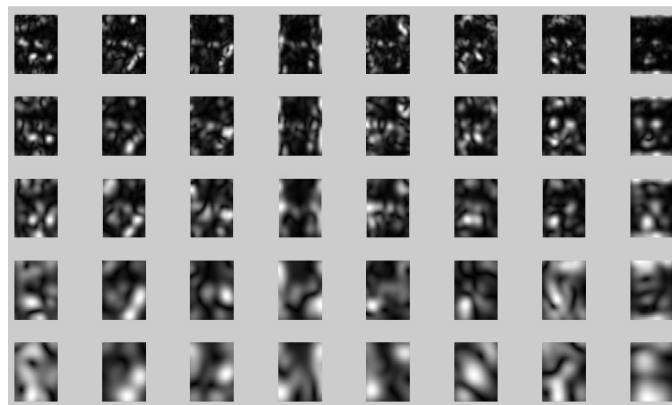**FIGURE 6.** Magnitude of the convolution output of the sample image.

Before the concatenation, each $O_{\mu,v}(z)$ is down sampled by a factor ρ to reduce the space dimension. Then constructing a vector out of the $O_{\mu,v}(z)$ by concatenating its rows (or columns). Now, let $O_{\mu,v}(\rho)$ denote the vector constructed from $O_{\mu,v}(z)$ (down sampled by ρ), the augmented gabor feature vector $X(\rho)$ is then defined as follows:

$$X(\rho) = (O(\rho)t_{0,0} \quad O(\rho)t_{0,1} \ ….. \ O(\rho)t_{4,7}) \tag{6}$$

where t is the transpose operator. The augmented gabor feature vector thus encompasses all the elements (down sampled) of the gabor wavelet representation set, S = { Oμ,v (z) : μ $\in$ {0,…..7} , v $\in$ {0,…..4}} as important discriminating information. Fig. 6 shows (in image form rather than in vector form) an example of the augmented Gabor feature vector, where the down sampling factor is 64, i.e. ρ = 64.

3.   PRINCIPLE COMPONENT ANALYSIS

The Principal Component Analysis (PCA) is a widely accepted and a successful technique that has been used for image compression and image recognition .The purpose of this method is to reduce the large dimensionality of the data (observed variables) to a reduced dimensionality of features (independent variables), which describe the data economically with a strong correlation between data and features [2]. These characteristic features are called eigen faces in the face recognition domain or principle components generally. In a face recognition system, PCA can transform each original image of the training set into a corresponding eigenface. An important feature of PCA is that reconstruction of original image is possible by combining the all eigenfaces from the training set. These eigenfaces are nothing but characteristic features of the training faces. Therefore summing up all the eigen faces in right proportion results in the original face image.

By using all the eigenfaces extracted from original images, exact reconstruction of the original images is possible. But for practical applications only certain part of the eigenfaces is used. Then the reconstructed image is an approximation of the original image. However losses due to omitting some of the eigenfaces can be minimized. This happens by choosing only the most important features (eigenfaces)[2].

3.1   Mathematics of PCA
 A 2-D face image can be represented as 1-D vector by concatenating each row (or column) into a long thin vector.

1. Assume the training sets of images represented by Γ_1, Γ_2, Γ_3,…, Γ_m, with each image $\Gamma(x,y)$ where (x,y) is the size of the image represented by p and m is the number of training images. Converting each image into set of vectors given by (m x p).
2. The mean face Ψ is given by:

$$\Psi = \frac{1}{m}\sum_{i=1}^{m}\Gamma_i \tag{7}$$

3. The mean-subtracted face is given by(Φ_i):

$$\Phi_i = \Gamma_i - \Psi \tag{8}$$

   where i = 1, 2, 3….m. and A = [Φ_1, Φ_2 … Φ_m,] is the mean-subtracted matrix with size Amp.
4. By implementing the matrix transformations, the vector matrix is reduced by:

$$C_{mm} = A_{mp} \times A_{pm}^{T} \tag{9}$$

   where C is the covariance matrix
5. Finding the eigen vectors V_{mm} and eigen values λ_m from the C matrix and ordering the eigen vectors by highest eigen values.
6. With the sorted eigen vector matrix, Φm is adjusted. These vectors determine the linear combinations of the training set images to form the eigen faces represented by U_k as follows

$$U_k = \sum_{n=1}^{m} \Phi_n V_{kn}$$,where , k = 1, 2…m. (10)

7. Instead of using m eigen faces, m′ eigen faces (m′<< m) is considered as the most significant eigen vectors provided for training of each individual.
8. With the reduced eigen face vector, each image has its face vector given by

$$W_k = U_k^T (\Gamma - \Psi)$$, k = 1, 2… m′. (11)

9. The weights form a feature vector given by

$$\Omega^T = [w_1, w_2, ...w_{m'}]$$ (12)

10. These feature vectors are taken as the representational basis for the face images with reduced dimension.
11. The reduced data is taken as the input to the next stage for extricating discriminating feature out of it.

4. LINEAR DISCRIMINANT ANALYSIS

The goal of PCA on choosing a lower dimensional basis is to minimize the reconstruction error. This is not the major concern in pattern recognition applications, whose goal is to maximize the recognition rate. The Linear Discriminant Analysis (LDA) provides a procedure to determine a set of axes whose projections of different groups have the maximum separation. The LDA searches for those vectors in the underlying space that best discriminates among classes (rather than those that best describe the data).

When the training data set are labeled for each identity, supervised training techniques like LDA are more profitable for feature extraction compared with methods of unsupervised learning. By applying the supervised learning, illumination variation and pose variation will be removed and retaining the identity information. The LDA provides a procedure to determine a set of axes whose projections of different groups have the maximum separation. This procedure can be described as follows.

Suppose that the sample consists of p face images from where m_j images are of individual j, for j = 1... g.

Let μ be the mean feature vector of images from individuals j. The samples between individual covariance matrix is defined as

$$S_B = \sum_{i=1}^{k} m_i (\mu_i - \mu)(\mu_i - \mu)^T$$ (13)

$$S_W = \sum_{i=1}^{k} (x_k - \mu)(x_k - \mu)^T$$ (14)

Where μ_i denotes the mean image of the class selected (individual) and μ denotes the mean image of the entire data set, and p denotes the number of images in the entire data set.

The eigenvectors of LDA are called "fisherfaces". LDA transformation is strongly dependent on the number of classes (c), the number of samples (m), and the original space dimensionality (d). It is possible to show that there are almost (c-1) nonzero eigenvectors. (c-1) being the upper bound of the discriminant space dimensionality. For Sw to be a non singular matrix, it is required to have (d+c) samples in the training set [10]. It is impossible to guarantee this condition in many real applications. Consequently, an intermediate transformation is applied to reduce the dimensionality of the image space. To this end, PCA is used. Now LDA derives a low dimensional representation of a high dimensional face feature vector of PCA. From Eqn. 15, the covariance matrix C is obtained as follows

C = S_W^{-1} * S_B (15)

The coefficients of the covariance matrix gives the discriminating feature vectors for the LDA method. The face vector is projected by the transformation matrix WLDA. The projection coefficients are used as the feature representation of each face image. The matching score between the test face image and the training image is calculated as the distance between their coefficients vectors. A smaller distance score means a better match. For the proposed work, the column vectors w_i (i = 1, 2…c-1) of matrix W are referred to as fisherfaces.

5.   INDEPENDENT COMPONENT ANALYSIS

The problem in PCA-based method is that it will decorrelates the input data using second order statistics, generating a compressed data with minimum mean squared projection error. But it will not account for the higher order dependencies in the input and attempts to find the basis along which the data is projected onto them. To overcome this problem, Independent Component Analysis(ICA) is used which minimizes both second order and higher order dependencies in the input and attempts to find the basis along which the data (when projected onto them) are statistically independent[11,12]. ICA of a random vector searches for a linear transformation which minimizes the statistical dependence between its components [13]. The principle feature vectors of the frontal image is obtained using PCA as given in Eqn. 12. This principle components are used as the input for the ICA and the most independent features are obtained. Since it is difficult to maximize the independence condition directly, all common ICA algorithms recast the problem to iteratively optimize a smooth function whose global optima occurs when the output vectors U are independent . Thus use of PCA as a pre-processor in a two-step process allows ICA to create subspaces of size m by m. In [14], it is also argued that pre-applying PCA enhances ICA performance by (1) discarding small trailing eigenvalues before whitening (linear transformation of the observed variable) and (2) reducing computational complexity.

Let the selected principle features be the input to ICA as given in Eqn. 12. i.e Ω be of the size m by z, containing the first m eigenvectors of the face database. The rows of the input matrix to ICA are variables and the columns are observations, therefore, ICA is performed on ΩT. The m independent basis images in the rows of U are computed as

$$U = W * \Omega T \qquad (16)$$

where W is the weight matrix from the PCA. Then, the n by m, ICA coefficients matrix B for the linear combination of independent basis images in U is computed as follows

Let C be the n by m matrix of PCA coefficients. Then,

$$C = I*\Omega \text{ and } I = C * \Omega T \qquad (17)$$

From U = W* ΩT and the assumption that W is invertible,
$$\Omega T = W^{-1} * U \qquad (18)$$
Therefore,

$$I = (C* W^{-1})* U \qquad (19)$$

$$= B * U$$

Each row of B contains the coefficients for linearly combining the basis images to comprise the face image in the corresponding row of I. Also, I is the reconstruction of the original data with minimum squared error as in PCA.

