PROCEEDINGS OF SPIE

Second International Conference on Digital Image Processing

Kamaruzaman Jusoff
Yi Xie
Editors

26–28 February 2010
Singapore

Organized by
International Association of Computer Science and Information Technology (Singapore)

Published by
SPIE

Volume 7546
Part Two of Two Parts

Proceedings of SPIE, 0277-786X, v. 7546

The papers included in this volume were part of the technical conference cited on the cover and title page. Papers were selected and subject to review by the editors and conference program committee. Some conference presentations may not be available for publication. The papers published in these proceedings reflect the work and thoughts of the authors and are published herein as submitted. The publisher is not responsible for the validity of the information or for any outcomes resulting from reliance thereon.

Please use the following format to cite material from this book:
Author(s), "Title of Paper," in *Second International Conference on Digital Image Processing*, edited by Kamaruzaman Jusoff, Yi Xie, Proceedings of SPIE Vol. 7546 (SPIE, Bellingham, WA, 2010) Article CID Number.

ISSN 0277-786X
ISBN 9780819479426

Published by
SPIE
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone +1 360 676 3290 (Pacific Time) · Fax +1 360 647 1445
SPIE.org

Printed in Singapore.

Publication of record for individual papers is online in the SPIE Digital Library.

SPIE Digital Library

SPIEDigitalLibrary.org

Paper Numbering: Proceedings of SPIE follow an e-First publication model, with papers published first online and then in print and on CD-ROM. Papers are published as they are submitted and meet publication criteria. A unique, consistent, permanent citation identifier (CID) number is assigned to each article at the time of the first publication. Utilization of CIDs allows articles to be fully citable as soon they are published online, and connects the same identifier to all online, print, and electronic versions of the publication. SPIE uses a six-digit CID article numbering system in which:

- The first four digits correspond to the SPIE volume number.
- The last two digits indicate publication order within the volume using a Base 36 numbering system employing both numerals and letters. These two-number sets start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B ... 0Z, followed by 10-1Z, 20-2Z, etc.

The CID number appears on each page of the manuscript. The complete citation is used on the first page, and an abbreviated version on subsequent pages. Numbers in the index correspond to the last two digits of the six-digit CID number.

Contents

Part One

SESSION 4

SESSION 5

SESSION 6

SESSION 7

Part Two

SESSION 8

SESSION 9

SESSION 10

SESSION 11

SESSION 12

SESSION 13

The use of images CBERS 2 and CBERS 2b in identification of areas affected by desertification

SILVA Claudionor Ribeiro, CASTRO Fabiana Silva Pires and CENTENO Jorge Antonio Silva

Federal University of Piauí, Bom Jesus/PI, 64900-000, Brazil
Brazilian Institute of Geography and Statistics (IBGE), Porto Alegre/RS, 90010-390, Brazil
Federal University of Paraná, Curitiba/PR, 90001, Brazil
Email: crs@ufpi.edu.br, fabiana.piresc@gmail.com, centeno@ufpr.br

ABSTRACT

The process of desertification, which extends from a long time ago, became a reality in Brazil. This phenomenon can be understood as land degradation, caused by factors including climatic changes and human activities. Besides being a threat to biodiversity, causes loss of soil productivity, threatening the lives of thousands of people living in affected regions. So, the identification of affected areas is essential to diagnose and prevent the problem. Satellite image has been a source of relatively low cost and widely used in this task. Therefore, is proposed in this study, a method to extract automatically areas heavily affected by desertification. The method is based on concepts of mathematical morphology, vegetation index and classification of digital images. Experiments are conducted separately, with images of CBERS 2 and 2B, and subsequently compared. The validation is done by crossing the results obtained with a reference image, created by a manual process.

Keywords: desertification, digital image, mathematical morphology, NDVI

1 INTRODUCTION

Desertification is defined as land degradation, affecting water resources, vegetation, biodiversity and reducing the quality of life of the population, and is directly related to human activities. This phenomenon is linked to climate changes and is related to the loss of soil productivity in arid, semi-arid and dry sub-humid areas, causing environmental and social problems such as hunger, forcing habitants to leave the region and causing immigration waves and sometimes causing death [1]. About one sixth of the world is already affected by desertification, which means about 33% of the surface.

Some activities, such as overgrazing and increase of the salinity of soils by irrigation turn about 6 million hectares into unproductive land per year. In addition to economic losses in terms of production, the cost of recovery of the affected areas is high. Asia and Africa are the most affected continents by desertification, but Latin America is already significantly affected.

The northeast region of Brazil and also the state of Rio Grande do Sul have areas that face the desertification problem. Most of these areas lie in the northeast region, with limited productive potential and arid, semi-arid and sub-humid climates. These areas are characterized by high evapotranspiration, the occurrence of periods of drought, shallow soils, high salinity, low fertility and low capacity to water retain. A total area of 223.000Km^2 is seriously affected

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754620 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856008

by the process in the northeast and is now useless for agriculture.

To solve problems associated to desertification, it is important to measure the location of the affected areas. However, traditional methods of measurement (geodetic and / or topographical surveys) are expensive and exhaustive. Processing remote sensing images appears as a useful tool to perform this task. The free availability of digital images, such as CBERS data are examples of more economic solutions for this task.

1.1 CBERS Satellite

The new series of CBERS-2B satellite, developed by the Brazil / China cooperation was launched in order to continue the imaging program, started in 1999 with the CBERS-1. The CBERS-2B mission is almost identical to CBERS-1 and CBERS-2, but some improvements were introduced. The main innovation refers to the payload, with the replacement of the IRMSS (Multispectral Scanner for High Resolution) and the High Resolution Panchromatic Camera (HRC). Other improvements are a new recording system on board and an advanced positioning system.

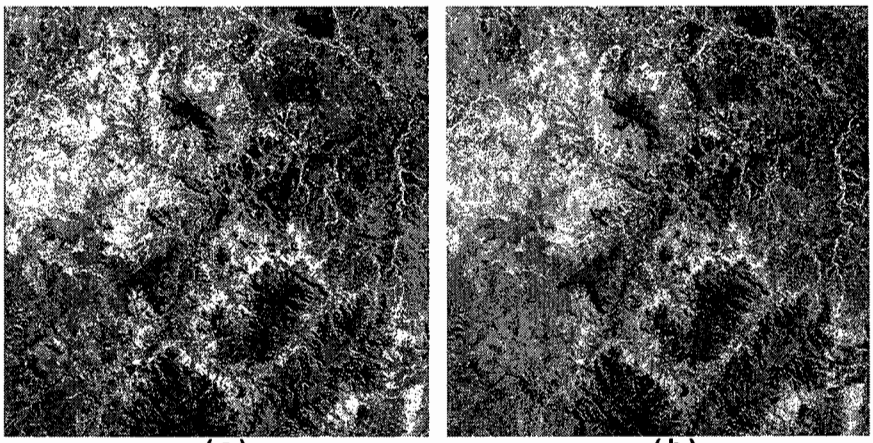
To accomplish the imaging, the CBERS-2B has three cameras: Camera of High Resolution Imaging (CCD), Panchromatic Camera of High Resolution (HRC) and Board Imager Wide Field Imager (WFI). The main one is the CCD camera, responsible for imaging a swath of 113 km wide with 20 m spatial resolution. The images generated by this camera are commonly used in tasks such as analysis of vegetation, water and soil, as well as mapping and related fields. The second camera (HRC) was designed to collect information with spatial resolution of 2.7 m. Finally, the third camera (WFI) collets image with lower spatial resolution, but higher coverage. Due to the low spatial resolution and larger swath, the applications of the images acquired by WFI are generally for macro-regional or state scales [2].

The main aim of this work is the identification of areas affected by desertification process through the use of satellite images. It is also aimed at the analysis of the potential of the images CBERS 2 and 2b in the identification of areas affected by desertification, as well as the behavior of this feature (desertification) in the images of both sensors in order to compare them.

2 MATERIALS AND AREA OF STUDY

The tests were performed using the software ENVI 4.1 and routines developed in MATLAB. The test area lies in southern Piauí (Brazil), in the municipality of district of Gilbués, Brazil. This area lies between the coordinates UTM (central meridian 23): E1 = 445460,00 m/ N1 = 8924820,00 m (left superior corner) and E2 = 485460,00 m/ N2 = 8884820,00 m (right inferior corner).

Two digital images was used, one CCD sensor of the satellite CBERS-2 (FIGURE 1a) and another corresponding to the same area, also the CCD sensor, but of the satellite CBERS-2B (FIGURE 1b), both with a spatial resolution of 20 meters. Although the images CCD / CBERS consist of four spectral bands related to the wavelengths of blue, green, red and infrared, only the latter three were used, by giving more information about the feature of interest.



(a) (b)

FIGURE 1 – Images (a) CBERS 2 e (b) CBERS 2B.

These images CBERS-2 and CBERS-2B have the same size, with 2000x2000 pixels and were taken at 29/07/2008 and 30/07/2008, respectively.

3 METHOD

The proposed method in this work involves three distinct stages. The first is the development of a reference image and the registration of the second image. The second step is the production of thematic maps derived from the application of NDVI and the classification with maximum likelihood (ML). Finally, the third step consists of the application of morphological operators to smooth the contours of the classes.

First, a reference image was geocoded. The reference image is a CBERS 2b scene, different from the other two images used in the classification step and in the generation of NDVI. This image was used as the basis of the registration process of the CBERS 2 and 2B images used in the experiments. A total of 09 control points were used in the image to image registration and the resulting mean square error (RMS) was lower than one pixel.

Then, a reference thematic map was obtained from the reference image (FIGURE 2) by a manual process, analyzing the image on the screen. The result is a thematic image that contains two classes: areas affected by desertification were labeled with the digital value 1 (one) and the remaining area were labeled with 0 (zero).

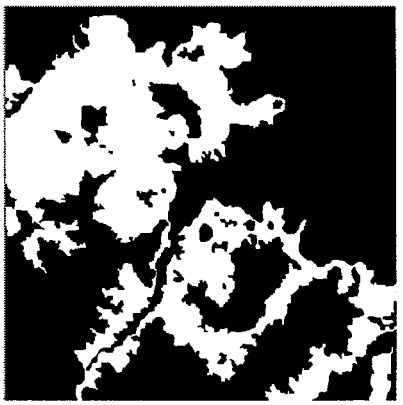


FIGURE 2 – Reference image.

3.1 The Normalized Difference Vegetation Index – NDVI

In the second step the NDVI was computed for the two georeferenced images. NDVI is an index that detaches the vegetation in relation to other existent features in the image. This index is determined by the ratio (EQUATION 1) between the spectral bands related to the wavelengths 0.63-0.69 μm (red - R) and 0.76-0.90 μm (infrared - IR), at the electromagnetic spectrum.

$$NDVI = \frac{IR - V}{IR + V} \quad (1)$$

Because the vegetation shows higher reflectance values in the near infrared and lower reflectance values in the visible region of the spectrum, the NDVI values corresponding to vegetation are high, which provides the highlighted of this feature in relation to the others in the image. The main feature of NDVI is to enhance the vegetation in relation to other features in the image. The images used in the experiments, were acquired in dates when the vegetation is more vigorous (green) and the desertification areas, due to the inherent properties of the process, have similar characteristics to the exposed soil. In the NDVI image, the areas containing vegetation have values close to 1 (one) and the remaining areas have lower digital values.

The identification of desertification areas was accomplished through thresholds. Three thresholds were defined empirically with values 0.18, 0.20 and 0.25. Each threshold generates a binary image, where the feature "desertification" has value 1 (one) and the rest of the image value 0 (zero). In the binarization the value 1 (one) is assigned to all pixels that have digital values smaller than the threshold, and 0 (zero) for the other pixels. Therefore, were generated three binary images for each scene.

3.2 Method of Classification

In addition to the binary NDVI images it was obtained two other thematic images classifying the multispectral CBERS images. The Maximum Likelihood (ML) classifier was applied on the two images, CBERS 2 and CBERS 2B, producing two thematic maps that were later binary coded by assigning the value 1 (one) for the class "desertification" and value 0 (zero) to the other existing classes in the thematic map.

Classification is defined as a method used to group the pixels of an image, based on their spectral characteristics[3]. The grouping is performed by a computer program enabled to identify spectrally similar pixels. The digital classification techniques require the implementation of a decision process so that the computer can assign a set of pixels to a particular class [4].

There are two classification procedures for the analysis of multispectral data [3]. In the first one, the unsupervised classification, the pixels are associated with a spectral class without the influence of the operator. The second method is the supervised classification, where the user is able to inform about the classes and its spectral properties.

In the present study it is used the Maximum Likelihood statistical classifier. This classifier describes each spectral class through a probability distribution. The distribution describes the chance of a pixel to belong or not to a particular class in any position of a multispectral image. This method is based on the criterion function, which evaluates the incorrect decisions, according to the decision rule presented in EQUATION 2. Thus, the pixel x_o will be allocated to the class which the rule is accepted [3].

$$x \in C_i \quad se \quad -\ln|S_i| - (x_o - \bar{x}_i)S_i^{-1}(x_o - \bar{x}_i) > -\ln|S_j| - (x_o - \bar{x}_j)S_j^{-1}(x_o - \bar{x}_j) \qquad (2)$$

$$\forall j \neq i \qquad i, j = 1,2,\ldots, m$$

where C_i are the existing classes, x_o is the vector corresponding to the pixel to be classified, $\bar{x}_i$, $\bar{x}_j$ are the mean vectors of classes i and j, respectively; S_i^{-1}, S_j^{-1} are the variance-covariance matrix of the observations and, m is the total number of existing classes.

3.3 Mathematical Morphology

In the third step, post-classification, procedures to reduce or eliminate isolated points and areas in the classified image were used. It is pointed out that the reduction or elimination of isolated points can eliminate areas correctly classified, however, if there are a large number of isolated points, the application of this technique can increase the final quality of the thematic map. The homogenization is performed applying mathematical morphology. In this work we used the morphological operators opening and closing. Several tests using different structuring elements were performed to find the best solution for the image used. The size and shape of the structuring elements are defined by the user, according to the area to be applied. The product of this processing is an image free, partial or totally, of unwanted features.

The mathematical morphology has been widely used in digital image processing. This technique allows the extraction of image components that are useful in the representation and description of the shape of a region, such as borders and skeletons [5]. The extraction of elements in image is accomplished with the help of a suitable structural element. The structural elements are matrices responsible for the removal or addition of labeled pixels in the image, which depend on their size and shape [6].

In this paper the mathematical morphology adopted is binary, thus only binary morphological operators were used. The basic operations of morphology are erosion and dilation. These operations are the basis for most of the operations used in mathematical morphology, in other words, they are combinations of these such as opening, closing, among others. The opening operator is used to remove parts of objects or even objects smaller than the structuring element [6]. Thus, the

opening operator can eliminate noise due to the erosion operator that is applied initially. The opening of a set A by B, denoted by A ° B is given by EQUATION 3.

$$A \circ B = (A \ominus B) \oplus B \quad (3)$$

where A represents the image being operated on and B is called structuring element.

The closing operator tends to join "islands" which the distance between them is less than the structuring element and closing holes smaller than this element. Being the same set A and a structuring element B, the closing of A by B, denoted by A • B, is given by EQUATION 4.

$$A \bullet B = (A \oplus B) \ominus B \quad (4)$$

4 RESULTS AND DISCUSSION

The NDVI and ML images modified by the morphological operators were compared with the reference image. The results can be seen in the following sections.

4.1 Analysis of NDVI images

The results for the CBERS-2 and 2B images are shown in FIGURES 3 and 4. It is observed that for each adopted threshold a different NDVI map was produced. Morphologic operators of opening and closing were applied in all the images obtained with NDVI, in order to reduce or to eliminate isolated points.

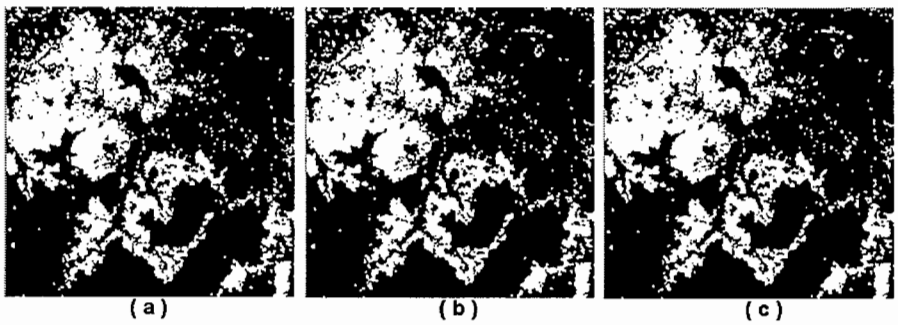


FIGURE 3 – Images from CBERS-2: (a) NDVI 0.18 (b) NDVI 0.20 and (c) NDVI 0.25.

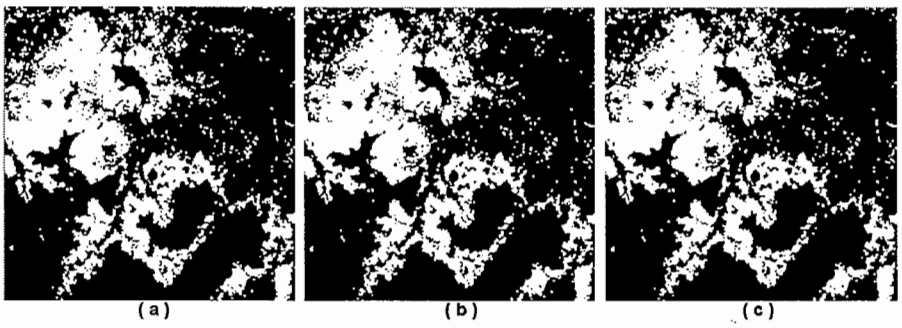


FIGURE 4 – Images from CBERS-2B: (a) NDVI 0.18 (b) NDVI 0.20 and (c) NDVI 0.25.

The results obtained in the comparison among the images NDVI and to reference images can be seen in the TABLES 1 and 2.

TABLE 1 -Accuracy in the classification for NDVI/CBERS2

	Threshold 0.18	Threshold 0.20	Threshold 0.25
Class accuracy	0.5696	0.6626	0.8520
Global accuracy	0.8227	0.8445	0.8524

TABLE 2 - Accuracy in the classification for NDVI/CBERS2B

	Threshold 0.18	Threshold 0.20	Threshold 0.25
Class accuracy	0.7373	0.8220	0.9478
Global accuracy	0.8639	0.8728	0.8143

The values of the Table 1 show that the best result, accuracy of 0.8520, was obtained when the threshold 0.25 was used. The same threshold (0.25) also leads to the best global accuracy of 0.8524. The values of Table 2 also show that the best result for the class accuracy is obtained using the threshold 0.25, but the best global accuracy, 0.8728, belongs to the threshold 0.20.

4.2 Analysis of the images classified with ML

The images classified by ML are illustrated in the FIGURE 5.

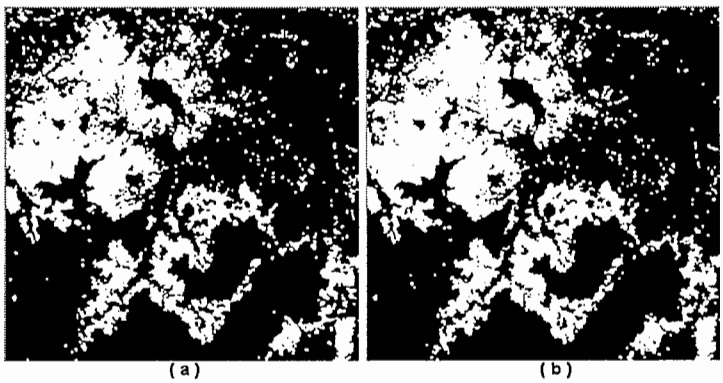


FIGURE 5 – The ML images from (a) CBERS-2 and (b) CBERS 2B.

The results obtained by the comparing of the ML images and the reference image can be seen in the TABLE 3, that shows that the higher accuracy, class accuracy and global accuracy, is obtained using the CBERS-2B images.

TABLE 3 - Precision in the classification for ML

	CBERS 2	CBERS 2B
Class accuracy	0,7643	0,8237
Global accuracy	0,8616	0,8761

5 CONCLUSIONS

Although the sensors of the satellite CBERS 2 and 2B theoretically have similar characteristics, we found different results for the same area. A possible explanation is that the sensors can be affected by external factors such as

the atmospheric effects. The images obtained by remote sensors are the product of interactions of radiation with the atmosphere and objects in the earth's surface. Such interactions can cause noises that interpose the obtaining information about the targets of interest. Among the bands used for the NDVI computation, the red band may suffer greater influence of the atmospheric effect. This influence can lead greater or lesser contrast between the scenes, regardless of the similarities between the characteristics of the sensors, thus affecting the results of digital processing techniques. Looking at the results of each sensors, it can be concluded that both sensors can be used to identify areas affected by desertification.

REFERENCES

[1] USGS, 2009. United States Geological Survey. http://pubs.usgs.gov/gip/deserts/desertification (accessed 15 ago. 2009).

[2] INPE, 2008. Instituto Nacional de Pesquisas Espaciais. São José dos Campos/SP, Brazil. http://www.inpe.br/crc/satelites.php (accessed 10 jun 2009).

[3] Richards, J.A.; Jia, X. [Remote sensing digital image analysis: an introduction]. Springer, Berlin, pp 363 (2006).

[4] Novo, E. M. [Sensoriamento Remoto: princípios e aplicações]. Edgard Blücher. São Paulo, pp 380 (1989).

[5] Gonzales, R. C., Woods, R. E. [Processamento de imagens digitais]. Edgard Blücher. São Paulo, pp 509 (2000).

[6] Dougherty, E. R. [An Introduction to Morphological Image Processing]. SPIE-International Society for Optical Engine, pp161. (1992).

Cell quantification and watershed segmentation in time lapse microscopy

RM.Suresh[1] , N.Jayalakshmi[2]

[1] Vice Principal, RMD Engineering College, Kavarapettai, Chennai, Tamilnadu
[2]Assistant Professor, Department of Computer Applications,Velammal College of Management and Computer Studies, Ambattur Redhills Road, Chennai, Tamilnadu, India.
(e-mail: jayamaithu@rediffmail.com)

ABSTRACT

Because of the complex nature of cells, the ability to distinguish a cell from the background of an image for automatic quantification remains a challenging task. Here, we describe a new technique for cell segmentation using an extended h-maxima transformation to find possible cell locations and a watershed algorithm. A novel method that is adopted to track the cells between image sequences is also discussed in this paper. The over segmentation problem of watershed algorithm is reduced by morphologic erosion, allowing for more accurate quantification, even in low contrast images. The number of cells and the average cell size could be determined in the image. Application of this method to a difficult dataset allowed us to identify 96% of the cells in the image and showed promising results for tracking cells between consecutive images.

Keywords: Stem cell detection, counting, image enhancement and segmentation.

1. INTRODUCTION

One of the most important and common tasks for biomedical researches is cell tracking. Manual tracking is a tedious and painful task due to the corrupted or blurred images. As a result, automated cell tracking systems are mandatory to further advance the study of biological cells.

Cell segmentation is the basis of cell identification and tracking. The cells under study may cluster or overlap with each other. So, the cell detection prior to segmentation is vital. The use of an intelligent computational tool that realizes the cells classification and counting is very useful to help in qualitative and quantitative aspect of final diagnosis.

Based on the observations a four-step algorithm is thus proposed. First cell shape information is obtained with a morphological processing and then enhanced using top hat and bottom hat filtering techniques to remove non-uniform illumination effects. Second, both intensity and shape information is used for extended h-maxima transform to identify the cell locations. Third, the detected cells were segmented via watershed algorithm. Finally the cells are tracked using modified Mahalanobis distance algorithm.

The next section describes the background of the system. The third section briefly describes the methods used in the study, while result section shows some results of the segmentation and tracking. Final section concludes the paper.

2. BACKGROUND

The segmentation process of medical images is a difficult task to be accomplished in digital image processing. This difficulty forces, in most situations, the use of very specific segmentation algorithms for each task or image being analyzed [1][2]. Thresholding methods have been used by Kittler et al [3], Otsu et al [4] and Wu et al [5]. Markiewicz et

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754621 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853427

al [6] have used watershed for segmentation of the bone marrow cells. Comaniciu et al [7] proposed a mean shift method for cell image segmentation for diagnostic pathology. Geusebroek et al [8] introduced a method based on nearest neighbor graphs to segment the cell clusters. Meas-Yedid et al [9] proposed a method to quantify the deformation of cells using snakes.

Combining segmentation and tracking of objects in image sequence results in faster and more reliable results. Zimmer et al [10] introduced a method using active contours/snakes. The drawback with this method is its inability to deal with splitting cells. Kirubarajan et al [11] demonstrates the segmentation and tracking algorithm where standard thresholding combined with clustering results in coarse segmentation, which is refined further using the result of association algorithm and tracking is done using modified Kalman filter. K.Althoff et al [3] used combination of nearest neighbour techniques and correlation matching techniques. Nearest neighbour techniques require low cell density and high temporal sampling rate, while correlation matching instead requires the individual cell features to be constant between two consecutive frames. Amalka Pinidiyaarachchi et al [13] used auction algorithm. Omar Al-Kofahi et al was tracking the cells by multiple object matching method using Gaussian distribution. Tomohiro Yasuda et al [14] used Euclidean distance algorithm for cell tracking.

3. MATERIALS AND METHODS

The data set we have used contains time-lapse microscopic images of Swiss Webster mouse at embryonic day E10-E12. The tissue was processed and collected for plating which contains 95% of single cells. The size of each image is 480 X 640 pixels, and the depth of each pixel is 8 bits [15].

3.1. Enhancement

Mathematical morphology is well suited to biological and medical image analysis. This system uses morphological techniques for shape analysis and filtering. Mathematical morphology is a non-linear process that operates the image with a sub image called morphological structuring element[16]. The morphological operations realized in this study are open, top hat and bottom hat. These are used to enhance the image and the they find the disk shaped objects with disk size 15. Top hat filtering with a disk-shaped structuring element is used to remove the uneven background illumination from the image. Then bottom hat subtracts the original image from a morphologically closed version of the image. The image is then applied to the inverse operation. Since the intensity image contains many local maxima there are numerous additional regions, not belonging to the cell. Those are merged to neighboring regions until only regions originated from the cells remain. The result of morphological enhancement is shown in fig.2 for the image in fig.1.

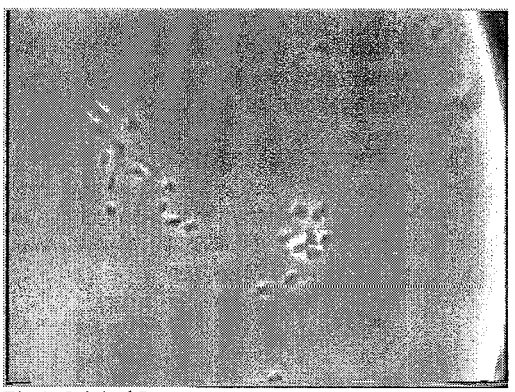


Fig.1. Original Image

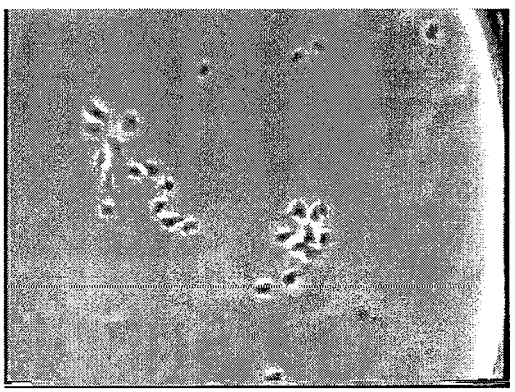


Fig.2. Enhanced Image

3.2. Cell Detection & Counting

Stem Cell counting is commonly used in various pathological and other biomedical researches. Generally we can roughly estimate the cell number with eyes under microscope. This method is usually tedious, time consuming and bad for eyes. In this paper we proposed a method to count the cells automatically.

The possible cell locations in an image were found using extended h-maxima transform method [17], i.e. to find seed points. The h-maxima transform suppresses all maxima whose height compared to the surrounding pixels are smaller than a given threshold h. The h-maxima is calculated using

$$HMAX_h(I) = R^{\delta}{}_I(I-h) \quad (1)$$

Where $R^{\delta}{}_I(I-h)$ is the geodesic dilation of the image I-h with respect to I iterated until the resulting image does not change. A geodesic dilation of image I with respect to image J is defined as

$$\delta_J{}^{(1)}(I) = \min(\delta^{(1)}(I), J) \quad (2)$$

Where $\delta^{(1)}(I)$ is the dilation of I. The extended h-maxima is defined as the regional maxima of an image I is defined as

$$RMAX(I) = I - R^{\delta}{}_I(I-1) \quad (3)$$

The h-maxima of an image with 20 as h value is shown in fig.3. The extended h-maxima binary image shown in fig.4 shows the places where the cells exist.

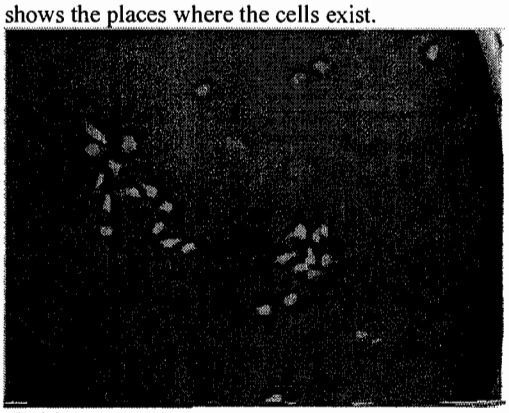


Fig.3 h-max

Fig.4 Extended h-maxima

Individual cell seeds are found by extracting isolated regions in the binary image using 8-connectivity. The final background seed is constructed by imposing the binary image on the original image followed by a removal of holes and isolated regions smaller than 20 pixels by considering them as noise.

3.3 Segmentation

The above section finds the suitable seed points for watershed segmentation through h-maxima transform, where h is fixed. We require the seeds to have intensity above the threshold h in the transformed image to use them for the watershed transform. The seeds are imposed on original image, which is shown in fig.5.

An intuitive model of watersheds is to see the image as a topographic surface, with both valleys and mountains. Assume that there is a hole in each maxima and the surface is immersed into a lake, the water will enter through the holes at the minima and flood the surface. To keep the water coming out of two different minima from meeting, a dam is built, wherever there would be a merge of the water. Finally, the only thing visible of the surface would be the dams. These dam walls are called the watershed lines [6]. Allowing the water to rise in all intensity minima will most often result in over segmentation. To overcome this problem, the water is allowed only to rise in certain points, called seeded

watershed segmentation. Fig.6 shows the result of seeded watershed segmentation using the seed points from the last section.

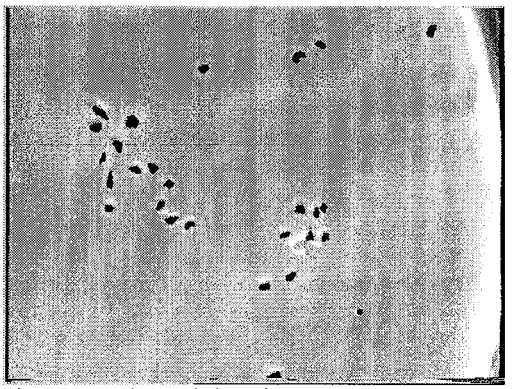


Fig. 5 Imposed extended maxima on Original.

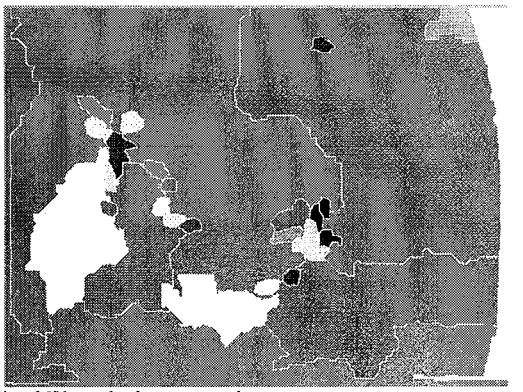


Fig. 6 Watershed segmentation

3.4 Cell Tracking

In the sequence of time-lapse microscopic images we have observed the appearance of new cells. The main reason is that there can be a biological cell division or the cell can be moved into the scene from outside. Another reason for cell split through over segmentation. Cell splits are of biological interest and therefore some efforts have been made to identify when the cell split has occurred. This has been found by comparing the characteristics of consecutive images before and after the split occurred. Different features like the centroid, major and minor axis length, eccentricity and orientation are calculated and used for associating the cells between images. The cells of the consecutive images are mapped and tracked using modified Mahalanobis distance algorithm. The result of splitting and mapping is given in Fig.7.

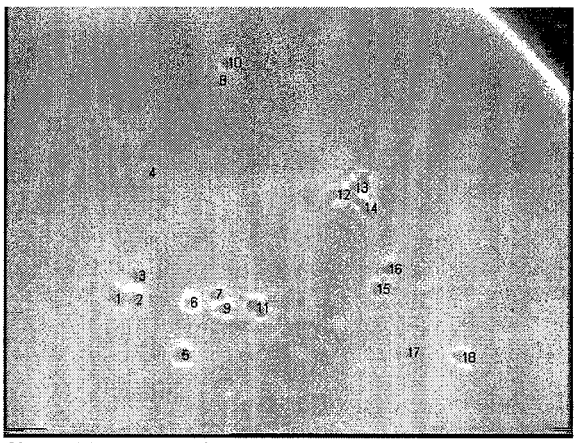
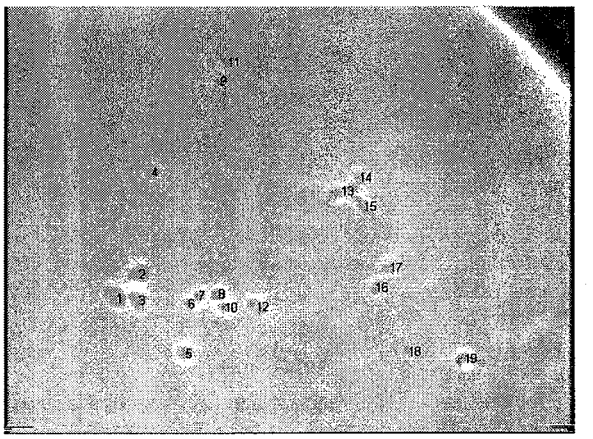


Fig. 7 (a) Image at time t



(b) Image at time t+1

4. RESULTS

The results obtained for finding number of cells, false positives and false negatives of segmentation with different images are shown in Table.1. The manual counting was made only to find the accuracy of the system. We have also evaluated the cell detection accuracy of our proposed method is 96%. Due to poor quality of the images the segmentation process is resulting 4% of over segmentation and 6% of under segmentation.

Table. 1

Image No.	Manual Counting	Automatic counting	Number of cells identified	False Positive	False Negative
1	13	14	13	1	0
2	23	24	21	1	2
3	34	36	33	1	3
4	43	44	41	0	2
5	13	17	12	2	0
6	22	19	21	2	3
7	28	27	28	0	1
8	39	38	37	1	3
9	35	35	34	0	1
10	29	29	27	0	2

By keeping the threshold for major and minor axis length we are able to remove the noise. When the cell splits one keeps its parent position and other moves for some distance. This helps in finding the sibling of the cell. If the distance is less than minimum sibling distance there is a possibility of cluster cells.

Fig.7 shows that the mapping of individual cells in two consecutive images and also we can observe that cell 6 of image t is split into cells 6 and 7 in image t+1. The cell mapping and splitting process proposed in this paper successfully works for the well segmented images. This system was implemented in Matlab 7, which has been convenient for the experimental development.

5. CONCLUSION

In this paper we proposed a method based on Watershed algorithm and Mahalanobis distance algorithm to segment and track the stem cells in time-lapse microscopy. Also our method can count the stem cells in an image. The morphological operations applied in the image reduce the noise and separates the cells that were together. However, the system failed to separate the overlapping cells. The quantitative method to count the cells is producing better results. The proposed method for tracking can be implemented for automatic lineaging in sequence of image frames to understand the cell behaviour, which in turn helps in understanding the neurogenic process.

References

[1] D.A. Langan, J.W.Modestino, and J.Zhang, "Cluster Validation for unsupervised stochastic model based image segmentation", IEEE transaction no image processing, vol. 7, 1998.

[2] A.A.B. Theodoros Mouroutis, Stephen J.Roberts, "Robust Cell nuclei segmentation using statistical modeling", 1998.

[3] J.Kittler and J.Illingsworth, "Minimum Error Thresholding", Pattern Recognition vol. 19, no. 1, pp. 41-47, 1986.

[4] N.Otsu, "A threshold selection method from gray-level histograms", IEEE Transactions on Systems, Man, and Cybernetics, vol. 9, no. 1, pp. 62-66, 1979.

[5] K.Wu, D.Gauthier, and M.Levine, "Live cell image segmentation", IEEE Transactions on Biomedical Engineering, vol. 42, no .1, pp 1-12, 1995.

[6] T.Markiewicz, S.Osowski, L.Moszczyski, and R.Satat, "Myelogeneous leukemia cell image preprocessing or feature generation" in 5th International Workshop on Computational Methods in Electrical Engineering, 2003, pp. 70-73.

[7] D.Comaniciu and P.Meer, "Cell image segmentation for Diagnostic Pathology", Advanced Algorithmic approaches to medical image segmentation: State-of-art applications in cardiology, neurology, mammography and pathology, pp-541-558, 2002.

[8] J.M Geusebroek, A.W.M. Smeulders, and F. Cornelissen, "Segmentation of vrll clusters by nearest neighbour graphs", in proceedings of the third annual conference of the advanced school of computing and Imaging, 1997, pp. 248-252.

[9] V.Meas-Yedid, F.Cloppet, A.Roumier, A.Alcover, J.C.Olivo-Marin, and G.Stamon, "Quantitative microscopic image analysis by active contours", in VI 2001 Vision Interface Annual Conference – Medical Applications, 2001.

[10] C.Zimmer, E. Labruere, V.Meas-Yedid, N.Guillen, and J.C.Olivo-Marin, "Segmentation and tracking of migrating cells in video microscopy with parametric active contours: a tool for cell based drug testing, IEEE Trans. Med. Imaging., 21(10) 1212-1221 (2002).

[11] T.Kirubarajan, Y.Barshalom, and K.Pattipati, "Multi assignment for tracking a number of overlapping objects", IEEE Trans Aerosp. Elect. Syst., 37(1) 2-21 (2001)

[12] J.Degerman, J.Faijerson, K.Althoff, T.Thorlin, J.H,Rodriguez, and T.Gustavsson, "A comparative study between Level set and Watershed image segmentation for tracking stem cells in Time-Lapse Microscopy," CiteULike

[13] Amalka pinidiyaarachchi, Carolina Wahlby, "Seeded watershed for combined segmentation and tracking of cells", ICIAP-2005 proceeding pp. 336-343.

[14] Tomohiro Yasuda, Hideo Bannai, Shuichi Onami, Satoru Miyano, Hiroaki Kitano; "Towards Automatic Construction of Cell-Lineage of *C. elegans* from Nomarski DIC Microscope Images", Genome Informatics 1999, Dec. 14-15, 1999.

[15] http://www.landesbioscience.com/journals/cc/supplement/alkofani.zip

[16] J.Serra and V.L., "An overview of morphological filtering", IEEE transactions on circuits, systems and signal processing, vol. 11, 1992.

[17] P.Soille, Morphological Image analysis: Principles and Applications, Springer-Verlag, October 1999.

On watermarking in frequency domain

Narendrakumar Ramchandra Dasre[1],
Hemraj Ramdas Patil[2]
E-mail: 1) narendasre@rediffmail.com (Corresponding author), 2) hrpatil73@rediffmail.com
Ramrao Adik Institute Of Technology,
Sector-07, Nerul, Navi Mumbai-400 706 INDIA

ABSTRACT

A wavelet-based image watermarking scheme is proposed, based on insertion of 'logo' image as watermark in mid-frequency domain. This new approach provides flexibility in determining the pixel to be watermarked and increases the data hiding capacity. It is easy to implement watermark embedding algorithm as well as the corresponding detection algorithm. The watermarking algorithm is tested under different attacks such as median filtering, image cropping and image compression. It is also robust. The experimental results prove that the method is more tamper proof and less perceptible for any type of images other than well known private methods in frequency domain. In the proposed approach, an original image is decomposed into wavelet coefficients then watermark is embedded through algorithm. The wavelet transform filters can be used as security key for the extraction of inserted watermark. The proposed watermark extraction technique is independent of the original image. The watermark embedded image is produced by taking the inverse 2-D discrete wavelet transform of the altered wavelet decomposition. Here we have given the relation between the area of the channel in which we insert the watermark and the area affected in original image.

Key words: Wavelet decomposition, watermarking, image processing and jpeg compression.

1. INTRODUCTION

In the past decade, digital watermarking has become an attractive method for data hiding and copyright protection [1-14]. The two traditional approaches for watermarking are spatial and spectral domain technique[13]. In spatial domain the watermark is embedded in selected regions chosen based on texture of the given image[13]. This approach is usually not robust to most signal processing attack. While in spectral domain, the watermark is embedded in the transform domain using methods such as DCT and DWT, in the mid-frequency range to ensure transferency and robustness of watermark simultaneously[14].

Some desirable properties of watermarking technique[10] include the following:

- The inserted watermark should not introduce visible artifacts.
- The watermark should not be easily removable.
- The watermark should be resilient to lossy data compression such as jpeg.
- The watermark should be resilient to image processing technique such as median filtering and image cropping.
- The original image is not required in watermark extraction.
- The watermark can only be extracted by privileged individuals for given the security key.

Since Spatial domain technique generally requires a lower computational cost compared to transform domain techniques. They are also generally easier to implement. A major concern of digital image watermarking is the trade-off between image degradation verses ease in removal of the inserted watermark via compression, filtering or cropping. Spatial domain techniques generally do not balance this trade-off well, hence transform techniques are preferred. This paper describes algorithm that inserts a watermark into a mid-frequency channel of a image's 2-D wavelet transform. Here we use non-zero scaling factor and other parameters as security keys. These security keys are only clues to extract watermark.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754622 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853227

2. HISTORY AND WATERMARKS INSERTION

In this section , we have discussed the history of wavelet , watermarking and selection of channel for insertion of watermark. The details are given below.

2.1 Wavelet history and selection of channel:

Let 'I' be the original digital image of size N_1 x N_1 and 'W' be the 3-level 2-D wavelet transform of I. The details of 3-level 2-D wavelet transform can be found in [8, 10]. Since it will be necessary to use the same wavelet transform for the watermark extraction. The filters used in these transformations can be determined by the owner of the image and used as a part of security key. The necessary condition for a transform is the wavelet analysis and subsequent synthesis transform gives an exact reconstruction system. A 3-level 2-D wavelet image decomposition W consist of 10 wavelet channels: LL3, LH3, HL3, HH3, LH2, HL2, HH2, NH1, HL1 and HH1. These channels correspond to sub-band frequencies of the image I given in the figure .We know that the energy of most natural images is concentrated on the lower frequencies [9]. DWT is considered to be a powerful signal process and analysis tool, especially for the character in time or frequency domain [12]. DWT decomposes an image into four coefficient sets i.e. LL, LH, HL and HH. In most of DWT watermarking schemes, watermark is embedded in low frequency domain i.e. in LL which is good against many signal attacks, including JPEG compression, median filtering and image cropping [12].Since low frequency LL_3 channel has more energy than other channels hence the modification of the coefficients of this channel can cause intolerable image degradation. Hence we never insert watermark in this channel. In [12], HH is chosen to embed watermark. While most cases, in order to improve the robustness, watermark is often embedded in low/middle frequency sub-bands.

2.2 Algorithm of insertion:

Let I be N_1 x N_1 digital image and W be 3-level, 2-D wavelet decomposition of I[10]. Let M be watermark image/logo, we wish to embed into image I and $M = gM$, $g \in R\backslash\{0\}$, where g is scaling factor which provides a trade-off between image degradation and resiliency to attacks which may remove the watermark [1]. With small value of g, yields less perceptible artifact where as choosing large value of g render more perceptible artifact and resiliency to attacks. Consider the rise of wavelet channel 'C' is $N_1/8$ x $N_1/8$ for 3-level 2-D decomposition. Let the size of watermark image M be N_2 x N_2. It is required that $N_2 \leq N_1/8$.

The watermark can be inserted into channel C by any one of the five ways. For simplicity consider the size of M is 16 x 16 and I = 256 x 256. If we use 3-level 2-D decomposition then channel 'C' (LH_3, HL_3, HH_3) size will be 32 x 32. Then the partition of channel can be done in any ways the reader like. For simplicity we are considering the 5 partitions as shown in Figure (1) and Figure (2).

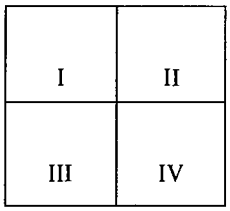


Figure (1)

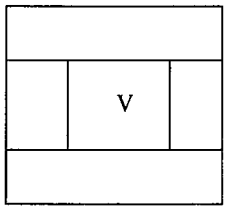
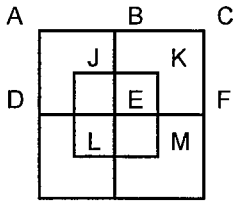


Figure (2)



Figure(3)

Area where the watermark is inserted	Affected area of Vegetable Image
I	ABED
II	BCFE
III	DEHG
IV	EFIH
V	JKML

Figure (4)

Figure (1) and Figure (2) gives the Partitions of the target Channel Where Watermark can be inserted . Where as Figure (3) gives the partition of the original image and Figure (4) gives the correspondence between the area of insertion of watermark and the affected area of original image after insertion of watermark. Consider the following cases:

I) We consider I[st] part of the channel. The logo is inserted by following equation.

$M = g.M$, where g is scaling factor. Therefore we have

$$C(n, m) = \begin{cases} M(n,m) & 1 \leq n \leq 16,\ 1 \leq m \leq 16 \\ C(n,m) & \text{otherwise} \end{cases}$$

II) Now we consider V[th] part (i.e. the most middle part). The logo is inserted by the following equations

$M = g.M$, where g is non-zero scaling factor.

$$C(n,m) = \begin{cases} M(n-8,m-8) & 9 \le n \le 24, 9 \le m \le 24 \\ C(n,m) & \text{otherwise} \end{cases}$$

The original image size, the wavelet transform filter, the watermark scaling parameter 'g' and the channel 'C' in which the watermark is inserted are components of the security key.

3. WATERMARK EXTRACTION

Watermark extraction requires that the original image size, the wavelet transform filter the scaling parameter g, and the wavelet channel in which the watermark is inserted are known. The same 3-level 2-D wavelet transform used in embedding algorithm is performed on the watermark embedded image J. The extracted watermark N is recovered from wavelet channel as below. Since we have considered two cases of insertion of watermark, these same cases we will consider for extraction of watermark.

Case I: $N(m,n) = \dfrac{C(m,n)}{g}$ for $1 \le m \le 16$, $1 \le n \le 16$ Case II: $N(m-8,n-8) = \dfrac{C(m,n)}{g}$ for $9 \le m \le 24$, $9 \le n \le 24$

4. EXPERIMENTAL RESULT

We have performed the experiments on vegetable image which is 256 x 256 pixels with 8 bit (256) gray scale level. In this paper we use a length Daubechies one (i.e. 'db1') or Haar (i.e. 'haar') wavelet which has been shown to be orthogonal system in [8]. Figure (5) shows the original vegetable image. Figure (6) shows the 3-level 2-D Wavelet Decomposition and the frequency regions. Figure (7) shows the cropped Image after watermarking. Figure (8) shows Watermarked image with scaling factor g = 1. Figure (9) gives the Median Filtered Image after watermarking and Figure (10) gives the Watermarked image with scaling factor g = π/100 .

Figure (5)

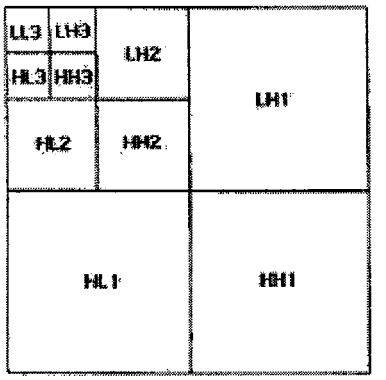


Figure (6)

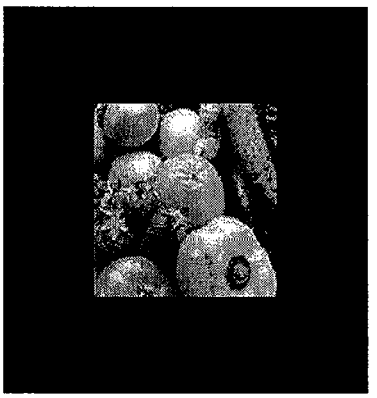


Figure (7)

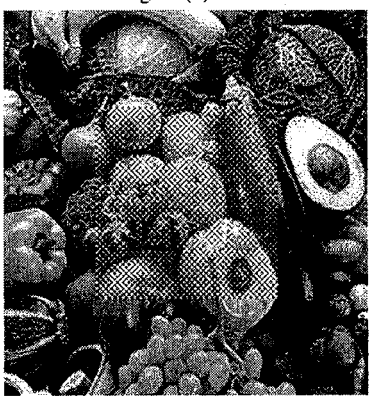


Figure (8)

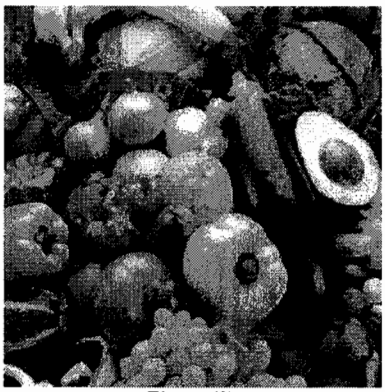


Figure (9)

Figure (10)

We use 16 x16 pixels 8 bit (256) gray level binary image as a watermark which is shown in Figure (11). Figure (12) shows the watermark extracted after image compression attack. Figure (13) gives the watermark extracted after image cropping attack and Figure (14) shows the extracted watermark without any attack.

Figure (11)

Figure (12)

Figure (13)

Figure (14)

4.1 Image Quality:

In Figure (8) , the watermark is embedded in the Vegetable image with scaling parameter g =1, the checker board image comes in the corresponding area. If we insert the watermark in HH_3 channel. When we take smaller scaling factor $g = \pi /100$, the checker board becomes less evident. These checker board is present in the form of ripples. The watermark embedded image of Vegetables with scaling parameter $g = \pi /100$ is shown in Figure (10).

4.2 Security Key:

The wavelet used, scaling parameter and level of 2-D decomposition are 3 security keys. Only the owner can extract the watermark using these security keys. For experiment, we have used *'db1'* wavelet in the embedding process .For extraction process, we use the same wavelet. If other than this wavelet is used Ex. *'db2'* then the watermark can not be extracted correctly from the channel.

4.3 Resiliency to Image compression:

Here we consider 'Vegetable' image of size 256 × 256 pixels with 256 gray levels (0 to 255) typically, such an image is represented using 8 bits pixel. These would require total 524288 bits to represent the image. The image compression or coding problem is concerned with finding ways of reducing the total number of bits required to represent the image while keeping the accompanying degradation in image quality as small as possible. Suppose the given image goes through one level of wavelet decomposition then we will have low resolution representation and three detail images each of the size 128 x 128 pixels. If we use 8 bits per pixel for each of these images, we would still be using a total of 524288 bits. However, it is not necessary to use 256 gray levels to represent each of these images. Most images have the property that their detail images have most values close to zero. Consequently, rather than allocating 256 levels to represent the detail images, we can have fewer levels and yet an acceptable representation of detail images. Suppose we represent HH_1 image with 1 bit and LH_1, HL_1 images with 2 bits then the original image can be represented by total 212 992 bits which is less than half as many bits as before. Again if we go for more levels of decomposition then more number of bits can be reduced. After reducing, still our watermark image remains there.

4.4 Resiliency to median filtering:

Here we have used 3 x 3 masks for median filtering the watermarked image. This process gives smooth watermark embedded image with preservation of major edges. After median filtering still we can regain our watermark. The robustness is directly proportional to the thickness of the watermark for median filtering attack. We have tested our thin watermark and several other watermarks with different thickness.

4.5 Resiliency to image cropping:

Here we cropped watermarked image of size 128 x 128 pixels which is $1/4^{th}$ of its original size. Then we regain the original size by assigning all the remaining pixels to be zero and then extracted the watermark. We observe that the watermark is still present.

5. CONCLUSION

We have presented a relationship between the area of watermark inserted in the channel and the area of the image affected. The details of this correspondence are given in Figure (4), Table of correspondence. We have presented the more number of choices of the regions in the channel and the partitions of the channel. Here we can choose any partition of the channel and embed the watermark. If we use proper security keys i.e. wavelet filter, scaling factor, channel and watermark size then we are easily able to extract the watermark after different attacks like median filter, image compression, cropping. Here we have given an idea of partitioning the channel as we wish. We can have the different partitions of the channel .This also increases the security of our watermark. Robustness of watermark image increases with increase in thickness.

6. ACKNOWLEDGEMENTS

We are thankful to Mr. Uday Shende, Director, R. A. I. T. and our Principal, Dr. S. R. Devane for their valuable support throughout the work.

7. LIST OF REFERENCES

[1] Hsieh M. S, Tseng D.C., Huang Y.H. "Hiding Digital Watermarks Using Multiresolution Wavelet Transform ", IEEE Transactions on Industrial Electronics, vol.48, No. 5, pp-875-882 , (Oct-2001)

[2] Wang Y., Doherty J.F., Van Dyck R.E. , "A wavelet-based watermarking algorithm for ownership verification of Digital Images " , IEEE Transactions on Image Processing, vol. II, No.2, pp-77-88, (2002)

[3] Chen P.C., Chen Y.S., Hsu W.H. "A Communication system model for digital image watermarking problems " , IAENG International Journal of Comp. Science, 34:2, IJCS_34_2_01.Advanced online publication (2007)

[4] Lee C.H. and Lee Y. K. "An adaptive digital image watermarking technique for copyright protection", IEEE Transactions on Consumer Electronics, vol. 45, No. 4, pp-105-1114 (1999).

[5] Ing. Petr Cika , "The improvement of the method for Digital Image Watermarking in Frequency Domain using BCH codes" , IJCSNS, International Journal of Comp. Science and Network Security, vol. 7, No. 3, pp-151-154 (2007).

[6] Wang X., Wu J., Niu P., "A new Digital Image Watermarking algorithm Resilient to Desynchronization Attacks" ,IEEE Transactions on Information Forensics and security, vol-2, No. 4, pp-655-663 (2007).

[7] Cox I. J., Kilian J., Leighton F. T., Shamoon T., "Secure spread spectrum watermarking for multimedia", IEEE Transaction on Image Processing, vol.-C, No.-12, pp-1673-1687(1997).

[8] Daubechies J., "Orthonormal basis of compactly supported wavelets" , Communications in Pure and Applied Mathematics. Vol.-51, pp-509-996 (1988).

[9] Hsu C-T , Wu J-L., "Hidden digital watermarks in images", IEEE Trans. Image Proc. ,Vol-8, No-1, pp-58-68 ,(1999).

[10] Tay P., Havlicek J. P., "Image watermarking using wavelet", Circuits and Systems,2002,MWSCAS-2002, 45th Midwest Symposium , pp-III-258 – III-261(2002).

[11] Chang C.C., Hwang K.F., Hwang M.S., "Robust authentication scheme for protecting copyrights of images and graphics" , IEEE Proc. Vis. Image Signal Process vol. 149, No. 1, pp-43-50 (2002).

[12] Lu W., Lu H.T., Chung F.L., "Chaos-based spread spectrum robust watermarking in DWT domain" , Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guang Zhou, pp-5308-5313 (2005).

[13] Khassawneeh M.A., Aviyente S., "Robust Watermarking on the joint spatial-spectral domain" IEEE 11th DSP Workshop and IEEE S P Education Workshop, pp-297-300 (2004).

[14] Lee C.H., Oh H.S., Back Y., Lee H.K., "Adaptive digital image watermarking using variable size of blocks in frequency domain" IEEE TENCON, pp-702-705 (1999).

Asymmetric Locating Position of Information Hiding against Tampering

Yulin WANG, Jian ZHAO

International School of Software, Wuhan University

Wuhan 430079, China

ABSTRACT

The paper provides a data hiding method of hiding media data in message data and a data extraction method of extracting the hidden data wherein message data is dispersively hidden in media data such as an image or sound to prevent a third person from modifying the message data easily. More specifically, the technique relates to a data hiding method in which media data is expressed as a media array while message data is expressed as a message array so that the array elements of the message array can be dispersively hidden in the message array based on state values specifying a particular array element of the media array, comprising the steps of: (a) determining the j^{th} ($j \geq 0$) state value S_j; (b) determining $(j+1)^{th}$ state value S_{j+1} based on the j^{th} state value, the array element of the media array indicated by the j^{th} state value, and the array element of the message array; and (c) hiding data with respect to the array element of the media array indicated by the $(j+1)^{th}$ state value S_{j+1}.

Keyword list Data Hiding, Content Security, Watermarking

1. INTRODUCTION

Due to the development of multimedia-oriented society, a large amount of digitalized image and sound information is distributed over Internet or as recording media[1-3]. Since anyone can simply make a complete copy of such digitalized information which is unlikely to be degraded, its unfair use is a problem. To prevent a third person from unfairly copying media data such as image or sound data, attention is being paid to a hiding technology that hides in original media data, information such as the sign of the author[4-5]. If the digital image data is illegally copied, the sign hidden in the copy can be checked to identify its source in order to determine whether it has been illegally made. Such a technology is called "data hiding"[6-7].

If the initial value has been published or known to a third person despite efforts to make it secret, the third person can use the value to identify the positions of messages easily and erase the message data or write a different message data over the first data. It has been difficult for conventional methods to effectively prevent a third person from erasing the original sign to make the source unknown or writing a different sign over the original sign to act as if he or she was the author.

Message data hidden in the image may include any information and may be, for example, a lattice pattern, a rule-like figure, or the sign of the producer of the image. Message data hidden in media data can be extracted by processing with a special program. Thus, based on the extracted message data, it can be determined whether the original media data has been modified.

Author Email:wangyulincn@yahoo.cn

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754623 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856010

2. PROPOSED TECHNIQUE

The following arrays and sequences are defined:

(1) Message array m:

If the message comprising J characters, the j^{th} character m[j-1], $0 \leq j \leq J-1$.

(2) Media array M:

The image is divided into **N** regions, with the i^{th} image region M[i], $0 \leq i \leq N-1$, and $N>J$.

(3) State sequence S:

The state sequence S containing J+1 elements determines the locations (media array values) to hiding, the j^{th} state value S_j, $0 \leq j \leq J$.

(4) Position sequence p:

The position sequence containing J+1 elements is used to identify the positions to hiding, with $p_j = S_j \bmod N$, $0 \leq j \leq J$, and p_j is an integer ($0 \sim N-1$). Since the number of image regions N is a constant, the value of the position pj can be uniquely determined once the state value S_j has been established. Thus, the positions that are subjected to hiding are virtually identified by the state sequence S.

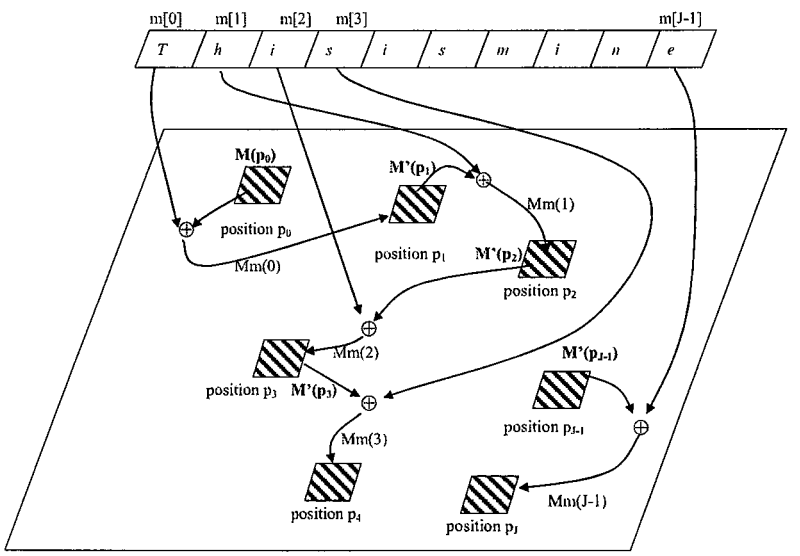


Fig. 1 A schematic drawing describing an object to be hidden

2.1 Data Hiding Algorithm

Step 1 Determine the state sequence S

The initial state S_0 is determined based on the characteristics of the message:

$S_0 = m[0] \oplus m[1] \oplus m[2] \oplus \ldots \oplus m[J-1]$

The state sequence S is determined based on the message and image data, and HK is the hiding position function. S_j is correlated with its forward state value S_{j-1} recursively.

$S_1=HK(S_0 \oplus m[0] \oplus M[p_0]) = HK(S_0 \oplus Mm[0])$

$S_2=HK(S_1 \oplus m[1] \oplus M'[p_1]) = HK(S_1 \oplus Mm[1])$

$S_3=HK(S_2 \oplus m[2] \oplus M'[p_2]) = HK(S_2 \oplus Mm[2])$

....

$S_J=HK(S_{J-1} \oplus m[J-1] \oplus M'[p_{J-1}]) = HK(S_{J-1} \oplus Mm[J-1])$

Step 2 Hide message data

Given S_j, $p_j = S_j \bmod N$, $0 \leq j \leq J$. The J characters will be hidden in position p_1 to p_J, and no character will be hidden in position p_0.

- At position p_1: $M[p_0]$ is simply taken from the position p_0. Let $Mm[0] = m[0] \oplus M[p_0]$. Mm[0] is hidden in the position p_1. This hiding operation changes the contents of the media array value $M[p_1]$ to $M'[p_1]$.
- At position p_2: Now the contents of the media array at position p_1 is $M'[p_1]$ not $M[p_1]$.
  Let $Mm[1] = m[1] \oplus M'[p_1]$. The contents of Mm[1] is hidden in the position p_2. This hiding operation changes the contents of the media array value $M[p_2]$ to $M'[p_2]$.
  ...
- At position p_J: Data hiding is completed by repeating such hiding processing until the position p_J.
  $Mm[J-1] = m[J-1] \oplus M'[p_{J-1}]$. The contents of Mm[J-1] is hidden in the position p_J. This hiding operation changes the contents of the media array value $M[p_J]$ to $M'[p_J]$.

The hidden data is not the message array value m[j] but the hiding array value Mm[j]. The number of the elements of each of the state sequence S and the position sequence (p) is larger than that of the message array because no data is hidden in the position p_0.

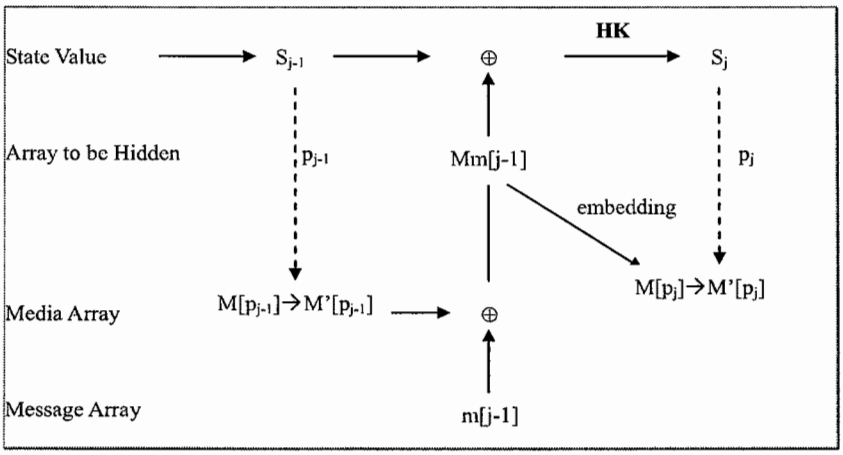


Fig. 2 Relations among the arrays during the message hiding

2.2 Data Extraction Algorithm

The following three pieces of information are provided to a third person to extract message data.

(1) Media array M' after hiding

(2) Final state value S_J

(3) Extraction position translation function EK

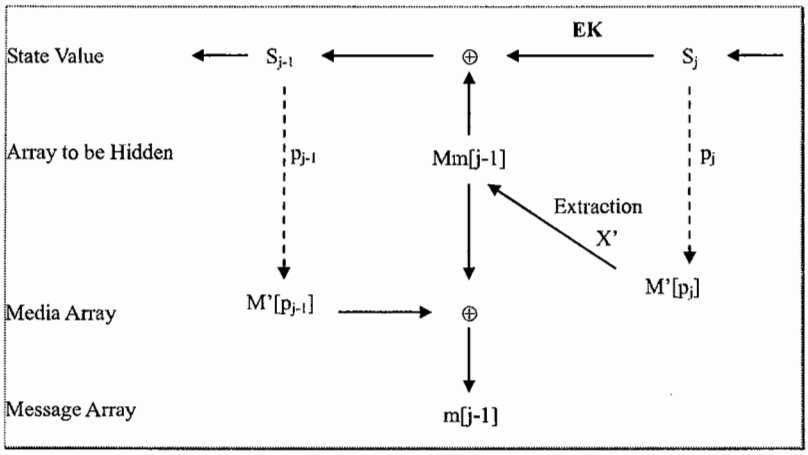


Fig. 3 Relations among the arrays during the message extraction

The initial position p_0 is identified from the initial state value S_0. Conventional methods have determined the initial state value S_0 as a particular constant regardless of the contents of the message to be embedded. In our technique, we set the seed $S_0 = m[0] \oplus m[1] \oplus m[2] \oplus \ldots \oplus m[J-1]$. This not only makes the determination of the initial state value complicated but also effectively prevents a third person from illegally writing data over the message.

When extracting the hidden message, the extractor does not need to know the state values S_0 to S_{J-1} as long as he or she knows only the S_J. This is because these state values can be identified by inversely converting the final state value.

Hiding position translation function HK is used to generate the *next* state value, and extraction position translation function EK is used to generate the *preceding* state value.

$$EK(HK(x))=x$$

To virtually prevent a third person from reproducing the hiding position translation function HK using the provided extraction position translation function EK, an enciphering function and a deciphering function in a public enciphering method are used.

Step 1 Derive preceding state value one by one from S_J

The extraction is recursively executed in the reverse order starting with the last message array value m[J-1], as shown in Fig. 3.

(1) Knowing S_J, we can get $p_J = S_J \bmod I$. In position p_J, Mm[J-1] is hidden in the media changed from original $M[p_J]$ to $M'[p_J]$.

(2) With a specific data extraction algorithm, Mm[J-1] can be extracted from media $M'[p_J]$.

Mm[J-1]=function X' ($M'[p_J]$), where function X' is a specific extracting algorithm.

(3)Since $S_J=HK(S_{J-1}\oplus Mm[J-1])$,

$EK(S_J)=EK(HK(S_{J-1}\oplus Mm[J-1]))$

$=S_{J-1}\oplus Mm[J-1]$.

Therefore, $S_{J-1}=EK(S_J)\oplus Mm[J-1]$. S_{J-1} can be solvable which is equal to $EK(S_J)\oplus Mm[J-1]$.

(4) Knowing S_{J-1} at (3), we can get p_{J-1}. Then we can locate the media $M'[p_{J-1}]$.

(5) Since $Mm[J-1]=m[J-1]\oplus M'[p_{J-1}]$,

$m[J-1]=Mm[J-1]\oplus M'[p_{J-1}]$.

Since we have obtained Mm[J-1]at (2) and $M'[p_{J-1}]$ at (4),

m[J-1] can be calculated.

Starting from S_{J-1}, repeat the above sub-step (1) --- (5), we can get m[J-2].

...

Each time m[j] is extracted, it is determined whether it is the beginning of the message array by determining whether $S_j = (m[J-1]\oplus m[J-2]\oplus \ldots \oplus m[j]$. The equation is established only when j=0 and not in the other cases ($j\neq 0$). Each time the final state value S_j is generated in the reverse order, it is compared to the output of the hash function based on all the extracted message array values.

Finally, we can extract the whole message sequence in the reverse order, m[J-1], m[J-2], …, m[2], m[1], m[0].

3. CONCLUSIONS

In the above public key method, the public key is published to third people but only the author who has produced the image has the secret key, so the third people cannot know the contents of the image. According to conventional techniques, the generated position sequence S has depended only on an initially provided constant regardless of the contents of the message. By simply knowing the constant, third people have been able to erase the message or overwrite a different message over an original message. According to the present algorithm, however, the generation of the state sequence also depends on the contents of the message, different messages result in different position sequences. Since the hidden position is based on the position sequence, a different message with different contents cannot be hidden in a position at which the original message is present, so the faked message will not be extracted. This is also true of media data.

Acknowledgments

This work is sponsored by National High Technology Research and Development Program of China (Grant No. 2009AA01Z412), and Natural Science Foundation of Jiangsu Province of China.

REFERENCES

[1] Long-Tae Park, Jong-Wook Baek, "Management of service level agreements for multimedia Internet service using a utility model," IEEE Communications Magazine, Volume 39, Issue 5, May 2001 Page(s):100 - 106

[2] R. Mohan, J.R.Smith, Chung-Sheng Li, "Adapting multimedia Internet content for universal access," IEEE Transactions on Multimedia, Volume 1, Issue 1, March 1999 Page(s):104 - 114

[3] K. Farrell, W. Mistretta, "VeriNet Web-speaker verification for the World Wide Web," IEEE International Conference on Multimedia and Expo, Volume 3, 30 July-2 Aug. 2000 Page(s):1497 – 1500

[4] Wen-Chung Kuo, Dong-Jin Jiang, Yu-Chih Huang, "A Reversible Data Hiding Scheme Based on Block Division," 2008 Congress on Image and Signal Processing, Volume 1, 27-30 May 2008 Page(s):365 - 369

[5] Zhicheng Ni, Y.Q. Shi, N.Ansari, "Robust lossless image data hiding," 2004 IEEE International Conference on Multimedia and Expo, Volume 3, 27-30 June 2004 Page(s):2199 – 2202

[6] V. Viswanathan, "Information hiding in wave files through frequency domain," Applied Mathematics and Computation, Volume 201, Issues 1-2, 15 July 2008, Pages 121-127

[7] Zhou Xin, Lai Dong, Yuan Sheng, "A method for hiding information utilizing double-random phase-encoding technique," Optics & Laser Technology, Volume 39, Issue 7, October 2007, Pages 1360-1363

Color Image Segmentation – a Review

Kanchan Subhash Deshmukh
M.G.M's College of Computer Sci. & IT, Nanded, India
Email- dkanchan_99@yahoo.com

ABSTRACT: Image segmentation is the process of dividing an image into homogenous regions. It is an essential step towards high-level image processing task such as image analysis, pattern recognition and computer vision. Processing of color images has become an important issue due to its huge usage in computer vision applications. It is observed that most of the color image segmentation techniques are derived from monochrome image segmentation. The techniques for segmentation of monochrome images are based on the principles of histogram thresholding, edge detection, region growing etc. Many color image segmentation algorithms using different color models and these principles are proposed. Extraction of objects within an image without a prior knowledge is one of the important issues in segmentation area. Novel approaches such as fuzzy set theory, neural network and neuro-fuzzy based segmentation are coming up to tackle this problem. This paper is an endeavor to review various algorithms and recent advances in color image segmentation.

Key words: Color image segmentation, Neural networks, Thresholding, Edge detection, Region-based approach, Fuzzy logic.

1. INTRODUCTION

Image segmentation plays an important role in the field of Computer vision to understand and analyze an image. Image analysis deals with the problem of processing of images by computer in order to find objects in it [1]. Several authors have defined image segmentation problem in variety of ways. Fu and Mui [2] defined image segmentation to be a psychophysical problem and therefore not simply a subject of analytical solutions. According to Spirkovska [3], image segmentation process is a bridge between low-level vision subsystem and high-level vision subsystem. Prior research work on image segmentation is concentrated on monochrome images. Therefore, it is observed that color image segmentation algorithms are often derived from monochrome image segmentation methods. The techniques for segmentation of monochrome images are mainly based on the principles of edge detection, region growing and histogram thresholding. Extraction of objects in the image without preceding knowledge is one of the significant issues in segmentation area. Novel approaches based on fuzzy theory, neural network are helpful in this regard. In this paper various advance techniques used for color image segmentation are elaborated.

The rest of the paper is organized as follows: Section 2 elaborates most recent techniques of color image segmentation. Section 3 concludes with summary.

2. COLOR IMAGE SEGMENTATION

It has long been recognized that the human eye can differentiate thousands of color shades and intensities but only two-dozen shades of gray. It is quite usual that the objects, which cannot be extracted using gray scale, are extracted using color information [4]. As compared to monochrome images, color image provides the additional information to

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754624 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856011

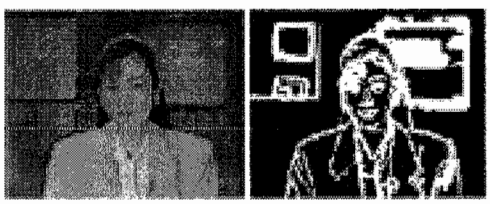
intensity. Color is a useful feature for pattern recognition and computer vision. Hence color image processing has become increasingly more attractive. Segmentation techniques for color images can be categorized into edge detection, region based, histogram thresholding, fuzzy techniques and neural network based approaches.

2.1 Edge detection techniques

Edge detection techniques are used to detect the boundaries of an object [5]. In color images, the information about edge is much richer than monochrome image. For example, an edge between two objects with the same brightness but different hues cannot be detected in monochrome images but can be detected in color images. In this section we have presented different color edge detection algorithm.

Fan [6] proposed an algorithm to find edges in YUV color space. The method uses second order neighborhood to describe the relationship between the current pixel and its neighboring pixels. One of the advantage of this method is that the obtained color edges can provide a simplified image that preserves domain geometric structures and spatial relationships in the original image but the weakness of this method is that since the edges are normally discontinuous or over detected, they cannot be used as an image content descriptors directly. However, they are very useful for providing the *structural seeds* of the image.

An algorithm for boundary detection technique using a predictive coding model is presented by Ma and Manjunath [7]. The system is able to recognize the direction of change in color and texture at any point and at a given scale. The system then forms an edge flow, which through propagation converges to image boundaries. Tao and Huang [8] proposed a method that finds clusters in RGB space. The method uses global color information to compute local gradient. The major chromatic components in an image are first computed using a clustering approach. According to these clusters, a set of chromatic transforms are generated. An appropriate chromatic transform is chosen for each pixel to maximize the gradient magnitude. In this way, edges are treated as a transition from one cluster to another. One of the advantages of this method is that the global information of a given color image is exploited to guide the local edge detector. Figure 1 shows the experimental result using this technique.



(a) (b)

Figure 1. (a) Original method (b) Output image.

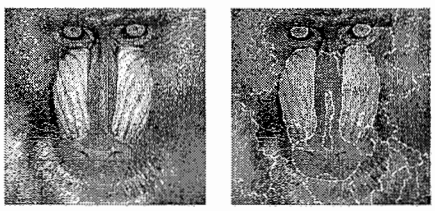
2.2 Region based techniques

As the name implies, *region growing* is a procedure that groups the pixels into different regions. In this technique, it is necessary to consider a set of seed points. The homogeneous regions are formed by attaching to each seed point those neighboring pixels, which have the associated properties such as (gray level, texture, color). Region based approach includes region growing, splitting, merging and their combination. However, the difficulty with this technique is the selection of an initial seed points.

Deng and Manjunath [9] proposed a JSEG algorithm to segment the image based on multiscale J images. The images, which correspond to the measurements of local homogeneities at different scales, are called J images. It

indicates the potential boundary locations. The system has the ability to segment color-textured images and video sequences without supervision. First in this technique the colors inside the images are quantized to several classes. This is helpful to distinguish various regions in the image. The pixels are then replaced by their corresponding color class label, which forms the class map of an image. The focus of this method is on spatial segmentation. Here a principle for "good" segmentation using class-map is proposed. J images are created by applying the criterion to local windows in the class-map. In the class map, high and low values correspond to the possible boundaries and interiors of color-texture regions. A region growing method is then used to segment the image based on the multiscale J-images.

A multilevel approach for color image segmentation is proposed by Junda and Chitsobhuk [10]. The segmentation is performed using a JSEG method followed by a region merging technique. The hierarchical mapping is performed using a local region growing technique to achieve final segmentation result. One of the limitations of this method is that, there is a difficulty in automatically selecting appropriate parameters for different types of images. Tremeau and Borel [11] presented a homogeneity criteria operating in RGB color space. In first step, system generates the connected regions using a region growing process. All the regions having similar color distributions are merged in a second stage. Lambert [12] proposed a quick and coarse color image segmentation method using cluster validity criteria. The clusters are calculated using fuzzy c-means (FCM) algorithm. The cluster merging is performed pair wise only if the overlapping between two clusters is comparable with their compactness. Figure 2 shows the experimental result using this technique.



(a) (b)

Figure 2 .(a) Original image (b) Output image.

2.3 Histogram thresholding

Histogram thresholding is one of the widely used techniques used for image segmentation. Many techniques using thresholding in different color spaces are developed.

Busin and Vandenbroucke [13] suggested a concept for color image segmentation using iterative analysis of 1-D histogram of color components. The method uses RGB color space for segmentation. At each iteration step, it constructs a class. The system finds multithreshold values by tracking zero crossings of the first order derivative of histogram. One of the drawbacks of this method is that, the histograms neglects the spatial interaction between the pixels and the selection of most relevant color space is only based on the analysis of color properties of the pixels. Navon [14] proposed an approach for color image segmentation using local threshold values. It divides an image into homogeneous regions using a local threshold values. The threshold values are calculated one by one during merging process. The advantage of this technique is that, the number of thresholds and their values are derived adaptively and the method does not require any statistical information of the image.

A new approach for extracting text within an image is proposed by Yingzi & Chang [15]. The method uses RGB color space for segmentation. A threshold value in each component is found out using Otsu's gray level thresholding approach. The multithreshold values are then combined using an unsupervised within class/ between class clustering

algorithm. The advantage of this method lies in its simplicity. It is also an effective method for text extraction. Liu [16] invent a threshold function for both RGB and HSV spaces using B-splines. One of the advantages of this method is that, the threshold values are calculated adaptively. Kurugollu [17] proposed a technique for multiband image segmentation using two-dimensional histogram. Segmentation is performed using multithreshold values followed by the union of resulting segmentation channels.

2.4 Fuzzy techniques

Fuzzy set theory provides a mechanism to represent and manipulate uncertainty within an image. In fuzzy subsets, each of the pixels has a degree to which it belongs to a region or a boundary characterized by a membership value. Recently, there have been number of fuzzy approaches for color image segmentation [18, 19].

Fuzzy c-means (FCM) is a well-known algorithm for clustering. Lim and Lee [20] introduced a segmentation algorithm for color images based on thresholding and FCM algorithm. It uses a scale space filter to analyze the histogram of three-color component. The segmentation is performed in I1-I2-I3 color space proposed by Ohta. The image is first segmented coarsely using multithresholding. The coarse segmentation is then polished by fuzzy c-means (FCM) clustering algorithm. Hall [21] uses FCM algorithm to segment MR images where each pixel is a 3-D vector consisting of different features yielded by the MRI system.

However, FCM algorithms are having some drawbacks:

1. FCM algorithm requires the prior knowledge about the number of regions existing with an image.
2. Adjacent clusters often overlap in color space, which causes the incorrect pixel classification.

(a) (b) (c)

Figure 3. (a) Original image (b) Segmented image obtained using HOB (c) Segmented image obtained using HIB.

A novel method for color image segmentation is proposed by Cheng and Jiang [22]. The method uses homogram based approach (HOB) to detect homogeneous regions within color images. Homogram considers both local as well as global information of a pixel. The threshold values in each plane are found out using fuzzy homogeneity approach. The method uses fuzzy entropy method to perform homogram analysis for finding all major homogeneous regions. Region merging process is carried out based on color similarity among these regions to avoid over segmentation. The segmentation results for corresponding color components are then combined.

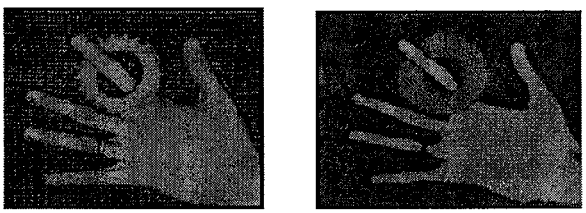
Neural network techniques

Neural networks are formed by several elements that are connected by links with variable weights. Artificial neural networks (ANN) are widely applied for pattern recognition. ANN is a physical cellular network that is able to acquire, store, and utilize experiential knowledge that has been related to network's capabilities and performance. Their processing capability and nonlinear characteristics are used for clustering. Self-organization of Kohonen Feature Map (SOFM) network is a powerful tool for clustering.

Li and Li [23] proposed an unsupervised approach for color image segmentation. The method uses neural network based approach to find automatic features of an image. First the multiple color features can be analyzed using a self-organizing feature map (SOFM) Then the useful feature sequence (feature vector) is determined. The encoded feature vector can be used in final segmentation. The method uses FCM algorithm for clustering. One of the advantages of this method is that using SOFM, a feature vector suitable for segmenting the input image can be extracted automatically using this algorithm. Consequently, this technique is an adaptive approach for segmenting different types of color images.

Moreira [24] proposed a neural network based technique for color image segmentation. The classification of pixels within an image is performed using SOM neural network. The SOM network has the ability to calculate main chromaticity present within an image. Each pixel in an image is classified according to the identified classes.

A novel technique for color image segmentation based on histogram thresholding and neural network is proposed by Deshmukh [25]. In this method, an adaptive neural network system for automatic color image segmentation (ANNSCIS) is presented. The method uses HSV space for segmentation. The system finds the multiple threshold values for clustering based on the first order derivative of histogram of saturation and intensity planes in HSV color space. One of the advantages of this method is that, the multiple objects within an image are found automatically and the system does not require the prior information about the number of objects present in an image. Figure 4 shows the experimental result using this technique.



(a) (b)

Figure 4. (a) Original image (b) Segmented image.

3 CONCLUSION

In this paper we have presented an overview of different kinds of algorithms for color image segmentation. Most of the monochrome image segmentation methods can be extended for color images such as histogram thresholding, clustering, region growing, edge detection, and fuzzy based methods. These methods can be applied on to the component of a color space. The results then combined to get a final segmentation result.

The fuzzy theory can be widely used in the area of image processing. Fuzzy theory provides a tool, which can represent the uncertainties arising within an image. Fuzzy operators, properties and inference rules (If—then) have more applications in image segmentation. Fuzzy segmentation results can be used in feature extraction and object recognition phases of image processing and computer vision.

REFERENCES

1. Jain A. K. Fundamentals of Digital Image Processing. Upper Saddle River, NJ: Prentic Hall, (1989).
2. Fu K.S. and Mui J.K. "A survey on image segmentation", Pattern Recognition, 13, 3-16 (1981).

3. Spirkovska L. "A summary of image segmentation techniques", NASA Technical Memorandum 04022, (1993).
4. Milan Sonka, Image Processing, analysis, and Machine Vision. Brooks/Cole Publishing Company (1999).
5. Carron T. and Lambert P. "Color edge detector using jointly hue, saturation and intensity", IEEE International Conference on Image Processing, 977-1081 (1994).
6. Fan J. and David. K. et al "Automatic image segmentation by integrating color edge extraction and seeded region growing", 10(10), 1454-1466(2001).
7. Ma W.Y. and Manjunath B.S. "Edge flow: A framework of boundary detection and image segmentation", CVPR'97, 744-749 (1997).
8. Tao H. and Huang T.S. "Color image edge detection using cluster analysis", ICIP, I 834-836 (1997).
9. Deng Y. and Manjunath B.S. "Unsupervised segmentation of color texture region in images and video", 1-25 (2001).
10. Junda A. and Chitsobhuk O. "Fast image segmentation based on multilevel approach", NCSCE, (2003).
11. Tremeau A. and Borel N. "A region growing and merging algorithm to color segmentation", Pattern Recognition, 30(7), 1191-1204 (1997).
12. Busin L. and Vandenbroucke N. et al "Color space selection for unsupervised color image segmentation by histogram multithresholding", IEEE Conference, 203-206 (2004).
13. Navon Ety and Often Miller et al "Color image segmentation based on adaptive local thresholds", Image and vision computing, 23, 69-85(2005).
14. Yingzi Du Eliza and Chein I Chang, "An unsupervised approach to color video thresholding", ICASSP, 373-376 (2003).
15. Liu L.J. et al "Efficient segmentation of nuclei in different color spaces", IEEE Computer society 773-778.
16. Kurugollu F. and B. Sankur et al "Color image segmentation using histogram multithresholding and fusion", Image and Vision Computing, 19, 915-928(2001).
17. Cheng H.D. and Sun Y. "A Hierarchical approach to color image segmentation using homogeneity", IEEE Trans. Image Processing, 2071-2082 (2001).
18. Huntsberger T.L. "Iterative fuzzy image segmentation", Pattern recognition, 18(2), 131-138 (1985).
19. Lim Y. W. and Lee Sang UK "On the color image segmentation based on the thresholding and the fuzzy c-means techniques", Pattern Recognition, 23(9), 379-396 (1989).
20. Hall L.O. and Bensaid A. "A comparison of neural network and fuzzy clustering techniques in segmenting magnetic resonance images of the brain", IEEE Transaction Neural network, 3, 672-682 (1992).
21. Cheng H.D. and Jiang X.H. "Homogram thresholding approach to color image segmentation", International conf. on computer vision, pattern recognition and image processing, (2000).
22. Pal N.R. and Pal S.K. "Object-background segmentation using new definition of entropy", IEE Proc. Part E-136 (4) 284-295 (1989).
23. Li N. and Li Y.F. "Feature encoding for unsupervised segmentation of color images", IEEE transactions on Systems Man and Cybernetics – Part B 33(3) 438-447 (2003).
24. Moreia Jander and Luciano da Fontoura costa, "Neural based color image segmentation and classification using self organizing maps", Outubro de, 47-54 (1996).
25. Deshmukh K.S. and Shinde G. N. et al "Multilevel approach for color image segmentation", Fourth Indian Conf. on Computer Vision, Graphics and Image Processing, 338-344 (2004).