6.   CLASSIFIER METHODS

Image classification systems analyze the numerical properties of various image features and organize data into categories. Classification includes a broad range of decision-theoretic

approaches to the identification of images (or parts thereof). All classification algorithms are based on the assumption that the image in question depicts one or more features and that each of these features belongs to one of several distinct and exclusive classes. In practice, the minimum (mean) distance classifier works well when the distance between means is large compared to the spread (or randomness) of each class with respect to its mean.

The minimum distance classifier is used to classify unknown image data to classes where minimum distance between the image data and the class in multi-feature space exists. The distance is defined as an index of similarity so that the minimum distance is identical to the maximum similarity. There exist various types of distance classifier techniques. Three different types of distance classifiers namely Euclidean distance, Cosine distance and Mahalanobis distance methods were considered for the proposed approach.

Euclidian Distance $\qquad$ $d(x, y) = kx - yk2$ $\qquad$ (20)

Cosine Metric $\qquad$ $d(x, y) = x.y/|x||y|$ $\qquad$ (21)

Mahalanobis Metric $\qquad$ $d(x, y, C) = squrt((x - y)TC^{-1}(x - y))$ $\qquad$ (22)

where the variables x and y in the above set of equations represents the train and the test features respectively with k and K1 as constant and C is the covariance matrix of the training set.

7.   PROPOSED METHOD

The block diagram of the proposed face recognition system is as shown in Fig. 7. This proposed work uses the ORL database acquired at the Olivetti Research Laboratory in Cambridge, U.K. The database is made up of 400 face images that correspond to 40 distinct subjects. Thus, each subject in the database is represented with 10 facial images that exhibit variations in terms of illumination, pose and facial expression. The images are stored at a resolution of 92 × 112 and 8-bit grey levels. Out of these 400 frontal images, 200 images (5 from each class) were considered for training and the remaining 200(remaining 5 from each class) were used for testing. Figs. 8&9 show the scaled training and test samples.
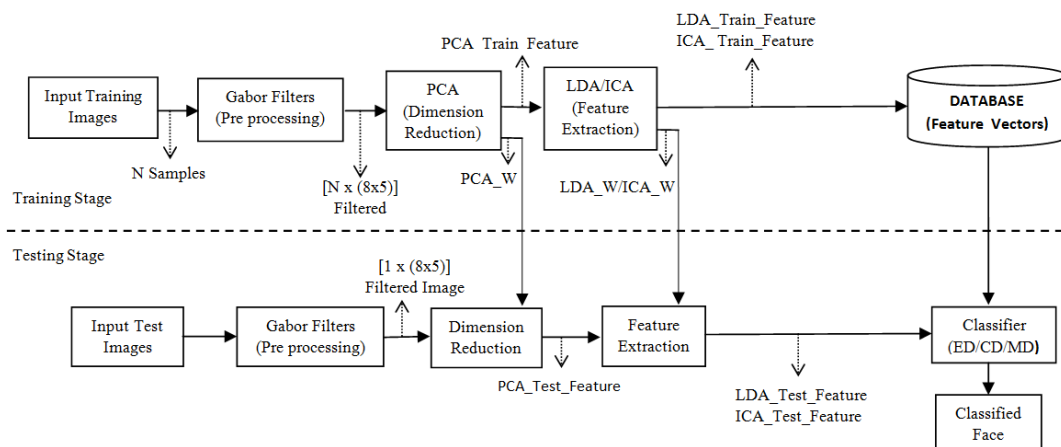
FIGURE 7. Proposed Block diagram

The system follows the image based approach and it consists of two stages namely the training and the recognition stage. In the training stage, gabor wavelet is used in the preprocessing stage which is robust to changes in illumination, pose and expression. To facilitate the convolution of the input image with gabor filters, the input images are scaled to 128x128 using a bicubic interpolation. The gabor kernels as defined in Eqn. 7 uses five different scales and eight

orientations which results in 40 different filters. The input image is convoluted with the gabor kernel and the convoluted real and magnitude responses are shown in Fig. 5 and 6 respectively. This convoluted feature vector is then down sampled by a factor of $\rho = 64$. To encompass all the features produced by the different gabor kernels, the resulting Gabor wavelet features are concatenated to derive an augmented gabor feature vector.

The resulting high dimension gabor feature vector is taken as input to the next stage. This stage uses the PCA for reducing the dimension of the feature vector and extract the principle features. Here the eigen values and the corresponding eigen vectors were calculated. The eigen values are sorted in the ascending order and the top eigen values (n-1) are used and the corresponding eigen vectors were selected for representation of feature vectors (PCA_Feat). Also the weight matrix is computed as PCA_W. This eigen projection is then used as input to the next stage of ICA using the FastICA algorithm. The algorithm takes the projected data and gives the independent features ICA_Feat and the independent weight matrix ICA_W. This is used to find the test_ICA features.

FIGURE 8. Scaled Train Samples

The same procedure is applied for feature extraction using LDA. Here PCA_Feat is used as input to the LDA block. The between class (SB) and within class scatter matrix (SW) is obtained using this projection matrix as given in Eqn 13 and 14. LDA gives the projected weight matrix LDA_W, which is used to find the LDA_test_features.

FIGURE 9. Scaled Test Samples

In the recognition stage, the test samples are preprocessed as done in the training stage. With the weight matrix from the PCA stage and the test image, the test features (PCA_Test_Feature) were obtained. This PCA_Test_Feature when convoluted with ICA_W gives the test features (ICA_test). The same procedure is used to obtain the test features for LDA method (LDA_test and LDA_W). The classification stage uses different distance measure methods. Here the training and test feature vectors are considered. For a given sample image, test feature vector is found and the distance between this test feature vector and the all the training feature vectors are calculated. From the distance measure, the index with minimum distance represents the recognized face from the training set.

8. RESULT AND DISCUSSION

The effectiveness of the proposed system is tested with ORL Face database and the results are compared for the following experimental setups.

1. Gabor with PCA.
2. Gabor with PCA and LDA.
3. Gabor with PCA and ICA.

For each of the above said feature extraction method, the extracted features is used as it is, followed by classification using these three distance measures. The number of features considered for recognition is varied and the corresponding recognition rate is given in Table 1-3. The successfulness of the proposed method is compared with some popular face recognition schemes like Gabor wavelet based classification [15], the PCA method [16], the LDA method [17], and ICA based systems[18].

S.No	Distance	Features				
		40	50	66	100	200
1.	Euclidean	87.0	89.3	93.2	93.7	94.0
2.	Cosine	87.5	91.0	95.0	95.9	96..0
3.	Mahalanobis	89.3	92.5	96.1	96.9	97.0

TABLE 1: Features Vs Recognition rate for different distance measure methods
(Gabor wavelets and PCA method for feature extraction)

S.No	Distance	Features				
		40	50	66	100	200
1.	Euclidean	87.8	90.1	94.3	95.2	95.3
2.	Cosine	88.4	91.4	93.6	96.8	97.1
3.	Mahalanobis	90.0	93.5	94.5	97.1	98.0

TABLE 2: Features Vs Recognition rate for different distance measure methods
(Gabor wavelets and PCA method for feature extraction)

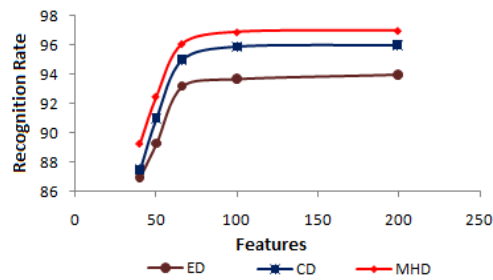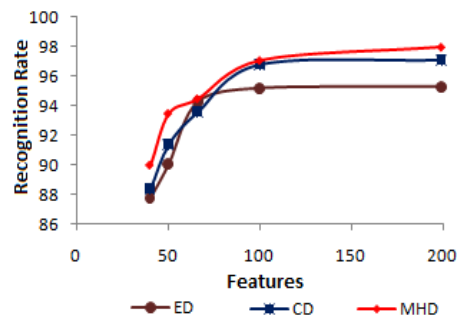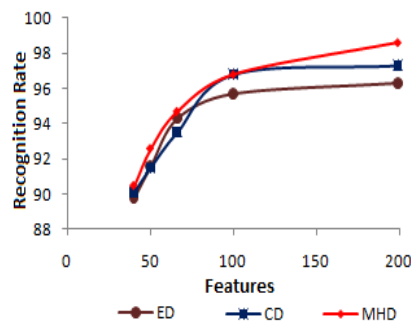| S.No	Distance	Features				
		40	50	66	100	200
1.	Euclidean	89.8	91.6	94.3	95.7	96.3
2.	Cosine	90.1	91.5	93.5	96.8	97.3
3.	Mahalanobis	90.5	92.6	94.7	96.8	98.6

TABLE 3:. Features Vs Recognition rate for different distance measure methods
(Gabor wavelets and PCA along with ICA for feature extraction)

10(a)

10(b)

10(c)

FIGURE 10. Performance characteristics of the three system considered
with different classifiers.

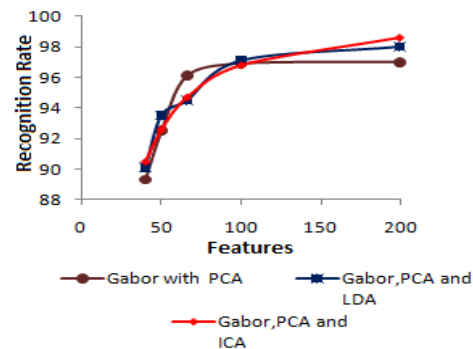10(a) – G + PCA   10(b) – G + PCA + LDA   10(c) G + PCA + ICA

FIGURE 11. Performance characteristics of the proposed system with other methods using Mahalanobis Distance Classifier

9. CONCLUSION

In this proposed work, a Gabor feature based method is introduced for face recognition system. For different scales and orientations of the Gabor filter, the input image is filtered. Using PCA, the high dimensionality of the Gabor feature vector is reduced followed by feature extraction using ICA and LDA. Three different types of distance measure methods namely Euclidean Distance, Cosine Distance, and Mahalanobis Distance are used for classification. From the simulation results, it has been found that the recognition rate for the selected database is high with features extracted from Gabor filters based on ICA than with LDA and simple PCA methods. This recognition rate is obtained by varying the number of features at the PCA stage. The results are tabulated in the Table .1 to 3. For all these three systems, the number of PCA features assumed for further processing is 40, 50, 66, 100, and 199.

From the results, it is obvious that as the number of features selected in PCA increase, then the more discriminating features is obtained from the LDA method and the more independent features is obtained from the ICA method. This helps to increase the recognition rate percentage for the proposed system. But this in turn increases the computational load. Also from the results, when compared to other classifiers tested in the proposed system, Mahalanobis distance classifier does a better classification for the Gabor features based on ICA method.

10. REFERENCES

1. W. Zaho, R.Chellappa, P.J.Philips and A.Rosenfeld, "Face recognition   A literature survey," ACM Computing Surveys, 35(4):399– 458,2003.

2. M. Kirby and L. Sirovich, "Application of the karhunenloeve procedure for the characterization of human faces", IEEE Trans. Pattern Analysis and Machine Intelligence, 12(1):103-108,1990.

3. M. Turk and A. Pentland, "*Eigenfaces for recognition*" J. Cognitive Neuroscience, 3: 71-86,1991.

4. D. L. Swets and J. J. Weng, "*Using discriminant eigenfeatures for image retrieval*", IEEE Trans. PAMI., 18(8):831-836,1996.

5. P. Belhumeur J. P. Hespanha and D. J. Kriegman, "*Eigenfaces vs. Fisherfaces; Recognition using class specific linear projection*", Pattern Recognition and Machine Intelligence, 17(9):711-720, 1997.

6. Chengjun Liu, and Harry Wechsler, "*Independent Component Analysis of Gabor Features for Face Recognition*", IEEE Transactions on Neural Networks, 14(4):919-928, 2003.

7. J. G. Daugman, "*Two dimensional spectral analysis of cortical receptive field profile*", Vision Research, 20: 847-856, 1980.

8. S. Marcelja, "*Mathematical description of the responses of simple cortical cells*" Journal Opt. Soc. Amer.,70:1297–1300, 1980.

9. M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, Wurtz R.P., and W. Konen, "*Distortion invariant object recognition in the dynamic link architecture*" IEEE Trans. Computers, 42:300–311, 1993.

10. Vitomir Struc, Nikola pavesi, "*Gabor-Based Kernel Partial-Least-Squares Discrimination Features for Face Recognition*", Informatica, 20(1):115–138, 2009.

11. M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski, "*Face recognition by independent component analysis*", IEEE Trans Neural Networks, 13:1450–1464, 2002.

12. Kresimir Delac, Mislav Grgic, Sonja Grgic , "*Independent comparative Study of PCA, ICA, and LDA on the FERET Data Set*", Wiley periodicals, 15:252-260, 2006.

13. P. Comon, "*Independent component analysis, a new concept*" Signal Processing, 36:287–314, 1994.

14. C. Liu and H. Wechsler, "*Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition*", In Proceedings of the International Conference on Audio and Video Based Biometric Person Authentication, Washington, DC, 1999.

15. A. J. O'Toole, H. Abdi, K. A. Deffenbacher and D. Valentin, "*A low- dimensional representation of faces in the higher dimensions of the space*", J. Opt. Soc. Am., A, 10:405-411,1993.

16. M. H. Yang, N. Ahuja, and D. Kriegman, "*Face recognition using kernel eigenfaces*," Proc. IEEE Int. Conf. Image Processing, 2000.

17. A. Hossein Sahoolizadeh, B. Zargham, *"A New feace recognition method using PCA, LDA and Neural Networks*", proceedings of word academy of science, Engineering and Tech, 31,2008.

18. Marian Stewart Bartlett, Javier R. Movellan, and Terrence J. Sejnowski,"*Face Recognition by Independent Component Analysis*", IEEE Trans on Neural Networks,13(6):1450-1464,2002.

A Novel Approach for Cancer Detection in MRI Mammogram Using Decision Tree Induction and BPN

Dr.V.Saravanan   tvsaran@hotmail.com

H.O.D,Department of M.C.A,
Karunya Deemed University,
Coimbatore, Tamil Nadu, India

S.Pitchumani Angayarkanni   pitchu_mca@yahoo.com

Assistant Professor, Department of Computer Science,
Lady Doak College, Madurai,Tamil Nadu , India

Abstract

An intelligent computer-aided diagnosis system can be very helpful for radiologist in detecting and diagnosing micro calcifications patterns earlier and faster than typical screening programs. In this paper, we present a system based on fuzzy-C Means clustering and feature extraction techniques using texture based segmentation and genetic algorithm for detecting and diagnosing micro calcifications' patterns in digital mammograms. We have investigated and analyzed a number of feature extraction techniques and found that a combination of three features, such as entropy, standard deviation, and number of pixels, is the best combination to distinguish a benign micro calcification pattern from one that is malignant. A fuzzy C Means technique in conjunction with three features was used to detect a micro calcification pattern and a neural network to classify it into benign/malignant. The system was developed on a Windows platform. It is an easy to use intelligent system that gives the user options to diagnose, detect, enlarge, zoom, and measure distances of areas in digital mammograms. The present study focused on the investigation of the application of artificial intelligence and data mining techniques to the prediction models of breast cancer. The artificial neural network, decision tree, Fuzzy C Means, and genetic algorithm were used for the comparative studies and the accuracy and positive predictive value of each algorithm were used as the evaluation indicators. 699 records acquired from the breast cancer patients at the MIAS database, 9 predictor variables, and 1 outcome variable were incorporated for the data analysis followed by the 10-fold cross-validation. The results revealed that the accuracies of Fuzzy C Means were 0.9534 (sensitivity 0.98716 and specificity 0.9582), the decision tree model 0.9634 (sensitivity 0.98615, specificity 0.9305), the neural network model 0.96502 (sensitivity 0.98628, specificity 0.9473), the genetic algorithm model 0.9878 (sensitivity 1, specificity 0.9802). The accuracy of the genetic algorithm was significantly higher than the average predicted accuracy of 0.9612. The predicted outcome of the Fuzzy C Means model was higher than that of the neural network model but no significant difference was observed. The average predicted accuracy of the decision tree model was 0.9635 which was the lowest of all 4 predictive models. The standard deviation of the 10-fold cross-validation was rather unreliable. The results showed that the genetic algorithm described in the present study was able to produce accurate results in the classification of breast cancer data and the classification rule identified was more acceptable and comprehensible.

S.Pitchumani Angayarkanni

1. INTRODUCTION

Breast cancer is one of the major causes for the increased moralitcause many among women especially in developed countries. It is second most common cancer in women. The World Health Organization's International estimated that more than 1,50,000 women worldwide die of breast cancer in year. In India , breast cancer accounts for 23% of all the female cancer death followed by cervical cancer which accounts to 17.5% in India. Early detection of cancer leads to significant improvements in conservation treatment[1]. However, recent studies have shown that the sensitivity of these systems is significantly decreased as the density of the breast increased while the specificity of the systems remained relatively constant. In this work **w**e have developed automatic neuron genetic algorithmic approach to automatically detect the suspicious regions on digital mammograms based on asymmetries between left and right breast image.

Diagnosing cancer tissues using digital mammograms is a time consuming task even highly skilled radiologists because mammograms contain low signal to noise ratio and a complicated structural background. Therefore in digital mammogram, there is still a need to enhance imaging, where enhancement in medical imaging is the use of computers to make image clearer. This may aid interpretation by humans or computers. Mammography is one of the most promising cancer control strategies since the cause of cancer is still unknown[2]. Radiologist turn to digital mammography as an alternative diagnostic method due to the problems created by conventional screening programs. A digital mammogram is created when conventional mammogram is digitized; through the use of a specific mammogram is digitizer or a camera, so it can be processed by the computer. Image enhancement methods are applied to improve the visual appearance of the mammograms. Initially the mammogram image is read from the dataset and partial filter (Combination of Low and high Pass filter) is applied to remove the noise from the image.