Pre-Processing for noise reduction in depth estimation

[1]Seong-O Shim, [2]Aamir Saeed Malik, [1]Tae-Sun Choi
[1]Department of Mechatronics, Gwangju Institute of Science and Technology, Korea
[2]Department of Electrical & Electronic Engineering, Universiti Teknologi Petronas, Malaysia

ABSTRACT

The objective of the 3D shape estimation from focus is to estimate depth map of the scene or object based on best focus points from camera lens. In shape from focus (SFF), the measure of focus – sharpness – is the crucial part for final 3D shape estimation. However the noise imposed during image acquisition process by imaging system prevents exact focus measure. The traditional noise filters remove not only noise but also sharpness information. In this paper, mean shift algorithm was applied to remove noise imposed by the imaging process while minimizing loss of informative edges. Experimental results show that the mean shift algorithm can be applied before computing focus measure from image sequence corrupted by Gaussian noise and Impulse noise. Applying mean shift filtering before computing focus measure is promising in case the noise type during image acquisition is not known.

Shape From Focus (SFF), Depth From Focus, Depth Map, Mean Shift, Focus Measure, Noise Reduction, Filtering, 3D Shape Recovery.

1. INTRODUCTION

There are a variety of 3D Shape estimation methods. Broadly they can be classified into mainly three types, namely, Contact, Transmissive and Reflective methods. The optical methods fall under the reflective model which can further be divided into "Passive" and "Active" Techniques. In active techniques, we actually are projecting light rays while in passive techniques, we simply capture the reflection of light rays without any projections. The passive methods can further be classified as Shape From X (Stereo, Motion, Shading, Focus etc).

The objective of the 3D shape estimation methods is to estimate depth map of the scene or object. The depth map processing is subject to various errors and factors that affect the 3D shape recovery and hence their effects should be incorporated by 3D shape recovery algorithms. It is important to get good estimate of the depth map because that estimate can further be used in approximation techniques and algorithms leading to recovery of three dimensional structure of the object which is required in many high level vision applications.

In this paper, we concentrate on the various types of noise that affect the depth map estimation and we propose to use mean-shift algorithm as a pre-processing stage before the application of 3D shape recovery algorithm. For 3D shape recovery, we used the algorithms based on Shape From Focus (SFF) method which is one of the optical passive 3D image reconstruction method. In addition, various focus measures are tested and their performance analyzed with and without noise.

In SFF, a sequence of images that correspond to different levels of object focus is obtained. A sharp image and the relative depth can be retrieved by collecting the best focused points in each image [1, 11]. The absolute depth of object

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754625 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853050

surface patches can be calculated from the focal length and the position of lens that gave the sharpest image of the surface patches. The depth or best focus is obtained by using some focus measure. The basic image formation geometry is shown in figure 1.

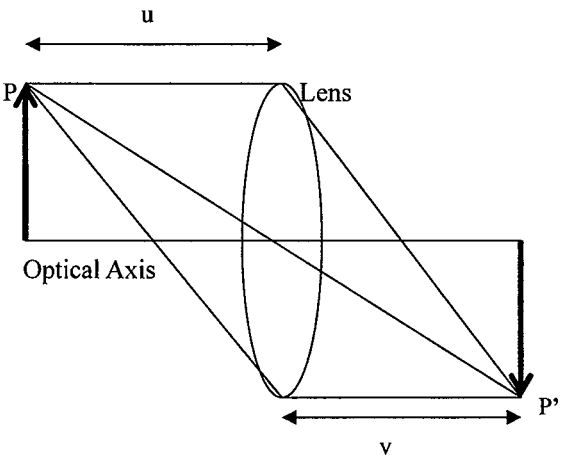


Figure 1. Image Formation of a 3D Object

In figure 1, the parameters related to the camera are already known. We need to calculate 'u', i.e., depth of object from the lens. We make a depth map by calculating 'u' for every pixel. We can use the lens formula to calculate 'u'. If the image detector (ID) is placed exactly at a distance v, sharp image P' of the point P is formed at v (see figure 1). Then the relationship between the object distance u, focal distance of the lens f, and the image distance v is given by:

$$\frac{1}{f} = \frac{1}{u} + \frac{1}{v} \tag{1}$$

Therefore, in SFF, a sequence of images that correspond to different levels of object focus is obtained. A sharp image and the relative depth can be retrieved by collecting the best focused points in each image. The depth or best focus is obtained by using some focus measure.

However, due to presence of noise during image acquisition because of imaging device, error is introduced and sharpness cannot be measured accurately. Mainly, Gaussian noise and Impulse noise arise during imaging process. Gaussian noise arises from electronic circuit and sensor due to poor illumination or high temperature, and Impulse noise arises from instantaneous transients such as faulty switching during imaging [2]. In section IV, we propose applying mean shift algorithm to remove these noises by minimizing loss of informative edges in an image.

2. RELATED WORK

Focus measure is defined as a quantity to evaluate the sharpness of a pixel locally. Malik [3, 12] has summarized various focus measures published in the literature. The focus measures include Tenenbaum (based on single derivative), Sum of Modified Laplacian (based on double derivative), Gray Level Variance (based on statistical variance parameter), Mean Method (based on the average values) etc.

After obtaining a robust focus measure, it is applied to a reconstruction scheme in order to construct a more accurate depth range image. Malik [3] has summarized various approximation techniques. In Traditional (TR) SFF, for each image in the sequence, focus measure at each pixel is computed by Sum of Modified Laplacian (SML) in a 2D neighborhood. For each pixel, the image frame with the maximum focus measure is determined. The camera parameters for this image frame are used to compute the distance of the object point corresponding to that pixel. Nayar [4] proposed using the best focus measure values obtained from the TR-SFF for interpolation. The best focus measure values are interpolated using a Gaussian fit in the vicinity of peak for final estimation of depth map.

Choi [5] proposed a new concept termed Focused Image Surface Method (FISM) based on planar surface approximations. First, an estimate of FISM is obtained using Traditional (TR) SFF method. This estimate is then refined by searching for a planar surface that maximizes focus measure computed over pixels on the FISM. Asif [3] used Neural Networks to learn shape of focused image surface by optimizing the focus measure over small 3-D windows. Ahmad [3] proposed the use of Dynamic Programming (DP) to handle the complexity of focused image surface. A direct application of DP on a 3D data is impractical due to computational complexity. Therefore, heuristic model based on DP was proposed. We have used optimum window size [6] for calculations.

3. MEAN SHIFT PROCEDURE

Mean shift is a non-parametric iterative procedure that shifts each data to local modes of density function without estimating the density. It was first proposed by Fukunaga and Hostetler [7] and its nice properties were rediscovered by Cheng [8]. Since then, mean shift has been widely used in computer vision and image processing areas for data clustering, segmentation, filtering, object tracking, etc [9, 10].

Given n measurement points $\mathbf{x}_i$, $i = 1,...,n$ in the d-dimensional space R^d with $d \times d$ identity bandwidth matrix $\mathbf{H} = h^2\mathbf{I}$, the *mean shift* is defined as:

$$\mathbf{m}_{h,G}(\mathbf{x}) = \frac{\sum_{i=1}^{n} \mathbf{x}_i g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)} - \mathbf{x} \tag{2}$$

$$= \frac{1}{2}h^2 c \frac{\hat{\nabla} f_{h,K}(\mathbf{x})}{\hat{f}_{h,G}(\mathbf{x})} \tag{3}$$

where,

$$\hat{f}_{h,K}(\mathbf{x}) = \frac{c_{k,d}}{nh^d}\sum_{i=1}^{n} k\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right) \quad \text{and} \quad \hat{f}_{h,G}(\mathbf{x}) = \frac{c_{g,d}}{nh^d}\sum_{i=1}^{n} g\left(\left\|\frac{\mathbf{x}-\mathbf{x}_i}{h}\right\|^2\right)$$

are multivariate kernel density estimators with Epanechnikov profile k and uniform profile g. $c_{k,d}$, $c_{g,d}$ are

normalization constants and $c = \frac{c_{g,d}}{c_{k,d}}$.

From (3), mean shift vector is aligned with gradient estimate of the density. Therefore, if the location $\mathbf{x}\left(=\mathbf{y}_0\right)$ is updated iteratively as:

$$\mathbf{y}_{l+1} = \frac{\sum_{i=1}^{n} \mathbf{y}_i g\left(\left\| \frac{\mathbf{y}_l - \mathbf{y}_i}{h} \right\|^2\right)}{\sum_{i=1}^{n} g\left(\left\| \frac{\mathbf{y}_l - \mathbf{y}_i}{h} \right\|^2\right)} \qquad l = 0,\ 1,\ 2,\ \ldots \tag{4}$$

then, it moves towards the local mode of density function.

4. MEAN SHIFT FILTERING

In SFF, sharpness measure, which is the level of focusing, is based on edges. But, various types of noise incorporated to image sequence prevent capturing informative edges. One way to remove noise is smoothing, but it blurs the entire image by removing not only noise but also informative edges.

We apply mean shift filtering proposed by Comaniciu and Meer [10] to remove noise by preserving edges. Let $\mathbf{x}_i$ and $\mathbf{z}_i$, $i = 1,\ldots,n$, be the d-dimensional input and filtered image pixels in the joint spatial-range domain, the mean shift filtering for each pixel is as follows:

1. Initialize $\mathbf{y}_{i,0} = \mathbf{x}_i$.

2. Compute $\mathbf{y}_{i,l+1}$, $l = 0, 1, \ldots$ according to (4) until convergence, $\mathbf{y} = \mathbf{y}_{i,c}$.

3. Assign $\mathbf{z}_i = \left(\mathbf{x}_i^s, \mathbf{y}_{i,c}^r\right)$.

Thus, the filtered data at the spatial location $\mathbf{x}_i^s$ has the range component $\mathbf{y}_{i,c}^r$ of the convergent point $\mathbf{y}_{i,c}$. In mean shift, uniform kernels defined on each pixels are windows centered on each pixels, and these windows have both spatial and range dimensions. The main difference between mean shift and other averaging filters is that mean shift takes into account the pixels having both spatial coordinates and range values inside the window. Thus, for the edge, only the pixels on the edge are averaged and the updated points remain on the edge.

After filtering every image in image sequence, Sum of Modified Laplacian (SML) [4] was applied to measure the sharpness of image. This is based on the Modified Laplacian (ML) operator which is a point and symmetric operator. It is obtained by adding squared second derivatives in the x and y directions.

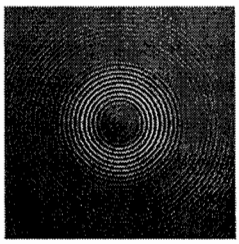
If the image has rich textures with high variability at each pixel, focus measure can be calculated considering single pixel. In order to improve robustness for weak-texture images, Shree Nayar and Yasuo Nakagawa [4] presented focus measure at (x,y) as sum of ML values in a local window.

$$SML(x_0, y_0) = \sum_{p(x,y) \in U(x_0,y_0)} \left(\frac{\partial^2 g(x,y)}{\partial^2 x} \right)^2 + \left(\frac{\partial^2 g(x,y)}{\partial^2 y} \right)^2 \quad (5)$$

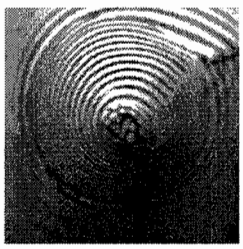
In next section, various types of noise were imposed on image sequence, and the result of mean shift filtered SML focus measure was compared with the result of SML focus measure with other filtering methods like wiener and median filters.

5. RESULTS

We applied the algorithm on a sequence of 97 simulated cone images and 97 real cone images. The resolution of the images is 360x360 pixels for cone images. One of frames of the above mentioned images is shown in figure 2.



Simulated Cone



Real Cone

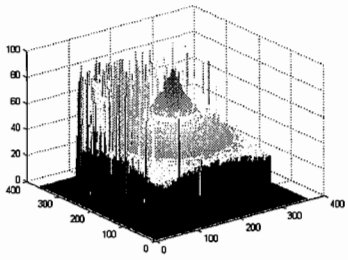
Figure 2. Test objects.

As described in earlier section, mainly Gaussian noise and Impulse noise arise during acquisition by imaging devices. Thus, in this section, mean shift algorithm was applied on image sequence corrupted by Gaussian noise and Impulse noise. After applying mean shift filtering, Sum of Modified Laplacian (SML) was computed for SFF. For comparison reason, instead of mean shift filter, same procedure was performed by applying wiener filter for Gaussian noise and median filter for Impulse noise. Wiener filter and median filter are commonly used filters to remove Gaussian noise and Impulse noise.

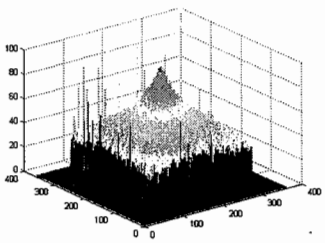
For Gaussian noise, SML with mean shift filtering shows better result than the result of SML without filtering from noise level at Variance = 0.005, and the performance enhances more as noise level gets higher. SML with wiener filtering gives comparable result to SML with mean shift filtering. Similar results are obtained for Impulse noise.

Figure 3 shows the effect of mean shift filtering in SFF for simulated cone object with Gaussian noise. As the noise

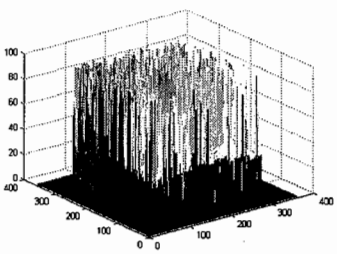
level grows, filtering effect becomes more apparent.



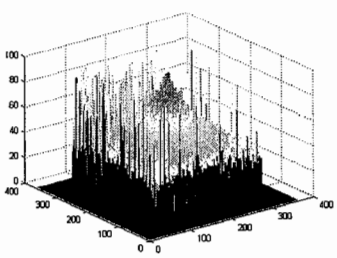
SML (Var = 0.005)



Mean Shift + SML(Var=0.005)



SML (Var = 0.01)



Mean Shift + SML(Var=0.01)

Figure 3. Comparison of SFF for simulated cone with Gaussian noise.

From the results, mean shift filter shows competitive performance to the filters commonly used for specific noise types – Gaussian noise and Impulse noise. Therefore, mean shift filter can be applied to remove noise introduced during image acquisition before computing focus measure in case the noise type is not known.

6. CONCLUSION

In SFF, selection of focus measure – which represents sharpness – is the most critical problem because it largely affects the eventual quality of the shape recovery. Traditional noise removal filters remove not only noise but also blurs the sharpness of the image. In this paper, mean shift filtering was applied as a preprocessor to remove noises imposed on image sequences during acquisition by imaging system while preserving the informative edges. Experimental result shows that mean shift filtering can be applied to remove Gaussian noise and Impulse noise. Applying mean shift filtering before computing focus measure is promising in case the noise type during image acquisition is not known.

ACKNOWLEDGEMENTS

This work was supported by the Korea Science and Engineering Foundation (KOSEF) grant funded by the Korean government (MOST) (No. R01-2007-000-20227-0).

REFERENCES

[1] Krotkov E., P., "Focusing", International Journal of Computer Vision 223-237 (1987).

[2] Malik, A., S., Shim, S.-O. and Choi, T.-S., "Depth map estimation using a robust focus measure", IEEE International Conference on Image Processing (ICIP) 564-567 (2007).

[3] Malik, A., S. and Choi, T.-S., "A Novel Algorithm for Estimation of Depth Map using Image Focus for 3D Shape Recovery in the Presence of Noise", Pattern Recognition 41(7), 2200-2225 (2008).

[4] Nayar S., K. and Nakagawa, Y., "Shape from focus", IEEE Transactions on Pattern Analysis and Machine Intelligence 16(8), 824-831 (1994).

[5] Malik, A., S. and Choi, T.-S., "Finding best focused points using intersection of two lines", IEEE International Conference on Image Processing (ICIP) 1952-1955 (2008).

[6] Malik, A., S. and Choi, T.-S., "Effect of noise and source illumination on 3D shape recovery", International Journal for Pattern Recognition & Artificial Intelligence 22(5), 945-958 (2008).

[7] Fukunaga, K. and Hostetler, L., D., "The Estimation of the Gradient of a Density Function, with Applications in Pattern Recognition," IEEE Trans. Information Theory 21,32-40 (1975).

[8] Cheng, "Mean Shift, Mode Seeking, and Clustering," IEEE Trans. PAMI 17(8), 790-799 (1995).

[9] Comaniciu, D. and Meer, P., "Mean Shift: A Robust Approach toward Feature Space Analysis," IEEE Trans. Pattern Analysis and Machine Intelligence 24(5), 603-619 (2002).

[10] Ramanan, D. and Forsyth, D., A., "Finding and Tracking People from the Bottom up," Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition 467-474 (2003).

[11] Malik, A., S. and Choi, T.-S., "Comparison of Polymers: A New Application of Shape From Focus", IEEE Transactions on Systems, Man, and Cybernetics--Part C: Applications and Reviews 39(2), 246-250 (2009).

[12] Malik, A., S. and Choi, T.-S., "Depth estimation by finding best focused points using line fitting", Springer Lecture Notes in Computer Science (LNCS) 5099, 120-127 (2008).

Towards Semantic Based Image Retrieval : A Review

Hui Hui Wang, Dzulkifli Mohamad, N.A Ismail

Department of Computer Graphics and Multimedia,
Faculty of computer science and information Technology, UTM, Skudai, Malaysia
hh8wang@gmail.com, dzulkifli@utm.my, azman@utm.my

ABSTRACT

This paper attempts to discuss the evolution of semantic based image retrieval. The explosive growth of image data leads to the need of research and development of Image retrieval. Image retrieval researches are moving from keyword, to low level features and to semantic features. Drive towards semantic features is due to the problem of the keywords which can be very subjective and time consuming while low level features cannot always describe high level concepts in the users' mind. Framework of semantic based image retrieval as well as the processes involved has been discussed. This paper also highlights both the already addressed and outstanding issues.

Keywords : Image retrieval, Semantic gap, semantic based image retrieval

1.0 INTRODUCTION

Image retrieval is the field of study concerned with searching and browsing digital images from database collection. This area of research is very active research since the 1970s [1, 2]. Due to more and more images have been generated in digital form around the world, image retrieval attracts interest among researchers in the fields of image processing, multimedia, digital libraries, remote sensing, astronomy, database applications and other related area. Effective and fast retrieval of digital images has not always been easy, especially when the collections grow into thousands. An effective image retrieval system needs to operate on the collection of images to retrieve the relevant images based on the query image which conforms as closely as possible to human perception.

2.0 EVOLUTION OF IMAGE RETRIEVAL

The purpose of an image database is to store and retrieve an image or image sequences that are relevant to a query. There are a variety of domains such as information retrieval, computer graphics, database management and user behavior which have evolved separately but are interrelated and provide a valuable contribution to this research subject. As more and more visual information is available in digital archives, the need for effective image retrieval has become clear [3,4] In image retrieval research, researchers are moving from keyword based, to content based then towards semantic based image retrieval as shown in Figure 1 below.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754626 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853332

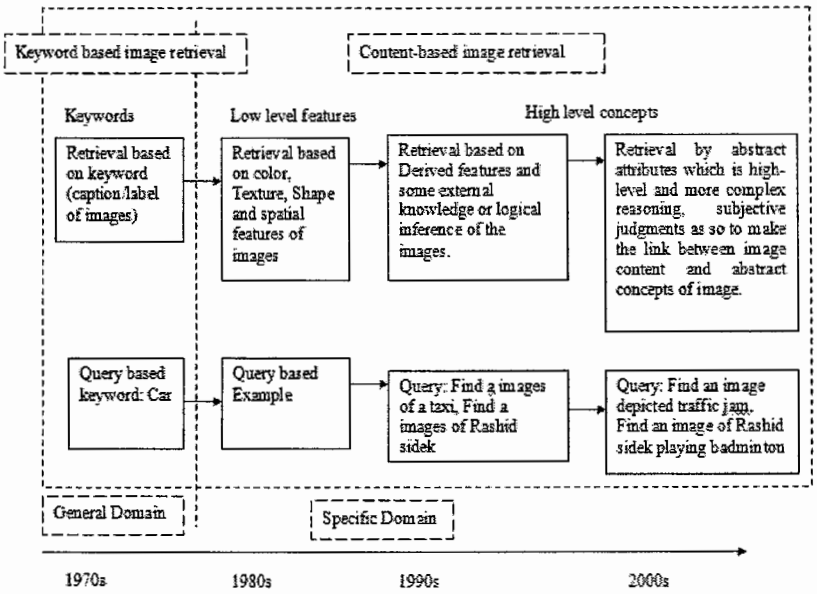


Figure 1: The evolution of image retrieval

In 1970s, the conventional image retrieval system used keywords as descriptors to index an image. However the content of an image is much richer than what any set of keywords can express. Text-based image retrieval techniques employ text to describe the content of the image which often causes ambiguity and inadequacy in performing an image database search and query processing. This problem is due to the difficulty in specifying exact terms and phrases in describing the content of images as the content of an image is much richer than what any set of keywords can express. Since the textual annotations are based on language, variations in annotation will pose challenges to image retrieval

At the 1980s, Content-Based Image Retrieval (CBIR) has then been used as an alternative to text based image retrieval. IBM was the first taking an initiative by proposing Query-By Image Content (QBIC). QBIC developed at the IBM Almaden Research Center is an open framework and development technology. Unlike keywords-based system, visual features for contents-based system are extracted from the image itself. CBIR can be categorized based on the type of features used for retrieval which could be either low level or high level features. At early years, low level features include color, texture, shape and spatial relations were used. Excellent survey on low-level image feature extraction in CBIR system can be found in [5].

Although there are many sophisticated algorithms to describe color, shape and texture features approaches, these algorithms do not satisfied and comfort to human perception This is mainly due to the unavailability of low level image features in describing high level concepts in the users' mind. For example, finding an image of a sport car in the middle of the road. The only way a machine is able to do this is by automatic extraction of the low level features that are represented in the low level features from the images with a good degree of efficiency.

Thus, the semantic gap which is the lack of correlation between the semantic categories that a user requires and the low-level features that CBIR systems offer has been explored. This is due to the fact that the visual image feature descriptors extracted from an image cannot (as yet) be automatically translated reliably into high-level semantics [4].

In 2000s, semantic based image retrieval has been introduced. This is due to neither a single features nor a combination of multiple visual features could fully capture high level concept of images. Besides, the performance of image retrieval system based on low level features are not satisfactory, there is a need for the mainstream of the research converges to retrieve based on semantic meaning by trying to extract the cognitive concept of a human to map the low level image features to high level concept (semantic gap). In addition, representing the image content with semantic terms allows users to access images through text query which is more intuitive, easier and preferred by the front end users to express their mind compare with using images.

3.0 TOWARDS SEMANTIC BASED IMAGE RETRIEVAL

3.1 Semantic Gap and Semantic Features

Bridging the semantic gap for image retrieval is a very challenging problem yet to be solved. The semantic gap is described 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. [6]

Semantic feature extraction and representation has been identified as an important issue to bridge the semantic gap in visual information access. It has been addressed as a good description and representation of an image and is able to capture meaningful contents of the image. Spatial relationship among these objects/regions can be used to further increase the confidence in image understanding. Besides, users preferred to express their information needs at the semantic level instead of the level of preliminary image features. Moreover textual queries usually provide more accurate description of users' information needs.

3.2 Framework of semantic based image retrieval

The framework of semantic based image retrieval is shown in Figure. 2

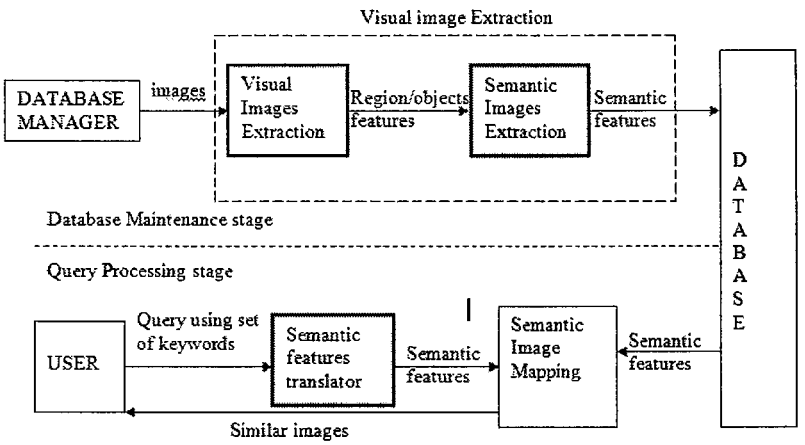


Figure 2 : The framework of semantic based image retrieval

The Figure 2 describes the framework and components of the semantic based image retrieval system and the communication between them.

The architecture consists of two stages which are the Database Maintenances stage that is responsible to get and store the images from Database Manager and the Query Processing Stage that performs user query processing. The Database stores the images from the Database Manager and also enables retrieval of the images in the user query processing.

3.2.1 Database Maintenance Stage

Database manager will load all images into visual image extraction process. The low level features of images that will be used as an input for semantic images extraction will be extracted to produce the semantic features that will be loaded into database. The details of the visual image extraction processes are discussed by the corresponding data flow diagram in Figure 3.

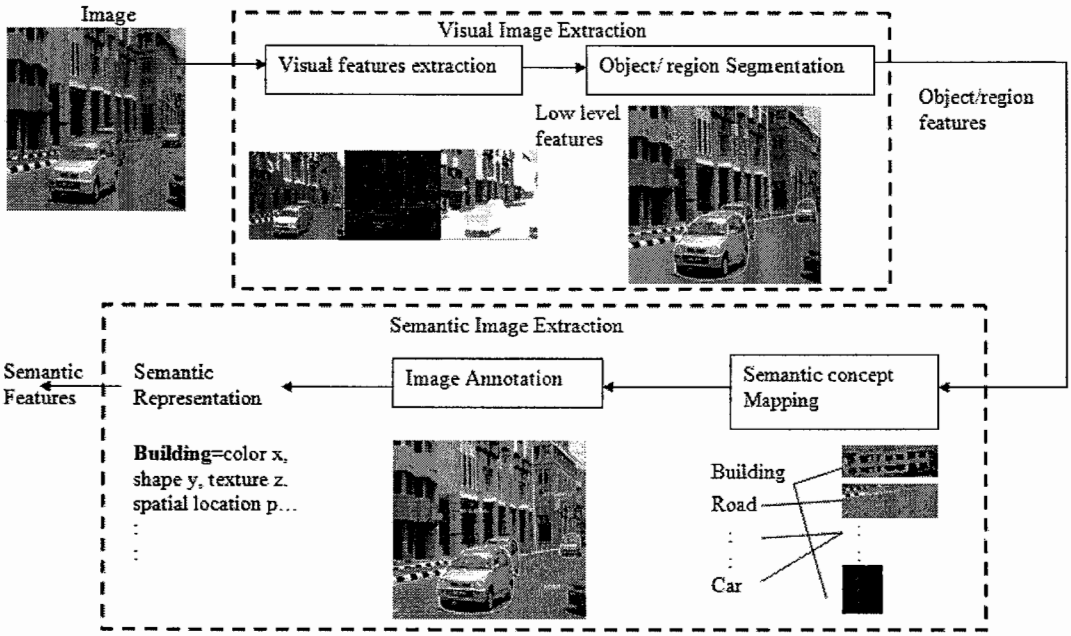


Figure 3 : The visual image extraction process data flow

The data flow diagram of the visual images extraction process is shown in Figure 3. All images need to go through the visual feature extraction process where it needs to extract the low level features of images such as color, shape and texture. The low level features will be used as an input for object/region segmentation process to extract meaningful regions based on the similar characteristics of the visual features and identification of interesting objects.

Next, the object/region features will go into Semantic Image Extraction process. First, it needs to go through semantic concept mapping process to find the best concept to describe the segmented region/objects based on the visual features. It usually will be done through supervise or unsupervised learning tools to associate the low level features with object concept. Secondly, the regions/objects representation will be annotated with textual word through image annotation process. Finally, images will be represented in semantic features that will be loaded and stored into database.

3.2.2 Query Processing stage

Image retrieval can be queried based on high level concept. User can have query based on a set of textual words. Then, it will go into semantic features translator to get the semantic features from the user query. The details of the semantic features translator processes are discussed by the corresponding data flow diagram in Figure 4

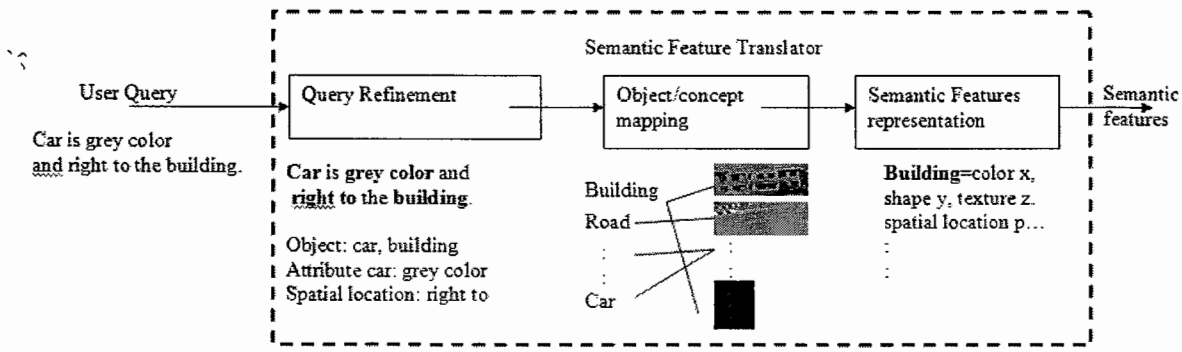


Figure 4 : The data flow of Semantic features translator

The data flow diagram of the semantic features translator process is shown in Figure 4. A set of textual words from user will be used as input for query refinement process. In the query refinement process, the meaningful words (concept) will be extracted and stop words (is, a, the and etc) will be removed. Next, the set of meaningful words will be used as input in object concept mapping, where the meaningful words (concept) will be assigned with list of possible low level features and region/object representation and semantic features will be extracted from semantic feature representation process. Finally, the semantic image mapping is responsible in estimating the similarity between the user query and database images and then retrieving images to the given query textual words.

Table I shows the common techniques used in bridging the semantic gap in image retrieval or in other words, mapping from low level to high level concepts.

Techniques		**Researchers**	**Approaches**
Annotation	Manual	Wikipedia image collection[8],	User annotated
		google image labeler [9]	User annotated
	Semi-Automatic	Bradshaw [10]	Bayer probability
		Ghoshal *et al* [11]	Co-Occurrence model
		Ilaria Bartolini and Paolo Ciaccia[12]	Graph based link
	Automatic	Huang *et al.*[13]	Decision Trees
		Feng and Chua [14]	Bootstrapping
		Gao et al. [15]	Latent Semantic Analysis
		Mori et al. [16]	hidden Markov model
Relevance feedback		Yang et al. [17]	Semantic feedback mechanism
		Rege et al. [18]	multi-user relevance feedback(user-centered semantic hierarchy)
Object Ontology		P.L.Standchey et al [19]	Color representation ontology
		V.Mezaris[20]	High level concept ontology
		Huang Wang [21]	Multi-modality ontology
Semantic template		S.F.Chang [22]	Semantic Visual Template

Table 1 : Research works on bridging the semantic gap in image retrieval

3.3 Issues and Trend

Even though the image retrieval research is moving towards semantic concept, much initial research in semantic image retrieval are focusing just on the simple semantic retrieval such as retrieval of objects of a given type. Little attention is on the retrieval by abstract attributes which involve a significant amount of high-level reasoning about the meaning and purpose of the objects or scenes depicted. In other words, the retrieval by abstract attributes was still not satisfied to human perception. Moreover, the retrieval involved human interference and is time consuming besides inconsistency. There is a need to further increase the confidence in image understanding and to effectively retrieve similar images that are conformed to human perception and without human interference.

There are still some spaces which need to be improved besides the challenges that are associated with mapping low level to high level concepts. Researchers are moving towards to intelligent image retrieval that are also supports more abstract in concept by understanding the image content in terms of high level concepts, which is closely related to the problem of computer vision and object recognition besides more intelligent system. The domain should be not specific but broad where all the extracted semantic features are applicable for any kind of images collection.

4.0 CONCLUSION

This paper provides a study of image retrieval work towards semantic based image retrieval. Recent works are mostly lack of semantic features extraction and user behavior consideration. Therefore, there is a need of image retrieval system that is capable to interpret the user query and automatically extract the semantic feature that can make the retrieval more efficient and accurate to bridge the semantic gap problem in image retrieval.

REFERENCES

[1] A. Rosenfeld. "Picture processing by computer". ACM Computing Surveys, 1(3):147-176(1969)

[2] H. Tamura and S. Mori. "A data management system for manipulating large images". In Proceedings of Workshop on Picture Data Description and Management, pages 45-54, Chicago, Illinois, USA, April 1977

[3] Ying Liua, Dengsheng Zhang, Guojun Lu, Wei-Ying M. "A survey of content-based image retrieval with high-level semantics". Pattern Recognition 40 262 – 282(2007)

[4] R.Datta, D.Joshi, J.Li, J.Z.Wang. "Image Retrieval: Ideas, Inuences, and Trends of the New Age". ACM Transactions on Computing Surveys, Vol. 40, No. 2(2008)

[5] Y. Rui, T.S. Huang, S.-F. Chang, "Image retrieval: current techniques, promising directions, and open issues", J. Visual Communication. Image Representation 10 (4) 39–62 (1999)

[6] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R. "Content-based image retrieval at the end of the early years". IEEE Trans. Pattern Anal. Mach. Intell. 22 1349–1380 (2000)

[7] Zhou, X, S., Huang, T.S. "Relevance feedback for image retrieval : A comprehensive review". Multimedia System, 8(6), 536-544(2003)

[8] http://wikipedia.org

[9] http://images.google.com/imagelabeler/

[10] Bradshaw B. "Semantic based image retrieval: a probabilistic approach". Proceedings of the ACM International Conference on Multimedia, Los Angeles, California, Oct. 30-Nov.4, 167-176(2000)

[11] Ghoshal A, Ircing P, Khudanpur S. "Hidden Markov models for automatic annotation and content-based retrieval of images and video". Proceedings of the 28th International ACM SIGIR Conference on Research and Development in Information Retrieval, Salvador, Brazil, 15-19: 544-551(2005)

[12] Ilaria Bartolini and Paolo Ciaccia. "Towards an Effective Semi-Automatic Technique for Image Annotation".SEBD.(2007)

[13] Huang J., Kumar SR., Zabih R. "An automatic hierarchical image classification". Proceedings of the ACM International Conference on Multimedia, Bristol, UK, Sep. 12-16, 219-228(1998)

[14] H. Feng and T.-S. Chua. "A bootstrapping approach to annotating large image collection". Proceedings of the Fifth ACM SIGMM international workshop on Multimedia information retrieval (MIR '03),pages 55-62, Berkeley (2003)

[15] S. Gao, D.-H. Wang, and C.-H. Lee. "Automatic image annotation through multi-topic text categorization". Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), volume 2(2006)

[16] Y. Mori, H. Takahashi, and R. Oka. "Image-to-word transformation based on dividing and vector quantizing images with words". In Proceedings of the First International Workshop on Multimedia Intelligent Storage and Retrieval Management. (1999)

[17] C. Yang, M. Dong, and F. Fotouhi. S-iras: "An interactive semantic image retrieval and annotation system". International Journal on Semantic Web and Information Systems, 2(3):37-54(2006)

[18] M. Rege, M. Dong, and F. Fotouhi. "Building a user-centered semantic hierarchy in image databases". Multimedia Systems, 12(4-5):325-338.(2007)

[19] P.L. Stanchev, D. Green Jr., B. Dimitrov, "High level color similarity retrieval", Int. J. Inf. Theories Appl. 10 (3) 363–369 (2003).

[20] V. Mezaris, I. Kompatsiaris, M.G. Strintzis, "An ontology approach to object-based image retrieval", Proceedings of the ICIP, vol. II, pp. 511–514(2003)

[21] Huan Wang, Song Liu, liang Tien Chia. "Image retrieval with a multi-modality ontology". Multimedia System (2008)

[22] S.-F. Chang, W. Chen, H. Sundaram, "Semantic visual templates: linking visual features to semantics", International Conference on Image Processing (ICIP), Workshop on Content Based Video Search and Retrieval, vol. 3, pp. 531–534(1998)

An approach for ordered dither using artificial neural network

Arpitam Chatterjee[1], Bipan Tudu[2], Kanai Chandra Paul[1]

([1]Department of Printing Engineering, Jadavpur University, Kolkata 700032, India
[2]Department of Instrumentation & Electronics Engineering, Jadavpur University, Kolkata 700032, India
Email: arpitamchatterjee@print.jusl.ac.in)

ABSTRACT

Ordered dither is one of the popular techniques for digital halftoning where the original continuous tone image is thresholded against an orderly generated screen matrix. This paper presents a technique to generate the screen matrix using three-layer back-propagation multi layer perceptron (BP-MLP) artificial neural network (ANN) model. The image raw data has been preprocessed prior feeding to the input layer. The output obtained at the hidden layer of the model has been considered as the screen matrix for ordered dither. The results achieved using this technique have been evaluated subjectively as well as objectively using commonly used quality indices like peak signal to noise ratio (PSNR), universal quality index (UQI) and structural similarity index measure (SSIM).

Ordered dither, digital halftoning, thresholding, artificial neural network (ANN), back-propagation multi layer perceptron (BP-MLP), PSNR, UQI, SSIM.

1. INTRODUCTION

Digital halftoning[1] is the process of rendering continuous tone images on bi-level media like printing, bi-level displays, etc. If an 8-bit continuous tone image is represented as $I(i,j) \in [0,1]$, where the image data is preprocessed in a range of (0, 1), then digital halftoning is the technique to represent $I(i,j)$ with a 1-bit image $H(i,j)$ such as $H(i,j) \in \{0,1\}$ where, '0'represents black and '1' represents white. The simplest way to perform this task can be fixed thresholding where each pixel intensity is thresholded against a fixed value (T) and any value higher or equal to T will be set to '1' and rest will be '0'. The major problem with such linear process is the lack of detail retention and it is not a parallel process. To generate a better quality and more detail retaining halftone image several techniques have been proposed [2,3,4]. The methods proposed can be segregated in two broad categories in terms of operation i.e. point processing approaches and neighborhood processing approaches[5]. In case of point processing the processing pixels do not involve the neighborhood pixels as in case of neighborhood processing. Ordered dither[6] being one of the popular methods under the point-processing category where in-stead of a fixed value, a screen matrix is employed. This parallel point processing method[7] can be divided in to two classes namely; clustered-dot and dispersed-dot ordered dither. Screen matrices for clustered-dot ordered dither are design to mimic the halftoning process used in conventional printing such as the dot grows around a pixel with the decrease in the pixel intensity of continuous tone image[2]. On the other hand dispersed-dot ordered dither can produce better illusion of the contone image but it can be used only on the system where individual pixel can be displayed[5]. The design of screen matrix for ordered dither has been reported in many literatures[8-10] but Bayer's proposition[11] is very well-known. Error-diffusion is one of the most popular paradigm under neighborhood processing approaches where the error resulted while processing an individual pixel i.e. difference between original pixel intensity and halftoned pixel intensity, is distributed in an weighted manner to the neighborhood pixels. Some of the most popular distribution weight kernels were proposed by Floyd-Steinberg[12], Jarvis[4] and Stucki[13].

In this paper improvement of the halftone image by non-linear generation and optimization of the screen matrix using BP-MLP neural network model has been presented. The proposed method has been tested with a set of grey scale images and performance evaluation of the technique has been carried out using PSNR, UQI and SSIM. The pictorial as well

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754627 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853179

tabular presentation of the result, presented at results and discussion section, confirms the ability of the proposed method to generate better quality halftoned image compared to conventional ordered dither screens.

2. PROPOSED METHOD

The method proposed in this paper constitutes of three major processing steps. First, preprocessing of raw data with suitable techniques. Second generation of screen matrix by optimizing the preprocessed data using BP-MLP for ordered dither and finally performing the thresholding to get the halftone image. A schematic diagram of the proposed method has been illustrated in figure I.

Step 1: Preprocessing of the image raw data

In the proposed method image raw data $I(i,j) \in [0,255]$ has been preprocessed[14] before processing. Preprocessing is an essential and important step in practices involving mathematical models since it significantly contributes towards the performance of the model[15]. Hence, for ANN based models also, preprocessing plays a vital role towards the efficiency of the model.

Step 2: Generation and optimization of screen matrix using BP-MLP

The preprocessed data is fed to the BP-MLP model. The architecture used in this paper consists of three layers, namely input, hidden and output layer. The input layer consists of same number of neuron as the number of pixels subjected to processing at a time. The number of neurons at hidden layer has been kept same as the input layer and the number of neuron at the output layer is one. The transfer function used for the hidden layer neurons is *'logsig'* which nonlinearly transfer the input data and transfer function for output layer neurons have been set to *'purelin'*.

For simplicity, the image has been partitioned using $N \times N(10 \times 10)$ window hence, at each processing step N^2 pixels have been processed in a parallel fashion. The target at the output neuron has been set to the *mean value* (μ) of the N^2 pixels under processing.

The training of the neural network has been accomplished using *error back-propagation technique*[16]. After completion of the training a non-linearly transformed dataset of same size as the image is obtained at the hidden layer output which has been used as the screen matrix $T(i,j)$.

Step 3: Thresholding

In this stage the preprocessed image data $I(i,j)$ is thresholded with the screen matrix $T(i,j)$ as shown in equation I.

$$H(i,j) = \begin{cases} 1, if T(i,j) \geq I(i,j) \\ 0, otherwise \end{cases} \quad \text{..... (I)}$$

where, $H(i,j)$ is the halftone image.