Fuzzy C Means clustering with texture based segmentation ,decision tree induction and genetic algorithm techniques were used in efficient detection of cancerous masses in mammogram MRI images . The selected features are fed to a three-layer Backpropagation Network hybrid with FCM,GA (BPN-FCM-GA) for classification and the Receiver Operating Characteristic (ROC) analysis is performed to evaluate the performance of the feature selection methods with their classification results.

In this paper various steps in detection of microcalcification such as i) Preprocessing and enhancement using Histogram Equalization and Parital filter ii) Fuzzy C Means clustering iii) Texture based Segmentation iv) Decision Tree Induction V) Genetic Algorithm V) System is trained using Back Propagation Network VI) FROC analysis is made to find how accurate the detection of cancerous masses in automatic detection system.

2.   PREPROCESSING & ENHANCEMENT TECHNIQUES:

One of the most important problems in image processing is denoising. Usually the procedure used for denoising, is dependent on the features of the image, aim of processing and also post-processing algorithms [5]. Denoising by low-pass filtering not only reduces the noise but also blurs the edges. Spatial and frequency domain filters are widely used as tools for image enhancement. Low pass filters smooth the image by blocking detail information. Mass detection aims to extract the edge of the tumor from surrounding normal tissues and background, high pass filters (sharpening filters) could be used to enhance the details of images. PSNR, RMS, MSE, NSD, ENL value calculated for each of 161 pairs of   mammogram images clearly shows that Partial low and high pass filter when applied to mammogram image leads to best Image Quality[3].
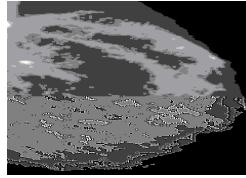
FIGURE 1: Image enhancement using histogram equalization and Partial Filters

3. FUZZY C MEANS CLUSTERING

This algorithm aims at detecting microcalcifications and suspicious areas. In the process of detecting, it may detect other areas that look like a microcalcification. It is up to the user to decide whether the resulting detection is a microcalcification or some other area[4]. The algorithm is simple and based on a fuzzy technique where the size of the microcalcification can be "about the size of a microcalcification." It uses a 8*8 window to scan over the entire digital mammogram and locate microcalcifications or other abnormalities:

WHILE entire 8*8 image has not been examined by 4 window
MOVE 8*8 window to next position
RECORD x,y position and grey level value of pixel with largest grey level in window
IF pixels **surrounding** the largest pixel are **as bright as** the largest pixel grey level value AND **outer** pixels are **darker** than the largest pixel grey level value
THEN largest pixel position is the center pixel of a microcalcification
area
END IF
END WHILE
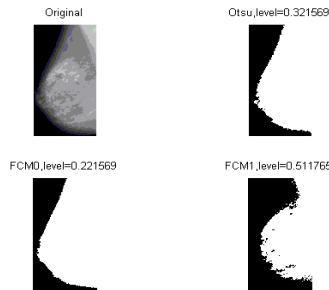
FIGURE 2: segmentation using Fuzzy C means

No. of iteration:36, Time:0.79Seconds

4.   FEATURE SELECTION

The main aim of the research method proposed was to identify the effectiveness of a feature or a combination of features when applied to a neural network[5]. Thus, the choice of features to be extracted was important. The following 14 features were used for the proposed method:

average histogram, average grey level, energy, odified energy, entropy, modified entropy, number of pixels, standard deviation,modified standard deviation, skew, modified skew, average boundary grey level, difference, and contrast. The formulas for entropy, energy, skew, and standard deviation were modified so that the iterations started with the first pixel of the pattern and ended at the final pixel. Traditionally, the formulas for these features have iterations starting with the lowest grey level possible and ranging to the highest grey level possible. This modification was done in an attempt to achieve a better classification rate than its traditional version.

Initially, the method determined the ranking of single features from best to worst by using each feature as a single input to the neural network. After this was completed, a combination of features was tested and a best feature or a combination of
features was determined.

1) First Feature Vector (Ten Features): Average histogram, average grey level, number of pixels, average boundary grey, difference, contrast , energy, entropy, standard deviation, and

skew.

2) Second Feature Vector (14 Features): Average histogram, average grey level, number of pixels, average boundary grey, difference, contrast, modified energy, modified entropy, modified standard deviation, and modified skew.The most significant feature or combination of features were

selected based on neural-network classification. It was done as follows. We started with a single feature by feeding it to the genetic algorithm and neural network and analyzing the classification rate. If it was increased or unchanged by adding a particular feature, then we included this feature to the input vector. Otherwise, we removed this feature and added another feature to the existing input vector and repeated the whole process again. The total 8190 combinations were investigated and the combinations with the best classification rate were selected for the development of the our CAD system.

Figure 1: A typcal benign microcalcification

Figure 2: A typical malignant microcalcification

Figure 3: A typical benign mass

Figure 4: A typical malignant mass
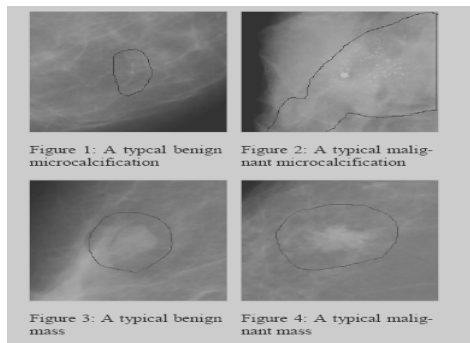
5. DECISION TREE INDUCTION METHOD

Decision trees are powerful classification algorithms that are becoming increasingly more popular with the growth of data mining in the field of information systems[6]. Popular decision tree algorithms include Quinlan's ID3, C4.5, C5, and Breiman et al.'s CART .A decision tree is a non-linear discrimination method, which uses a set of independent variables to split a sample into progressively smaller subgroups. The procedure is iterative at each branch in the tree, it selects the independent variable that has the strongest association with the dependent variable according to a specific criterion (e.g., information gain, Gini index, and Chi-squared test. In our study, we chose to use J4.8 algorithm as our decision tree method and all subjects according to whether or not they were likely to have breast cancer.

Decision rules	Fitness value(train data, 453 cases)	Accuracy(train data, 453 cases)	Accuracy (test data, 246 cases)
If 5.6 < clump Thickness < 7.2 AND 1.8 < Marginal adhesion < 4.0 AND 3.2 < single Epithelia , 8.6 AND 2.1<normal nucleoid <3.1 THEN class=benign else Malignnant	1	0.9993	0.9878

TABLE 1: Decision Rule with corresponding Accuracy

6. GENETIC ALGORITHM

A genetic algorithm is an iterative procedure until a pre-determined stopping condition (usually the number of generation). Genetic algorithm involves a population of individuals, each one represented by a finite string of symbols, known as the genome, encoding a possible solution in a given problem space. The simple genetic algorithm as a pseudo code is:

Step 1. Generate an initial population of strings randomly.
Step 2. Convert each individual of the current population into If–Then rule.
Step 3. Evaluate each of If–Then rules from training dataset.

Step 4. Select parents for the new population.
Step 5. Create the new population by applying selection, crossover and mutation to the parents.
Step 6. Stop generation if a stopping condition is satisfied, otherwise go to step 3.

a.   FITNESS EVALUATION OF RULES

The role of the fitness function is to encode the performance of the rule numerically. In our study, the objective of the GA method is to find the accurate and general rules among all the rules in the population. Thus, the GA method uses the composite fitness function consisting of accuracy and coverage. To measure the accuracy and coverage of the rule, we use the following definitions: when a rule is used to classify a given training instance, one of the four possible concepts can be observed: true positive (tp), false positive (fp), true negative (tn) and false negative (fn). The true positive and true negative are correct classifications, while false positive and false negative are incorrect classifications. For a two-class case, with class 'yes' and 'no', the four concepts can be easily understood with the following descriptions:

true positive(TP): the rule predicts that the class is 'yes' (positive) and the class
of the given instance is in fact 'yes' (true);
false positive(FP): the rule predicts that the class is 'yes' (positive) but the
class of the given instance is in fact 'no' (false);
true negative: the rule predicts that the class is 'no' (negative) and the class of
the given instance is in fact 'no' (false);
false negative: the rule predicts that the class is 'no' (negative) but the class of
the given instance is in fact 'yes' (true). Using these concepts, we present a very simple fitness function defined as:

Maximize Fitness Function =TP/(TP+FP)