3. RESULTS AND DISCUSSIONS

The proposed method has been tested with a set of test images. The results with three images namely *mandrill, monalisa* and *lena* have been presented in figure II. The figure II shows subjective comparison between the results obtained using proposed technique and standard screening matrix for ordered dither. The visual appearance shows that the halftone image generated using proposed method is not offering the visually objectionable structures and also false contouring problem is not present like other images.

The quality of the halftone images have been evaluated with peak signal-to-noise ratio PSNR [17], UQI [18] and SSIM [19]. PSNR calculated the peak signal-to-noise ratio in *dB* as shown in equation II, but it alone fails to interpret the image quality since it does not involve any parameter related to the vision. Thus for image quality evaluation metrics like UQI and SSIM add more significance since they are based on the *human visual system* (HVS) parameters. The mathematical expressions for UQI and SSIM have been shown in equation III and IV respectively. The common feature between these three quality metric is that for all these three higher value indicates better quality.

$$PSNR = 10\log_{10}\left(\frac{L^2}{MSE}\right) \quad \text{.....(II)}$$

where, 'L' represents the allowable dynamic range of the image (typically 256 for a 8-bit system) and 'MSE' is the *mean square error*.

$$UQI = \frac{4\sigma_{xy}\mu_x\mu_y}{\left(\sigma_x^2+\sigma_y^2\right)\left[\left(\mu_x\right)^2+\left(\mu_y\right)^2\right]} \quad \text{.....(III)}$$

$$SSIM = f\left(\left(l(x,y)\right),\left(c(x,y)\right),\left(s(x,y)\right)\right) = \left(\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1}\right)\left(\frac{2\sigma_x\sigma_y+C_2}{\sigma_x^2+\sigma_y^2+C_2}\right)\left(\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3}\right) \quad \text{.....(IV)}$$

For both equation III and IV σ_x and σ_y denote the standard deviation of the input image, resulting halftone image. The average intensity value for input image and output image have been expressed as μ_x and μ_y respectively. σ_{xy} can be calculated as $\frac{1}{N-1}\sum_{i=1}^{N}\left(x_i-\mu_x\right)\left(y_i-\mu_y\right)$. The SSIM expression as shown in equation IV clearly shows the consideration of luminance, contrast and structure components as represented by $l(x,y)$, $c(x,y)$ and $s(x,y)$ respectively which are independent and highly important parameters in HVS model. C_1, C_2 and C_3 are small constants used to ensure the stability and as shown in equation III, $C_1 = C_2 = 0$ in case of UQI.

The comparison between the results obtained with the method described above and with the conventional ordered dither screen while performed with three test images has been shown in table I. The results confirm that the purported method gives a higher PSNR value compared to the images found by performing ordered dithering with (9×9) and (16×16) screen matrix. The UQI values also show the same improvement. The SSIM values are almost equal to that obtained using ordered dither. The implementation of the proposed process has been accomplished with MATLAB® and the quality measures have been carried out using halftoning toolbox[20].

Table I: Comparison between proposed method and conventional ordered dither

Test Images	**Ordered dither (9×9)**			**Ordered dither (16×16)**			**Proposed method**		
	PSNR (in dB)	UQI	SSIM	PSNR (in dB)	UQI	SSIM	PSNR (in dB)	UQI	SSIM
mandril	55.5534	0.1202	0.9964	55.4429	0.1108	0.9963	56.1828	0.2701	0.9925
lena	57.2178	0.0687	0.9965	56.8922	0.0654	0.9965	58.0242	0.1440	0.9870
monalisa	55.5584	0.0622	0.9975	55.4604	0.0586	0.9975	56.3214	0.1125	0.9912

4. CONCLUSION

In this paper a neural network based technique for ordered dither has been proposed. The results show that the presented method can address the major drawbacks of conventional ordered dither, mainly, repetitive objectionable patterns and false contouring. The ANN model introduces non-linearity in the system which helps to avoid the objectionable pattern caused by periodic screen used in conventional ordered dither. The results also show that the false contouring which is very prominent for images obtained using conventional screens is not present in the images obtained using proposed technique. The false contouring becomes more visible and objectionable when the halftone image is subjected to zooming, but the images obtained using presented technique will not suffer from such problems. Apart from the subjective or appearance based comparison the mathematical or objective comparison also support the improvement in resulting halftone image. In objective measures, along with PSNR, UQI also indicated significant improvement which validates the method's betterment ability in terms of HVS system. Hence, this technique can be successfully implemented in the field of digital halftoning.

REFERENCES

[1] Lau D.L. and Arce G. R., [Modern Digital Halftoning], Marcel Dekker, New York, (2001).

[2] Mese M. and Vaidyanathan P.P., "Recent advances in Digital Halftoning and Inverse Halftoning Methods", IEEE Trans. on Circuits and Systems – I, Fundamental theory and applications, Vol. 49(6), 790 - 805 (2002).

[3] Ulichney R. A., "Dithering with blue noise", Proc. IEEE, vol. 26, 2788 – 2798 (1987).

[4] Jarvis J. F., Judice C. N. and Ninke W. H., "A Survey of Techniques for the Display of Continuous Tone Pictures on Bilevel Displays", Computer Graphics Image Processing, Vol.5, 13 – 40 (1976).

[5] Ulichney R., [Digital Halftoning], MIT Press, Cambridge, MA, (1987).

[6] Judice C. N., Jarvis J. F. and Ninke W. H., "Using ordered dither to display continuous tone pictures on an AC plasma panel", Proc. SID. 15 4 (Fourth Quarter), 161–169 (1974).

[7] Jing-Ming Guo and Yun-Fu Liu, "Improved Dot Diffusion by Diffused Matrix and Class Matrix Co-optimization", IEEE Trans. on Image Processing, Vol. 18(8), (2009).

[8] Limb J. O., "Design of dither waveforms for quantized visual signals", Bell Sys. Tech. J., 2555 – 2582 (1969).

[9] Lippel B. and Kurland M., "The effect of dither on luminance quantization of pictures", IEEE Trans. Comm., Vol. COM – 19(6), 879 – 888 (1971).

[10] Lippel B., "Two and three-dimensional ordered dither in bi-level picture displays", Proc. SID 17/2, 115 – 121 (1976).

[11] Bayer B. E., "An optimum method for two level rendition of continuous tone pictures", Rec. IEEE Conf. ICC' 73, 26 – 11 – 26 – 15 (1973).

[12] Floyd R. and Steinberg L., "An adaptive algorithm for spatial grayscale", in Proc. SID'76, 75 – 77 (1976).

[13] P. Stucki, "MECCA—A Multiple-Error Correcting Computation Algorithm for Bilevel Image Hard-Copy Reproduction," Research Report RZ-1060, IBM Research Laboratory, Zurich, Switzerland, 1981.

[14] Johnson R. A., Wichern D. W. [Applied Multivariate Statistical Analysis], Prentice-Hall International Englewood Cliffs, New Jersey, (1982).

[15] Jurs P. C., Bakken G. A.and McClelland H. E., "Computational Methods for the Analysis of Chemical Sensor Array Data from Volatile Analytes", Chem. Rev. 2000, 100, 2649-2678.

[16] Haykins S., [Neural Networks: A Comprehensive Foundation], Pearson Education Asia, Hong Kong, (2001).

[17] Wang Z. and Bovik A. C., "Mean Squared Error: Love It or Leave It?" IEEE Signal Processing Magazine, 98 – 99 (2009).

[18] Wang Z. and Bovik A. C. "A Universal Quality Index", IEEE Signal Processing Letters 9(3),81-84 (2002).

[19] Wang Z., Bovik A. C., Sheikh H. R. and Simoncelli E. P., "Image Quality Assessment: From Error Visibility to Structural Similarity", IEEE Trans. on Image Processing 13(4), 600 – 612 (2004).

[20] Evans B. L., Monga V. and Venkata D. N., "Halftoning toolbox for MATLAB," Version 1.1 (2002) 2002, [http://www.ece.utexas.edu/~bevans/projects/halftoning/].

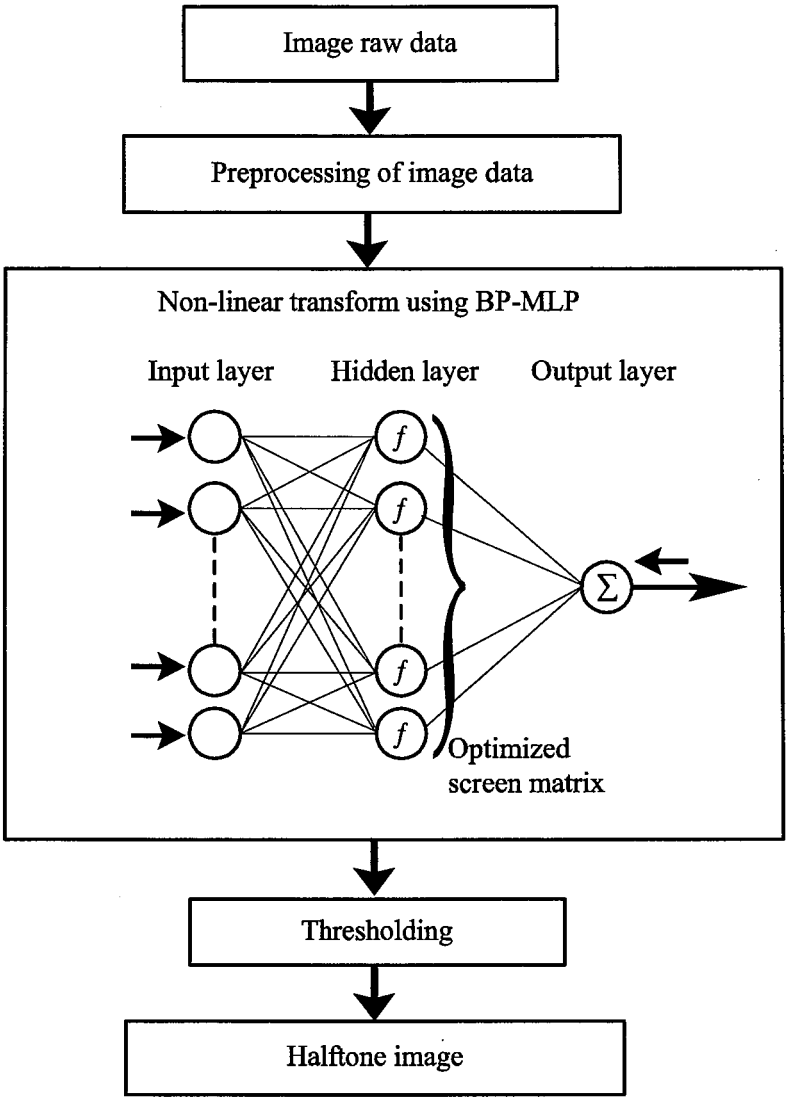


Figure I. Flow-chart for proposed ordered dither method

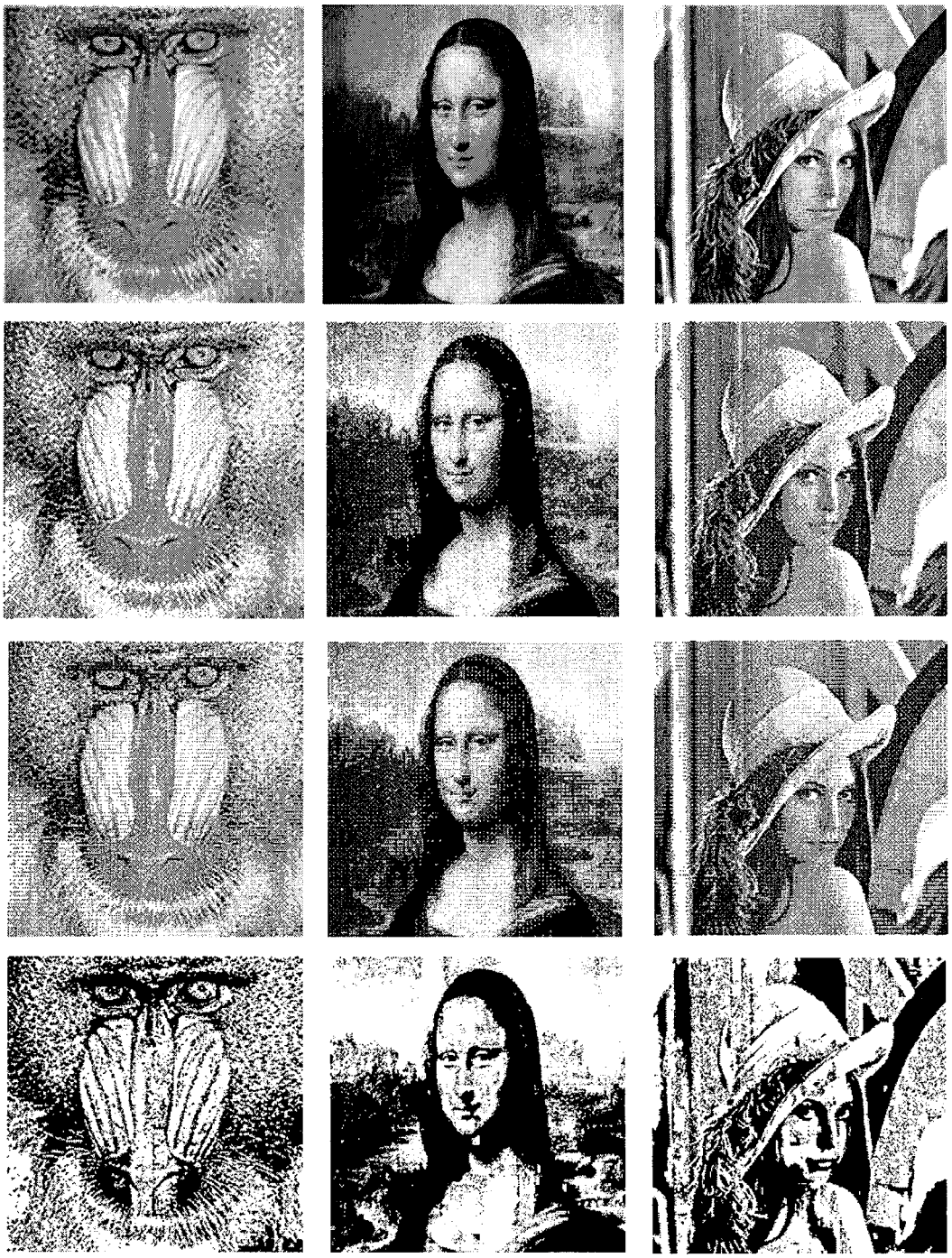


Figure II. The first row shows the original test images. The second and third row represents ordered dither halftoning with (9×9) a dispersed screen matrix and (16×16) dispersed screen matrix respectively (performed with[20]). Finally the fourth row shows the results obtained using proposed method.

Iris Detection based on Pupil Prospect Point and Horizontal Projections

Muhammad Shahid*, Tabassam Nawaz**
Department of Computer Engineering, University of Engineering & Technology, Taxila, Pakistan

ABSTRACT

Iris is one of the most discriminating human physiological traits being used for personal human identification. The success of iris based system is highly relied on the accurately captured and precisely segmented iris image. Falsc rejection rate has been a major challenge in the success of such system which primarily results from inaccurate iris segmentation. Most of the presented algorithms on iris segmentation considers pupil as perfect circle. However, according to observation this is not true in all the cases. In addition, a little angular shift in position of subject iris can further deteriorate the performance of algorithms based on circular assumption. To improve the quality of segmentation, an effective algorithm is proposed for iris segmentation which takes into consideration issues related to irregular pupil boundary as well as computation intensive nature of the prevailing algorithms. Contrary to all the previous approaches, the proposed algorithm is based on detection of pupil prospect point within the pupil region and utilizes bi-directional horizontal projections and distance parameters to detect the pupillary as well as limbus boundaries. The processing involved is linear in nature. Simulation of the proposed algorithm is done in Matlab.

Keywords: Iris detection, iris segmentation, false rejection rate, horizontal projections, pupil prospect point, biometrics

1. INTRODUCTION

Traditionally, identity cards and passwords had been major instruments of personal identification and access control. However, such methods were prone to loss, stolen and forgotten. With the phenomenal growth in the electronic field over a last decade or so, such techniques are no more considered adequate. Biometrics, as a more secure option, takes into account human attributes (physiological or behavioral) for authentication, removing altogether the inherited constraints of traditional methods. Further, they have relieved the user from memorizing, hiding or carrying them along. Biometrics involves different techniques. Few of them focus on hand signatures, voice signatures, fingerprints, palm prints, facial geometries and iris patterns. Based on numerous trials and tests performed by the researchers to determine the biometric method that provides more viable source of identification yield that iris patterns are most reliable and robust for human identification. There are several reasons for these claims:

- Iris patterns cannot be changed or modified with relative ease as compared to other techniques.
- Every person has unique iris patterns which remains stable throughout life time.

Iris based systems also accompany some constraints that need to be addressed:

- A small occlusion in the iris can cause high false rejection rates.
- Eye lids may impose some iris region or semi open images.
- Eye lashes may affect iris patterns.
- Many techniques consider pupil as a perfect circle which may end up in inaccuracies in iris segmentation.
- Problem caused by a spot reflection of light.

2. PRESENTED WORK ON IRIS

A lot of research work has been done over the years in a bit to enhance the efficiency of iris detection phase. The pioneer promising algorithm in this direction was proposed by Daugman[1] which was based on Integro-differential operator for iris detection. The operator assumes pupil and limbus as circular contours and performs circular edge detector.

*muhammud_shahid@yahoo.com; **tabassam@uettaxila.edu.pk

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754628 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853387

$$\max_{r,x_o,y_o} \left| G_\sigma(r) * \frac{\partial}{\partial r} \oint_{r,x_o,y_o} \frac{I(x,y)}{2\pi r} ds \right| \tag{1}$$

where $I(x,y)$ is the source image, $G_\sigma(r)$ is a Gaussian smoothing function, and s is the contour of the circle defined by parameters r, x_o, y_o. The operator, by varying radius and centre x and y position of the circular contour, searches for the circular curve with maximum change in pixel values. Many other techniques were purposed after Daugman. Wildes[2] purposed a technique in which detection was based on Hough transform and used circular detection techniques instead of Daugman's contour operator. Boles et al.[3] utilized gray levels of images for detection. He obtained zero crossings using wavelet transforms for iris representation. Qi-Chaun et al.[4] purposed algorithm which makes use of black pupil to locate center and self adaptive window for iris detection. This method was quite similar to that used by Wildes. Tisse et al.[5] purposed algorithm which combines the traits of integro-differential operator and Hough transform to increase the performance of iris detection.

3. PROPOSED ALGORITHM

3.1 Image Preprocessing

Iris recognition systems have to process million of images for recognizing each subject. The method for iris detection should be such that minimum processor utilization must be required. Based on such consideration, iris images are first preprocessed to reduce the size of input images in a bit to cut down the processing effort.

$$G(x,y) = resize[I(x,y) * [M/4, N/4]] \tag{2}$$

where $I(x,y)$ is the input image, $G(x,y)$ is the resized image and M=1,2,.....320, N=1,2.....280 are the image dimensions.

3.2 Edge Map

Edge map, $EdgeMap(x,y)$ is obtained by applying canny edge detector on the resized image $G(x,y)$, with uniform threshold parameter of 0.1.

$$f\nabla = F\left(\frac{\partial^2 \ G * I}{\partial n^2} \right) \tag{3}$$

Pupillary and limbus edges are based on comparatively larger connected components. Edges of relatively smaller pixel areas don't qualify for such boundaries. A threshold of ten pixels is chosen. Edges having connected pixels of less than or equal to ten pixels are dropped from the edge map.

$$EdgeMap(x,y) = \begin{cases} 1 \rightarrow EdgeMap_{EdgeArea} > 10 \\ 0 \rightarrow otherwise \end{cases} \tag{4}$$

The edges in the resultant $EdgeMap(x,y)$ are labeled to yield $LblEdgeMap(x,y)$.

3.3 Pupil Prospect Point Detection

The foremost step in pupil detection is to locate the darkest pixel in the image through histogram analysis with the consideration that the same must be within pupil region. Obviously, the complete eye image contains other regions too, whose pixels gray level values will be in vicinities to those retained by pupil pixels. Eye brows, eye lids and eye lashes have been some of them.

Histogram of the image is calculated to find distribution of the image pixels across 256 gray levels.

$$h(r_k) = n_k \text{ for } k = 0,1,2,.....255 \tag{5}$$

where r_k is the k[th] gray level and n_k is the number of pixels in the image having gray level r_k.

Pupil is among the darkest region of the eye image. Therefore, the pupil pixels will be concentrated along the lower histogram bins. The lower most bins are analyzed for pupil prospect point. A threshold based on the number of pixels retained by each histogram bin is applied and the first bin from the lower gray levels containing more than five pixels is considered qualified. The threshold is applied since pixels contained in the lower bins having less than five pixels might be odd stranded pixels due to noise factors.

The position of each pixel of the selected bin is located in the gray image. The neighborhood information of each of these pixels is taken into account to find the pupil prospect point. Gray scale values of pixels in 3 x 3 neighborhood of each pixel of the selected bin is obtained and averaged. Such averaging is applied to 3 x 3 neighborhood of all the pixels contained in the selected bin. The pixel with lowest neighborhood averaged gray scale value is picked as pupil prospect point, (PPP).

$$Avg(i) = \frac{1}{9}[\sum_{x=-1}^{x=1}\sum_{y=-1}^{y=1} G(x,y)] \tag{6}$$

$$ProspectPnt = \min(avg(i) \rightarrow i = 1,2,....) \tag{7}$$

3.4 Pupil Detection

Bi-directional horizontal outward projections are drawn from PPP, in the labeled edge map as illustrated in figure 1 and looked for both the projections striking pixels on the same labeled edge which would be the pupil boundary. It is common observation that the pupil edge is continues and ring shaped. If the location of obtained PPP is within the pupil region, it will definitely double strike the same labeled edge, once on either side of PPP. In case if, it doesn't strikes twice, then the obtained PPP will be assumed of non-pupil. The next lowest average pixel is considered as PPP for outward projections and assessed on similar parameters. The process is recursive in nature and repeat until it strikes twice the same labeled edge.

$$PupilPnt1 = match(\sum_{j=n}^{j=y_{PP}} EdgeMap(x,j),1) \tag{8}$$

$$PupilPnt2 = match(\sum_{j=0}^{j=y_{PP}} EdgeMap(x,j),1) \tag{9}$$

Where $EdgeMap$ is the binary image of the subject. Y_{PP} is the y-component of Prospect Point location and n is the total number of pixels in each row.

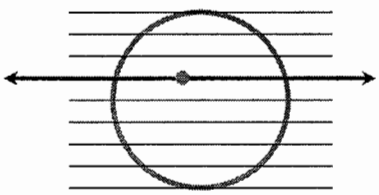


Figure 1. Horizontal outward project from PPP.

All the connected pixels are detected in the edge map and the corresponding pixels in the original image $G(x,y)$ are marked. It will result in a closed, continuous and smooth curve of connected components of pupil boundary.

$$G(x,y) = \begin{cases} 255 \rightarrow LblEdgeMap(x,y) = Label1; \\ G(x,y) \rightarrow otherwise \end{cases} \tag{10}$$

3.5 Pupil Center Detection

The utility of determination of pupil center is in the detection of limbus boundary. The approach adopted is illustrated in figure 2 and is as under:

- Horizontal distances of the pupil edge pixels are calculated on each pixel row. Row having maximum pixel distance is obtained. Dividing the pixel distance by two will yield the x-component of the center co-ordinate.

$$C_x = Pupil_{LowestCol} + Pupil_{HighestCol} \; /2 \tag{11}$$

- Vertical distances of the pupil edge pixels are calculated on each pixel column. Column having maximum pixel distance is obtained. Dividing the pixel distance by two will yield the y-component of the center co-ordinate.

$$C_y = Pupil_{LowestRow} + Pupil_{HighestRow} \; /2 \tag{12}$$

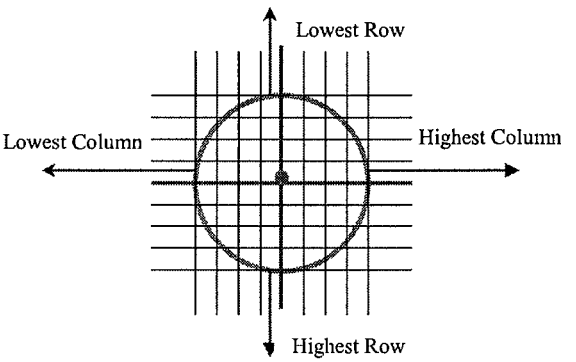


Figure 2. Illustrates pixel rows and columns consideration for pupil center detection

3.6 Limbus Detection

Limbus boundary is not as distinct as pupil. A lot of varying size edges can also come across during outward projection. Method for limbus detection needs to be robust enough to distinguish between ordinary edge and the limbus edge.

Pupil edge co-ordinates, both left and right, of the pixel row holding the pupil center co-ordinate are calculated. Outward horizontal projection starting from these pupil edge pixels are drawn in *EdgeMap* in a bit to strike limbus edge. X-component is kept constant while y-component is decremented by 1 in leftward projection and incremented by 1 in rightward projection and checked for edge pixel. *EdgeMap* is a binary image representing edge pixels by 1 and rest of the image by 0. All the non-zero pixels found during outward projections represent edges. Y-components of all the edge pixels found during left and rightward projection are stored in two separate arrays to be able to equate their respective distances from the pupil edge at later stage.

$$IrisPnts1 = match(\sum_{j=n}^{j=PupilPnt1} EdgeMap(x, j), 1) \tag{13}$$

$$IrisPnts2 = match(\sum_{j=0}^{j=PupilPnt2} EdgeMap(x, j), 1) \tag{14}$$

Where *IrisPnts*1 and *IrisPnts*2 are arrays for storing y-components of all the non-zero pixels of left and rightward projections respectively. *EdgeMap* is the binary image of the subject. *PupilPnt*1 and *PupilPnt*2 are the starting location for left and rightward projections respectively. X-component remains static during processing. Distance (y-components) of each obtained edge pixel from starting pupil edge pixel is obtained and store separately.

$$Dist1(i) = PupilPnt1_{y-comp} - IrisPnt1(i)_{y-comp} \tag{15}$$

$$Dist2(i) = PupilPnt2_{y-comp} - IrisPnt2(i)_{y-comp} \tag{16}$$

The limbus boundary is at equidistance from either side of the pupil boundary. No other edge on the path resembles such feature. Although there might be many edges in the region between the pupil and limbus boundaries but it would be very rare phenomenon of two edges, other than limbus maintaining such geometrical configuration of having equidistance from either side of pupil boundary.

Besides this, limbus boundary has another feature that in most of the cases it can be one complete edge having circular contour around pupil and both the outward projections striking the same edge twice – once during each projection. There may be many other edges but they will be of smaller area and probably none of them is maintaining even near circular contour around the pupil boundary. So even such edges might fulfill the equidistance factor in a rear cases but may not represent uni-edge. The first test in a bit to find limbus is to check whether the edge pixels found on both sides belong to the same edge and whether they maintain some mathematical relationship between them in term of distance factor.

The distance maintained by the edge pixels found on left and right-side of pupil boundary is compared. Two edge pixels, obtained out of different projection, having equidistance are picked. A tolerance of two pixels is allowed to cater for the scenarios of angular shift in the iris position in the eye region. To enhance the performance of algorithm, distance of more than fourteen pixels is considered for the purpose of comparison. During thorough testing of the algorithm on CASIA database[6], it has been observed that the distance between pupil and iris edge is relatively larger than fourteen pixels. Unitizing this assumption will reduce complexity involved in the assessment of many edge pixels.

Based on the fact that pupil and iris are not always co-centric, the next step was to find the iris centre co-ordinates. The mean location of the two iris edge pixels is computed, as shown below:

$$C_{iris} = IrisEdgePt_{Left} + IrisEdgePt_{Right} \; /2 \tag{17}$$

Generally, the set of pixels at the limbus edge seldom characterizes an edge completely because of breaks in edge due to non-uniform illumination and other effects that introduce spurious intensity discontinuities. Thus the algorithm needs procedures to assemble edge pixels into meaningful iris-sclera boundary. Now, since we have obtained the center co-ordinates of the iris-sclera boundary as well as two boundary edge pixels locations, a simple method that is employed using this information is drawing circle which works on center co-ordinate and radius.

Finally, the image is to be resized to its original dimension (i.e. 280 x 320 pixels).

$$G(x,y) = resize[G(x,y) * [M*4, N*4]] \tag{18}$$

The process of reducing the image size for performance enhancement and then regenerating original size result in degradation and loss of lots of pixel characteristics in term of gray scale information. To preserve the original information, pixel locations of all the marked pupillary as well as limbus boundaries (turned to grayscale 255) in the resized image are located and the respective pixels in the corresponding original image are marked.

$$f(x,y) = \begin{cases} 255 \rightarrow G(x,y) = 255; \\ f(x,y) \rightarrow otherwise \end{cases} \tag{19}$$

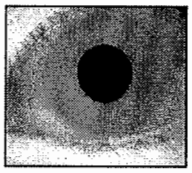
4. PERFORMANCE ANALYSIS

This research work has focused on detection accuracy and computational complexity in iris detection and has proposed a new scheme for iris segmentation to increase performance and accuracy of iris recognition system. The success of iris recognition system is highly dependent on iris segmentation. A poor segmentation can lead to inaccuracy and ultimate failure of the recognition system. Precisely detecting such boundaries, especially, while considering the irregular nature of pupil boundary was one of the considerations of research to develop a new algorithm. The proposed algorithm takes into consideration that an actual pupil boundary is a near-circular contour rather than a perfect circle. Further, a minor angular shift in iris position at the time of capturing the image further distorts the near circular dimension. The proposed algorithm has successfully detected almost all the pupil boundaries while detection of limbus boundary too has attained very high success results. The second motivation behind the work was to reduce the computation complexity of the algorithm. Algorithm proposed by Daughman[1], Wildes[2], Kong et al.[7], Tisse et al.[5], and L. Ma et al.[8], all look for circular contour, both in pupillary and limbus boundaries detection and involves global search methods. The circular assumption of pupillary boundary can lead to imperfect pupil detection as shown in figure 3(b). The approaches are computation intensive, especially those using Hough transform operates on 'brute-force' approach which is computationally too

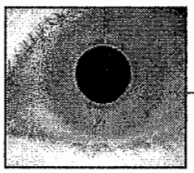
intensive to be suitable for real-time applications. The proposed scheme involves linear processing. Rather then looking for pupillary and limbus center co-ordinate, it takes into account the darkest region strategy, looks for pixels in the pupillary boundary and then draws outward projections in image edge map to mark both the boundaries. The algorithm has been run on CASIA database[6] version 1.0 and attained detection precision of 98.68% and 87.70% in pupillary and limbus boundaries. The effect of irregular detection in pupillary boundary due to considering it circular contour has also been successfully addressed and proven as shown in figure 3(c).

Table. 1. Edgewise performance of algorithm

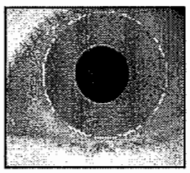
Database	Pupillary Edge	Limbus Edge
CASIA 1.0	98.68	87.70



(a) Original Image



(b) Imperfect marking of pupillary boundary based on circular assumption



(c) Perfectly marked pupillary boundary based on near circular assumption

Figure 3. Marking of boundaries

5. CONCLUSION

False rejection rate has been a major challenge in the success of iris recognition system which primarily results from inaccurate iris segmentation. To improve the quality of segmentation, a novel algorithm has been proposed for iris segmentation which takes into consideration issues related to irregular pupil boundary as well as computation intensive nature of the prevailing algorithms. Contrary to all the presented algorithms based on circular path search, this approach is based on detection of pupil prospect point within the pupil region and utilizes bi-directional horizontal projections and distance parameters to detect the pupillary as well as limbus boundaries. The processing involved is linear in nature. Simulation of the proposed algorithm is done in Matlab.

REFERENCES

[1] Daugman, J., "How iris recognition works," IEEE Transactions on Circuits and Systems for Video Technology 14(1), 21-30 (2004).

[2] Wildes, R.P., "Iris recognition: an emerging biometric technology," Proc. IEEE, 85(9), 1348-1363 (1997).

[3] Boles, W.W., and Boshash, B., "A human identification technique using images of the iris and wavelet transform," IEEE Transactions on Signal Processing 46(4), 1185-1188 (1998).

[4] Qi-Chuan Tian, Quan Pan, Yong-Mei Cheng and Quan-Xue Gao, "Fast algorithm and application of hough transform in iris segmentation," Proc. 3rd International Conference on Machine Learning and Cybernetics, Shanghai, 26-29, 3977-3980 (2004).

[5] Tisse, C., Martin, L., Torres, L. and Robert, M., "Person identification technique using human iris recognition," Proc. 15th International Conference on Vision Interface, 294-299 (2002).

[6] Chinese Academy of Sciences–Institute of Automation, Database version 1.0 of 756 Grayscale Eye Images. www.sinobiometrics.com.

[7] Kong, W.K., and Zhang, D., "Accurate iris segmentation based on novel reflection and eyelash detection model," Proc. International Symposium on Intelligent Multimedia, Video and Speech Processing, 263-266 (2001).

[8] Ma, L., Wang, Y., and Tan, T., "Iris recognition using circular symmetric filters," Proc. 16th International Conference on Pattern Recognition, 2, 414-417 (2002).

FACE DETECTION IN COLOR IMAGES USING SKIN COLOR, LAPLACIAN OF GAUSSIAN AND EULER NUMBER

Shylaja S S*, K N Balasubramanya Murthy, S Natarajan
Ramya S, Radhika S, Nikita J and Prateek R

Department of Information Science and Engineering, P E S Institute of Technology, INDIA

Abstract: In this a paper, a feature based approach to face detection has been proposed using an ensemble of algorithms. The method uses chrominance values and edge features to classify the image as skin and nonskin regions. The edge detector used for this purpose is Laplacian of Gaussian (LoG) which is found to be appropriate when images having multiple faces with noise in them. Eight connectivity analysis of these regions will segregate them as probable face or nonface. The procedure is made more robust by identifying local features within these skin regions which include number of holes, percentage of skin and the golden ratio. The method proposed has been tested on color face images of various races obtained from different sources and its performance is found to be encouraging as the color segmentation cleans up almost all the complex facial features. The result obtained has a calculated accuracy of 86.5% on a test set of 230 images.

Keywords: Digital Image Processing, LoG, Euler Number, Golden Ratio, Receiver Operating Characteristics (ROC).

1 INTRODUCTION

Face detection is an important process in many applications such as face recognition, person identification and tracking, and access control. The technique used for face detection depends on how a face is modeled. Human face detection plays an important role in applications such as video surveillance, human computer interface, face recognition, and face image database management.

***Corresponding Author:** shylaja.sharath@pes.edu

Face detection is the essential front end of numerous Digital Image Processing applications.Some recent digital cameras use face detection for auto focus. Also, face detection is useful for selecting regions of interest in photo slideshows that use a pan-and-scale Ken Burns effect.

Many researchers have proposed different approaches for face detection, the main ones being feature based and image based techniques. The feature-based techniques use edge information, skin color, motion, symmetry measures, feature analysis, snakes, deformable templates and point distribution models (PDMs). Image-based Techniques include Neural Networks, Linear Subspace methods like Eigen faces, Fisher faces, and statistical approaches like Support Vector Machines, higher order statistics.

In [1] the face detection in cluttered images is implemented using Canny edge detector. This detector uses tangential orientation and hence gives discontinuous edges. The discontinuity creates complexity in the image segmentation. To tackle this problem of discontinuity, LoG is found to be more suitable.

The LoG filter is invariant to orientation and therefore can have profound application in face detection system and hence has been used in the method proposed in this paper.

The paper presented here is a feature based approach which gives fast and encouraging results. The initial focus is on segmenting the regions of the image which basically classify them as probable face or non face. The method uses the chrominance information pertaining to skin regions as well as the edge information. LoG operator

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754629 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853491

separates the merged faces that are too close and suppresses the sensitivity to noise. Local features are used to refine the probability measure of the face regions. Euler numbers as well as the percentage of skin serve as useful candidates in such cases. The regions thus extracted may not be of correct dimension which is normally found in faces. Thus to enhance accuracy, golden ratio principle is applied to classify the regions.

Rest of the paper is organized as follows: section 2 covers the related work, section 3 describes about the proposed methodology used in this paper, experimental results are illustrated in section 4 and section 5 focuses on conclusions & suggestions for further work.

2 RELATED WORK

Digital Image Processing is an area characterized by the need for extensive experimental work to establish the viability of proposed solutions to a given problem. Applications include processes that allow manipulation of digital images in numerous ways such as segmentation, description of those objects to reduce them to a form suitable for computer processing, and classification of individual objects. Face is an important object that can be extracted from these images for detection and recognition purposes.

Yang *et al* [2] makes the distinction between detection and recognition. Given an arbitrary image, the goal of face detection is to determine whether or not there is a face in the image and, if present, return the image location and extent of each face. Face recognition involves comparing a face that has been located in an image with a database of faces with the intent of finding a match. The greatest challenge facing computer scientists is face detection.

A lot of research is been going on in the field of human face detection but there is still scope for improvement. Though there has been a great deal of progress in this area, there is yet to be a solution with 100 percent accuracy. Some of the most recent methods tried are fast face detection using a cascade of detectors[3], rapid object detection using a boosted cascade of simple features, using antifaces, statistical learning of multi-view face detection and computationally efficient face detection. Haar-like features was used in fast face detection and rapid object detection[4], anti-faces employs the template matching techniques, while, the use of statistical learning algorithms for object localization is followed in multi-view face detection. In [5] an approach to automatic detection of human faces in color images is proposed. This has human skin segmentation to identify probable regions corresponding to human faces; adaptive shape analysis to separate isolated human faces from initial segmentation results; and a view-based face detection to further identify the location of each human face.

The human skin segmentation employs a model-based approach to represent and differentiate the background colors and skin colors. The adaptive shape analysis uses a series of morphological operations based on the prior shape knowledge of upright human faces. The view-based face detection is constructed based on principal component analysis and neural network classification.

In [6], a face is defined as a skin region and a lips region that meet certain geometrical criteria. The face detection system has skin detection, lips detection, and face verification modules. The Multi-Layer Perceptron (MLP) neural networks were used for the skin and lips detection modules.

In [7], a face detection algorithm for color images in the presence of varying lighting conditions as well as complex backgrounds is addressed. Based on a novel lighting compensation technique and a nonlinear color transformation, the method detects skin regions over the entire image and then generates face candidates based on the spatial arrangement of these skin patches. The algorithm constructs eye, mouth, and boundary maps for verifying each face candidate.

In [8, 9] color-based approach to segment human skin pixels from background, using a 2D histogram-based approach as a preprocess stage for human face detection is

discussed. Derivation of skin classifier rules from the lookup tables are based on how often each attribute value (interval) occurs, and their associated certainty values.

3 PROPOSED METHOD

The method uses color space of skin and edge information which is simple, easy to implement, gives good results and quickly locates faces in a given image. This face detection system consists of several steps. First, the image is binarized into skin and non-skin pixels based on color information of skin pixels by making use of the fact that different human skin colors from different races have been found to fall in a compact region in color spaces. However, there are limitations to using just the skin color. When faces are too close, the skin regions can get merged. Thus, connectivity of skin and edge information is used to segregate faces that are close, into separate regions and also to separate background from faces. Once the different skin regions are obtained, the last step is to decide whether each of the regions identified is a face or not. This is done using various parameters. They are the height to width ratio of the skin region, number of holes in each region and the percentage of each skin region in the skin-detected image. A detailed step-by-step explanation of the method used is discussed in the following sections.

3.1 Skin Pixel Classification

The first step of this face detection system is to classify each pixel in the given image as a skin pixel or a non-skin pixel. Experimental analysis on images of human faces show that the Cb and Cr component value of skin pixels fall between the range of 90-130 and 137-177 respectively. Using this information, the pixels in the YCbCr image is classified as skin or non skin. A skin detected binary image is obtained, where only the skin pixels are shown. A general RGB image (Fig 3.1.a), its equivalent in YCbCr space (Fig 3.1.b) and its skin detected image (Fig 3.1.c) are shown.

3.2 Edge Detection

The skin pixels should be categorized into different groups so that they will represent something meaningful as a group, for example a face, a hand etc. The different skin regions are determined using edge detection and 8-connectivity analysis. Edges are pixels where brightness changes abruptly. The information about edges is used to find region boundaries. There are many filters available to

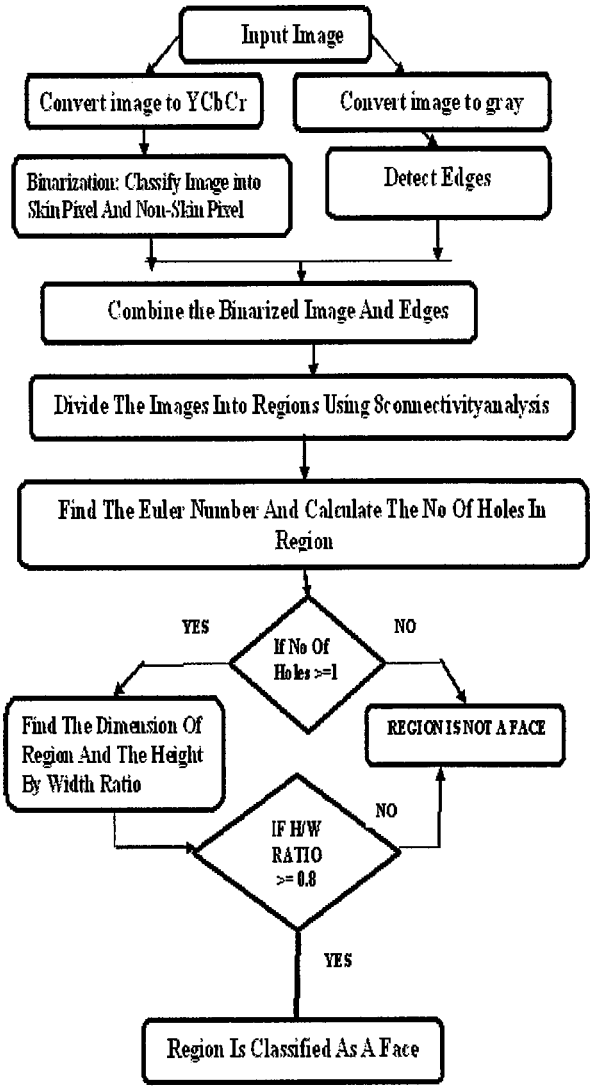


Fig. 1: Block diagram showing various stages

detect edges of an image. Few of them are Sobel, Roberts, Prewitt, Robinson, Zero-Crossing and LoG operators.

LoG operator is used for edge detection here. Laplacian focuses only on edge magnitudes without regard to their orientation, i.e. it has the same properties in all directions and is therefore invariant to rotation in the image. The Laplacian is approximated in digital images by a convolution sum. The Laplacian operator has a disadvantage: it responds doubly to some edges in the image. Due to this disadvantage a smoothing operator (eg: Gaussian operator) is used.

Smoothing aims to suppress noise or other small fluctuations in the image. Gaussian smoothing effectively suppresses the influence of the pixels that are up to a distance 3 sigma from the current pixel. With this the Laplacian operator becomes an efficient and stable measure of changes in the image. The advantage of this approach compared to classical edge operators of small size is that a larger area surrounding the current pixel is taken into account; the influence of more distant points decreases according to the standard deviation of the Gaussian.

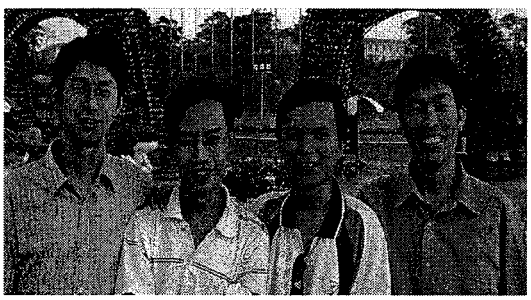


Fig 3.1.a: Original RGB image

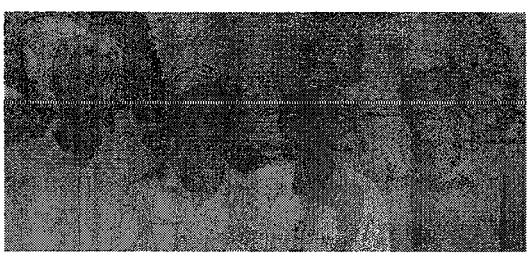


Fig 3.1.b: Equivalent image in YCbCr color space

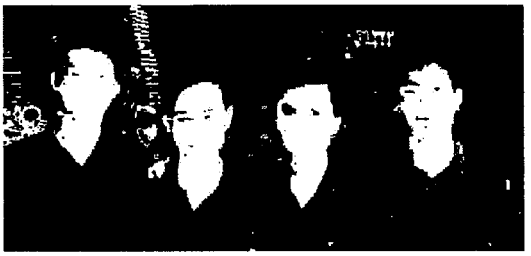


Fig 3.1.c: Skin Detected Image

3. 2.1 Laplacian of Gaussian Operator

The Laplacian is a 2-D isotropic measure of the 2nd spatial derivative of an image. The Laplacian of an image highlights regions of rapid intensity change and is therefore often used for edge detection. The Laplacian is often applied to an image that has first been smoothed with something approximating a Gaussian smoothing filter in order to reduce its sensitivity to noise. The operator normally takes a single gray level image as input and produces another gray level image as output.

The Laplacian $L(x, y)$ of an image with pixel intensity values $I(x, y)$ is given by:

$$L(x,y) = \frac{\partial^2 I}{\partial x^2} + \frac{\partial^2 I}{\partial y^2} \quad \text{.....(1)}$$

This can be calculated using a convolution filter. Since the input image is represented as a set of discrete pixels. These kernels are approximating a second derivative measurement on the image but they are very sensitive to noise. To counter this, the image is often Gaussian smoothed before applying the Laplacian filter. This pre-processing step reduces the high frequency noise components prior to the differentiation step. The 2-D LoG function centered on zero and with Gaussian standard deviation has the form:

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

The LoG operator (Fig. 3.2.1.1) calculates the second spatial derivative of an image. This means that in areas where the image has a constant intensity (*i.e.* where the intensity gradient is zero), the LoG response will be zero. In the vicinity of a change in intensity, however, the LoG

response will be positive on the darker side, and negative on the lighter side. This means that at a reasonably sharp edge between two regions of uniform but different intensities, the LoG response will be: zero at a long distance from the edge, positive just to one side of the edge, negative just to the other side of the edge, Zero at some point in between, on the edge itself. By itself, the effect of the filter is to highlight edges in an image.

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

Figure 3.2.1.1 Discrete approximation to LoG function with $\sigma = 1.4$

The RGB of size m*n*3 image is converted to a grayscale image m*n and the log filter is applied to it by the process of 2D convolution. The corresponding matrix obtained is complemented and a binary image with highlighted black edges is obtained as shown in Fig 3.2.1.2

Fig 3.2.1.2: Image with Edges detected using LoG filter

The information present in the skin detected image (Fig 3.1.a) and edge detected image (Fig 3.2.2) are combined into a single image using morphological AND operation. This results in a single binary image with edges and detected skin pixels as seen in Fig 3.2.3.

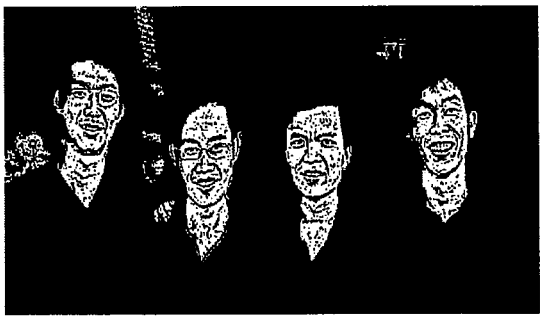


Fig 3.2.3: Skin detected Image with Edges

3.3 Region Separations

The various regions found by combining the edges and the skin pixels are now separated out using the concept of 8-connectivity analysis. The skin pixels in the image are grouped based on a 8-connected neighborhood i.e. if a skin pixel has got another skin pixel **in** any of its 8 neighboring places, then both the pixels belong to the same region. At this stage, regions are separate as probable face or nonface (Fig. 3.3). Now each region is processed individually to detect the regions that are faces. The number of holes in each region is found in order to eliminate regions with no holes, since the eyes and nose in a face manifest themselves as holes.

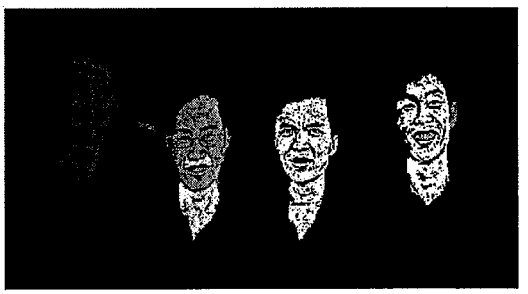


Fig 3.3: Labeled regions

To determine the number of holes inside a region, we compute the Euler number [3] of the region, defined as:

Euler number = Number of connected components - number of holes

$$E = C - H \quad \text{..... (3)}$$

In this paper the number of connected components (i.e. the skin region) is set to one since one skin region at a

time is considered. The number of holes : **H = 1 – E**
Once the system has determined that a skin region has more than one hole inside the region, few characteristics in that particular region mentioned below are used in further refinement.

To eliminate very small regions, which cannot be a face, the percentage of skin occupied by that region in the skin-detected image is: **P = (A1/A2) x 100,** Where, A1 is the area of the region, A2 is the area of the skin detected image. If the percentage of skin is greater than **9 percent** then the dimension of the particular region is found, as it could probably be a face. To determine if it is so, the height by width ratio is calculated.

3.4 Golden Ratio

Average width of the region is calculated by determining the left most coordinate (l) where the region starts and the right most coordinate (r) of the region where it ends. Hence, **Width = l – r.** Similarly, average height of the region is calculated by determining the top most coordinate (u) where the region starts and the bottom most coordinate (d) of the region where it ends.
So, **Height = d – u.**

The height to width ratio of human faces falls within a small range on the real axis, using the parameter along with percentage of skin in a region, the algorithm should be able to reject most of the non-face regions. So if the height to width ratio falls within the range of well-known :

golden ratio(=(1+sqrt(5))/2)+- tolerance ..... (4)

If the percentage of skin is higher than a threshold called percentage threshold which is 9% here, then that region is considered a face region. Finally, a rectangle, with the regions height and width, is drawn on the original image to represent the detected face.

4 EXPERIMENTAL RESULTS

This algorithm has been applied on a large database consisting of a variety of still color images of single and multiple faces. The images are taken under different lighting conditions and with different backgrounds. It works accurately for images with uniform illumination and non-skin color background. The algorithm is implemented in MATLAB 7 on a Pentium Dual core 256MB RAM Windows XP workstation on a total test set of 230 images. The average time taken for detection is less than a second. After testing on a wide variety of people of different races it can be concluded that it works accurately except for the false alarms sometimes being shown in hand regions and neck regions when being excessively exposed.
An ROC graph is plotted to visualize and analyze the working of this face detection classifier. It is a two dimensional graph in which TP rate, true positive rate, is plotted on the Y axis and FP rate, false positive rate, is plotted on the X axis. Given the face detection classifier and an image, there are four possible outcomes. If an image contains a face and is classified a face, it is counted as true positive; and if it is not classified as a face, it is counted as false negative. If an object in the image, which is not a face, is classified as a face, it is counted as false positive; and if it is classified as not a face, it is counted as true negative. Given a set of test images a two by two contingency table is constructed representing the dispositions of the set of images.

Hypothesized class		True class: p	True class: n
	Y	True Positives = 251	False Positives = 27
	N	False Negatives = 11	True Negatives
Column totals:		P = 262	N = 27

$$\text{fp rate} = \frac{FP}{N} \qquad \text{tp rate} = \frac{TP}{P}$$

$$\text{precision} = \frac{TP}{TP+FP} = 90.2\%$$

$$\text{accuracy} = \frac{TP+TN}{P+N} = 86.85\%$$

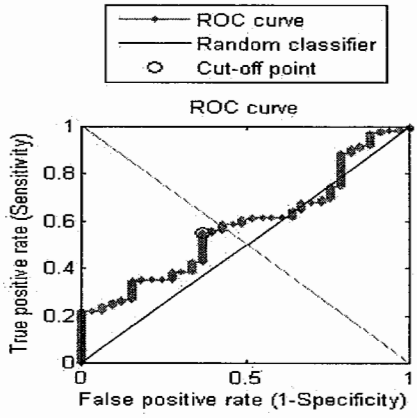
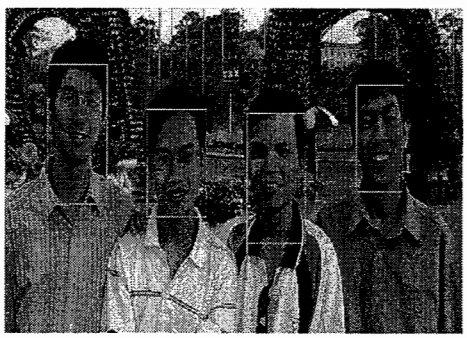


Fig 3.4: ROC Curve

As seen from the Fig. 3.4, the true positive rate is higher than the true negative rate; hence this classifier can be considered as conservative and shows that it gives good results. The calculated accuracy is 86.85% with a precision of 90.2%, which is quite satisfactory.

SNAPSHOTS OF IMAGES WITH FACE DETECTED



5 CONCLUSIONS & FURTHER WORK

The approach proposed in this paper for face detection minimizes computation time while achieving high detection accuracy and can be used in real-time applications. This paper brings together new algorithms, representations, and insights, which are quite generic and may well, have broader application in computer vision and image processing. The algorithm is tested on natural images taken under uncontrolled conditions such as varying illumination, pose and alignment and the performance is very satisfactory. Some kind of post-processing such as extraction of eyes will help to capture the face contour exactly for use by a face recognition system. The challenges to be addressed further involve detection of occluded and overlapping faces, faces with non-frontal views, and images with noisy backgrounds.

REFERENCES

[1] K. Sandeep and A.N. Rajagopalan,"Human Face Detection in Cluttered Color Images using Skin Color and Edge Information", In the Proceedings of International Conference on Computer Vision, Graphics and Image Processing (ICVGIP), Article 161, Dec 2002.

[2] Ming-hsuan Yang , Narendra Ahuja , David Kriegman, "Detecting Faces in Images :A Survey", EEE Transactions on Pattern Analysis and Machine Intelligence, Volume 24, Issue No. 1, pg. 34-58, 1999.

[3] Viola, P. Jones, M., "Rapid object detection using a boosted cascade of simple features", In the proceedings of Computer Vision and Pattern Recognition, 2001 (CVPR 2001), Volume 1, pg. I-511- I-518, 2001.

[4] Rainer Lienhart, Jochen Maydt, "An Extended Set of Haar-Like Features for Rapid Object Detection", In the proceedings of IEEE International Conference on Image Processing (ICIP), pg. 900-903, 2002.

[5] Jean-Christophe Terrillon , Martin David , Higeru Akamatsu, "Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments", In the proceedings of Third International Conference on Automatic Face and Gesture Recognition", pg. 112- 117, 1998.

[6] Jamal Ahmad Dargham, Ali Chekima, "Component-Based Face Detection in Colour Images", In the proceedings of the 10th WSEAS international conference on Mathematical methods, Computational Techniques and Intelligent Systems", Corfu, Greece, pg. 101-105, 2008.

[7] R.L. Hsu, M. Abdel-Mottaleb, A.K. Jain, "Face Detection in Color Images", In the proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 24, no. 5, pg. 696-706, May 2002.

[8] Ihab Zaqout, Roziati Zainuddin, Apian Baba, "Human Face Detection In Color Image", In the proceedings of the Journal of Advances in Complex System, Volume 07, Issue 03, pg. 369-383, 2004.

[9] Face detection in color images Export by: Rein-Lien Hsu, M. Abdel-Mottaleb, A. K. Jain Pattern Analysis and Machine Intelligence, IEEE Transactions on, Vol. 24, No. 5, pg. 696-706, 2002

Feature selection for facial expression recognition using deformation modeling

Ruchir Srivastava[a], Terence Sim[b], Shuicheng Yan[a], Surendra Ranganath[a]

[a]Department of Electrical and Computer Engineering, National University of Singapore, Singapore 117576;

[b]School of Computing, Computing 1, 13 Computing Drive, National University of Singapore, Singapore 117417

ABSTRACT

Works on Facial Expression Recognition (FER) have mostly been done using image based approaches. However, in recent years, researchers have also been trying to explore the use of 3D information for the task of FER. Most of the time, there is a need for having a neutral (expressionless) face of the subject in both the image based and 3D model based approaches. However, this might not be practical in many applications. This paper tries to address this limitations in previous works by proposing a novel technique of feature extraction which does not require any neutral face of the subjects. It has been proposed and validated experimentally that the motion of some landmark points on the face, in exhibiting a particular facial expression, is similar in different persons. Separate classifier is made and relevant feature points are selected for each expression. One vs all SVM classification gives promising results.

Keywords: Facial expression recognition, deformation modeling, feature extraction

1. INTRODUCTION

Approaches to Facial Expression Recognition (FER) can be either image based or 3D model based. For both of these approaches, feature extraction plays a prominent role. Researchers use either geometric features or appearance based features. But there are a few problems that need to be addressed. Many works using geometric features (Eg. Tang and Huang;[1] Srivastava and Roy[2]) require a neutral (expressionless) face of the person. However, having the neutral face of a person is only possible when the person is known and his neutral face has been previously captured. Even if a person is monitored using a video camera, it is difficult to ascertain when the face is actually neutral. This limits the application of these methods. Work by Soyel and Demirel[3] use six facial feature distances without any reference to the neutral face. This method is very prone to interpersonal variations as shown by Srivastava and Roy. [2]

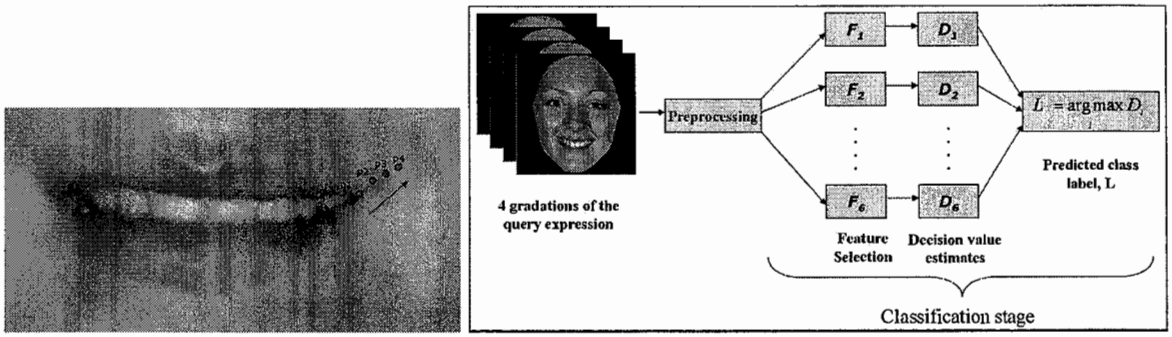
Using appearance based features (E.g. Bartlett et al.[4]), information from every point on the face is utilized. However, information about the expression is mainly present in the feature points, so there is a lot of redundant information in taking whole of the face.

In this paper, a novel feature extraction technique for FER is proposed. Proposed technique is based on motion of the landmark points when expressions are exhibited. This motion is used to model deformation of the face when a person exhibits a certain expression. Deformation modeling has also been utilized by other researchers (5, 6, 7) in order to handle facial variations but each of these works require a neutral face of the test subjects which is impractical as mentioned above.

The main contribution of this technique is that it does not need any neutral face of the test subjects. Other advantages of the proposed approach are that it does not use the redundant information as in appearance based approaches.

Section 2 tells us about the proposed approach. Explanations on the approach along with experimental details are presented in section 3. Finally section 4 concludes the paper, mentioning scope for future work.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462A · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853220



a. b.

Figure 1. a. Movement of right lip corner while exhibiting happy expression b. Classification scheme for the proposed algorithm. F_is represent the relevant feature set for each of the six expressions while D_is represents the decision value estimates of the query expression for each of the six expressions

2. PROPOSED APPROACH

Proposed approach models the deformations in the face caused by expressions. Deformation of a face is given by the movement of various landmark points on the face. To understand how an expression is modeled, consider Fig 1a. where the right lip corner position moves from point p_1 to p_4 from a low gradation of the 'happy' expression to the highest gradation of the same expression. Gradation refers to the intensity of the expression. *Our proposition is that movement of this point will be similar in persons when they exhibit 'happy' expression.* Deformation model is a representation of that similarity. In this work the temporal variation of 3D spatial position of the landmark points is taken as the measure of similarity.

Let $f_p^i(n)$ represent the p^{th} landmark point on the i^{th} person at a temporal instant n. Here $p = 1$ to 83 (corresponding to 83 facial landmark points); $i = 1$ to 100 (corresponding to 100 persons in the database) and $n = 1$ to N_s where N_s is the number of temporal samples found out by interpolation. Pseudocode of the proposed algorithm for feature extraction followed by classification is as follows:

Pseudocode of the proposed algorithm:

1. Inputs:
   - 3D models of the test face, with (x, y, z) coordinates given at 83 landmark points corresponding to four gradations (intensities) of the expression.
   - N training examples X_i^e, $i = 1, 2, 3...N$ and e refers to the expression. Each example has four gradation of the expression.
2. For each X_i^e, find 3D position at each facial point
   - Set nose tip as the origin.
   - Using Cubic spline interpolation find the (x, y, z) coordinates of the point at intensities intermediate in between gradations j and $j + 1$, $j \in [1, 3]$. Thus (x, y, z) coordinates $f_p^i(n)$ are obtained with $n = 1$ to N_s.
   - Transform the coordinates of $f_p^i(n)$ from Cartesian $((x, y, z))$ to spherical coordinate system (r, θ, φ).
   - Find $\vec{\theta}_{p_0}^{i_0}$ and $\vec{\varphi}_{p_0}^{i_0}$ as per equation 1. $\vec{\theta}_{p_0}^{i_0}$ and $\vec{\varphi}_{p_0}^{i_0}$ represent 3D position of the p_0^{th} landmark point of i_0^{th} person.
3. For i^{th} person, the feature vector is given by Eq. 2

Further author information: (Send correspondence to Ruchir Srivastava)
Ruchir Srivastava: E-mail: ruchir@nus.edu.sg, Telephone: 65 6516 6332

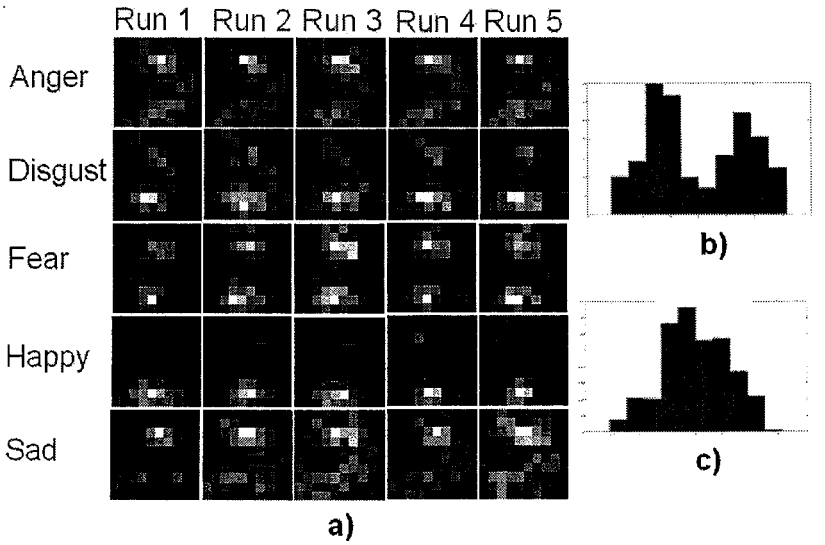


Figure 2. a. Histograms for the distribution of $\theta^i_{p_0}(n)$ and $\varphi^i_{p_0}(n)$ over 5 runs b. Separate histogram for $\theta^i_{p_0}(n)$ c. Separate histogram for $\varphi^i_{p_0}(n)$ corresponding to the 2D histogram in run1 and anger expression

4. Select relevant features for each expression, using the significance ratio test.
5. Classify the test data X_{tes} using one vs. all scheme of SVM.

Details of the algorithm are given in section 3.

3. EXPERIMENTS AND RESULTS

3.1 Interpolation

3D face models from BU3DFE database[8] are selected for 60 subjects in training dataset and 22 subjects in the testing dataset.(x, y, z) position of the 83 landmark points given for all these models is utilized. It should be noted that there are four gradations of expressions available which correspond to only four temporal samples from neutral to peak of the expression. However, only 4 temporal samples are insufficient to have a reasonable estimation of the motion direction. Intermediate samples are found using cubic spline interpolation. After interpolation at a landmark point; say the left cycbrow cornor; wc havc 61 tcmporal samplcs for each of the 60 subjccts uscd for training and 22 subjccts uscd for tcsting.

After interpolation, the spatial positions of the points are transformed from Cartesian to the spherical (r, θ, φ) coordinate system. This is because the parameters θ and φ directly give the required information about the 3D position of a point with respect to the origin by just two parameters θ and φ instead of three parameters, x, y and z. r parameter is not considered here since we are concerned with the direction of motion only. Let the θ and φ values for $f^i_p(n)$ be represented by $\theta^i_p(n)$ and $\varphi^i_p(n)$, respectively.

3.2 Modeling the motion at a landmark point

For the purpose of modeling the motion direction, 3D position of the points $f^i_p(n)$ for a fixed value of p say, p_0, are considered. Out of the 60 subjects in the training dataset, a subset I composing of 30 subjects was selected at random. 2D histograms of $\theta^i_{p_0}(n)$ and $\varphi^i_{p_0}(n)$ values were plotted for $f^i_{p_0}(n)$ where $i \in I$. This procedure was repeated 5 times. Histograms found thus, are shown as images in Fig. 2a. The figure shows that the histograms are consistent over all the 5 runs. This means that the temporal samples at the landmark point for different subjects, belong to a specific nature of distribution. This supports our proposition about the motion of landmark point being similar in different persons. Consistency in the histograms for the same expression and the difference in between the histograms of different expressions, clearly shows that the motion directions are discriminative and can be used as features for FER.

By projecting the 2D histogram on each of the θ and φ axes, the 2D histogram is decomposed into two 1D histograms showing the distribution of $\theta^i_{p_0}(n)$ and $\varphi^i_{p_0}(n)$, separately (Fig. 2 b. and c.). Deformations are modeled as these 1D distributions.

For each person, instead of using all of $\theta^{i_0}_{p_0}(n)$ and $\varphi^{i_0}_{p_0}(n)$ we represent them by $\vec{\theta}^{i_0}_{p_0}$ and $\vec{\varphi}^{i_0}_{p_0}$ which are given by:

$$\begin{aligned} \vec{\theta}^{i_0}_{p_0} &= \left[\theta^{i_0}_{(1)p_0}\theta^{i_0}_{(2)p_0}\theta^{i_0}_{(3)p_0}\theta^{i_0}_{(4)p_0}\right] \\ \vec{\varphi}^{i_0}_{p_0} &= \left[\varphi^{i_0}_{(1)p_0}\varphi^{i_0}_{(2)p_0}\varphi^{i_0}_{(3)p_0}\varphi^{i_0}_{(4)p_0}\right] \end{aligned} \tag{1}$$

where, $\theta^{i_0}_{(1)p_0}, \theta^{i_0}_{(2)p_0}, \theta^{i_0}_{(3)p_0}$ and $\theta^{i_0}_{(4)p_0}$ represent the mean, variance, skewness and kurtosis of $\theta^{i_0}_{p_0}(n)$ with n varying from 1 to N_s. Similar notation is used for φ as well.

For the i_0^{th} person, the feature vector is given by

$$\vec{x}_{i_0} = \left[\begin{array}{cccccccccccc} \vec{\theta}^{i_0}_1 & \vec{\theta}^{i_0}_2 & \dots & \vec{\theta}^{i_0}_p & \dots & \vec{\theta}^{i_0}_{83} & \vec{\varphi}^{i_0}_1 & \vec{\varphi}^{i_0}_2 & \dots & \vec{\varphi}^{i_0}_p & \dots & \vec{\varphi}^{i_0}_{83} \end{array}\right] \tag{2}$$

3.3 Classification

There are total 83 landmark points used for feature extraction. However, for each expression, only a subset of these 83 feature points is relevant. These subsets are chosen using the significance ratio test.[9] Six individual classifiers are implemented using the one vs. all scheme of SVM. Classification scheme is given in Fig.1b. For a query expression, each classifier gives a decision value estimate. Predicted expression corresponds to that expression for which classifier gives the maximum decision value.

3.4 Results

Results of individual classification are given in the form of the Receiver Operating Characteristic (ROC) curves along with areas under the curves in Fig. 3. It can be seen that the AUCs are very close to 1 for all the individual binary classifiers. However, in the case of fear vs. other expressions, AUC is not that good. This shows that there is less discrimination in between the expression fear and the other expressions. Results of the final classification are presented in Fig. 4 in the form of a confusion matrix.

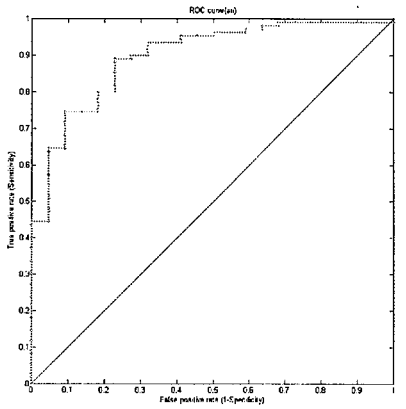
We see from the classification results that highest recognition rate of 100% has been achieved for 'disgust'. The average recognition rate is 87.3% if we ignore the results for the Fear expression. Even when we include results of Fear expression, the average recognition rate is 80.3%. As was evident by the ROC curve, there was less discrimination in between fear and the other expressions. This is also substantiated by the results.

4. CONCLUSION AND FUTURE WORK

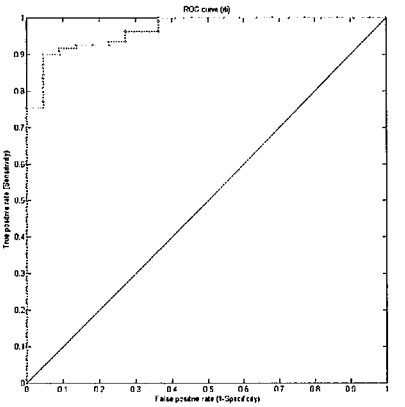
A novel feature extraction technique for facial expression recognition using 3D models was presented. Advantage of this method is that during testing, there is no need to have a 3D model corresponding to the neutral (expressionless) face. This makes the method very relevant in the real life situations where it might not be always possible to have the neutral facial model corresponding to a test expression.

One vs. all scheme of classification was implemented using Support Vector Machines (SVM). For each individual classifier, a feature selection was performed using the significance ratio test. Promising classification results support the proposed feature extraction technique.

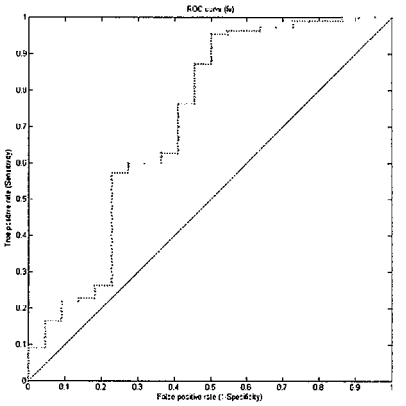
There are few improvements that are possible in the proposed work. Presently, interpolation has been used to get the temporal samples of the motion of the landmark points. However, if we have a 3D temporal data, then there will be an actual information about the motion of the landmark points, instead of a cubic spline approximation. Presently, the results are not very good for the fear expression and a technique needs to be devised to accurately classify fear expression as well. Also, an automatic feature extraction technique needs to be developed in order to extract landmark points on the face.



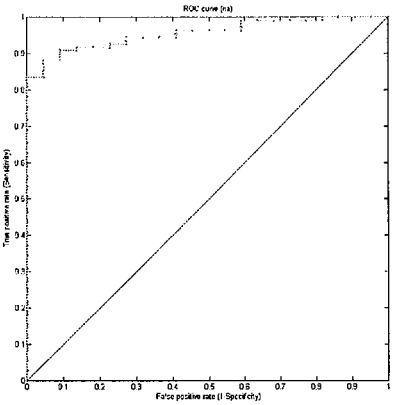
a.Anger
AUC=0.8988



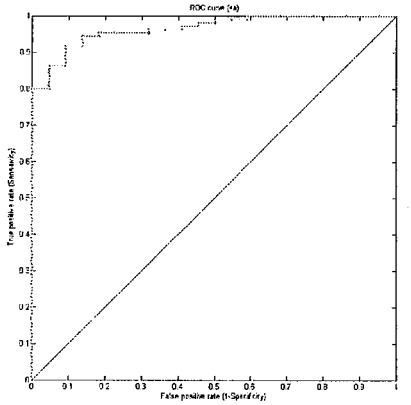
b.Disgust
AUC=0.9678



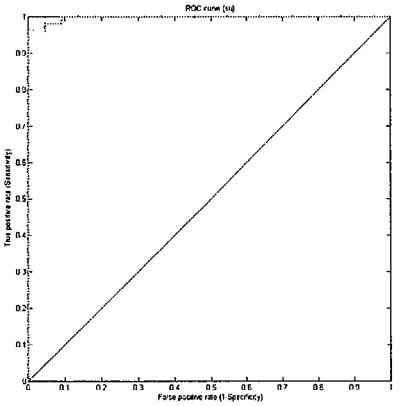
c.Fear
AUC=0.7182



d.Happy
AUC=0.9562



e.Sad
AUC=0.9661



f.Surprise
AUC=0.9975

Figure 3. ROC curves for individual classifiers along with the areas under ROC curves (AUCs)

	AN	DI	FE	HA	SA	SU
AN	**72.7**	9.1	4.5	0	13.6	0
DI	0	**100**	0	0	0	0
FE	4.5	9.1	**45.5**	31.8	9.1	0
HA	0	4.5	0	**95.5**	0	0
SA	9.1	4.5	4.5	0	**77.3**	4.5
SU	0	0	9.1	0	0	**90.9**

Figure 4. Confusion matrices of the classification results. An: Anger, Di: Disgust, Fe: Fear, Ha: Happy, Sa: Sad, Su: Surprise. Training: 60 subjects, Testing: 22 subjects

ACKNOWLEDGMENTS

This work is partially supported by NRF/IDM grant NRF2008IDM-IDM004-029.

REFERENCES

[1] Tang, H. and Huang, T., "3D facial expression recognition based on automatically selected features," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 1–8 (2008).

[2] Srivastava, R. and Roy, S., " 3D Facial Expression Recognition Using Residues.," in *TENCON 2009 (Accepted)*, (2009).

[3] Soyel, H. and Demirel, H., "Facial expression recognition using 3d facial feature distances," in *ICIAR 2007,Lecture Notes in Computer Science*, **4633**, 831–838, Springer Berlin / Heidelberg (2007).

[4] Bartlett, M., Littlewort, G., Frank, M., Lainscsek, C., Fasel, I., and Movellan, J., "Recognizing facial expression: machine learning and application to spontaneous behavior," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **2** (2005).

[5] Blanz, V. and Vetter, T., "Face recognition based on fitting a 3D morphable model," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI* **25**(9), 1063–1074 (2003).

[6] Lu, X. and Jain, A., "Deformation modeling for robust 3D face matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI* **30**(8), 1346–1357 (2008).

[7] Al-Osaimi, F., Bennamoun, M., and Mian, A., "An Expression Deformation Approach to Non-rigid 3D Face Recognition," *International Journal of Computer Vision* **81**(3), 302–316 (2009).

[8] Yin, L., Wei, X., Sun, Y., Wang, J., and Rosato, M. J., "A 3d facial expression database for facial behavior research," in *Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition, IEEE Computer Society*, 211–216 (2006).

[9] Weiss, S. M. and Indurkhya, N., [*Predictive Data Mining: A Practical Guide*], Morgan Kaufmann Publishers (1998).

Uses of software in Digital Image Analysis: a Forensic Report

Mukesh Sharma* and Shailendra Jha[+]
*In-charge, Mobile Forensic Science Unit, Pratapgarh
[+]Assistant Director, Physics Division, State Forensic Science Laboratory, Jaipur -302 016
*corresponding Author: e-mail: mksphy@gmail.com

ABSTRACT

Forensic image analysis is required an expertise to interpret the content of an image or the image itself in legal matters. Major sub-disciplines of forensic image analysis with law enforcement applications include photo-grammetry, photographic comparison, content analysis and image authentication. It has wide applications in forensic science range from documenting crime scenes to enhancing faint or indistinct patterns such as partial fingerprints.

The process of forensic image analysis can involve several different tasks, regardless of the type of image analysis performed. Through this paper authors have tried to explain these tasks, which are described in to three categories: Image Compression, Image Enhancement & Restoration and Measurement Extraction. With the help of examples like signature comparison, counterfeit currency comparison and foot-wear sole impression using the software Canvas and Corel Draw.

Key words: Digital Image Processing, Corel DRAW and Forensic sciences

1. INTRODUCTION

In early days of computing, data was in the form of numerical. Later, textual data became more common after, 1990-2000. Today, many other forms of data: voice, music, speech, images, computer graphics, etc. each of these types of data is reported in terms of signals. It is generally defined as a signal is a function that conveys information. Digital image is a finite collection of discrete samples (pixels) of any observable object. The pixels represent a two- or higher dimensional "view" of the object, each pixel having its own discrete value in a finite range. The pixel values may represent the amount of visible light, infra red light, absortation of x-rays, electrons, or any other measurable value such as ultrasound wave impulses. The image does not need to have any visual sense; it is sufficient that the samples form a two-dimensional spatial structure that may be illustrated as an image. The images may be obtained by a digital camera, scanner, electron microscope, ultrasound stethoscope, or any other optical or non-optical sensor [1]. Digital images have been widely used in our community. Such massive amount digital images have been recently applied in forensic science, such as one can figure out features of

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462B · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856292

suspects or characteristic marks of criminal vehicles.

However, with proper computer software, we can modify or duplicate those image data easily. If those modification or duplication is unauthorized, it will make us doubtful when submitting digital images as evidence in court [2].

During the last decade, inexpensive yet powerful digital computers have become widely available and have been applied to a multitude of tasks. By hitching computers with imaging detectors and displays, very capable systems for creating and analyzing imagery have been constructed and are being applied in many arenas. For example, they now are used to reconstruct x-ray and magnetic resonance images in medicine, to analyze multi-spectral aerial and satellite images for environmental and military uses, to read Universal Product Codes that specify products and prices in retail stores, just to name a few [3-4].

Through this paper authors have investigated the basic principles and introductory applications of digital imaging systems, and include many simple examples to illustrate the concepts using some software like Canvas, Photoshop and Corel Draw [5-6] and how these tools can be used in the Forensic examinations. Most of the images used to illustrate the concepts are rather "crude", consisting of only 4096 individual picture elements (pixels) in a 64 × 64 array. To guarantee the reproducibility, all processes or steps taken by forensic examiners have to be documented or recorded in detail, such as dodging, burning, color balancing, contrast adjustment, unsharp masking, multi-image averaging, integration, or Fourier analysis, and so on, when dealing with digital image evidence to provide integrity protection [7].

2 Methodology:

The general digital image processing system may be divided into three components: the input device (or digitizer), the digital processor, and the output device (image display).

1. The digitizer converts a continuous-tone and spatially continuous brightness distribution f [x, y] to an discrete array (the digital image) $f_q[n, m]$, where n, m, and fq are integers.
2. The digital processor operates on the digital image fq[n, m] to generate a new digital image $g_q[k, l]$, where k, l, and g_q are integers. The output image may be represented in a different coordinate system, hence the use of different indices k and l.
3. The image display converts the digital output image $g_q[k, l]$ back into a continuous tone and spatially continuous image g [x, y] for viewing. It should be noted that some systems may not require a display (e.g., in machine vision and artificial intelligence applications); the output may be a piece of information. For example, a digital imaging system that was designed to answer the question, Is there evidence of a cancerous tumor in this x-ray image? Ideally would have two possible outputs (YES or NO), i.e., a single bit of information.

Used tools in the present case:

In the present paper, three different types of cases using the Canvas software namely Canvas 9 and Corel software using Corel Draw 9 have been reported. These tools are very user friendly and having wide range of applicability

like in the Research Labs, University Lab (Ph.D. scholars) and in the Photo Developing Lab in India. Through this paper the applicability of these tools in the forensic sciences has been reported.

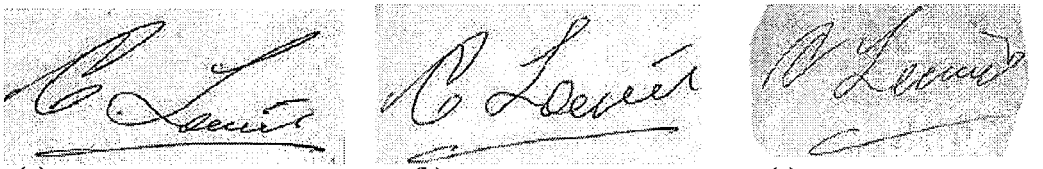
Case I: Signature Examination:

For all handwriting and signature comparison cases, the main hypotheses to explain the observations must be followed:

1. The writer of the specimens wrote the questioned the material.
2. A person other than the writer of the specimens wrote the questioned materials, any similarities to the questioned writing having arisen by chance coincidence.
3. A person other than the writer of the specimens wrote the questioned materials, any similarities to the questioned material having arisen because of the simulation process.

Complications may be further arise are considerations of distortion, disguise or self-simulation of the writing. Through this paper, the applicability of the Canvas, Photoshop software and Corel Draw in Signature examination is explained.

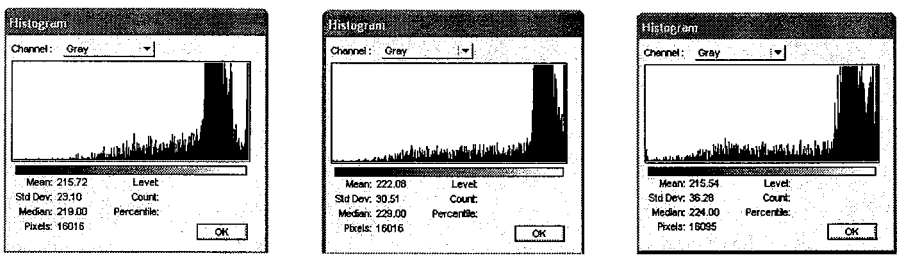
In this case, signature of C. Lewis in three different modes was examined and the results are discussed as below:



(a) **(b)** **(c)**

Fig. 1: (a) genuine signature of "C. Lewis"; (b) slowly written freehand simulation of "C. Lewis"; and (c) quickly written freehand simulation of "C. Lewis"

Comparisons of these signature and remark:

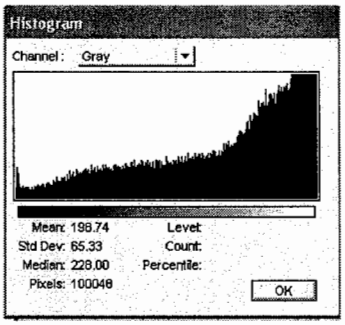
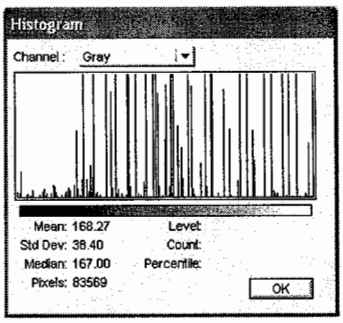


(a) **(b)** **(c)**

After analysis of these histogram (a), (b) and (c), it is clearly seen that the sign are showing similar characters but having some difference in the number of pixels, mean of the histogram and std. deviation point. It means that the speed adopt during the creating a sign is also play a crucial role for examination in terms of forensic point of view.

Examination of counterfeit currency note using Canvas and Corel Draw software:

A comparison has been summarized in the Table 1.

S. No.	Original 500 Rupees Note	Fake 500 Rupees Note
1.	Inverted Image in the Canvas Tool 	Inverted image in the Canvas Tool 
2.	Histogram of the Image on grey scale 	Histogram of the Image on grey scale 
3.	Number of the pixels in the same size examination 100048	Number of the pixels in the same size examination 83589
4.	St. Deviation about 65.33	St. Deviation about 38.40
5.	This banknote contains the Mahatma Gandhi watermark with a light and shade with a light shade effect and multi direction lines in the watermarks window.	This banknote does not contain the Mahatma Gandhi watermark with a light and shade with a light shade effect and multi direction lines in the watermarks window.
6.	The number panels of this banknote are printed in fluorescence ink. The banknotes also have optical fibres. Both can be seen when the banknotes are exposed to UV light.	The number panels of this banknote are not printed in fluorescence ink. The banknotes also have not that much optical fibre.

From the Table 1, one can clearly examine the counterfeit currency note of Indian rupees using these software (like Photoshop 5, Corel DRAW 9 and Canvas 9) and hardware like scanner, digital camera and PC.