7.   NEURAL NETWORK MODEL

We use the popular BPN architecture in building the neural network diagnostic model. Cybenko indicated that one-hidden-layer network is sufficient to model any complex system with any desired accuracy, the designed network model will have only one hidden layer. We use 9 input nodes in the input layer, the initial number of hidden nodes to be tested was chosen to be 6, 7, 8, 9, 10 and the network has only one output node status of the breast cancer or without breast cancer. As Rumelhart [**33**] concluded that lower learning rates tend to give better network result, we decided to incorporate the lower learning rates in our test models. The prediction results of the BPN networks with combinations of different hidden nodes learning rate are summarized in Table 2. As illustrated in Table 2, the {9,6,1} topology with a learning rate of 0.1 gives the best result (i.e. the lowest testing RMSE). The diagnostic result using the designed BPN model stratified by 10-fold cross-validation can be summarized in Table 4. From the result in table 9, we can observe the average correct classification rate is 95.02%. It can be observed that BPN has the highest average correction classification rate in comparison with the decision tree.
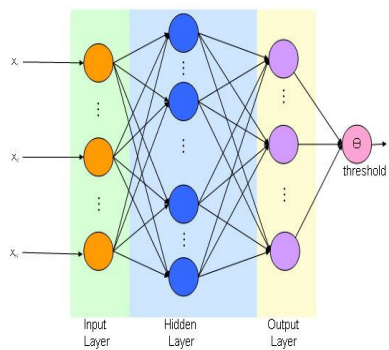
FIGURE 3: Graphical Representation of ANN

Number of Hidden Nodes	Learning Rate	Training RMSE	Testing RMSE
6	0.01	0.1146	0.0600
	0.03	0.0917	0.0433
	0.05	0.0749	0.0210
	0.1	0.1082	0.0089
	0.5	0.5082	0.6878
7	0.01	0.1164	0.0768
	0.03	0.0089	0.0434
	0.05	0.1015	0.0201
	0.1	0.0872	0.0111
	0.5	0.6976	0.5213
8	0.01	0.1342	0.0692
	0.03	0.0726	0.0455
	0.05	0.0947	0.0209
	0.1	0.0653	0.0128
	0.5	0.6976	0.5213
9	0.01	0.1133	0.0762
	0.03	0.0704	0.0430
	0.05	0.0472	0.0210
	0.1	0.1082	0.0092
	0.5	0.5082	0.5213
10	0.01	0.1169	0.0633
	0.03	0.0761	0.0416
	0.05	0.1069	0.0228
	0.1	0.1082	0.0146
	0.5	0.5082	0.6878

TABLE 2: ANN MODEL PREDICTION RESULT

Prediction Variables used in cancer detection:
1. Clump thickness
2. Uniformity of cell size
3. Uniformity of cell shape
4. Marginal adhesion
5. Single epithelial cell size
6. Bare nuclei
7. Bland chromatin
8. Normal nucleoli
9. Mitoses
10. Class (benign/malignant)

Stability Test Results of GA Model:

Group	1-5					6-10				
Evolution Generation	1	2	3	4	5	6	7	8	9	10
Mean Fitness Value	0.93	0.83	0.89	0.81	0.86	0.85	0.77	0.90	0.96	0.90
Optimal Fitness Value	1	0.976	0.983	0.988	1	0.988	0.984	1	1	1

FIGURE 2: Convergence Graph of GA

8.   CONCLUSION

In the present study, various data mining models including Fuzzy C Means , decision tree, and artificial neural network 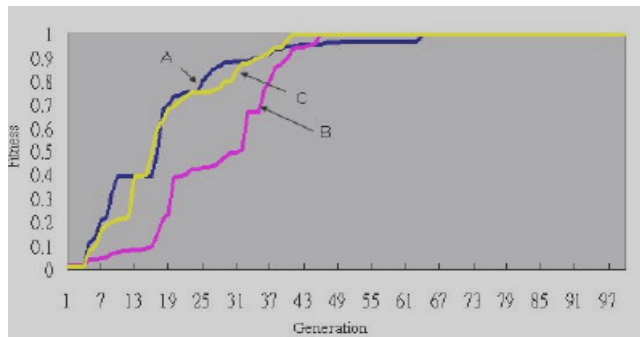were used to compare with the genetic algorithm model by evaluating the prediction accuracy, sensitivity, and specificity. Ineach model, the 10-fold crossover validation was used to compare the results of these 3 models. After testing 10 times, accuracies of Fuzzy C Means was 0.9534 (sensitivity 0.98716 and specificity 0.9582), the decision tree model 0.9634 (sensitivity 0.98615, specificity 0.9305), the neural network model 0.96502 (sensitivity 0.98628, specificity 0.9473), the genetic algorithm model 0.9878 (sensitivity 1, specificity 0.9802). The accuracy of the genetic algorithm was significantly higher than the average predicted accuracy of 0.9612.

In the future, genetic algorithm can be used in combination with many different modifications such as: (1) the use of a different selection method like the steady state selection; (2) the development of adaptive genetic algorithm (AGA) for the parameters varying with populations; (3) the modification of chromosomal encoding; and (4) the development of exclusive fitness functions for different diseases.

Fold	Logistic Regression					Decision Tree				ANNs				
No	Confusion Matrix		Accuracy	SEN	SPE	Confusion Matrix		Accuracy	Sen	Spe	Confusion Matrix		Accuracy	Sen	Spe
1	447	12	0.9628	0.9696	0.9495	441	18	0.9399	0.9483	0.9230	444	15	0.9506	0.9568	0.9387
	14	226				24	216				20	230			
2	446	13	0.9628	0.9716	0.9458	435	24	0.9443	0.9645	0.9104	439	20	0.9442	0.9585	0.9170
	13	227				16	244				19	221			
3	447	12	0.9628	0.9696	0.9497	437	22	0.9384	0.9541	0.9087	443	16	0.9542	0.9651	0.9333
	14	226				21	219				16	224			
4	447	12	0.9670	0.9759	0.9502	437	22	0.9413	0.9583	0.9094	443	16	0.9585	0.9714	0.9341
	11	229				19	221				13	227			
5	447	12	0.9642	0.9717	0.9497	429	30	0.9284	0.9554	0.88	440	19	0.9427	0.9544	0.9201
	13	227				20	220				21	219			
6	446	13	0.9642	0.9737	0.9460	438	21	0.9470	0.9647	0.9142	439	20	0.9442	0.9585	0.9170
	12	228				16	224				19	221			
7	447	12	0.9628	0.9696	0.9495	437	22	0.9442	0.9625	0.9102	443	16	0.9556	0.9672	0.9336
	14	226				17	223				15	225			
8	447	12	0.9628	0.9696	0.9459	437	22	0.9427	0.9604	0.9098	442	17	0.9570	0.9714	0.9303
	14	226				18	222				13	227			
9	447	12	0.9642	0.9717	0.9497	437	27	0.9446	0.9732	0.8941	440	19	0.9484	0.9628	0.9214
	13	227				12	228				17	223			
10	446	13	0.9642	0.9737	0.9460	440	19	0.9642	0.9737	0.9460	442	17	0.9470	0.9567	0.9282
	12	228				20	220				20	220			
Mean			0.9637	0.9716	0.9482			0.9435	0.9615	0.9105			0.9502	0.9628	0.9273
S.D.			0.0013	0.0021	0.0019			0.0089	0.0080	0.0171			0.0057	0.0062	0.0078

TABLE 3: Relationship between Logistic Regression, Decision Tree and ANN

REFERENCES

[1] Barr E, *The handbook of artificial intelligence*, vol. 1-3, William Kaufmann,Los Altos 1982.

[2] Laurikkala J, Juhola M, *A genetic-based machine learning system to discover the diagnostic rules for female urinary incontinence*, Comput Methods
Programs Biomed **55** (1998), no. 3, 217-228.

[3]. N. Karssemeijer, *Computer-Assisted Reading Mammograms, European Radiol.*, vol. 7, pp. 743–748, 1997.

[4] L. Mascio, M. Hernandez, and L. Clinton, "Automated analysis for microcalcifications in high resolution mammograms," *Proc. SPIE—Int.Soc. Opt. Eng.*, vol. 1898, pp. 472–479, 1993.

[5.] L. Shen, R. Rangayyan, and J. Desaultels, *Detection and Classification Mammographic Calcifications, International Journal of Pattern Recognition and Artificial Intelligence*. Singapore: World Scientific, 1994, pp. 1403–1416.

[6.] F. Aghdasi, R.Ward, and B. Palcic, "Restoration of mammographic images in the presence of signal-dependent noise," in *State of the Art in Digital Mammographic Image Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 42–63.

S.Pitchumani Angayarkanni

[7.]   Y. Chitre, A. Dhawan, and M. Moskowtz, "Artificial neural network based classification of mammographic microcalcifications using image structure features," in *State of the Art of Digital Mammographic Image, Analysis*. Singapore: World Scientific, 1994, vol. 7, pp. 167–197.

[8.] Bosch. A.; Munoz, X.; Oliver.A.; Marti. J., *Modeling and Classifying Breast Tissue Density in Mammograms*, Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on Volume 2, Issue , 2006 Page(s): 1552 – 15582.

 [9] Dar-Ren Chen*a*, Ruey-Feng Chang*b*, Chii-Jen Chen*b*, Ming-Feng Ho*b*, Shou-Jen Kuo*a*, Shou-Tung Chen*a*, Shin-Jer Hung*c*, Woo Kyung Moon*d*,*Classification of breast ultrasound images using fractal feature, ClinicalImage*, Volume 29, Issue4, Pages 234-245.