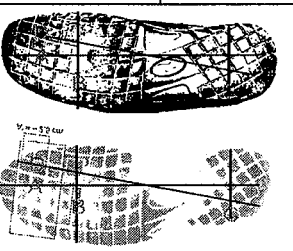
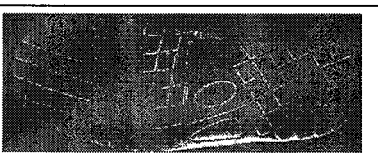
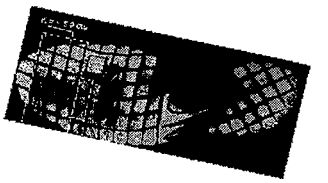
Examination of Shoe prints using Canvas and Corel Draw software:

Many researchers have recently been examining how automated computer systems could help police search and match shoeprint in order to link suspects to crime scenes. Bodziak [8] is one of the leading authorities in the area

of manual human shoeprint classification. He described the process of detection, recovery and examination of footwear impression evidence.

The subject of shoe prints examinations are images of soles left by the perpetrator on surfaces such as floor, carpet, soil, pavement and street. In Table 2 summary of the two shoes print comparison with the help of Corel Draw Software and Canvas 9 software. From Table 2, one can easily conclude that the shoe print belongs to the same shoe.

Table 2: Summary of the shoe comparison

Name of the drawn line on the shoe print using software	Original shoe print on Grey scale in Corel Draw		Shoe print lifted from the scene of occurrence on grey scale on Corel Draw	
	Selection width/length	Angle rotation	Selection width/length	Angle rotation
A	6.97158"	90.728°	6.97158"	90.728°
B	2.93682"	0.0°	2.93682"	0.0°
C	2.10562"	0.0°	2.10562"	0.0°
D	6.54660"	85.023°	6.54660"	85.023°
Photo graph of the shoe print				
Image of the shoeprints using CorelDraw software				
Wear and tear marks	Wear and tear marks on the sole of the shoe exhibit can easily be determined using the software.		Similar positions have been observed for the sole impression under similar size.	
Image used in the Canvas 9				

The performed shoeprint analyses showed that there is a possibility of per forming comparative shoe print analyses even on such difficult material as on floor using software like CorelDraw and Canvas. How ever, it should be emphasized that most of them only a limited group identification is possible, i.e. establishing consistency between the trace and the (comparative) sole in terms of size and shape and dimensions of particular geometrical figures making up the pattern of the investigated sole. Because of the mechanism of formation of these tyres of trace, it is unlikely that individually characteristic features, which are usually small, will be revealed within the traces. For forensic scientists, it is often interesting to establish whether the investigated shoeprint has already been photographed. The results of this study could therefore be exploited in the future to test the ability of the method to find the same or similar shoeprint in the images database and the evaluation would in this case be done using different photographs of shoeprints to test the retrieval in the database.

3 Discussions:

All latent examiners should in common is the need to prepare court displays of latent fingerprint comparisons, counterfeit currency examinations, shoe print and signature comparisons. The traditional process of film based photography and manual drafting techniques can take hours. Using a digital camera or scanner, Corel DRAW, Canvas and Adobe Photoshop, the whole procedure can be accomplished in an hour. Lines and numbers are drawn on screen in any of up to 16.7 million colors. If a line or number is misplaced, it can be erased and redrawn easily without leaving a trace of the original error. Once completed, the chart can be printed in color or black and white. It can also be printed on acetate for projection on an overhead projector. The savings in time and materials can be significant. Any application that requires the combining of images, text and line art can be performed using a digital capture device and a computer.

Reference:

[1] Russ, John C., The Image Processing Handbook, 2nd ed., CRC Press (1995); Also Bernd, J., Digital Image Processing: Concepts, Algorithms, and Scientific Applications, 2nd ed., Springer-Verlag (1993)

[2] Berb E., The Digital Future of Investigations, Law Enforcement Technology, pp. 38 – 40, (1995).

[3] Berg E., Latent Image Processing - A changing technology , The Pacific NW IAI Examiner, pp. 12 – 15,(1994).

[4] Dalrymple, B., Computer Enhancement of Evidence through Background Noise Suppression, Journal of Forensic Sciences, 39(2), 537 – 546 ().

[5] http://www.corel.com

[6] http://www.adobe.com/downloads/; also
Bouton G D and Bouton B, Inside Adobe Photoshop 3, NewRiders Publishing, New York (1995).

[7] Scientific Working Group on Imaging Technology (SWGIT) , Draft Recommendations and Guidelines for the Use of Digital Image Processing in the Criminal Justice System, Version 1.1, (2001).

[8] Bodziak W J, Footwear Impression Evidence Detection, Recovery and Examination
Second CRC Press (2000).

A new image fusion method based on Curvelet transform

Binbin Chu[1] Xiushun Yang Dening Qi Congli Li Wei Lu
New Star Research Institute of Applied Technology, Hefei 230031, Anhui Province, China

ABSTRACT

A new image fusion method based on Multiscale Geometric Analysis (MGA), which uses the improved fusion rules, is put forward in this paper. Firstly, the input low-level-light image and infrared image are decomposed by Curvelet transform, which is realized by Unequally-Spaced Fast Fourier Transforms. Secondly, the decomposed coefficients in different scales and directions are fused by corresponding fusion rules. At last, the fusion image is acquired by recomposing the fused coefficients. The simulation results show that this method performs better than the conventional wavelet method both in the subjective vision aspect and the objective estimation indices.

Keywords: Image Fusion; MGA; Curvelet Transform; Wavelet Transform

1 INTRODUCTION

With the rapid development of image sensor technology, multi-sensor image fusion technology has attracted growing attention. Especially in recent years, development of digital image processing technology, data fusion technology and multi-scale analysis, have made image fusion technology research to become hotter and hotter. Image fusion is a process of combining multiple input images of the same scene into a single fused image, which preserves relevant information and also retains the important features from each of the original images and makes it more suitable for human and machine perception. The goal is to enhance the image quality and information so that it provides more detail information than the information available in single image. Image fusion plays an important role in many areas,such as remote sensing, automatic target recognition, computer vision, intelligent robotics, network security, industrial inspection, etc. Particularly, in the military command areas, multi-sensor image fusion ,as the core content of the battlefield sensing technology, has became influential in modern warfare[1,2]. A typical digital image fusion system structure model is shown in Figure 1.

The wavelet transform has achieved good results in the fields of data compression and denoising. But wavelet theory with the limitations itself, it is not the best representation for an image contained abundance edges and textures. Working for an optimal representation of high-dimensional function is precise the purpose and motivation of the development of multi-scale geometry analysis. Curvelet, as a revolutionary method of Multiscale Geometric Analysis, is more suitable for analyzing the property of 2D curves and straight line like edge than the conventional wavelet method. Also, it is revolutionary in the sense of higher approach accuracy and better sparse representations[3-5]. The theory fundament is described in following section. A new image fusion method based on Curvelet transform, which uses the improved fusion rules, and image fusion evaluation parameters are introudced in section 3. The simulation results and assesemtnt are described in Section 4. It is followed by the conclusion.

[1] Binbin Chu, postgraduate student in New Star Research Institute of Applied Technology, Email: cbbpaper@126.com

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462C · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853126

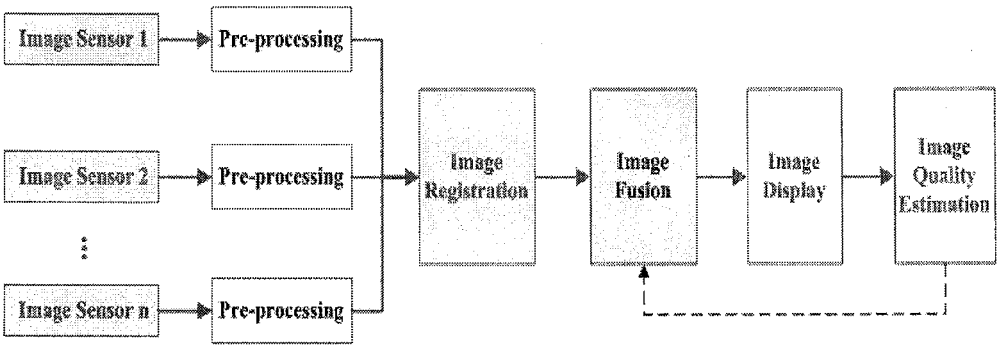


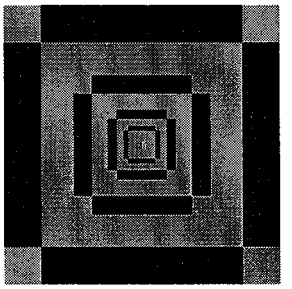
Figure 1. Image fusion system model

2 PRELIMINARIES

The main benefit of curvelets is their capability of representing a curve as a set of superimposed functions of various lengths and widths. The curvelet transform is a multiscale transform, but, unlike the wavelet transform, contains directional elements. Curvelets are based on multiscale ridgelets with a bandpass filtering to separate an image into disjoint scales. The side length of the localizing windows is doubled at every other dyadic subband, hence maintaining the fundamental property of the curvelet transform, which states that elements of length $2^{j/2}$ serve for the analysis and synthesis of the jth subband. The algorithm used to implement the digital curvelet transform (CT) is described in literature [4]. As the digital realization of the first generation Curvelet transform is complex and requires a series of steps such as sub-band decomposition, smooth block, regularization, and Ridgelet analysis. These operations would induce enormous data redundancy. Emmanuel J. Candès and David L. Donoho proposed to get a simpler and easier algorithm of Curvelet transform, called the second-generation Curvelet transform. It has two kinds of implementations, respectively based on unequally-spaced fast fourier transforms (FDCT_USFFT), and the other based on wrapping of specially selected fourier samples (FDCT_WARP). In this paper, the first scheme is selected[3-5].

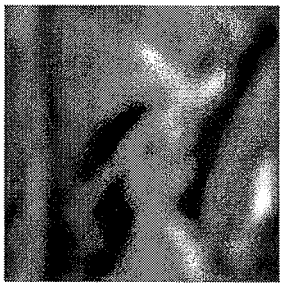
As showned in figure 2, one Lena picture sized 512×512 pixels is decomposed into 6 layers by Curvelet transform. The innermost (first floor), or calld coarse layer, is composed of low-frequency coefficients from the 32 × 32 matrix that contains overview of the image. The most outer layer (sixth layer), or calld fine layer is formed by the high frequency coefficient of 512 × 512 matrix, reflecting the details of the image such as edge features. The middle layer (second to the fifth floor), or calld detail level, that is composed of the high-intermediate-frequency coefficient matrix, reflecting the characteristics of the image information in different directions, whose coefficients on each floor are divided into four broad direction, and each direction is divided into 8, 8, 16, 16 small directions. Images are recomposed by the each single scale coefficients, simultaneity the remaining layers of coefficients are set to zero. They are shown in figure 3.

(a) lena



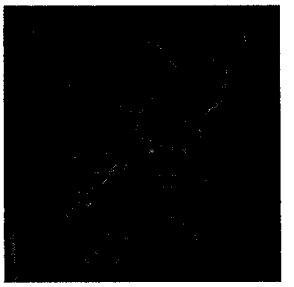
(b) lena decomposed by Curvelet transform

Figure 2. An example of image decomposed by Curvelet



(a)Coarse layer

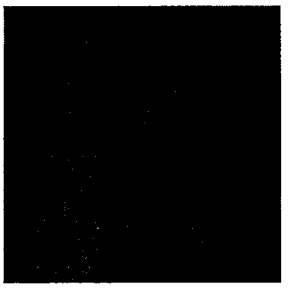
(b)2nd layer



(c) 3rd layer

(d) 4th layer

(e)5th layer



(f) Finelayer

Figure 3. An example of image recomposed by each single scale coefficients

3 IMAGE FUSION BASED ON CURVELET TRANSFORM

Sec. 3.1 Fusion rule

At present, fusion rules in transform domain are classified mainly into two categories: (i) pixel based (eg. transform coefficients direct replacement or additional, maximum selection, the weighted average, etc.) and (ii) region based(such as method based on local variance, based on local energy, based on gradient method, etc.). Pixel based rules generally deal with pixel level information directly. Methods using pixel based rules are affected by blurring effect which directly affect on the contrast of the image. The fundamental idea of region based is that we human begins interpret images at the

region or object level rather than pixel level. Region based rules has many advantages over pixle based like it is less sensitve to noise, better contast, less afected by misregistration. Considering the neighborhood character of selected coefficient, it has a more widly range of applicability[5-7].The follow rules are used in this paper.

(i) fusion rule 1, low-frequency coefficients using the average of the two source images, high-frequency coefficients selecting the value of the maximum modulus

(ii) fusion rule 2, low-frequency coefficients using edge-preserving strategy, high-frequency coefficients selecting the largest region energy

Sec. 3.2 The proposed image fusion method flow chart[3,4,6]

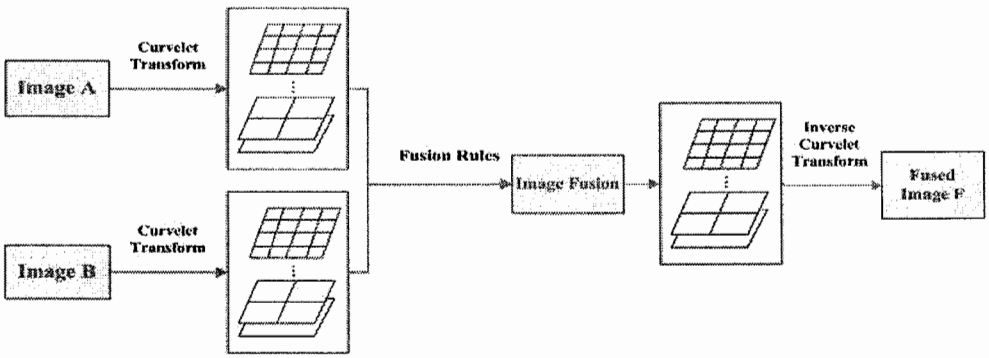


Figure 4. Block diagram of image fusion method based on Curvelet Transform

Sec. 3.3 Fused image quality evaluation[3,6,8]

The evaluation parameters of image fusion include subjective and objective indices. Subjective indices rely on the ability of people's understanding and are hard to come into application. While objective indices can overcome the influence of human vision, mentality and knowledge, and make machines automatically select a superior algorithm to accomplish the mission of image fusion. Any image fusion algorithm can be assessed using subjective and objective indices which may further divided into reference and non reference based fusion quality assessment parameters. Usually there is no standard reference image, so non reference quality evaluation according to the image's statistical properties is of more practical significance. Most widely used non reference base image fusion measurement parameters are average gray, standard deviation, definition and the information entropy. These indices are described detailedly in literature [3]. And we selecte these four indices to evaluate the proposed image fusion method in this paper.

4 SIMULATION AND ASSESSMENT

Sec. 4.1 Experimental conditions and contents

The proposed image fusion scheme has been implemented on Matlab 7.0 using PC with the configuration of the Pentium(R)4 CPU 2.66GHz, 2.67GHz, 512MB RAM, WindowsXP system. These experiments has been tested on low-level-light image fused and infrared image. All set of images results of tested images are shown in the Figure 5. The

tested image of size 512 x 512 shown in the Fig.5 (a) is low-level-light image and other tested image of size 512 x 512 shown in the Fig.5 (b) is infrared image. Experiments has been operated as follows:

(1) image fusion based on wavelet transform(WT), using the Db2 wavelet, decomposed three times, fused by rule 1 and rule 2 respectively

(2) image fusion based on curvelet transform(CT), using the second generation of Curvelet realized by unequally-spaced fast fourier transforms(FDCT_USFFT), decomposed once, fused by rule 1 and rule 2 respectively

Sec. 4.2 Simulation results and assessment

The resultant fused images of all simulations are shown in Fig.5 (c)-(f), and the objective assessment indices of each image are listed in Table 1. Obviously the fused images could provide more detail information than the two source image. The visual quality of fused image generated from this scheme is better than it generated from the wavelet transform based scheme, especially in improving the blocking effect of defects. But in the time-consuming aspect, proposed scheme based on curvelet transform performs not well in comparison with wavelet transform based.

(a) low-level-light image (c) fused image based on WT by rule 1 (e) fused image based on CT by rule 1

(b) infrared image (d) fused image based on WT by rule 2 (f) fused image based on CT by rule 2

Figure 5. Image fusion experiment

5 CONCLUSION

The image fusion method based on the second generation Curvelet transform, which is realized by unequally-spaced fast fourier transforms, has been implemented in this paper. The results of fusion experiments has shown that this scheme performs well, and the objective evaluation indices of fused image achieved higher precision. It significantly improved

fusion performance in comparison to the conventional wavelet based method, which overcome the wavelet based method of introducing false information and the presence of blocking effect of defects. In terms of the algorithm complexity and fusion time-consuming, method in this paper is still to be further improved. This method is suitable for high precision image occasions, and may provides direct reference for other similar multi-source image fusion.

Table 1 The objective estimation parameters of Images in experiment

Evaluation indicators	Fig.5(a)	Fig.5(b)	Fig.5(c)	Fig.5(d)	Fig.5(e)	Fig.5(f)
Average gray	155.578	41.026	98.200	98.204	97.635	97.687
Standard deviation	50.338	64.367	42.468	42.550	49.107	49.321
Definition	4.416	6.648	8.003	8.330	7.869	8.078
Entropy	5.271	4.074	5.009	5.005	5.052	5.050
Time-consuming(seconds)			3.1492	6.9207	88.2816	100.1093

REFERENCES

[1] C.Y. Wen and J.K. Chen, "Multi-Resolution Image Fusion Technique and Its Application to Forensic Science," Forensic Science International, 140, 217-232(2004).

[2] G.Piella,[A general framework for multiresolution image fusion:from pixels to regions], CWI, Amsterdam, Research Report PNA-R0211(2002).

[3] JIAO Li-Cheng and HOU Biao,[Image Multiscale Geometric Analysis: Theory and Applications], Xidian University Publishers, Xi'an,101-156(2008).

[4] D.L Donoho and M.R.Duncan, [Digital Curvelet Transforms: strategy, implementation and experiments], Stanford Univ., Stanford, CA, Technical Report(1999).

[5] Candès E J and Donoho David L, [DCTvUSFFT : Digital Curvelet Transforms via unequispaced fast fourier transforms], California Institute of Technology, Technical Report(2004).

[6] Yingzi Song, "Research on the Theory of the Second Generation Curvelet Transform and Its Applications in the Image Fusion," HoHai University, Degree Thesis(2007).

[7] WU Yi-Quan, LI Xiao-Yan and CHEN Sa, "Image fusion method based on contourlet-domain ICA and SVM," Journal of Optoelectronics. Laser, 2009(6), 839- 842(2009).

[8] Li Huihui, Guo Lei and Liu Hang, "Research on Image Fusion Based on Second Generation Curvelet Transform," available at http://www.paper.edu.cn

Pattern Recognition based on Multi-Agent

CHENG Xian-Yi[1,2], ZHU Qian[2] and WANG Lili[2]

1. College of Computer Science, Nantong University, Nantong Jiangsu 226019, China

2. College of Computer Science and Communication Engineering, Jiangsu University, Zhjiang Jiangsu 212013, China

ABSTRACT

Traditional method of pattern recognition confuse difference tow procession of pattern memory (microcosmic layer) and pattern classifying (macroscopic layer), it is main cause which the pattern methods are difficulty applied to solve really problem. A new frame of APRF (Agent-based Pattern Recognition Frame) is proposed based on Agent theory. The study goal of APRF are cognizing pattern from integral view, erecting the bridge between microcosmic layer and macroscopic layer and uncovering perplex of pattern emerge.

Key words: pattern recognition; Agent; APRF; pattern memory; pattern emerges.

1. INTRODUCTION

Pattern recognition is a basic intelligence of human being. People often perform many "pattern recognition" in their daily lives. For example, many of us can recognize faces we have not seen for many years, even in disguise, and recognize voices over a poor telephone line. In recent fifty years, the new methods and technology of pattern recognition emerge in endlessly, such as statistical pattern recognition, syntactic pattern recognition, neural network pattern recognition, synergetic pattern recognition, bionic pattern recognition, sub-space pattern recognition, fuzzy pattern recognition and multi-classifier combined pattern recognition, etc [1]. These methods have been applied in a variety of disciplines like industry, agriculture, national defense, scientific research, medical treatment, meteorology, astronomy area etc. successfully.

But at our practical applications, we often encounter some dissimilar recognition situations: multi-angle recognition at same pattern (face recognition, remote sensing pattern recognition); multi-pattern recognition at same pattern (CT image, MR image, PET image of a same organ which medical image recognition always need to deal with); multi-postural recognition at same pattern (expression recognition) and pattern recognition which depending on environment or other subjects (remote synergetic diagnosis, text recognition),etc. All of these pattern recognition methods need to investigate distributed pattern recognition, but hitherto only few research results published at this field. So, it does urgently need a new theory, means and technology to solve the pattern recognition task, which can only be resolved high efficient by collaboration.

This paper puts forward new pattern recognition framework-APRF bases on multi-agent theory. The essential thought of this framework is to divide the pattern recognition into two stages: pattern memorizing (microcosmic level) and pattern emerging (microscopic level); separates the pattern feature into three sections: sample feature, quantitative feature and qualitative feature (see definition 1-3 at section 5). The study goal of APRF are cognizing pattern from integral view, erecting the bridge between microcosmic layer and macroscopic layer and uncovering the perplex of pattern emerge.

In section 2, the function of agent at pattern recognition is presented. Then the new pattern recognition framework APRF is proposed in section 3.The next section 4 and 5 conclude with discussion of the pattern memorize principle and pattern merge theory of APRF respectively. Simulation results and analysis is described in section 6. Concluding remarks are presented in the final section.

2. THE APPLICATION OF MULTI-AGENT SYSTEM IN PATTERN RECOGNITION

Agent computation is integration and development of many technologies like expert system, neural network, object oriented, visible technical and grid computing etc. Each agent has its own control thread, can sense other agent's

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462D · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853020

intention actively. It just likes the imagery "object is a free service, agent serves for money". The main characteristic of agent is as follows:

(1) Each agent only has incomplete information ability about the problem, which need to be solved. Thereby, agent's view is partial and its structure is extremely simple;

(2) No global control;

(3) Data are stored distributed;

(4) Computation is asynchronous and evolutional.

Multi-agent system has the feature of information-distributed representation, local operation with global sharing, actively running, mutually interactive, etc. Its kernel admits accumulative effect of knowledge, achieves system's complicated behavior emerging through interaction among agents [3].

Complex system is usually made up of a large number of components that interact with complicated ways. One classical solution to deal with such system is the microscopic level analysis method. According to this method, we decompose complex system into more and more essential components, and then keep resolving until the primitive level that we can comprehend. This method builds on the assumption that the system is closed or system's structure is fixed in advance. Actually, complex system is an open system, system's structure is created by self-organizing and generates new feature at microscopic level by interacting between various integrants. Consequently, there's a huge gap between microcosmic level and microscopic level. The target of APRF is to erect a bridge for this gap.

3. AGENT-BASED PATTERN RECOGNITION FRAME

In order to perform pattern recognizing, human being need to distinguish the specific feature from all sensed signal acted on us (from bottom to top). When the specific features are selected, we still need to explain these features (from top to bottom). The framework of pattern recognition based on multi-agent system is illustrated in fig.1.

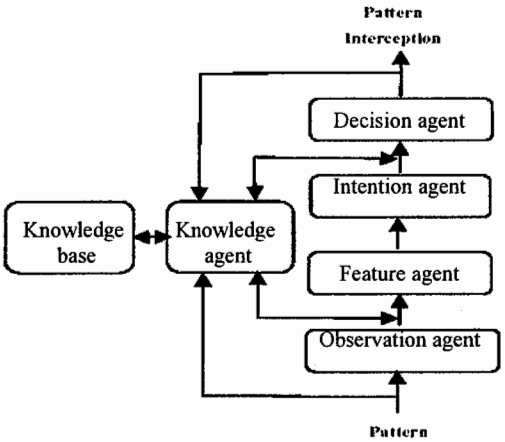


Fig.1 Frame of APRF

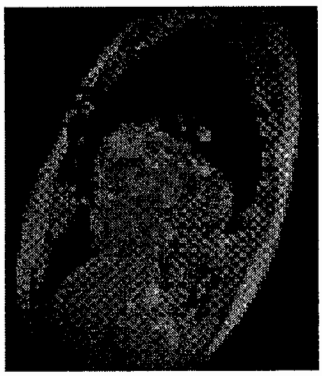


Fig.2 MRI of cardiac

The framework is made up of a series of agents. Each agent serves as different role.

Observing agent's task is the simplest. They only record signal's original state and changing of pattern, just function as an observer. Feature agent searches the features such as line and angel feature and so on. Goal agent is in charge of detecting reaction of concept agent. Each goal agent is responsible to distinguish a specific pattern. For example, one goal agent goes to identify letter "A", another goal agent charge for letter "B", so analogize. Goal agent responsible for "A" makes great efforts to search the feature agent related to specific pattern "A". If goal agent finds a suitable feature agent, its energy will increase. More feature agent has found, more energy the goal agent will has. Finally, intention goal will aware of energy changes of goal agent, and then select the agent with biggest energy as the interpretation of pattern. Intention agent is tantamount to a classifier during the process. The process is from bottom to top, also called data driving. Because at this moment, all activities of systems start with arrived perceptive data, and

continue going forward by all levels analysis in turns, following some kind of logic. To the data driving system, nothing will happen before data inputs the lowest layer. If have had inputs, the process will give the answer quickly and directly. That's why data driving lack of flexibility.

But in fact, pattern recognition is not so simple like this. It is very difficult to explain figure 2 using data driving system when we know nothing about it. It will be helpful and accelerate the explain process if we're been told the figure is a cordis MRI image. When you know what you want to see, you will see it easily. In other words, conceptual knowledge about one object will be helpful to interpret this object. This interpret mechanism is top- bottom or driving by concept. Concept driving starts with the problem what's concept it maybe, then seeks affirmative evidence and makes the processor incline to the anticipate result. Concept driving system is the reverse side of data driving system from some point of view. Mankind's pattern recognition process needs concept driving. Therefore, we try to improve reliability and efficiency of system's analysis and interpretation effectively in APRF, and combine domain knowledge into pattern recognition. We divide knowledge into two categories: one is background knowledge (truth, characteristic, method about world object), equivalent to environment; another is control knowledge (also called primitive knowledge or the knowledge using knowledge). The introduction of Knowledge agent is in order to implement pattern recognition's concept driving. Knowledge agent sends signals to observe agent, feature agent, goal agent and decision-making agent randomly, and help them carry on correct statistics and decisions.

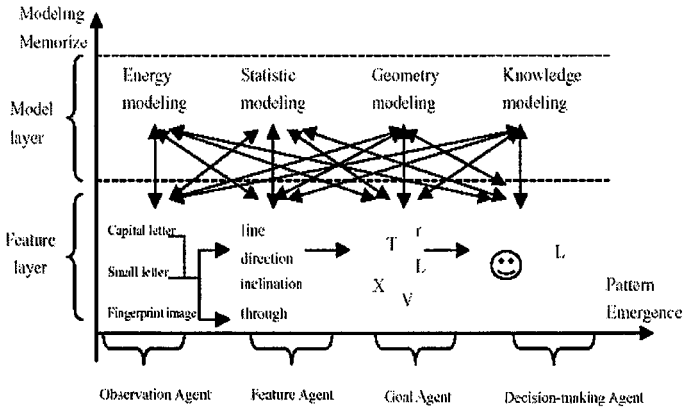


Fig.3 The process of recognition alphabet " L"

Fig.3 show the recognition process of character "L". The relationship between feature and model is equivalent to the relationship between object and class in fig.3. Granularity of feature or included knowledge will begin more and more abundant with the penetration of pattern emerging process. Feature layer impenetrate from left to right gradually till pattern classifies successfully. Model Layer is parallel at every stage, equivalent to knowledge agent.

Knowledge agent mainly includes: (1) energy model related to physical quantity, for example, gray level, chromatic, saturation, texture and so on; (2) statistic model related to statistic, also called algebra model, includes: mean value, variance, information entropy, all steps moment and so on;(3) geometry model related to geometry values such as: distance, angle, shape, curvature, smooth degree, topological relation and so on;(4) background knowledge related to application;(5) method and knowledge related to specific domain etc.

4. PATTERN MEMORIZING

Essence of Pattern memorizing is pattern expression and bottleneck of pattern recognition. All the intelligence system needs memory system, so does the pattern recognition system. Memory system can instruct pattern's evolution direction, store-learning experience[5]:

In fact, a lot of information of pattern conceal in people's efforts for construct memory type. We want to find the model controlled by simple rules to explain pattern, and achieve the long-term memory purpose. As to memory, growing complex is a basic thought. Generated endless novelty restrict by simple rules provide powerful theory foundation for

model constructing and verifying. Through neglect the details and highlight the features occurred frequently, we can establish a rule with following style:

"If current situation demonstrates X, and then carry out action Y".

Of course, to make certain which is important feature and which is detail relate to art of pattern recognition.

The goal of pattern memory is to find the element which can describe the pattern and form all the elements into a memory network. We can consider every point of the memory network as an agent, so the memory network is a multi-agent system or called agent influence map, marked as AIM. Because it's impossible to know the pattern's detail very clearly from the beginning of pattern memory, so we need a pattern memory way which grows in proper sequence.

Because pattern memorizing need feature extraction, and the extracting process maybe come from the pattern directly or indirectly. Feature granularity has difference too. For demonstrate those features of pattern memory, we distinguish simple feature, quantitative feature and qualitative feature from feature.

Definition 1: *in two neighbor layers, the upper layer feature is quantitative feature comprise with the lower layer feature, abbreviated as quantitative feature.*

Definition 2: *in two neighbor layers, the lower layer feature is qualitative feature comprise with the upper layer feature, abbreviated as qualitative feature.*

Definition 3: *the feature extracting directly from the pattern called simple feature.*

According to the definition 3, simple feature is the same with traditional pattern recognition feature. It conforms to people's cognitive process that divided feature into simple feature, quantitative feature and qualitative feature. Quantitative feature and qualitative feature is opposite. Simple feature must be quantitative feature because it's in the bottom of feature layers. The collection consists of the features from same layer, called quantitative feature. For this reason, we build the pattern memory model AIM based on qualitative feature. From the following discussion we can know that, the features of pattern memory such as multi-dimension network structure with multi-layer, multi-semantic, fuzzy, dynamic interconnecting and so on, all of them can be demonstrate maturely at AIM.

The pattern memory process (Fig.3, the Y axis) means people recognize the inputting pattern with an agent (can be method, template, feature, angle and so on) from the beginning. Because lacking knowledge, one agent can't recognize the inputting pattern by itself. It will reproduce or recombine agent according to some rules in order to increase the knowledge about the inputting pattern. As the time goes on, the number of agent will increase until complete the pattern constructing. As far as the numbers of agents are concerned, this process forms above triangular shape, which are only a few at the beginning and increasing gradually by the level.

5. PATTERN EMERGENCE

The process of pattern mergence (pattern classify) is extreme similar to a scene in the television entertainment show: one person (suppose as agent A) back stands in front of a picture, meanwhile, another person (suppose as agent B) faces the picture directly. The play rule is, agent A puts forward some hypothesis, agent B gives answer to let agent A guess what's the picture is as soon as possible. Agent B can't reply include concept related to the picture, agent B had better use "yes" or "no" to answer agent A's question.

Agent B functions as pattern memory, and agent A function as pattern classify. Therefore, after complete pattern constructing, pattern classify just like an expert system.

Chaos is the randomness inside of definite system. Emergence is the certainty inside of random system. Both chaos and emergence are remarkable feature of pattern recognition. In pattern memorizing stage, the main feature is chaos. In pattern emerging stage, the main feature is emergence. Like the research about chaos, people haven't sufficient investigation about emergence yet. Chaos theory still imperfectly and haven't have an accurate definition. However, many scholars carry on their chaos research from their own views. USA scholar Holland have more Systematic thinking in emergence [4]. It defined emergence as "emerge is behavior of whole system which come from interactions between many participators, and it can not be predict from individual behavior knowledge of system element". Through a lot of research at complicated adaptation system, Holland find that, even if the structure, character, behavior rules of system element is quite simple, it's also possible to generate extremely complicated structure, character and behavior. Therefore, he defines the system phenomenon which generates complex entirety from simple part as emergence, in other words, "emerging is big result produced by small reason". It's equivalent to "chemical reaction", and a case which

investment bigger than outcome. This is essential difference between intelligence system and non-intelligence system. Complexity of emergence not only originates from the rules which decide the system's changing, but also closely related to the way which transforms parts into entirety. Emerging is achieved through its components transformation caused by small accumulating experiences. It's just like at a painter's work, the painter not only see the picture's composing points, but also aware of some specific pattern.

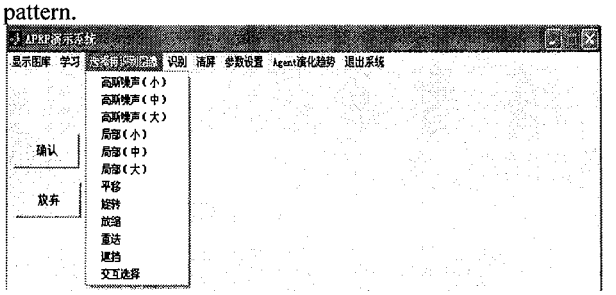


Fig. 4 the main interface of APRF simulation

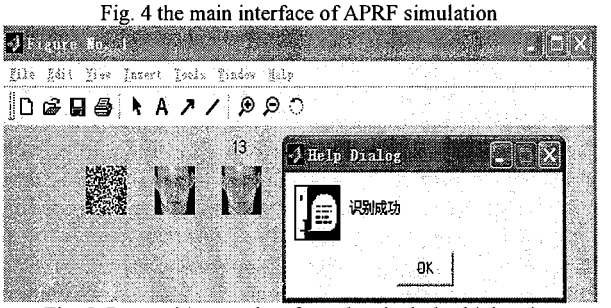


Fig. 5 Recognition results of test data includes high noise

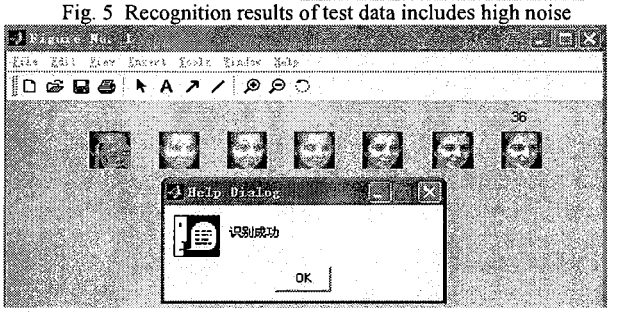


Fig. 6 Recognition results of test data includes superposing noise

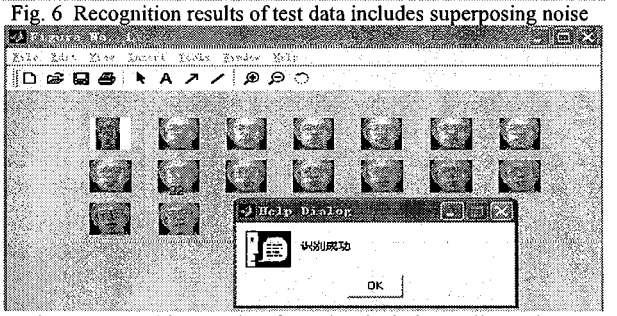


Fig.7 Recognition results of test data includes scaling noise

Emergence behavior occurs without controlled by an executor, it relates to some components interacting mechanism. For instance: when consider an image as a network consists of agents, each agent performs behavior about pattern recognition (pattern observation, feature extraction, object description, decision classify, etc.). Those behaviors compete and collaborate with each other to complete pattern emerging.

Pattern emergence (Fig.3, the X axis) starts with some goal agents produced from constructing process. Then keeping the goal agent with strong survives ability alive and killing the goal agent with weak survives ability through negotiation, collaboration, coordination, competition and interaction. As time goes by, numbers of goal agent decreases gradually till the end of correct classify.

6. SIMULATION RESULTS AND ANALYSIS

According to the frame of APRF in section 4, we perform the face recognition simulation on ORL face base using Matlab, after entering the software system the main interface as in fig.4.

The system mainly consists of modules including: learning, recognition, input model selecting, parameter setting, and Agent evolvement trend. Learning module perform the standard pattern selecting, normalizing, orthogonal, dimensional reduction, prior knowledge selecting and etc; recognition module can perform single and multi- pattern recognition, recognition rate calculating and etc; input pattern selecting module will make certain the defined input mode or interfacing input mode. In Fig.5 is the recognition results if there exists high Gaussian noise, which obtained after two evolving steps. In Fig.6 is the recognition result if there exists superposing noise, which obtained after 6 evolving steps. And in fig.7 is the recognition result if there exists scaling noise, which obtained after 17 evolving steps. Recognition of all figures from Fig.5 to Fig.7 using PCA will fail.

7. CONCLUSIONS

The traditional pattern recognition research takes the simple feature as its basic foundation, but APRF focuses on the activities happened at the whole feature network. Therefore, the APRF research takes qualitative feature as its foundation. Table 1 demonstrates the comparison among five pattern recognition methods.

Table 1. APRF compare with other pattern recognition methods

	Traditional means	Synergetic Means	Bionic Means	Sub-space Means	Agent-based Means
Evolution way	Static	Dynamic	Dynamic	Static	Dynamic
Control mode	Centralized	Centralized	Centralized	Centralized	Distribute
Element	Direct control	Pixel	Quantitative feature	Vector	Qualitative feature
Pattern demonstration	Feature Vector	Grey level Vector	Feature Vector	Sub space	AIM
Means	Bottom-top	Top-bottom	Top-bottom	Bottom-top	Bottom-top and Top-bottom
Principle	Pattern distinguish	Pattern formation	Pattern cognition	Pattern distinguish	Pattern memory and emergence

References

[1]Wang Shou-jue. Bionic Pattern Recognition——A New Model of Pattern Recognition Theory and Its Applications, ACTA Electronic a Sinica.2002,30(10),1417 ~ 1420.

[2] Best J B. Cognitive Psychology [M].Beking: China Light Industrial Publishing House,2000.

[3]Cheng Xian-yi. Agent Computing Heilongjiang Scientific & Technological Publishing House , Publisher, Haebin 2003.

[4]J.H.Holland. Emergence—from Chaos to Order, Shang Hai Scientific & Education Publishing House. Publisher, ShangHai,2001.

[5] Feng Jia-Li, Dong Zhan-Qiu. A Mathematical Model of Perception Schema Generation and Recognition Based on Attributive Integration, Journal of developing of computing and research,1997,34(7),487 ~ 491.

Content-Based Image Retrieval

Yasir Zaheer

National University of Computer and Emerging Science,
Computer Science Department, Pakistan
yasir.zaheer@yahoo.com

ABSTRACT

Large collection of information is being created in many areas of modern life on daily basis. This information exists in many forms from plain text to high resolution multimedia. Today computers are many times faster than human in text based searching using keywords and indexing but the story is totally different in case of multimedia. In context of image retrieval, acquiring storing, sorting and transmitting photos is now trivial, but it is significantly harder to manipulate, index, sort, filter or search through them. The research presents an overview of different techniques used in content-based image retrieval (CBIR) systems and what are some of the proposed ways of querying such searches that are useful when specific keywords for the object are not known. Advances, applications and problems in content-based image retrieval are also discussed. Moreover a system is also developed for content based image retrieval and tested with two database containing 1,000 and 10,000 images respectively.

Content Based Image Retrieval, Image retrieval, Image Descriptors, Semantic Gap, Image Database, Object Recognition, Image features.

1 INTRODUCTION

Multimedia content has become an integral part of the huge repositories of information today exist for example the World Wide Web. Images are the most primitive and widely-used form of media. They contain loads of information as it is rightly said; *'A picture is worth thousand words'*. There is a large volume of unstructured digital media coming from both military and civilian equipments and its volume is getting day by day larger and larger. According to the press release [3] of Photo Bucket there were 34 million unique users using top five photo-sharing sites only in US in 2006. One can imagine with more community sites present nowadays than ever how much this number has increased all over the world. It is nearly impossible to make use of such data unless it is organized. There is a need for effective and additional techniques for image retrieval. Image retrieval incorporates two major research areas i.e. computer visualization and database management (DBMS) [7].

Computer vision is one of the most rapidly emerging technology giving dependable vision competences. It deals with the immense complexity of the visual input and associated degrees of freedom in image understanding. Today we have technology to search the textual information using famous web searches that use text mining approach but image search engines requires human to tag every image in the database. This technique fails for very large databases or images that are automatically generated from many sources of military and civilian equipments e.g. from observatory cameras. The search might not retrieve similar images tagged with synonyms in their explanations.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462E · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853482

Some of the early proposed framework [7] for image retrieval includes the technique of annotating the image first by natural language and then usage of text based database management for image retrieval but with the exponential growth in multimedia information flowing in especially due to World Wide Web it was very difficult to tackle the challenge of manual image annotation. To catch up with the trend Content Based Image Retrieval (CBIR) was proposed. Content-Based Image Retrieval is actually the application of computer vision to the image retrieval problem. This means that the search will analyze the actual content of the image, content might refer to any attribute the image has for example; color, shape, texture or any other information derived from image itself. There exist a lot of application domains in which CBIR is very important [11]. Examples of some areas are; Weather forecasting, Military, GIS systems, Criminal Investigation, Bio-Medical Imaging, Scientific database, Surveillance systems, Remote Sensing (Satellites).

CBIR extracts many features from each image to be used by the retrieval and indexing procedures [2]. Such features delineate the content of the image and that is why they must be suitably selected according to the context. There comes a new space i.e. feature space in which all the visual content of the images are mapped based on the feature extracted from them. The feature space obtained must represent the objects present in the image. The feature space can be made from a number of features extracted from the image but to build a successful retrieval system, one must choose the features which uniquely represent the image. Essential foundations [7] for CBIR are *Extraction of visual characteristics*, *Classification scheme & System design for retrieval.*

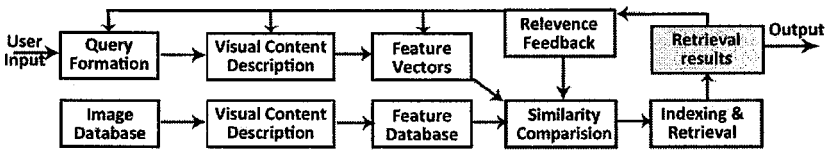


Figure 1: General framework of a content-based image retrieval system [13]

Figure 1 shows a general framework of a CBIR system. First visual content and feature vectors are extracted from query image. These visual contents and feature vectors are then compared with pre-calculated features and contents from the image database. Relevance feedback is provided by the user for the retrieved results that assist in further similarity comparisons.

2 RELATED WORK

The phenomenal growth and the participation of people, research papers in the field of CBIR at beginning of this century is one of the reasons that motivated me to write this paper. It originated in February 1992 [8], at a workshop organized by US National Science Foundation (USNSF) in redwood, California. Since then, it has been used to describe mechanism of retrieving images using its features as desired from huge collections of databases. Since 1997, the number of research publications on the extraction of visual information, its classification, user interaction and management of the database systems has multiplied a lot. Similarly, many military, academic and commercial systems has also been developed by research based, government, private and military organization.

To validate the above theory a simple exercise has been done [1], this exercise was based on a search for publications that includes "Image Retrieval" as a keyword using Google Scholar. The search was made over the most authenticate digital libraries IEEE, ACM and Springer. The publications searched were dated between1995-2005. This exercise gives an overview of the amount of research been done in this area. The result can be seen in Figure 2. This shows that the trend has a nearly exponential growth, from 1998 to 2005, in the interest in image retrieval and its related topics.

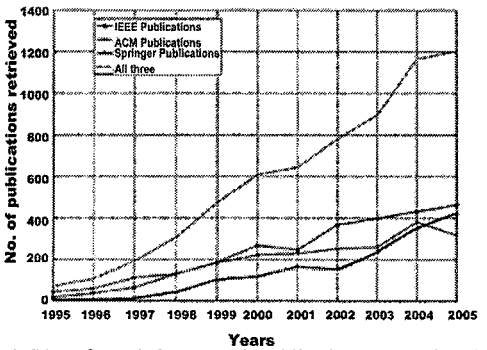


Figure 2: Plot of trends in research publications containing 'image retrieval'[1]

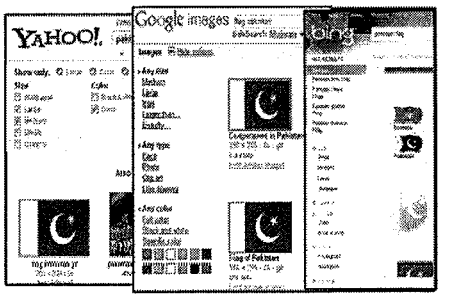


Figure 3: Content-based image retrieval in famous search engines.

This trend is also visible on the World Wide Web. As almost all the renowned search engines are now providing content based image searches one way or other as displayed in Figure 3. Google® for the last two months has been providing searches on the bases of colors or contents like faces, clip-arts, photos and line drawing. Similarly, Yahoo® is giving image search with automatically annotated text images, AltaVista® and Bing® search are also facilitating the end user with such techniques. This shift in image retrieval is not only targeted for the end user but also for internal management of such huge amount of media and to increase the efficiency of retrieval.

3 SCOPE

While using a system based on context based retrieval a user has generally three basic aims in mind [8], a) *Search by relationship* i.e. search by the similarity measure, finding objects of interest starting from an example, more user interaction is involved. b) *Aim the seek* in which user tries to search some object in mind there is no such sketch or image available to be searched these are generally helpful in component based searches, catalogues and art searches. c) *Group Search* that includes retrieval of images that represents a specific class. However these are not the only types of searches available, there are distinctive patterns of use such as, searches of; one unique image, general interactive search, extensive story picture search, searches to exemplify a document and seeking for some values within the picture.

The *Sensory Gap* is the difference between the description of an object in the real world and as it is described in the computational model. Ambiguity due to the sensory gap is present in the viewpoint as well as the in the illumination, disorder and closure. *Semantic gap* is the gap between the results and information retrieved by visual data on the bases of low-level features [8] and the understanding a human user has for the same scene. An example [3] for this is a query image that shows a 'train' and because of this contains a leading arch-like arrangement; such a query can be compared to a 'bridge' or an 'arch' from a firm illustrative point. Any retrieval system can match it with images that contain bridges or some arch like structures. However, people may only compare it with some other images of train.

To align image databases in such a manner that users can search the images using natural language *image-annotation* or *linguistic-indexing* is used [8]. Here the focus is on automatic annotation with a little help from a knowledge expert as manual annotation of such huge amount of repository available is impossible. Computers are very far behind from the human in this respect as they have learned to recognize objects using different viewpoints that can be changed any time. There are two types of approaches to address this problem. *Joint Text-Image representation approach:* Many methodologies [1] to image annotation have been enthused by research in the text mining field. Thoughts from word representation have

been successfully traded in to mutually model textual and visual data. Latent Dirichlet allocation (LDA) model is proposed for portraying relations between words and images. *Monitored classification approach* is to deal image annotation [6] as a monitored classification problem. Concept discovery using monitored categorization, involving plain ideas such as natural sceneries can be attain with high precision.

4 IMAGE PROCESSING: FEATURE EXTRACTION

Content (feature) extraction is the foundation of CBIR systems. There are two classifications of features within the visual feature domain; Color, shape and texture are include in the general feature class while biometric (faces, finger prints etc) and other domain specific features are the other class.

Color Information is the mostly widely used features in CBIR systems [7], as human also distinguish objects on the basis of their color combination. *Color histogram* depicts the probabilities of all the three color channels in a joint manner. The histogram intersection was first proposed by Swain and Ballard as the resemblance evaluation for the color histogram. Pixels in the pictures are made up of points in the 3-D *Colour Space*. Many colour schemes are used to for CBIR, including RGB (Red, Green, and Blue), HSV (Hue, Saturation and Value) and opponent colour space. Any of these can be used for effective image retrieval. *Colour Moments* that are very effective in image retrieval for describing the colour distribution of images are Mean, Variance and Skewness which are first, second and thirst moments respectively by order. These are defined as below (1) represent mean, (2) and (3) represent Variance and Skewness respectively;

$$\mu i = \frac{1}{N} \sum_{j=1}^{N} ij \tag{1}$$

$$\sigma i = \left(\frac{1}{N} \sum_{j=1}^{n} (fij - \mu i)^2\right)^{1/2} \tag{2}$$

$$Si = \left(\frac{1}{N} \sum_{j=1}^{n} (fij - \mu i)^3\right)^{1/3} \tag{3}$$

Here, the value of *i-th* color component is represented by f*ij*. N is the total numbers of pixels in the image. To improve the image retrieval performance generally all these three color moments are being used. *Colour correlogram* technique is used to classify the spatial correlation between colour of pixels and also the pixel distribution. An entry for each colour defines a value in distance between two pixels of an image. *Colour auto- correlogram* is a simplified characteristic of colour correlogram as it only stores the spatial correlation between similar colours. *Texture* is one the most important possessions of images. Some examples are Fourier transformation, shift-invariant principal component analysis, fractal models etc. The autocorrelation function [5] (Eq. 4) of a scene is used to compute the promptness and the stiffness of a texture. This function is defined for an image I as:

$$\rho(x,y) = \frac{\sum_{u=0}^{N} \sum_{v=0}^{N} I(u,v) I(u+x),(v+y)}{\sum_{u=0}^{N} \sum_{v=0}^{N} I^2(u,v)} \tag{4}$$

The texture is a significant feature of human visual awareness, and it is the second significant feature [7] taken out automatically from image segments present in visually all surfaces, including plants, bricks, textiles, hairs etc. Wavelet transformation, Tamura, Gabor transformation filters are one of the generally used features used in many CBIR applications. A texture is distinguished by a set of characteristics called energy, contrast, entropy, homogeneity and contrast. These can be represented by the following set of equations as follows;

$$Energy = \sum_i \sum_j P^2 d\,(i,j) \tag{5}$$

$$Entropy = \sum_i \sum_j Pd(i,j) \log Pd\,(i,j) \tag{6}$$

$$Contrast = \sum_i \sum_j (i,j)^2 Pd\ (i,j) \tag{7}$$

$$Homogeneity = \sum_i \sum_j \frac{Pd(i,j)}{1+|i-j|} \tag{8}$$

Some context-based image retrieval systems use *Shape Information* embedded in an object. These features are described after the pass of color filters and segmentation as they require much complex algorithms to compare. Boundary-based polygonal estimation, limited element models, Fourier-based shape signifiers are some state of the art methods for shape comparisons but they are quite complex. Fourier descriptors [13] and moment invariants are the two successful techniques used nowadays [7].

Edge information is also used in information retrieval. An edge histogram is computed [5] for this purpose; it is used normally in the fields of computer vision, especially in keeping track of moving objects. Local edge distribution [13] in an image can be obtained by computing the *Edge Histogram*, by dividing the image into 16 sub images. For each sub-image there is a separate histogram if every histogram has 5 bins then it means that there are 16 x 5 bins in total. Now the histogram is classified into four guiding edges called vertical, horizontal, 135 degree, 45 degree, and one non-directional edge. Now the edge strength is detected coefficients for the filter are applied. The edge density of each sub-image is calculated using the equation (9). Let, ($x2$, $y2$) and ($x1$, $y1$) are the bottom right and the top left corner of the sub image then the edge density can be written as.

$$f = \frac{1}{ar} \sum_{x=x1}^{x2} \sum_{y=y1}^{y2} e(x,y) \tag{9}$$

A simple but greedy approach *Similarity Comparison and Greedy Method* [13] is used to overcome the issue of increasing dimensions of the characteristic vector. Take some segments I1 (S1, S2, S4), I2 (S2, S5, S8, S7), and I3 (S1), of 3 database images I1, I2, and I3. Likewise, let (S7, S2) are the segments of the query image QI. In figure 4 an algorithm is illustrated using greedy approach comparing the similarity between database image and the query image.

```
Algorithm ImageSimilarity
// I[N] - Image DB with N images
// Q1- Query Image
foreach (Image I in I[N])
   foreach (Segment s in SegmentSet)
     if(Euclidean(Q1[s], I[s]) < threshold)
        // continue to check other segments
     else
        // no need to check other segments
end.
```

Figure 4: Algorithm for Similarity comparison based on greedy strategy [13]

5 IMAGE RETRIEVAL SYSTEMS

There is some extensive research going on in the field of Content-Based Image Recognition since the early 1990s [7]. Both commercial and research organizations have been active in building image retrieval systems. All-most all such systems support features like Search by drawing/sketch, Stochastic browsing, Navigate by example, Search by keywords/speech, Browse with tailored image categories.

QBIC was the first commercial system that stands for Query based Image Content [10]. Developed by IBM, it was the first commercial CBIR. QBIC has very deep effects on later content-based image retrieval systems. It supports queries based other images as example, sketches constructed by the user and drawings, and selected visuals and patterns. It also provides an improved version of texture features like Tamura texture illustrations. Shape area, eccentricity, circularity and major axis orientation and some moment invariants are the part of QBIC shape feature. QBIC online demo is present

at http://wwwqbic.almaden.ibm.com/. *VIR Image Engine* which is also known as Virage, the system was built by Virage Inc. It is more customizable as user can adjust the weights related to the features. Some abstract data types for image content are also defined in this system like local and global value histograms, graphs over shapes and textures. A technique known as relevance feedback was also provided for user for later efficient image retrieval. Some applications of this technology are AltaVista Photofinder and Illustra's Visual Intelligence. Virage demos are present at http://www.virage.com/cgi-bin/query-e. *RetrievalWare* was developed by Excalibur Technologies. Its basic idea was based on neural nets to retrieve images. It also allow user to associate weights to each feature and also to combine the above said features and querying on example based on some HSV histogram, texture properties and dissimilarity in lines of image. This technique was used in the search powered by Yahoo! and InfoSeek and also in a software names Image Surfer System. RetrievalWare demo page is at http://vrw.excalib.com/cgi-bin/sdk/cst/cst2.bat. MIT Media Lab provided a set of interactive tools for searching and browsing of image contents in a project named as *Photobook*. It comprised of three sub-books out of which texture, shape and face information is retrieved. A new motivation of bringing the user in the retrieval loop was based on a general note that human observation is subjective. This technology has been used in several Police Departments by Visage technology for face identification. Photobook demo page is at http://vismod.www.media.mit.edu/vismod/demos/photobook/index.html

VisualSEEk is a search engine meant for the visual features developed by Columbia University students with its version *WebSEEk* for World Wide Web. Its key features are retrieval of features even from a compressed domain. Texture feature using wavelet transformation and color layout features were also supported in this technology. VisualSEEk allows queries based on both spatial relationships and visual features of image. VisualSEEk demos are at http://www.ctr.columbia.edu/VisualSEEk/. *SYNAPSE* stand for SYNtactic APpearance Search Engine. In SYNAPSE the visual queries are developed on the appearance of the object in a particular image. These queries are defined as the form of intensity plane and characterized using yield of a set of Gaussian filters. A separate index is created on each of the co-ordinates of the constant vector. Online demo of SYNAPSE is at http://cowarie.cs.umass.edu/~demo/Demo.html. *MARS*, developed by University of Illinois, stands for Multimedia Analysis And Retrieval System. The research characteristic of MARS is the addition of DBMS and IR (Information Retrieval), incorporation of classification and retrieval and integration of computer and human. Main focus of MARS was to categorize diverse visual features into significant retrieval architecture. It also contains some attractors: one for each turn of the picture (background attractors) and an object attractor placed at the center of the image. Fourier Descriptors are used to calculate the shape of the boundary of the focused object. Online Demo of MARS is at http://jadzia.ifp.uiuc.edu:8000

6 IMPLEMENTATION OF A PROPOSED DESIGN

The CBIR framework implemented is shown in Figure 5. The image database contains all the images. After preprocessing, images are some filters are applied to the image separately and all of their data is stored in a separate manner.

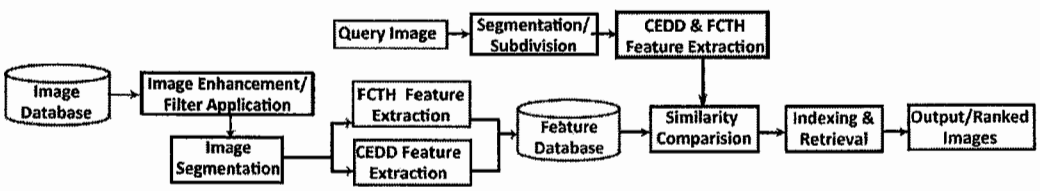


Figure 5: Implemented CBIR model

While populating the database first of all five filters have been applied to the image and then color and edge directivity descriptor [2] and Fuzzy color and Texture Histogram [4] approach is applied to fetch both of these descriptors from the images. The fetched filter is then stored in the database as a separate entity. While matching for results; same steps are applied to the query image to get its descriptors and then best matches are retrieved. As the choice of feature to be extracted from the image change with the scene some filter are applied before feature extraction so that some extreme effects can be negated and a universal approach can be used independent of the context. *Negative* of the image RGB color space is used. This technique is used to identify properly those areas that are faded out due to brightness level. *Log & Maxmin* filters were used to occlude the effect of change in light intensity due to change in view-point. *Emboss laplacian & Prewitt edge detection* filters help in further edge detection in images where edges of object are not clear they can be useful in the texture extraction and in the edge histogram.

6.1 Local feature extraction in the implemented system

To extract colour information (*Color segmentation*) a system is used that produces a fuzzy linking histogram. First the RGB space is converted to the HSV space and then a 10 bin histogram was produced from the three channels of HSV taken as input. Each of the 10 bins illustrate a dominant colour which are; Magenta, Blue, Black, Green, Cyan, White, Gray, Red, Orange and Yellow. The mapping of these 10 predetermined colours is found by the limit calculations, having the foundation on the location of the vertical edged of images that represent the HSV space channels. Figure 6 illustrates the procedure of shaping the location of member values using the vertical edges of channel H. Edges in the form of selected Hue lines [2] have been extrapolated as seen in the image. Using the same procedure we can find the membership limit values for Saturation and Value channels too.

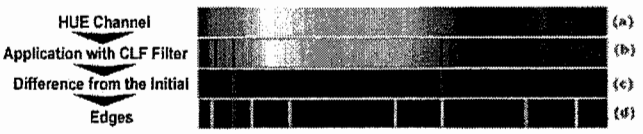


Figure 6: Edge extraction of channel Hue [2]

For detection of the vertical edges coordinate logic filters has been used as they proved to be most suitable for extraction with fine differences. The outcomes of Coordinate logic filter AND function on its 3X3 surroundings were replaced by the pixels in the image. In such operation the edges of image is found. Hue is partitioned into the following fuzzy areas. Orange, Green, Yellow, Cyan, Magenta, Blue, Red to Orange and Blue to Red. Saturation is partitioned into 2 fuzzy regions and Value into 3 areas. First fuzzy area will define the black color and second gives gray in combination with channel S. After this each color is broken down to 3 hues to give a 24 bin histogram as output. Each bin characterizes one of the preset colors. Each color has a light shade, original and a dark shade for example; Yellow is distributed into Light yellow, yellow and Dark Yellow.

Texture Segmentation is done in two different ways in the two implementations. Fuzzy Colour and texture Histogram uses the 'Fuzzy Texture Segmentation' while Colour and Edge Directivity Descriptor uses the 'Edge Histogram Descriptor' approach. *Fuzzy Texture Segmentation* [4]: Energy in high frequency bands of wavelet transform can be characterized by three properties which are very useful in classifying the texture information. These properties are the square root of second order moment of wavelet constant in elevated frequency bands. These properties can be obtained alternatively by Haar transformation applied to the YIQ colour space of the image. The three features are figured out from the LH and HH bands using wavelet coefficients $\{Ckl, Ck,l+1, Ck+1,l, Ck+1,l+1\}$ and from the formula

$$f = \left(\frac{1}{4}\sum_{i=0}^{1} \sum_{j=0}^{1} C^2_{k+i,\, l+j}\right)^{1/2} \tag{10}$$

These features are very useful for the reflection of the texture properties and have been proven effective in over various frequency bands. The reason behind it is the fact that coefficients indicate in different direction in various frequency bands. If an image has high energy in the HL band and low energy in the LH band then it means that the image posses vertical strips and vice versa. *Edge Histogram Descriptor*: It is another way of extracting the texture information. By using this, the algorithm can characterize the edges of object as; horizontal, vertical, diagonal 45-degree, diagonal 135-degree and non directional. Each image is divided into 4 sub-blocks. And then the edge magnitudes are calculated using the gray level of each sub-block and the filter coefficient of the directional edges described above. After calculating edge magnitudes it is normalized by dividing these with the maximum edge magnitude. The resultant edge magnitudes describe the texture information using a six area histogram. A threshold is defined to put the image block in either a Texture block or a Linear Non-Texture Block.

6.2 Fuzzy color and texture histogram & Color and edge directivity descriptor - implementation

In FCTH decisions are taken according to the texture information retrieved by the image and the histogram is made up of 8 regions. Each of these sections has 24 unique sections as an output of the second fuzzy system. There are 192 bins (8X24) as an output. The image is first divided into 1600 blocks to determine the histogram. Once each block goes through the entire fuzzy scheme then we can describe the bin that gives output about the fuzzy system of texture detection. *Haar* wavelet transformation is applied on each block and it is converted to the YIQ color space. The coefficient in LH, HL and HH are calculated and the image block is classified into one of its 8 resultant bins. Then the 10 bin output is obtained using the second fuzzy system that takes the mean values of S and V as input and then calculate the hue of the color and shape of the 24 bin histogram. This process is repeated again and again for all the blocks of image. At the end the histogram is normalized in the space {0/1} with a 3-bit bin quantization.

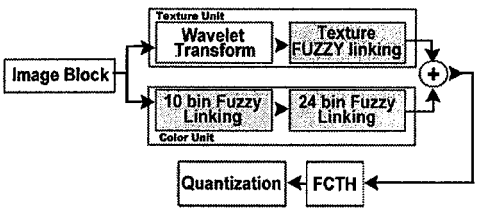


Figure 7: FCTH Flowchart [4]

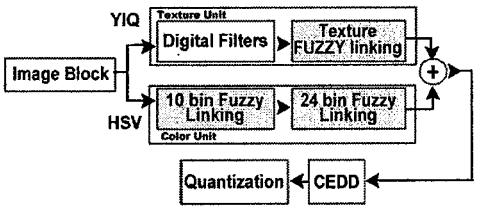


Figure 8: CEDD Flowchart [2]

In *CEDD* decisions are taken according to the texture information retrieved by the image and the histogram is made up of 6 regions. Each of these sections has 24 unique sections as an output while emanating the color information. There are 144 bins (6X24) as an output. The image is first divided into 1600 blocks to determine the histogram. Each block in CEDD is converted to the HSV color space. The coefficient in H, S and V are calculated and the image block is classified into one of its 8 resultant bins. Then the 10 bin output is obtained using the second fuzzy system that takes the mean values of S and V as input and then calculate the hue of the color and shape of the 24 bin histogram. This process is repeated again and again for all the blocks of image. At the end the histogram is normalized in the space {0/1} with a 3-bit bin quantization. Figure 7 and 8 shows the FCTH and CEDD flowchart respectively.

6.3 Experimental setup and results

C# .Net® 2.0 was used to program the application with SQL Server 2000® database running on Dell® Optiplex® with Intel® Pentium® Core 2 Duo workstation and 1 GB RAM. Database consist of 2 image sets of 1000 categorized images of resolution 350X250 pixels & 10000 images with resolution 125X85. While populating feature table this application processes 2-3 images/sec on average for 1000 image set. Data population speed decreases to 4-5 images/sec 10,000

image set. Time taken for image retrieval varies with the increase in image database. It takes 1.8 sec approx. to search through 1,000 images on the above stated configuration and increased to 9-11 seconds approx. for searching in a feature database of 10,000 images. This rate of retrieval is not much good and can be fine tuned using clustering models [12] in which search space gets reduced and high dimensional indexing. Recall and Precision or the Retrieval Efficiency was calculated on two set of images; 1000 natural color images (100 in each category) and 10,000 images widely dispersed using the standard formula.

$$Precision = \frac{No.of\ relevant\ images\ retrieved}{Total\ No.of\ images\ retrieved} \tag{11}$$

$$Recall = \frac{No.of\ relevant\ images\ retrieved}{Total\ No.of relevant\ images\ in\ Database} \tag{12}$$

Figure 9 and 10 shows two histograms populated using precision and recall values obtained by making classified images as queries. A set of queries from different classes of images like, people, animal, locomotive etc were placed in the proposed CBIR system the precision and recall values were calculated on using the standard formulas [5] given above. Quite different results were obtained on the two image sets. These results were not satisfactory for image set containing 10,000 images. This is because of the reason that images in this set were noisy due to compression and have very low resolution. While images in the small set are quite sharp and clear with high resolution.

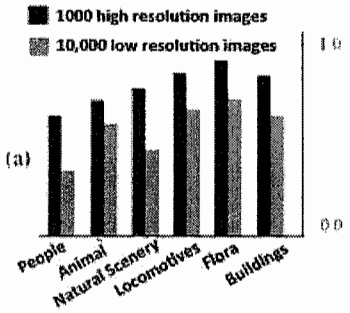


Figure 9: Precision value histogram for retrieved results

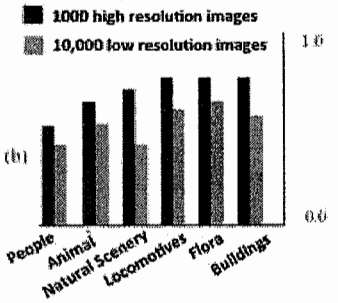


Figure 10: Recall value histogram for retrieved results

Figure 11 shows some of the retrieved results by the application. The resemblance between the retrieved images can be clearly seen. Query Image is on extreme left side along with 6 best retrieved results from top right to bottom left in a row wise fashion

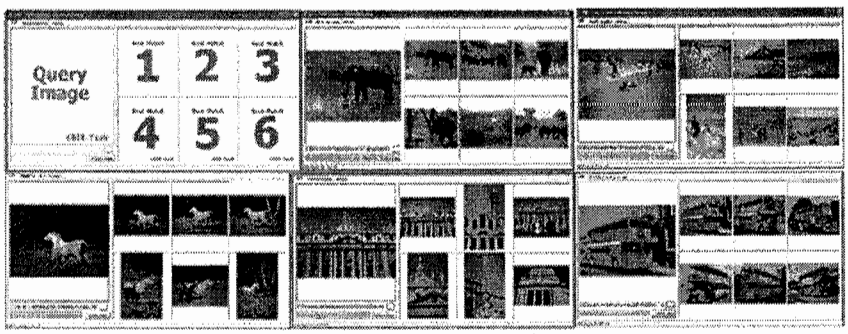


Figure 11: Retrieved Result

7 CONCLUSION AND FUTURE WORK

In this paper we tried to presents an overview of content-based image recognition, its scope, and its fundamentals different techniques used in CBIR systems. The semantic gap between user and machine based retrieval was also discussed and what is image-annotation in a CBIR system. Image describing features on the basis of color, texture, shape and edges were also explained and how are they helpful in such systems. We also provided brief information about some of the well known CBIR system and what are the key features they have, their ways of querying objects and their application in commercial or research fields. An implementation was also provided using two techniques recently known in the field of image retrieval; its results were also analyzed on precision and time basis. A lot much has been done in the field of Content-Based Image Retrieval but this field is still new and active with many open issues. Effective Clustering in image database are required not only to reduce the search space but also to get the association between objects of same classes. Initially CBIR systems were completely automated but now interactive systems are developed to minimize the semantic gap, as Human perceive this visual world in a subjective manner, MIT Media Lab is doing some research in this regard. Integration of High-level Concepts and Low-level visual feature for fast computation is also required this also helps in narrowing down the semantic gap. High-Dimensional indexing is needed as retrieval speed is now becoming a bottle neck due to large image databases.

8 REFERENCES

[1] Datta, R., Joshi, D., Li, J., Wang, J.Z. "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Computing Surveys, Vol. 40, No. 2, Article 5, (April 2008)

[2] Chatzichristofis, S.A., Boutalis, Y.S. "CEDD- Color and Edge Directivity Descriptor- A Compact Descriptor for Image Indexing and Retrieval", ICVS 2008, Springer-Verlag Berlin Heidelberg, (2008)

[3] Vasconcelos, N. "From Pixels to Semantic Spaces: Advances in Content-Based Image Retrieval", Published by the IEEE Computer Society, 0018-9162/07, (2007)

[4] Chatzichristofis, S.A., Boutalis, Y.S. "FCTH- Fuzzy Color And Texture Histogram" - A Low Level Feature For Accurate Image Retrieval, Ninth International Workshop on Image Analysis for Multimedia Interactive Services, IEEE, (2008)

[5] Nandagopalan, S., Adiga, Dr.B.S, Deepak, N. "A Universal Model for Content-Based Image Retrieval", Proceedings Of World Academy Of Science, Engineering And Technology, ISSN 2070-3740, Volume 36, (December 2008)

[6] Yeh, T., Grauman, K., Tollmar, K., Darrell, T. "A Picture is Worth a Thousand Keywords: Image-Based Object Search on a Mobile Platform", ACM CHI 2005, Portland, Oregon, USA. ACM 1-59593-002-7/05/0004, (April 2004)

[7] Rui, Y., Huang, T.S. "Image Retrieval: Current Techniques, Promising Directions, and Open Issues", Journal of Visual Communication and Image Representation, (January 1999)

[8] Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R. "Content-Based Image Retrieval at the End of the Early Years", IEEE Transactions On Pattern Analysis And Machine Intelligence, Vol. 22, No. 12, (December 2000)

[9] Veltkamp, R.C., Tanase, M. "Content-Based Image Retrieval Systems- A Survey", Department of Computing Science, Utrecht University, (October 2002)

[10] Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafher, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.: "Query by Image and Video Content", The QBIC System, IEEE Computer, (1995)

[11] Djeraba, C., Savory, I., Briand, H. "Retrieve of Images by Content", Proceedings of the 1996 IEEE International Joint Symposia on Intelligence and Systems (IJSIS), (1996)

[12] Pengyu, L., Kebin, J., Zhuoyi, L. "An Effective and Fast Retrieval Algorithm for Content-based Image Retrieval", Congress on Image and Signal Processing, (2008)

[13] Long, Dr.F., Zhang, Dr.H., Feng, D.D.: "Fundamentals of Content Based Image Retrieval", (2001)

Automatic Annotation of Image and Video using Semantics

A R Yasaswy[1], K Manikanta[2], P Sri Vamshi[2], and Shashikala Tapaswi[3]

Atal Bihari Vajpayee - Indian Institute of Information Technology and Management, Morena Link Road, Gwalior, India, 474010.

ABSTRACT

The accumulation of large collections of digital images has created the need for efficient and intelligent schemes for content-based image retrieval. Our goal is to organize the contents semantically, according to meaningful categories. Automatic annotation is the process of automatically assigning descriptions to an image or video that describes the contents of the image or video. In this paper, we examine the problem of automatic captioning of multimedia containing round and square objects. On a given set of images and videos we were able to recognize round and square objects in the images with accuracy up to 80% and videos with accuracy up to 70%.

Automatic image annotation, image retrieval, image segmentation, object recognition, image database.

1. INTRODUCTION

With an exponential growth in personal digital image libraries, the need of image understanding continues to be one of the most exciting and fastest-growing research areas in the field of computer vision. Labeling the semantic content of image or videos with a set of keywords is a problem known as image or video annotation [3] [4]. Annotation is required primarily for image or video database management, especially for image or video retrieval. Annotated images or videos can usually be found using keyword-based search, while those which are not annotated can be extremely difficult to find in large databases. Earlier efforts in image or video search has been mainly focused on content-based image retrieval (CBIR), which retrieves images by comparing and analyzing low-level image information[1][2]. Currently, most of the image or video database systems employ manual annotation [5]. That is, users enter some descriptive keywords when the images or videos are loaded into the database or when the user is browsing the images or videos. Although manual annotation of image or video content is considered to be highly accurate, since keywords are selected based on human determination of the semantic content of image or videos, it is a very laborious and tedious process consuming large amount of time. In addition, manual annotation may also introduce retrieval errors due to users forgetting what descriptors they used when annotating their image or videos after a lengthy period of time [8]. Because it requires time and effort to annotate photographs manually even with improved interfaces, automatic image or video annotation techniques may be desirable as explained in [6] [7]. In this paper, we propose an automatic annotation strategy for semantically annotating images or videos that optimizes the efficiency of automatic annotation. The strategy is to create and refine annotations by encouraging the user to provide feedback while examining retrieval results. When the user provides some feedback about the retrieved image or videos or videos, indicating which image or videos are relevant or irrelevant to the query keywords, the system automatically updates the association between the keywords and each image or video based on this feedback. The annotation process is accomplished behind the scenes except for relevant/not relevant gestures required by the user. As the process of retrieval and feedback repeats, more and more images will be annotated through a propagation process and the annotation will become more and more accurate [8]. The result is a set of keywords associated with each individual image or video in the database. As we show in our experiments, the strategy is practical and fairly easy to use, although we are still iterating the design of the user interface. We describe the strategy in detail in the following sections.

[1]{yasaswy_iiitm@yahoo.co.in}
[2]{(manikanta.ks, vamshisri) @students.iiitm.ac.in }
[3]{stapaswi@hotmail.com}

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462F · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853434

2. OUR APPROACH

The following is the strategy we implemented on images. This can also be implemented on videos by extracting significant frames from a video sequence.

2.1 Segmentation

The aim of segmentation is to partition images into recognizable regions, each pixel of which belongs to one object in the real world. In order to identify the shape of an object first it has to be segmented into regions. Firstly we converted the image into a binary image (black and white).This can be done by specifying a threshold value to a pixel .The pixels with value less than this threshold are assigned zero while those above this threshold are assigned the value one. This helps in differentiating darker and lighter regions. We have used image processing toolbox in MATLAB[4] for this purpose.

2.2 Property Extraction

In our work we have used MATLAB to get the properties of a region and then calculate the parametric values of perimeter, area which are useful in determining the shape of the region. Each object will be described by its respective properties. The properties represent, rather roughly, major visual properties (e.g. Area, Perimeter, Eccentricity etc). MATLAB is capable of producing a large amount of low-level data about the features and objects within an image e.g., area, location, pixel color, centroid, eccentricity, orientation etc. These extracted values for each region of an image are stored in MySQL database.

2.3 Object Recognition

After the extraction of visual properties of the regions in the image we proceed to recognize the objects i.e. relating the regions to the objects in the real world. In our work we have used the regional properties like area, perimeter, eccentricity, orientation, major axis length and minor axis length of the regions to determine round and square objects. In case of round objects, their area and perimeter are related as 4 x pi x Area = (Perimeter) ^2. And in case of square or rhombic objects, their area and perimeter can be related as 16 x Area = (Perimeter) ^2. In both the object types the ratio of their major axis length to their minor axis length is unity. Taking these parametric relations as a base we can recognize round and square objects.

2.4 Storing In Database

Now, after recognizing the objects in the image we have to store the object names (keywords) in a database file. So when object recognition is complete, an instance of the image with the objects recognized is included in the database file .The instance specifies the type and number of objects present in the image. This is helpful in better ranking and successful retrieval of the image. The storage can be in XML format since most web based queries are preferred in this format.

3. EXPERIMENTAL RESULTS

For experimental work, a sample database consisting of 100 images and 10 videos related to various domains such as sports, space, daily usage items etc have been considered. We implemented our automatic annotation strategy to recognize the objects and annotate the images and videos using MATLAB.

We have used the relations between parameters, which were extracted for every region, as metrics to determine if the region is circular or square shaped. The parameters we have used are the area, perimeter, major axis length and minor axis length. The metrics which we used are as follows in Table1.

[4] http://www.mathworks.com/products/matlab/

	Parametric relations
Metric1	4 x pi x Area = $(\text{Perimeter})^2$
Metric2	16 x Area = $(\text{Perimeter})^2$
Metric3	Major axis length / Minor axis length

Table 1.Metrics

In case of a perfect circle the Metric1 and Metric3 of the region are equal to unity. So we determined the regions with their Metric1 and Metric3 falling between 0.95-1.05 (with an error approximation of 1- 0.05 to 1+ 0.05) as round shaped. In case of a perfect square or rhombus the Metric2 and Metric3 of the region are equal to unity. So we determined the regions with their Metric2 and Metric3 falling between 0.90 - 1.1(i.e. with an error approximation of 1- 0.1 to 1+ 0.1) as square shaped or rhombic.

Some experimental results of the process which we implemented on the images in our database have been shown below in The Figures 1 to 4.

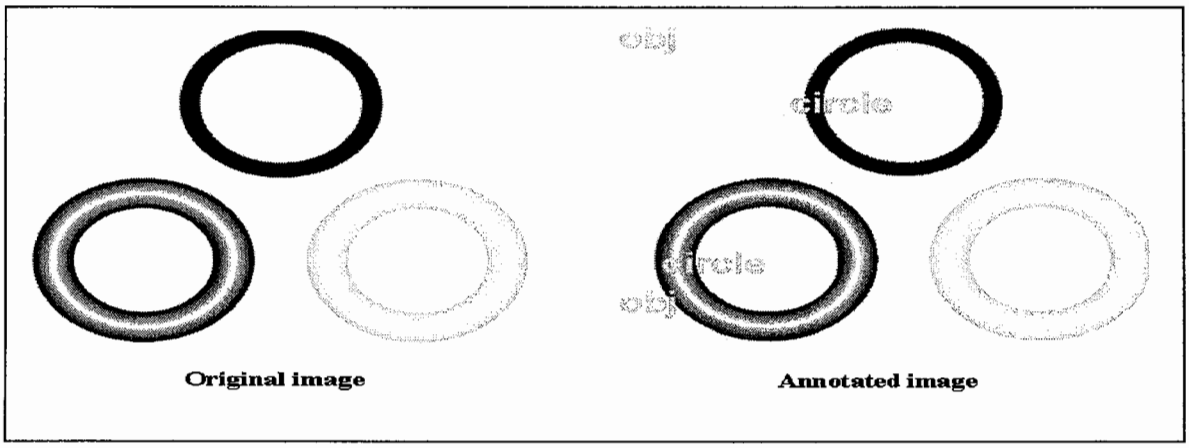


Figure1: Pointing out to the two circular rings as circular objects.

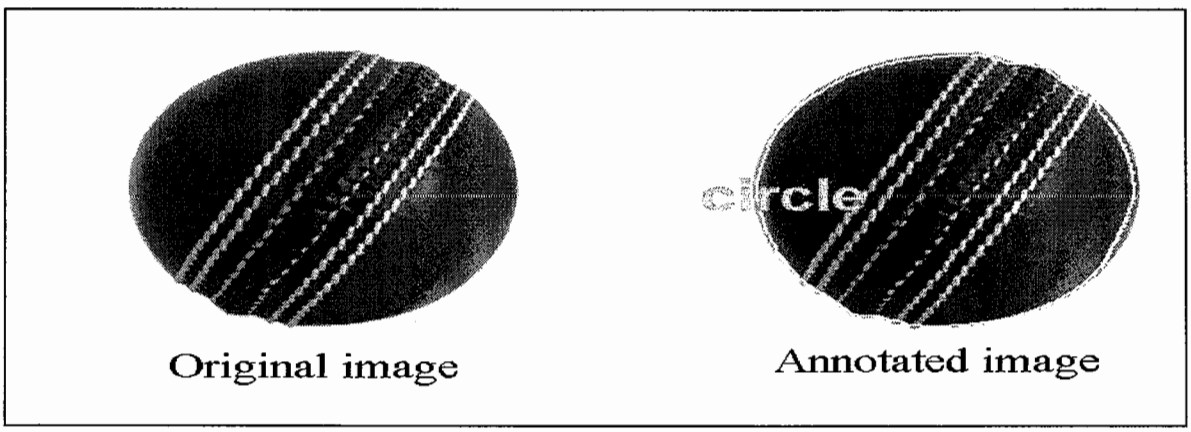


Figure 2: Pointing out to the cricket ball as circular object

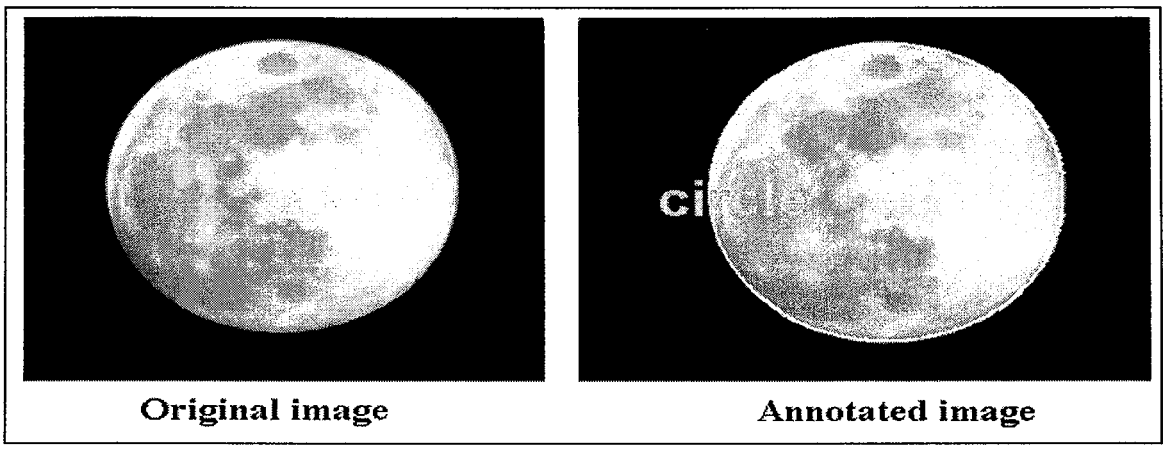


Figure 3: Pointing out to the moon as a circular object.

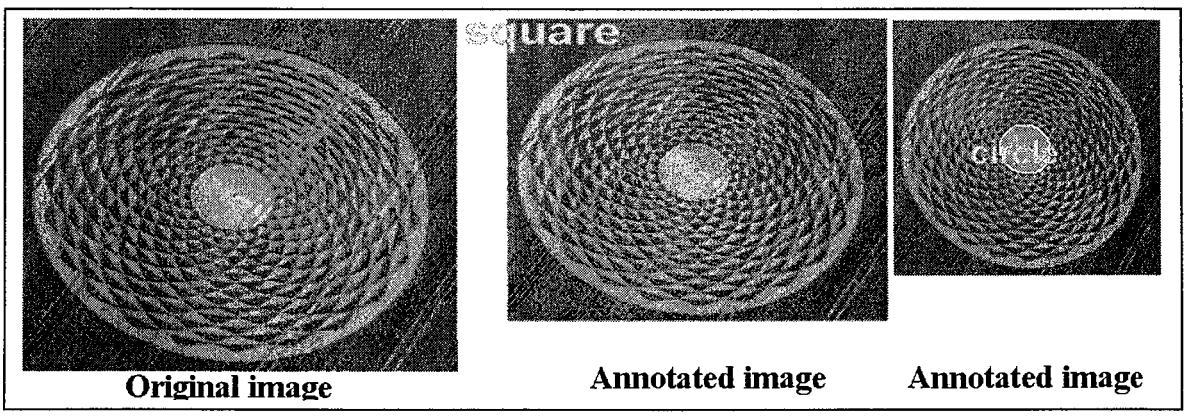


Figure 4: Pointing both circular and square shapes in the image.

This is a snapshot of our database storage of the annotated images.

filename	obj_name
img1	circ
img2	circ
img3	sqr
img4	circsqr
img5	sqr

Table 2.Annotated image table

For every image in the database there is another table describing the regions.

A sample snapshot of the image table is:

region	circ_metric	sqr_metric	regn_name
reg1	**0.76**	**0.66**	**obj**
reg2	**0.96**	**0.77**	**circ**

Table 3.Image table.

The experimental results on Images are tabulated below:

	Total images considered	Actual No of circular objects present	Number of circular objects recognized	Actual no of square objects present	Number of square objects recognized
Images	100	65	53	45	37

Table 4.Results

4. CONCLUSION & FUTURE WORK

In this paper we present an automatic annotation strategy which can be implemented on images and videos to recognize objects which are circular and square shaped. This automatic annotation strategy can be implemented on a very large database of images and videos for regular use of the system. This strategy is also useful in a dynamic database system where new images are continuously imported. Of the 100 images we have considered, we were able to identify objects and hence annotate 79 images accurately. In the remaining images either the objects were not recognized or were wrongly interpreted. And of the 10 videos we considered, we were successful in extracting significant frames and annotating 7 videos correctly. Thus this strategy is very efficient compared to manual annotation. Therefore our strategy avoids tedious manual annotation and hence saves time. The work we have presented in this paper is still preliminary. We plan to extend our work towards annotation of domain specific images with the help of a well developed ontology. We also plan to implement a learning method for initial images using neural networks and then automatically annotate the rest of the images in our database.

5. REFERENCES

[1] Smeulders Arnold W. M., Worring Marcel, Santini Simone, Gupta Amarnath and Jain Ramesh, "Content-based image retrieval at the end of the early years", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, iss. 12, pp. 1349-1380 (2000).