[10] Suri, J.S., Rangayyan, R.M.: *Recent Advances in Breast Imaging, Mammography,and Computer-Aided Diagnosis of Breast Cancer*. 1st edn. SPIE (2006)

[11] Hoos, A., Cordon-Cardo, C.: *Tissue microarray pro.ling of cancer specimens and cell lines: Opportunities and limitations*. Mod. Pathol. 81(10), 1331–1338 (2001)

[12] Lekadir, K., Elson, D.S., Requejo-Isidro, J., Dunsby, C., McGinty, J., Galletly, N.,Stamp, G., French, P.M., Yang, G.Z.: *Tissue characterization using dimensionality reduction and .uorescence imaging*. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4191, pp. 586–593. Springer, Heidelberg (2006).

Method for Real Time Text Extraction of Digital Manga Comic

Kohei Arai   arai@is.saga-u.ac.jp
Information Science Department
Saga University
Saga, 840-0027, Japan

Herman Tolle   emang@ub.ac.id
Software Engineering Department
Brawijaya University
Malang, 65145, Indonesia

Abstract

Manga is one of popular item in Japan and also in the rest of the world. Hundreds of *manga* printed everyday in Japan and some of printed *manga* book was digitized into web *manga*. People then make translation of Japanese language on *manga* into other language -in conventional way- to share the pleasure of reading *manga* through the internet. In this paper, we propose an automatic method for detect and extract Japanese character within a *manga* comic page for online language translation process. Japanese character text extraction method is based on our comic frame content extraction method using blob extraction function. Experimental results from 15 comic pages show that our proposed method has 100% accuracy of flat comic frame extraction and comic balloon detection, and 93.75% accuracy of Japanese character text extraction.

Keywords: E-comic, Manga, Image Analysis, Text Extraction, Text Recognition

1. INTRODUCTION

Manga is one of popular item in Japan and also in the rest of the world. Hundreds of *manga* book is printed everyday in Japan, and some of printed *manga* book is digitized into web content for reading comic through the internet. People then make translation of Japanese language in manga into other language to share enjoy of reading *manga* for non Japanese reader. However, people make translation of the text on printed comic book (they call it *scanlation*) in manually because there is no automatic method for translate comic text image into any other language. The challenge in extracting Japanese character in *manga* is how to detect comic balloon and extract text in vertical direction as Japanese classic writing direction is top down and right to left.

Several research projects [1-4] proposed method for text extraction from images but not specific for extraction from comic image. There are two base methods for text extraction, using region based method and texture based method. In [5], propose the concept of automatic mobile content conversion using semantic image analysis that include comic text extraction, but this paper did not explain the details for text extraction. Also, Yamada [6] proposed method for comic image decomposition for reading comic on mobile phone that including comic text extraction but not details on comic text extraction. The conventional method assuming extraction process in offline way and using scanned comic image. In the internet and mobility era, we need advance method for extraction text in online way and automatically make translation using online translation feature on internet like Google language translation.

In this paper, an approach for automatically extract Japanese character from *manga* comic is presented. We propose a new method for automatically extract text inside comic balloon from digital comic (e-comic) image in online way. Comic frame contents such us balloon and text inside balloon is extracted for further purpose, for example for language translation, multimedia indexing or data mining. This research work is improvement of our previous research on automatic e-comic content adaptation [7] that designed for extraction comic content from existing comic image in comic web portal and adapting it to mobile content for reading on mobile phone. Our propose method is an efficient and effective text extraction method that sufficient for real time online implementation. The experimental results of our method had shown the better results on accuracy and processing time comparing with other methods.

The reminder of this paper is organized as follows. In section 2, a detail description of comic content extraction method is given, including balloon detection method. Section 3 describes the detail process on text extraction method. Experimental results with comparing to conventional method are presented in Section 4. Finally, conclusions are drawn in Section 5.

2. COMIC FRAME CONTENT EXTRACTION

Extraction of the text from digital comic image page assuming only text situated within a comic balloon which is located in a comic frame. That is why comic frame detection and comic balloon detection has to be done clearly before text extraction as shown on flow diagram of comic text extraction method in Figure 1. We use the same process for comic content extraction [8] on frame extraction but improvement on balloon detection and text extraction. In comic frame content extraction, we extract comic frames and then checking if any overlapped frames situated on extracted frames. If there any overlapped frames are detected, then processed with overlapped frame division. After every frame extracted, all balloon inside a frame and texts inside balloons are extracted. All process is done base on modified connected component labeling (CCL) algorithm [9] as comic blob extraction function.
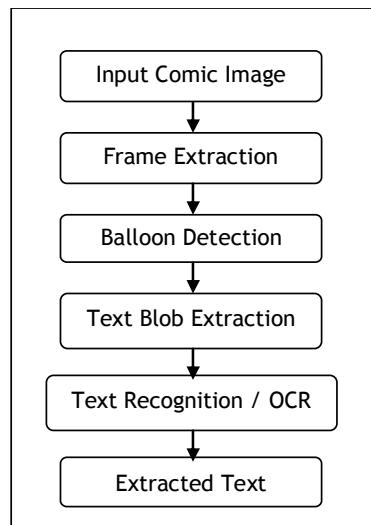
```
┌─────────────────────────┐
│   Input Comic Image     │
└─────────────────────────┘
            │
┌─────────────────────────┐
│    Frame Extraction     │
└─────────────────────────┘
            │
┌─────────────────────────┐
│   Balloon Detection     │
└─────────────────────────┘
            │
┌─────────────────────────┐
│  Text Blob Extraction   │
└─────────────────────────┘
            │
┌─────────────────────────┐
│  Text Recognition / OCR │
└─────────────────────────┘
            │
┌─────────────────────────┐
│     Extracted Text      │
└─────────────────────────┘
```

FIGURE 1: Flow diagram of digital comic text extraction method,

2.1 Comic Frame Extraction

Common comic frames are separated by white pixel line or white region, so the rest of white pixel region must be the frames. We investigated many traditional and conventional comics those in case of there is no balloon or comic art is overlapped on frames - it is called 'flat comic' hereafter-, each frame can be detected as a single blob object. In our propose method, we define all

connected white pixels as a single blob object, and then each comic frames can be identified as an individual blob object.

We modify connected component labeling algorithm [9] for specific function on comic frame blob extraction. Figure 4.a show the flow diagram of the process of modified CCL for comic frame blob extraction function and Figure 4.b shows the results in step-by-step basis. Firstly, binarization is applied to converting color comic images to black and white images. Binarization with an appropriate threshold number produces each frame as separate blobs. The heuristic value of threshold is 250 (for the images with quantization bits of 8 bits) that chosen empirically based on experiments and produced good results. Then color inversion is done to switch color between blobs and background, because our blob extraction method assume black pixel as background color. After that, blob detection process will produce connected pixels as separate blob object. Last process is frame blob selection to select only blob with minimal size that determine as comic frame. The minimal size of selected frame blob is [*Image.Width*/6] x [I*mage.Height*/8].

The proposed methodology has 100% of success rate for extract comic frames from complete flat comic page like comic "Nonono Volume 55" and Dragon Ball Chapter 195 that we use in our experiment. The modified CCL for comic frame blob extraction method is, however, not perfect because comic image includes not only 'flat' frame but also more complicated frame images those are overlapped with other comic balloons or comic arts. Then we improved our comic frame extraction method with overlapped frame checking and extraction using division line detection method [7], but this paper currently focus on the improvement of text extraction method.
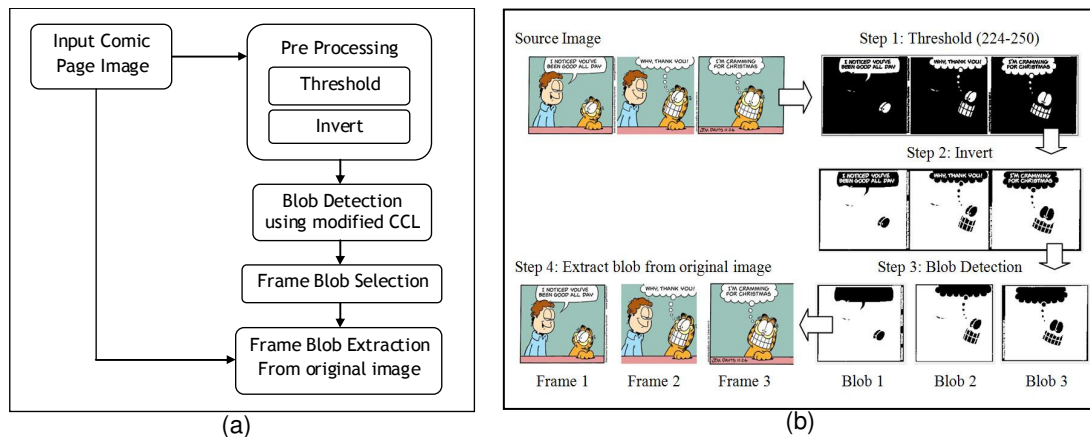

FIGURE 2: Flow diagram of comic frame extraction using comic blob extraction method,
(b). Step-by-step process and result on frame extraction

2.2 Comic Balloon Detection

Comic balloon detection is one of the important processes in comic text extraction, because we only extract text inside a comic balloon. The accuracy of balloon detection is correlated with successfully of text extraction. We use same basic blob extraction function for frame extraction in balloon detection process. In typical comic images, balloon text usually has a white background. So, using base blob extraction method without inversion can detect comic balloon which has white pixel area. Flow diagram of balloon detection is shown in Figure 3.

Blob detection process will detect many blobs, then balloon blob selection is needed to classify and select only balloon candidate. Balloon blob selection is base on 4 rules for classification as follows:
1. *Blob Size (width & height)*: Minimal size of the blob is about [*Image.Width*]/10 and [*Image.Height*]/8 of frame image size.
2. *White pixel occurrence*: Minimal number of white pixels in blob is 45% of blob area.

3. *Straight Line*: At least two straight vertical white lines with 70% long of blob height size is detected. One line in the half left and one in the half right.
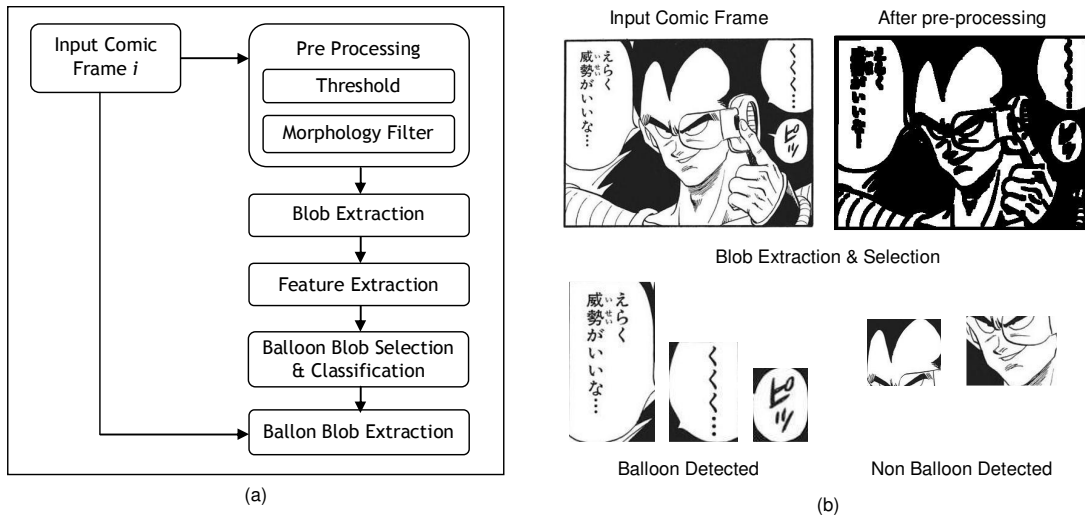4. *Width to Length Ratio*: Blob width size is less than 1.5 of blob height size.

FIGURE 3: Flow diagram of comic balloon detection using comic blob extraction method, (b). Step-by-step process and result on comic balloon detection

3. BLOB EXTRACTION BASED METHODOLOGY FOR TEXT EXTRACTION

The objective of text extraction method is to extract Japanese text from comic balloon as single blob object for same vertical sentence or word. This vertical direction word will useful for character recognizing using optical character recognition (OCR) method which is out of this scope at this time. The proposed methodology is blob extraction based or region based method, and operates on digital comic image which has frame extraction and balloon detection processed previously. The text extraction method comprises of 5 phases; pre processing with binarization and morphology filter, blob extraction, blob feature extraction, text blob selection and classification, and the last process is text blob extraction. The block schematic of the text extraction methodology is given in Figure 4. The detailed description of each phase is presented in the following subsections.
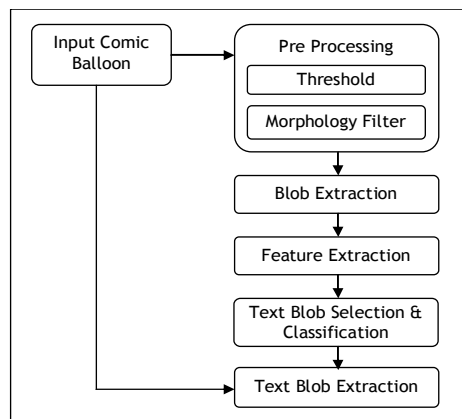
FIGURE 4: Text Extraction Algorithm

3.1 Pre-Processing

Before extracted text blob from comic balloon, pre-processing filtering has to be done for preset image into binary color. First, adaptive threshold method is applied to comic page image. Threshold value T is obtained from average pixel value of comic image μ. We set threshold value is 30% above from the average empirically.

$$\mu = \frac{1}{N}\sum_{i=0}^{N-1} I_i \qquad (1)$$

$$T = \mu / 0.30 \qquad (2)$$

Then we implement a morphology filter in *pre-processing* to merge neighboring character blob image as a single blob. In pre-processing, erosion filter and opening filter is applied with top and bottom side priority more than left and right side because Japanese character reading order in *Manga* is from top to bottom. The block diagram and results sample of pre processing is given in Figure 4.

3.2 Blob Extraction

After pre-processing, we detect all connected pixels as a separate blob object. Blob extraction is done using same blob extraction method with 2 previous processes to extract text blobs candidate from a comic balloon. This process produces text blobs and also non text blobs. To classify a blob as a text blob or non text blob, we extract some features from text blob candidate for classification. In text blob extraction using blob extraction function, we select only blob with minimal size that selected as candidate of text blob. The minimal size of the text candidate blob width is [*Image.Width*]/20 and the text blob height is [*Image.Height*]/40. Parameter of *[Image]* corresponds to balloon blob image size as the input of process.

3.3 Text Blob Feature Extraction and Classification

Text blob classification is needed to classify text blob from extracted blobs into vertical (column) text region. Classification is based on the features of blob size and blob position. Classification process classify text blob with the same column into same cluster as a representative of Japanese classic sentence or word in column. To classify and combine text blob in vertical direction, we extract some features from detected text blob. Features to extract are *average text blob width* and *text blob x center point* (*TB.XCenter*). Average text blob width is needed to determine distance between two columns of vertical text. Average text blob width is chosen from average width of three blobs in the *mean* of blob after sorting all blob based on blob's size. *Text blob x center point* is described in Equation 3. Text blob size of each blob is actual parameter of blob object which is containing *top* and *left* position parameter, and *width* and *height* parameter.

$$\text{TB[i].XCenter} = \text{TB[i].Left} + (\text{TB[i].Width} / 2) \qquad (3)$$

Where;
TB[i] corresponds to i^{th} blob of detected text blob.
TB[i].Left corresponds to left point position parameter of detected text blob
TB[i].Width corresponds to width parameter of detected text blob

For selecting and classifying text blob into vertical text region, we use 2 rules for classification as follows:
1. Blob position is not related with balloon blob border. Assuming minimal distance is d_{min}.
2. Classify blob with same vertical center point in same cluster. Assuming difference distance between centers is BD_{min}.
   If (TB[i].XCenter +- BD_{min}) $\in$ Column[j] Then TB[i] member of Column[j]

Minimal distance from border d_{min} is approximately 5% pixels from balloon edge and minimal distance BD_{min} is approximately less than half of average text blob width.

3.4 Text Blob Extraction

Last step on text extraction method is text blob extraction from original balloon image based on text blob cluster from previous process. In classification process, each blob in same cluster is combining as a new blob as representative of text in vertical. Then using the position of new blobs, text candidate is extracted from original balloon text image from input process. Figure 5 show step by step process and results of *manga* text extraction algorithm. The final result is set of text blob that represent Japanese sentence from comic text balloon in vertical direction. Those image blobs than store in database or directly proceed to OCR application for text recognition.

FIGURE 5: Sample of Text Extraction Process
(a) Original balloon text image; (b) Threshold; (c) Morphology-Erosion filter;
(d) Morphology-Opening filter; (e) Extracted blob from original image

4. EXPERIMENTS

The proposed methodology for extracted Japanese character from digital comic image has been evaluated using various comic image pages. We implement the proposed method using Microsoft.Net environment with C# as native language and AForge image processing framework on desktop computer with Pentium Dual Core processor and 1 mega byte of RAM. Experiment is conducting through 15 comic pages to evaluate the success rate (accuracy) of text extraction. Common comic image size that we use in our experiment is 800x1200 pixels. The results of the experiment then reported and compared with other methods.

4.1 Experimental Results

Performance evaluation of proposed methods on frame extraction, balloon detection and text extraction is evaluated for 15 comic pages from Dragon Ball Chapter 195 Japanese version. This comic contains 15 pages with 60 frames, 70 comic balloons and 160 columns of Japanese sentences. Experimental result of balloon detection and text extraction method has shown in Table 1. The results were classified into 3 groups such as "*correctly extraction*" (*T*), "*missed detection*" (*M*) and "*false detection*" (*F*). In the case of frame extraction, the term *correctly extraction* means the number of true detection of frames; *missed detection* means that there are frame that not detected by system; and the terms *false detection* means that some non frames are detected as frame. We also use the same interpretation on *T*, *M* and *F* for balloon detection and text extraction process. In Table 1, there are 2 methods, base method and modified method. Base method is our method which is explained in previous chapter, while *modified method* is base method with modification. Detail of modified method explained in *Discussion* on next sub chapter. From the experimental results, 100% of success rate of frame extraction method, 100% of success rate of comic balloon detection method and 93.75% of success rate of manga text extraction methods is achieved.

Table 1. Comic Balloon and Text Extraction experimental results

Comic Content	Total	T	M	F	%
Frame Extraction	60	60	0	0	100

Balloon Detection based method	70	70	0	53	100
Balloon Detection with modification	70	70	0	7	100
Text Extraction	160	150	10	6	93.75

Time consuming in processing is main issue in real time online application. We evaluate processing time of our method in offline simulation for each process on comic text extraction without text recognition process. The result of processing time evaluation is shown in Table 2, with comparison with other method. Processing time experimental results shows that our proposed method is faster that other method. Comparing with conventional method, processing time of our method is about 90% faster than conventional method. Online situation need more processing time consuming rather than offline situation because of another processing within the systems, but still acceptable as online application.

Table 2. Processing time experimental results and comparison

Comic Page	Processing Time (minutes)	
	Conventional Method	Our Method
15 comic page offline with 70 balloon	Approx. > 30	2.5

4.2   Discussion

Although our base method for balloon detection and text extraction has performed good results on *correctly extraction*, it still has a lot of false detection. In balloon detection method, there are a lot of non balloon detected as balloon candidates. It causes by the form of comic image that looks like balloon and pass our selection criteria. To reducing false detection, we make a simple modification for balloon detection criteria by adding one process. The new process is performing text detection on balloon candidate. If a candidate of text detected inside a balloon candidate, then classify it as a new balloon candidate. If there is no text blob candidate detected inside a balloon, then classify it as a non balloon. Implementing of this modified method, reduce about 90% of false detection. Figure 6 shows the sample of false detection of balloon candidate, which eliminated on text detection. Actually, the occurrence of false detection in text extraction process is not a serious problem because in the next process while we implement OCR for extract text from text blob image, the false detected of balloon or text produces nothing after OCR processing. In the other hand, the accuracy of text extraction method is 93.75% while 6.25% of texts are not detected or system fail to detect and extract. Most of the failure on text extraction is cause by non standard of position of the character within a balloon as shown in Figure 6.b. System also fail to extract 2 or more text column that close to another.

Implementation of our method is possible in offline or online real time basis because the simplicity of the program code and fast processing time. By using our method, vertical Japanese sentence in a comic balloon is easy to extract then recognized for language translation. We can create a mobile application for reading comic with language translation process by implementing of our method combine with OCR feature and *Google Language Translation* tool using mobile programming *framework* like *Android* or another framework.
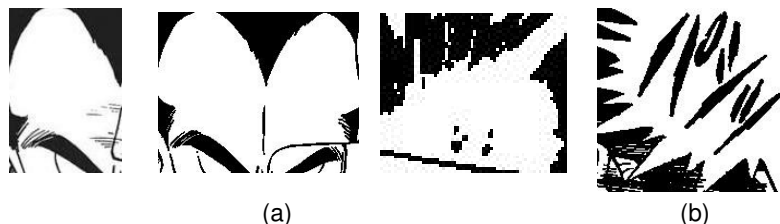

(a)   (b)

FIGURE 6: (a) Sample of non balloon detected as a balloon candidate (false detection);
(b) sample of miss detection text

5. CONCLUSION & FUTURE WORK

We implemented a system for automatically extract Japanese character from digital manga comic image. Our proposed method can automatically detect comic frame and detect all balloon within a comic frame. After all balloon detected, then system will extract Japanese character text inside a balloon as comic dialog in Japanese language for further purpose especially for language translation. The proposed method on Japanese character text extraction is possible to implement in offline or online real time basis so it is possible to become an online application for language translation. From the experimental results, our comic frame extraction method has 100% accuracy for flat comic, balloon detection method achieves 100% accuracy and text extraction method achieves 93.75% accuracy.

Combining the proposed method with our e-comic content adaptation system, will create a robust system for automatically convert digital comic image from comic web portal into mobile content with language translation features. We will improve this current method with implementing character recognition function in the systems. Also for further work, we will implement an application for automatically convert Japanese comic into international comic for enjoy reading comic anywhere anytime using mobile phones.

6. REFERENCES

1. L. Fletcher and R. Kasturi, A robust algorithm for text string separation from mixed text/graphics images. IEEE Trans. Pattern Anal. Mach. Intell. 10 (1988), pp. 910–918

2. A.K. Jain and B. Yu, Automatic text location in images and video frames. Pattern Recognition 31 12 (1998), pp. 2055–2076

3. O. Iwaki, K. Kubota and H. Arakawa, A character/graphic segmentation method using neighborhood line density. IEICE Trans. Inform. Process. J68 4 (1985), pp. 821–828.

4. L.A. Fletcher and R. Kasturi, A robust algorithm for text string separation from mixed text/graphics images. IEEE Trans. Pattern Analysis Mach. Intell. 10 6 (1988), pp. 910–918

5. Eunjung Han, et.al. "Automatic Mobile Content Conversion Using Semantic Image Analysis", Human-Computer Interaction HCI Intelligent Multimodal Interaction Environments, LNCS 4552, Springer, Berlin, 2007

6. Yamada, M., Budiarto, R. and Endoo, M., "Comic image decomposition for Reading comics on cellular phones". IEICE transaction on information and systems, E-87-D (6):1370-1376, June 2004.

7. Kohei Arai, Tolle Herman, "Method for Automatic E-Comic Scene Frame Extraction for Reading Comic on Mobile Devices," itng, pp.370-375, 2010 Seventh International Conference on Information Technology, 2010

8. Kohei, A., Tolle, H., "Automatic E-Comic Content Adaptation", International Journal of Ubiquitous Computing IJUC Volume (1): Issue (1), May 2010.

9. F. Chang, C-J. Chen and C-J. Lu. "A Linear-Time Component-Labeling Algorithm Using Contour Tracing Technique", Computer Vision and Image Understanding, 93(2):pp. 206-220, 2004.

10. R. Gonzalez and R. Woods. "Digital Image Processing", Addison-Wesley Chap.2., Publishing Company (1992)

CALL FOR PAPERS

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

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

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

IJIP List of Topics

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

- Architecture of imaging and vision systems
- Character and handwritten text recognition
- Chemistry of photosensitive materials
- Coding and transmission

- Color imaging
- Data fusion from multiple sensor inputs
- Document image understanding
- Holography

- Image capturing, databases

- Image processing applications

- Autonomous vehicles
- Chemical and spectral sensitization
- Coating technologies
- Cognitive aspects of image understanding
- Communication of visual data
- Display and printing
- Generation and display
- Image analysis and interpretation
- Image generation, manipulation, permanence
- Image processing: coding

- Image representation, sensing

- Implementation and architectures
- Materials for electro-photography

- New visual services over ATM/packet network
- Object modeling and knowledge acquisition
- Photographic emulsions
- Prepress and printing technologies
- Remote image sensing
- Storage and transmission

- analysis and recognition
- Imaging systems and image scanning
- Latent image
- Network architecture for real-time video transport
- Non-impact printing technologies
- Photoconductors

- Photopolymers
- Protocols for packet video
- Retrieval and multimedia
- Video coding algorithms and technologies for ATM/p

IMPORTANT DATES

Volume: 5
Issue: 1
Paper Submission: January 31, 2011
Author Notification: March 01, 2011
Issue Publication: March /April 2011

CALL FOR EDITORS/REVIEWERS

CSC Journals is in process of appointing Editorial Board Members for **International Journal of Image Processing (IJIP)**. CSC Journals would like to invite interested candidates to join **IJIP** network of professionals/researchers for the positions of Editor-in-Chief, Associate Editor-in-Chief, Editorial Board Members and Reviewers.

The invitation encourages interested professionals to contribute into CSC research network by joining as a part of editorial board members and reviewers for scientific peer-reviewed journals. All journals use an online, electronic submission process. The Editor is responsible for the timely and substantive output of the journal, including the solicitation of manuscripts, supervision of the peer review process and the final selection of articles for publication. Responsibilities also include implementing the journal's editorial policies, maintaining high professional standards for published content, ensuring the integrity of the journal, guiding manuscripts through the review process, overseeing revisions, and planning special issues along with the editorial team.

A complete list of journals can be found at http://www.cscjournals.org/csc/byjournal.php. Interested candidates may apply for the following positions through http://www.cscjournals.org/csc/login.php.

Please remember that it is through the effort of volunteers such as yourself that CSC Journals continues to grow and flourish. Your help with reviewing the issues written by prospective authors would be very much appreciated.

Feel free to contact us at coordinator@cscjournals.org if you have any queries.

Contact Information

Computer Science Journals Sdn BhD
M-3-19, Plaza Damas Sri Hartamas
50480, Kuala Lumpur MALAYSIA

Phone: +603 6207 1607
      +603 2782 6991
Fax: +603 6207 1697

BRANCH OFFICE 1
Suite 5.04 Level 5, 365 Little Collins Street,
MELBOURNE 3000, Victoria, AUSTRALIA

Fax: +613 8677 1132

BRANCH OFFICE 2
Office no. 8, Saad Arcad, DHA Main Bulevard
Lahore, PAKISTAN

EMAIL SUPPORT
Head CSC Press: coordinator@cscjournals.org
CSC Press: cscpress@cscjournals.org
Info: info@cscjournals.org