[2] Al-Khatib W, Day Y. F, Ghafoor A, Berra P. B, "Semantic Annotation of Images and Videos for Multimedia Analysis", Proc. 2nd European Semantic Web Conference (ESWC), Herakleion, Greece, vol. 3532, pp. 592-607 (2005).

[3] Dejan Depalov, Thrasyvoulos Pappas, Dongge Li, Bhavan Gandhi, "Perceptually based techniques for semantic image classification and retrieval", Human vision and electronic imaging No.11,Proc. SPIE vol. 6057, pp. 60570Z.1-60570Z.10 (26 ref.) (2006).

[4] Aleksandra Mojsilovic, Jose Gomes, "Semantic Based Categorization, Browsing and Retrieval in Medical Image Databases", Proc. International Conference Image Processing, pp. 145-148 (2002).

[5] Gong Y, Zhang H, Chuan HC and Sakauchi M, "An image database system with content capturing and fast

image indexing abilities", Proc. IEEE International Conference on Multimedia Computing and Systems, pp. 121 - 130 (1994).

[6] Lei Wang, Li Liu, Latifur Khan, "Automatic image annotation and retrieval using subspace clustering algorithm", Proc. 2nd ACM international workshop on Multimedia databases, session: Multimedia data mining, pp. 100 – 108 (2004).

[7] Rong Jin, Joyce Y. Chai, Luo Si, "Effective automatic image annotation via a coherent language model and active learning", Proc. 12th annual ACM international conference on Multimedia, session: Technical session 13: managing images, pp. 892 – 899 (2004).

[8] LiuWenyin, Susan Dumais, Yanfeng Sun, Hong Jiang Zhang, Mary Czerwinski, Brent Field, "Semi-Automatic Image Annotation", In INTERACT2001, 8th IFIP TC.13 Conference on Human-Computer Interaction, pp. 326-333 (2001).

Offline Signature Verification and Skilled Forgery Detection using HMM and Sum Graph Features with ANN and Knowledge based Classifier

[1]Mohit Mehta, [2] Vijay Choudhary, [3] Rupam Das, [4] Ilyas Khan
[1] MTECH (IT), TIT Bhopal, India
[2] Asst.prof, IT TIT, Bhopal, India
[3] Member, Technical Team, Integrated Ideas, Gulbarga, India
[4] Asst.prof, IT TIT, Bhopal, India
E-mail: [1] mohit23.mehta@gmail.com, [2]
vijay_ashish@yahoo.com,[3]rupam@integratedideas.co.in,
[4] mklbpl_2003@yahoo.co.in

ABSTRACT

Signature verification is one of the most widely researched areas in document analysis and signature biometric. Various methodologies have been proposed in this area for accurate signature verification and forgery detection. In this paper we propose a unique two stage model of detecting skilled forgery in the signature by combining two feature types namely Sum graph and HMM model for signature generation and classify them with knowledge based classifier and probability neural network. We proposed a unique technique of using HMM as feature rather than a classifier as being widely proposed by most of the authors in signature recognition. Results show a higher false rejection than false acceptance rate. The system detects forgeries with an accuracy of 80% and can detect the signatures with 91% accuracy. The two stage model can be used in realistic signature biometric applications like the banking applications where there is a need to detect the authenticity of the signature before processing documents like checks

Keyword list Nearest Mean Classifier, PNN, Biometric, Sum Graph, HMM, Forgery Detection

1. INTRODUCTION

n *off-line* signature verification system is one where the signatures are not verified on the fly(At The time Aof writing), rather they will be stored in a database, analyzed and processed. A signature is acquired as an image. This image represents a personal style of human handwriting, extensively described by the graphometry. In such a system the objective is to detect whom the signature belongs to and if there are any forgeries. Each signature is related to intra and inter-personal variability. A forgery is usually represented by a signature sample that belongs to a different writer of the signature model. A skilled forgery is represented by a suitable imitation of the genuine signature model by. This paper detects the skilled forgery with best possible efficiency. We have taken the help of Creative writers and Writing imitators to forge the signature and the level of correctness affiliated by the person whose signature is forged. A database of hundred persons is collected for the experiment.

Figure 1: Represents the signature of the author.

Figure 2: Represents a Signature Forgery of the author.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462G · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853308

Signature verification is a challenging area as even the signature of the same person differs from time to time, where as in other cases the forged signature matches closely to the actual signature. The very nature of this time dependent variability is unique in comparison to other biometrics like finger print identification. Hence it requires techniques that have a high false rejection (Rejecting the signature of the appropriate person if it deviates from the original) than False Accepting (Verifying the false signature correctly).

The study of signature verification involves following independent Studies

1. General Image Processing Techniques
2. Image Preprocessing (Filtering and Noise Removal)
3. Feature Extraction
4. Classification
5. Statistical analysis.

The study is a distinct case of document analysis and biometric. Each step will be discussed in detail in methodology.

2. RELATED WORK

[1] Elaborates the signature verification for banking applications. Most of the banks issue the chequebooks to their customers having different background colors where as the original specimen signature is always taken on the white paper at the time of the opening of the account. When these signatures are verified digitally the results are not accurate. In this paper, we have proposed a mechanism that automates the offline signature verification for bank cheques even with different background colors. Work present in [1] is focused, to examine whether an input signature of colored bank cheque is a genuine signature or a forged. This task is performed by comparing the collected signature samples (white background) with input signatures (colored background). The Signature Verification of Colored Cheques (SVCC) system of verifying the signatures having different background colors in spite of white paper specimens through variation of color intensity is discussed and presented.

[2] Reports the contribution to signature verification considering different forgery types in an HMM framework. It also elaborates that the signature verification is a function of HMM state transition

[3] Reports on a global method based on only two classes of models, the genuine signature and the forgery. The main objective of the proposed method is to reduce the number of signature samples required by each writer in the training phase. For this purpose, a set of graphometric features and a neural network (NN) classifier are used.

A hidden Markov model (HMM) based approach is proposed in [4] for recognition of offline handwritten Devanagari words. The histogram of chain-code directions in the image-strips, scanned from left to right by a sliding window, is used as the feature vector.

In [5], a novel offline signature verification scheme has been proposed. The scheme is based on selecting 60 feature points from the geometric centre of the signature and compares them with the already trained feature points. The classification of the feature points utilizes statistical parameters like mean and variance. The suggested scheme discriminates between two types of originals and forged signatures. The method takes care of skill, simple and random forgeries..

[6] Presents a new feature extraction method based on the intensity of the coefficients of the Gabor transform. This method first uses the multichannel Gabor transform, then the transformed image of Gabor was equally divided into some non-overlapping boxes, and the angle features of the position of the maxima intensity of the Gabor transform coefficients were extracted.

3. PROPOSED WORK

We propose a method for signature verification which is based on Perception and Probability. It means first the system roughly determines to which class a signature belongs to and then it finally decides whether the signature can be accepted or not.

Perception presents the class which a signature "possibly" belongs and the pattern classification based on state transition determines if at all it belongs to that class. Beside it defines an accurate closeness function. We propose such a system where we combine the spatial features of sum graph and HMM and classify them separately by a PNN and Knowledge based classifier.

The detailed mathematical elaboration of the process is presented in the methodology section.

Is a known fact that the rate of correct classification decreases with the increase in number of classes we propose a novel approach to overcome this shortfall. First we divide the entire database into some classes. We have divided 100 images into 20 base classes with each class containing 5 signature each. The classifier now first assumes the signature belongs to set 1. and it tries to identify the signature amongst all the five signatures of the class. If it is not matched then the set two is loaded and so on. Hence number of classes at any given instance of time is constant and is five. Therefore we increase the performance of the classifier

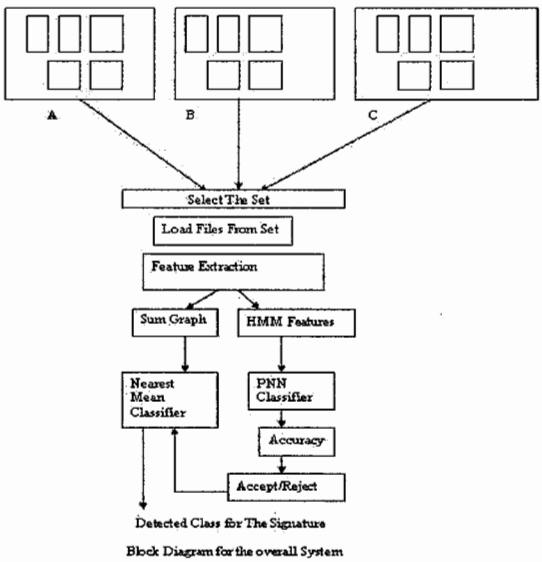


Figure 3: Overall System Block Diagram

A. Preprocessing

A signature needs to be preprocessed before using it for the verification process. As it is a contour based feature, results are much dependent on the shape, size and orientation of contour.

Binarization

All gray scale images are binarized with the help of Natural Threasholding. In this process the mean gray level value of the images are extracted and values greater than the mean are mapped to white where as the values lesser than that are marked as black. This algorithm works fine for most of the handwritten text images with homogeneous or semi homogeneous backgrounds. Hence it is adopted in the current context.

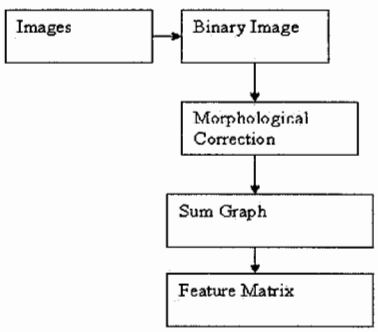


Figure 4: Sample block diagram depicting Sum Graph Feature Extraction

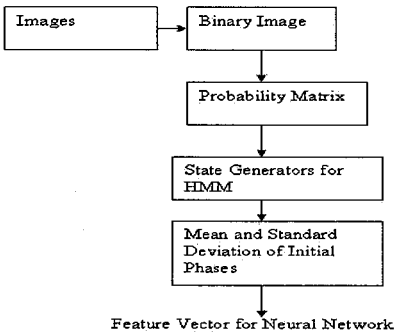


Figure 5: Feature Extraction for Neural Network

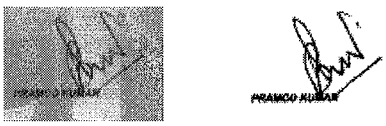


Figure 6: Result of Preprocessing

Noise Removal

Once the image is binarized, a simple 5X5 morphological filter is applied to clean the spatial noise characterized by small components of size (1 to 3 pixels). This clean binarized image is cropped from both the directions. The image is further padded on all the four sides.

Rotation invariance is not considered in this work because every person signs at a different slant. Hence slants are also important preserving features.

B. Feature Extraction

We consider two different types of features viz: 1) Sum graph and 2) HMM features.

Sum Graph

Let us consider that our Input image be I (i, j);

Then the sum graph represents the column wise summation of the color values.

$$S(j) = \sum_{J=1:N} I(I, j) \quad (1)$$

Where N=number of columns.

Sum graph is an important feature in representing signature properties. The area under each peak represents a distinct feature.

HMM

First we determine the probability matrix of appearance of 1 w.rt 0 and 1 and probability of occurrence of 0 wrt 1 and 0. This pattern is unique for all the signatures. As the signatures are considered random in nature, we consider a hidden markov model that can generate such a pattern.

The resulting series of the velocity values for each probability serves as the sequence of observation Symbols for training a discrete HMM. As outlined in Section 1, a HMM based signature verification system typically involves two phases – training and verification. The training phase consists of learning the features (*R V*) of a particular user from a set of signature samples of that user termed as the training sample set. The system then builds a reference HMM model for the user from this training set and stores it in the database. a HMM is quantitatively described by Equation 2.

$$\lambda=\{A,B,\Pi\} \quad (2)$$

Where, A is the state transition matrix and gives the probability of transiting from one state to another.B is the observation probability matrix and gives the probability that a particular

Observation symbol occurs in a particular state.

π is the initial probability distribution for the states. At the end of the training phase, the system evolves the most optimum values of A and B based on the input observation sequences (in this case, the *R V* values from each sample in the training set)

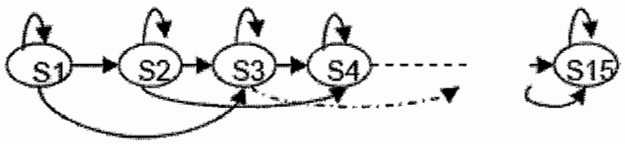


Figure7: State diagram of HMM model

Since a Left-to-Right topology is used therefore, π is fixed at {1, 0, 0....., 0} and is not re-estimated in the training phase. this Left-to-Right topology accounts for the time-dependent dynamic property of the signature data. Furthermore, the presence of single skips in the LTR model gives more freedom in the number of transitions in the model and hence yields better performance in signature verification systems since it better incorporates the intra-user variance inherent in the training sample set of a particular user. The large number of states (at 15) is used because of large observation sequence lengths.

Generally HMM is associated with online signature verification where the states are the sequence of the strokes. But in this paper we have adopted the HMM model successfully for even offline signature verification. Here we have obtained the co-occurance matrix of the binary signature which gives four probabilities. Hence we construct a set of probabilities, S={ {1,1},{1,0},{0,1},{0,0} }. Initial probability π is considered as 1. State generated by giving this set as input to the HMM model represents the variability in signature pattern.

C. Classification

Each state now represents a matrix which can define a probability matrix as calculated prior to HMM. The statistical mean and standard deviation of each value is taken as the training vectors for a neural Network. Neural network is trained based on input statistical features. Output vector is the class representing the signature person. The result of the neural network is further mapped with the results of the knowledge based classifier. A knowledge based classifier (Nearest mean) is designed to classify the signature based on sum graph features. Differential of the sum graph points from each sample wr.t the input vector is calculated. Then the smallest of them is calculated. The smallest set represents the detected class. The accuracy of classification is measured. In case if the accuracy is less than 80% than the class is rejected. For calculating the distance between the training class vectors and test vectors we use following distance measure.

$$dk = \sum (x_S - k_S) \qquad (3)$$

Where dk represents the distance of test sample from kth class. x_S is the sum graph of test sample and k_S is the sum graph of kth class.

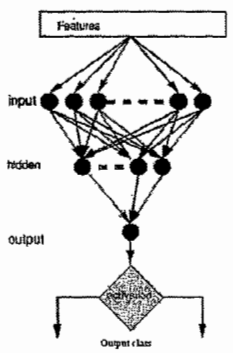


Figure 8: Neural Network Classifier

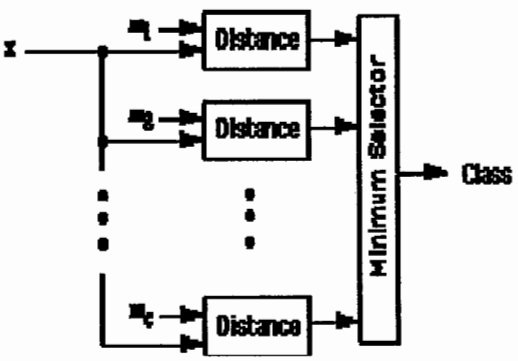


Figure 9: Knowledge based Minimum Distance classifier, m1, m2..represents the training features, x represents the test sample.

4. RESULTS AND ANALYSIS

The results show that the system maintains a good correct classification rate for any number of classes. The lowest rate of correct classification is 86% and the highest is 92%. We define correct classification as follows.

Rate of Correct Classification= (Total Number of Input Classes-(False Accepted Test+ False Rejected Test))*100/ Total Number of Input Classes **(4)**

Where, (False Accepted Test+ False Rejected Test) presents the total false classification.

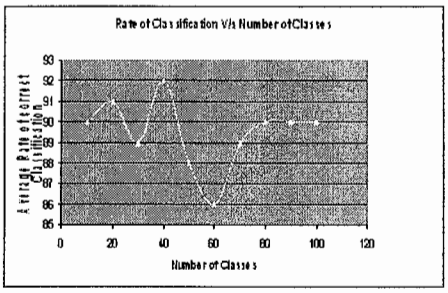


Figure 10(Result1): Number of classes V/s Rate of Classification

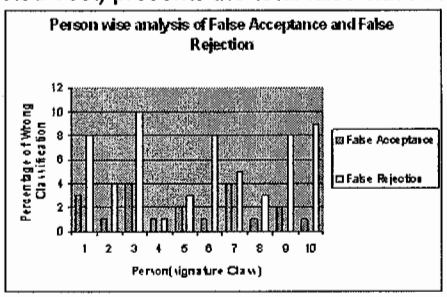


Figure 11(Result 2): Person wise analysis

Result 2 Shows the Graph of independent false acceptance and rejection rate for 10 different classes. The graph shows that for most of the user the false rejection rate is higher than the false acceptance rate. It is the most desired condition in biometric authentication and signature techniques. Therefore we can say that our system performance is satisfactory.

5. CONCLUSION AND FUTURE WORK

The technique presented here, meets the prerequisites and detects the forgeries with good efficiency. This is due to the reason that a model where features are combined outperforms any independent method as a whole in biometric or pattern classification. The accuracy can be further improved by integrating more statistical moments like variance, skewness etc. Here we have not analyzed the time variant analysis of the persons. i.e. samples of the persons were collected at the same time. By nature of signatures, the signature of same person shows slight variations. The algorithm can be further developed to detect such variability by combining more features.

6. REFERENCES

[1]Javed Ahmed Mahar, Mohammad Khalid Khan, Prof. Dr. Mumtaz Hussain Mahar," Off-Line Signature Verifications of Bank Cheque Having Different Background Colors" ,IEEE/ACS International Conference on Computer Systems and Applications 2007, 738-745(2007)

[2] Edson Justino ,Edson J.R. Justino, Flavio Bortolozzi, Robert Sabourin," Off-Line Signature Verification Using HMM for Random, Simple and Skilled Forgeries", Sixth International Conference on Document Analysis and Recognition, 1031-1034(2001)

[3] Cesar Santos, Edson J. R. Justino, Flavio Bortolozzi, Robert Sabourin, "An Off-Line Signature Verification Method based on the Questioned Document Expert's Approach and a Neural Network Classifier", Ninth International Workshop on Frontiers in Handwriting Recognition, 498-502(2004)

[4] Bikash Shaw, Swapan Kumar Parui, Malayappan Shridhar "Offline Handwritten Devanagari Word Recognition: A holistic approach based on directional chain code feature and HMM", Proceedings of the 2008 International Conference on Information Technology, 203-208(2008)

[5] Debasish Jena, Banshidhar Majhi, Saroj Kumar Panigrahy, Sanjay Kumar Jena, "Improved Offline Signature Verification Scheme Using Feature Point Extraction Method" 7th IEEE International Conference on Cognitive Informatics 2008, 475-480(2008)

[6] Jing Wen, Bin Fang , Yuan-Yan Tang , Tai-Ping Zhang , Heng-Xin Chen , "Offline signature verification based on the gabor transform." International Conference on Wavelet Analysis and Pattern Recognition 2007, 1173-1176(2007)

Image retrieval using Feature Extraction based on Shape and Texture

T.Tharani [1] and M.Sundaresan [2]

1. Lecturer, School of CS, RVS College of Arts and Science, Sulur, Coimbatore . ttharani@gmail.com
2. Associate Professor, School of CSE, Bharathiar University, Coimbatore-641046. bu.sundaresan@gmail.com.

ABSTRACT

Data mining refers to the process of extracting knowledge that is of interest to the user. Traditional data mining techniques have been developed mainly for structured data types. The image data type does not belong to this structured category, suitable for interpretation by a machine and hence the mining of image data is a challenging problem. Accordingly, in image mining, an image retrieval system is a computer system that can browse, search and retrieve images from a large database of digital images. This research work is aimed at compression and retrieval of images from large image archives. A Kohonen Self Organization Map approach using content categorization, including feature level clustering, is developed to provide a differential compression scheme. It ensures that the visual features are mapped to codebooks, which significantly speed up content-based retrieval. The interaction between compression and content indexing are proposed, which include techniques for feature extraction, indexing, and categorization. K-means clustering algorithm is used to build the feature cluster. This approach leads to the similarity matching based on shape and texture, which supports functions like "query by example". Experimental results demonstrate that the proposed method can improve the compression ratio compared to VQ. The average retrieval time is less than 2seconds, which is proved to be efficient.

KEYWORDS : Image Retrieval, , Feature Extraction, Indexing, Content Categorization, Compression

1. Introduction

An exponential growth in the generation and storage of digital images in different forms is seen because of the advent of electronic sensors like Complementary Metal Oxide Semiconductor (CMOS) or Charge-Coupled Device (CCD) and image capture devices such as digital cameras, camcorders, scanners, etc. Therefore, image databases are typically very large in size. It leads images to a compression and mining process in order to reduce the space and to gather user-preferred information.

The first process in image mining is to maintain large collections of images and the second is the combined data mining of large collections of image. Image mining deals with the extraction of implicit knowledge, image data relationships or other patterns not explicitly stored in the images [7]. It is more than just an extension of data mining to the image domain. Image mining is the discovery of patterns from within images or a collection of images. Image mining is an interdisciplinary endeavor that draws upon expertise in computer vision, pattern recognition, image processing, image retrieval, data mining, machine learning, database, artificial intelligence and compression.

Conventional image mining techniques include object recognition, image storage, image retrieval, image indexing, feature extraction, classification, clustering and knowledge discovery. Intelligently classifying an image by its content is an important way to mine valuable information from a large image.

2. Content based image retrieval

Content-based image retrieval, also known as Query By Image Content (QBIC) and Content-Based Visual Information Retrieval (CBVIR) is the application of computer vision to the image retrieval problem, that is, the problem of searching for digital images in large databases. "Content-based" means that the search will analyze the actual contents of the image. The term 'content' in this context refers to colors, shapes, textures, or any other information that can be derived from the image itself. This research work is based on shape and texture features.

Retrieval of a query image from a large database of images is an important task in the area of computer vision and image processing [1]. The general model of CBIR system is shown in the fig 1.

Content based retrieval is one of the functions of image mining [4]. Moreover, image mining inherently deals with spatial information and often involves multiple interpretations for the same visual pattern. Hence the image mining algorithms are different from the traditional algorithms of data mining.

Standards development relevant to CBIR can be grouped under two headings;

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462H · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853481

- Content based image compression and
- Query specification

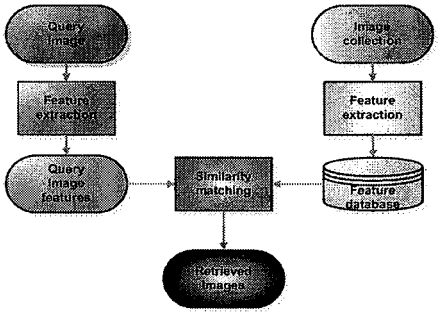


Fig 1. General model of CBIR System

This research work focus on content-based image compression and retrieval, it is described in the section 4.

3. Related work

QBIC [3, 2, 6, 5], standing for query by image content, is the first commercial content-based image retrieval system. Its system framework and techniques have profound effects on later image retrieval systems. QBIC supports queries based on example images, user-constructed sketches, drawings, selected color and texture patterns, etc. The color feature used in QBIC are the average (R,G,B), (Y,i,q), (L,a,b), and MTM (mathematical transform to Munsell) coordinates, and a k-element color histogram. Its texture feature is an improved version of the Tamura texture representation; i.e. combinations of coarseness, contrast, and directionality [2]. Its shape feature consists of shape area, circularity, eccentricity, major axis orientation, and a set of algebraic moment invariants [6]. QBIC is one of the few systems which takes into account the high dimensional feature indexing. In its indexing subsystem, KLT is first used to perform dimension reduction and then R¤-tree is used as the multidimensional indexing structure [5]. In its new system, text-based key word search can be combined with content-based similarity search. The on-line QBIC demo is at http://www.qbic.almaden.ibm.com.

4. SYSTEM ARCHITECTURE

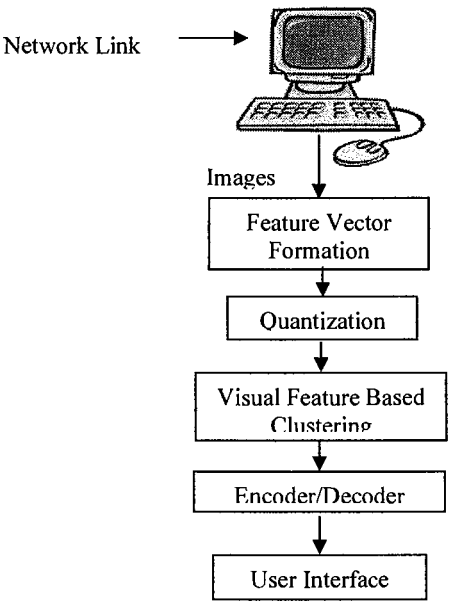


Fig 2. Client-Server Architecture for Image Search System

In figure 2, network link is used by the user to access the server using a limited bandwidth, say W bits per second. User interface is the place where the results of the CBIR are displayed. The research work focuses on compressing 28-bit high resolution images with 120 x 120 sizes. Two visual features, shape and texture, are used for creating feature vectors. Compression technique that uses two quantization processes is used for image compression. They are, (i) Vector Quantization and (ii) Self-Organizing Maps

Clustering of quantized vector is performed using K-Means algorithm and the result of feature indexing is used for accurate image retrieval. Query-by-Example method is used to retrieve images that are visually similar to the query. The following section explains each module of the architecture in detail.

4.1 Visual Feature Extraction

The construction of CBIR begins with the extraction of suitable features from the image database. In the proposed work texture and shapes are used to retrieve images from the image database.

4.2 Shape

Shape does not refer to the shape of an image but to the shape of a particular region that is being sought out. Shapes will often be determined by first applying segmentation or edge detection to an image. In this paper the edges are detected to find the shapes of the images.

4.3 Texture

Texture measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image. These sets not only define the texture, but also where in the image, the texture is located. Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated.

4.4 Quantization

Encoding of images takes place in the server. The resulting codebook, indices are transmitted to the client, where decoding procedure takes place and images are displayed in the user interface. Quantization is the process of representing a large set of values with a much smaller set. VQ is a lossy data compression method based on the principle of block coding. It is a fixed-to-fixed length algorithm. VQ can be thought of as an approximator where the idea is very similar to "rounding-off" to the nearest integer.

In VQ, codebooks are created for each image which is used to encode an image, each image is represented by a codeword in the codebook. As a result codebook contains the lookup table for coding and decoding. The general architecture of an image VQ scheme is shown in Figure 3.

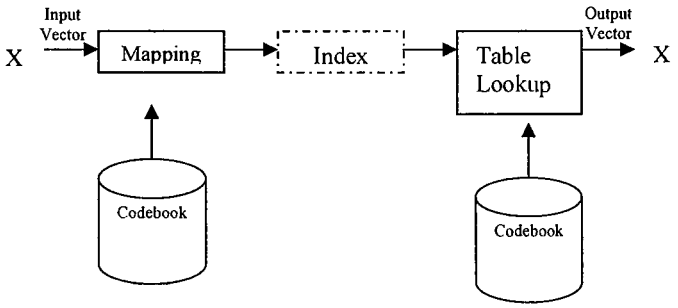


Fig 3. General Scheme of Image VQ

An input image is partitioned into non-overlapping blocks. Each block of pixels is encoded as the index of its nearest codeword in the VQ codebook, according to distortion measures. The decoder simply looks up the indices and outputs the corresponding codewords. The average distortion of the VQ coding scheme is

$$D = E[d(F_i, F_i')]$$

where F_i is the input block and F_i' is the decoded block. Thus the optimal codebook is the one that yields the least average distortion. The K-Means clustering algorithm is widely used to find an optimal codebook.

4.5 K-Means Clustering

The K-means clustering algorithm is a simple and straightforward method for clustering image data. In the K-means there are always K clusters. Thus the features will be grouped into at least one group. The K-means algorithm is implemented using the following three steps

- The dataset is partitioned into K clusters and the data points are randomly assigned to the clusters resulting in clusters that have roughly the same number of data points.
- For each data point:
  - ✓ Calculate the distance from the data point to each cluster.
  - ✓ If the data point is closest to its own cluster, leave it where it is. If the data point is not closest to its own cluster, move it into the closest cluster.
- Repeat the above step until a complete pass through all the data points results in no data point moving from one cluster to another. At this point the clusters are stable and the clustering process ends.

The main advantage of K-means algorithm is its simplicity and speed which allows it to run on large datasets.

4.6 Self Organization Map

The Self-Organizing Feature Map is an ingenious Neural Network built around a one or two dimensional lattice of neurons for capturing the important features contained in the input. The aim of the Self-organizing Feature Map algorithm is to store a large set of input vectors by finding a smaller set of weights so as to provide a good approximation to the original output space, which is the motivation for adopting Self-Organizing Feature Map for image compression.

The Self-Organizing Feature Maps are most simple to implement and hence a well known algorithm for compression of images. As image sample tends to have redundancy in nature, the Self-organizing feature maps are capable of grouping nearby similar pixel.

4.7 Initializing weights for the network

The architecture of the Self Organizing Feature Maps for Vector Quantization depends on the size of the codebook. Each code vector in the codebook is the weight matrix of each node of the competitive layer of the Self Organizing Feature Maps. The codebook size is chosen to be 120. Thus 120 code vectors that represent the input image block are produced as output by this network. There is no separate output layer. Only the weights of the neurons of the competitive layer are taken as output. Since in image compression the weight values after training has to represent the input gray levels, assigning random values ranging from 0 to 255 does initializing of weights.

The input layer takes as input the gray level values from all the 16 pixels of the gray level block. The weights assigned between node j of the Kohonen layer and the input layer represents the weight matrix $W_{ji} = [w_{j0}, w_{j1}, w_{j2},\ldots w_{j15}]$. Similarly all the 256 nodes have W_{ji} for j = 0, 1...255 and i = 0, I...15. Once the weights are initialized randomly network is ready for training.

4.8 Training

The image blocks are given as input to the network. The input vector, say x is simultaneously applied to all nodes. The similarity between x and weight w_i is measured in terms of spatial neighborhood N_m. The weights affecting the currently winning neighborhood undergo adoption at the correct learning step other weights remain unaffected. The neighborhood, Nm is found around the best matching node m such that

$$\|x - w_m\| = \min [\|x - w_i\|]$$

The radius of N_m will be decreasing as the training progresses. Towards the end of training the neighborhood may involve no cells other than the central winning one. After training the weight vectors of each neuron of the Kohonen layer acts as code vectors.

4.9 Encoding image using code book

After the codebook is generated the same Self Organizing Feature Map is used for encoding and decoding the input image. The steps for encoding of the input image is given below.

- ✓ The input image is divided into blocks
- ✓ The image blocks are given as input to the Self Organizing Feature Map.
- ✓ Inner neuron is selected as the neuron having the minimum Euclidean distance.
- ✓ The index of winner neuron for each input block is stored consequently.

- ✓ The set of indices of all the winner neurons for the blocks along with the codebooks gives the compressed form of the image.

4.10 Decoding image using code book and index

The codebook as well as the index of the winner node for all the blocks is the compressed form of the image. The image is decoded using the following steps

- ✓ For each index value find the neuron in the Self Organizing Feature Map.
- ✓ Generate the weight vector of that neuron to the input layer neurons. This value of the neuron weights is the gray level for that block.
- ✓ Display the gray level value thus obtained as pixels

5. Image retrieval using query by example

Query by example is a query technique that involves providing the CBIR system with an example image that it will then base its search upon. The underlying search algorithms may vary depending on the application, but result images should all share common elements with the provided example. Options for providing example images to the system include:

- A preexisting image may be supplied by the user or chosen from a random set.
- The user draws a rough approximation of the image they are looking for, for example with blobs of color or general shapes.

This query technique removes the difficulties that can arise when trying to describe images with words. In Query by example method, example images are given as input to the content-based image retreival system, which then retreives "similar" images as output.

6. Experimental results

The results of the proposed system are compared with that of VQ based system and are presented in Tables 1,2,3 and figure 4. The experiments were conducted with codebook size two for 256 images.

Codebook Size	VQ Based	Proposed method
2x2 Codebook size (N=256)	89	92
4x4 Codebook size (N=256)	91	94.5

Table 1. Codebook Size

IMAGE	VQ	Proposed method
Image01	1.55	1.37
Image02	1.56	1.46
Image03	1.51	1.32
Image04	2.10	1.76

Table 2. Retrieval Time in secs

The Mean Square Error (MSE) and the Peak Signal to Noise Ratio (PSNR) are the two error metrics that are used to compare image compression quality. The MSE represents the cumulative squared error between the compressed and the original image, whereas PSNR represents a quality measurement between the original and a compressed image.

IMAGE	MSE		PSNR	
	VQ	Proposed method	VQ	Proposed method
Image01	22.62	17.05	31.91	34.84
Image02	23.05	17.63	31.08	33.03
Image03	21.98	18.05	31.94	34.68
Image04	22.72	16.98	32.01	34.65

Table 3. MSE and PSNR

Accordingly retrieval time, precision and recall measures show the CBIR efficiency achieved by the proposed system.

- **Retrieval Time**

Retrieval time is the time taken by the CBIR system to match a given query example image in the server, create codebooks, perform compression, transmit them to server, perform decompression, time taken to reconstruct the retrieved image. The results obtained are compared with the VQ as presented in the table 2.

- **Precision and Recall**

The deviations from this can be measured using two metrics Precision and Recall. Any CBIR system should have as few retrieval errors as possible. The two types of errors are calculated as follows : Precision is calculated by dividing the no of retrieved relevant in answers by no of retrieved and Recall is calculated dividing no of retrieved relevant in answers by total no of relevant in the collection.

The interpretation of the results in terms of precision is that a high precision means few false alarms and for recall (point 2), a high recall value indicates few false dismissals (point 1). Precision tells how many of all retrieved images belong to the correct class. Recall tells how many of the retrieved images do not belong to the correct class Figure 5 shows the screen shot with the sample query image and retrieved images with results.

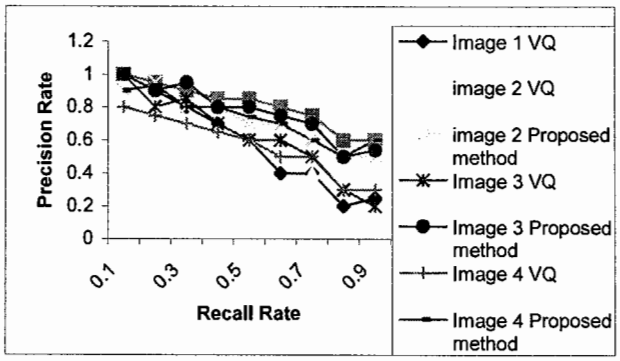


Fig 4. Precision-Recall Chart

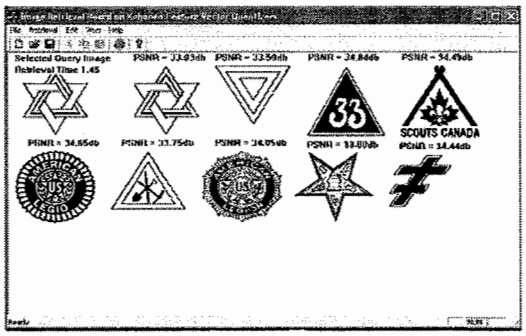


Fig 5. Screen Shot

7. Conclusion

In this paper we propose a Self Organization feature map used for content categorization based on its visual features. The experimental results show that this scheme is effectively compress the image database and it speed up the retrieval time compared to VQ.

8. References

[1] Acharya, T. and Ray, A.K. "Image Mining & Content-Based Image Retrieval", John Wiley & Sons, Inc., 2005

[2] Equitz, W. and Niblack, W. "*Retrieving Images from a Database using Texture-Algorithms from the QBIC System*", Technical Report RJ 9805, Computer Science, IBM Research Report, 1994

[3] Haralick, R.M., Shanmugam, K., and Dinstein, I. "*Texture features for image classification*", IEEE Trans.On Sys.Man. and Cyb.SMC-3(6), 1973

[4] Haralick, R.M., Shanmugam, K., and Dinstcin, I. "*Texture features for image classification*", IEEE Trans.On Sys.Man. and Cyb.SMC-3(6), 1973

[5] Lee, D., Barber, R., Niblack, W., Flickner, M., Hafner, J. and Petkovic, D., "*Indexing for complex queries on a query-by-content image database*", in Proc.IEEE Int.Conf.on Image Proc.., 1994

[6] Scassellati, B., Alexopoulos, S. and Flickner, M., "*Retrieving images by 2d shape:Acomparison of computation methods with human perceptual judgments*", in Proc.SPIE Storage and Retrieval for Image and Video Databases, 1994.

[7] Zheng, N., Zhang, Z., Zheng, H., and Gang, S., "*Deterministic annealing learning of the radial basis function nets for improving the regression ability of RBF networks*", Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks.IJCNN Vol.3, pp.601–7., 2000

User Region Extraction from Dynamic Projected Background for Virtual Environment System

Taeyoung Uhm[a] , Hanhoon Park[b] , Moon-Hyun Lee[a] and Jong-Il Park[a*]

[a]Division of Electrical and Computer Engineering, Hanyang University, Seoul, Korea;
[b]NHK STRL, Tokyo, Japan

ABSTRACT

There has been considerable interest in immersive and realistic virtual environment system and how to improve human and computer interaction has been a main challenge. For vision-based human-computer interaction, extraction of user region from camera images is an essential part. In this paper, we propose a background subtraction method that segments dynamic projected background in a rear-projection-based virtual environment system. In the projector-camera system, the projected background is inherently known by the projector input images although its appearance is changed by the geometric and radiometric transformation between projector and camera. Therefore, we can compute the expected background location and appearance based on geometric and radiometric calibration of projector-camera system and thus separate user region from dynamic projected background by simple subtraction between camera images and the computed background. Experimental results are given for verifying the usefulness of the proposed method.

Keywords: Dynamic background subtraction, projected imagery segmentation, virtual environment system, geometric calibration, radiometric calibration, projector-camera system

1. INTRODUCTION

For last decades, virtual environment system based on projector-camera system has gained consistent attention and recently commercial systems have been emerging. However, virtual environment systems still suffer from the lack of natural interface. For being natural, vision-based interfaces will be most desirable but the dynamic background has been a great obstacle. Therefore, this paper proposes an effective method for segmenting the dynamic background and extracting user region. The method is based on the fact that the geometry and appearance of dynamic background in camera images are inherently predictable from the projector input images. The prediction is done by formulating the geometric and radiometric transformation between projector and camera. Therefore, we can extract user region precisely by simply subtracting the computed background from camera images.

The organization of this paper is as follows. In section 2, we review the extraction method of the user region and how it can be used effectively. Then, we explain the formulation of transformation and the extraction method of user region in section 3 and 4. In section 5, we present the result of a variety of experiments about Korean traditional dance sequence in virtual environment system by adapting the method from section 3 and 4. Finally, we conclude this experiment and discuss the future work.

2. RELATED WORKS

The study of natural interface for user interaction in projection environment has been steadily done and fundamentally based on user region extraction [1]. The classic approaches for extracting interest region is mainly based on chroma keying, difference keying, depth keying, thermal keying, infrared keying etc.

Chroma keying separates user region from a single colored (primarily blue or green) or fully-known colored background. It is the most classical method for extracting the region of interest but due to its high performance it is also the most-

* Further author information: (Send correspondence to Taeyoung Uhm.)
Taeyoung Uhm.: E-mail: uty02@mr.hanyang.ac.kr, Telephone: +82-02-2299-4368
Hanhoon Park: E-mail: hanuni@mr.hanyang.ac.kr, Telephone: +81-3-5494-3342
Moon-Hyun Lee: E-mail: flymoon@mr.hanyang.ac.kr,
Jong-Il Park : E-mail: jipark@hanyanga.ac.kr

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462I · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853010

widely used method, especially in the broadcasting field where high quality segmentation is required [2]. However, the condition that background color is given and known in advance is too strong in ordinary environments.

Difference keying is a technique that extracts the region of interest from trained static background [3, 4, 5]. That is, background appearance is trained from the temporally static pixel values in camera images. Background can have arbitrary colors but should be static.

Depth keying is a technique that extracts foreground region using depth information. Therefore, it can be used for arbitrary colored and dynamic background. However, special stereo camera or multiple calibrated cameras are required and the low resolution of depth map is a critical obstacle [6].

Thermal keying and infrared keying extract the region of interest based on the thermal [7] and spectral difference [8], respectively. The thermal keying or infrared keying is robust to environmental changes such as illumination change compared to the preceding methods using color information. However, they are available in the limited environments, e.g. the thermal difference between foreground and background should be large, the background should be made of infrared retro-reflective material, etc. Moreover, they require the additional device such as thermal imager and infrared camera.

The existing approaches are hard to be applied to our virtual environment system which has dynamically projected background and uses a single RGB camera. Consequently, to the best of our knowledge, there is no vision-based method for segmenting user region from dynamically projected background using a single RGB camera without additional devices.

3. FORMULATION THE GEOMETRIC AND RADIOMETRIC TRANSFORMATION BETWEEN PROJECTOR AND CAMERA

The geometric relationship between projector input image (I) and camera image (C) is represented as

$$C = \begin{bmatrix} x_c & y_c & 1 \end{bmatrix}^T = H_{p-c} \begin{bmatrix} x_p & y_p & 1 \end{bmatrix}^T = H_{p-c} I. \quad (1)$$

Here, I, C, and H represent a projector input image, a camera image, and their 2D homography via a projection screen. Then, the radiometric relationship between projector input image and camera image is represented as

$$C = g(VAf(I) + F). \quad (2)$$

Here, V, A, and F indicate the color mixing matrix between projector and camera, screen reflectance, and environmental light. '$f(\cdot)$' and '$g(\cdot)$' indicate the nonlinear response function of projector and camera.

For the derivation of two equations (1) and (2) and the computation of the parameters in the equations, refer to [9].

4. USER REGION EXTRACTION

In a camera image (C), user region consists of two regions according that the background is the projection screen (R_O) or not (R_X) as

$$C = R_O + R_X + N \quad (3)$$

$$\text{Where } R_O = F_O + B_O, \; R_X = F_X + B_X$$

Here, F, B, N indicate the foreground, background, and environmental noise, respectively. This is also illustrated in Fig. 1. The separation between two regions can be done by the geometric calibration [9] that informs us where the projected background is in the camera images. From the two regions, user region is extracted by two cooperative methods.

If the background is not the projection screen (it must be static), the existing difference keying with shadow elimination [3], which computes the background image from camera images in a statistical way, can be used as it is for extracting F_X. If the background is the projection screen, the background (B_O) will be changed dynamically due to image projection and thus the statistical method is not useful anymore. For adapting for dynamic background, we take advantage of the geometric calibration and radiometric calibration method for projector camera system [9].

For extracting F_O from R_O, we can predict the background geometry and appearance using the geometric and radiometric calibration [9]. Then, we subtract the predicted background image (B_O') from R_O and extract F_O where the difference is larger than a threshold.

The user region is the sum (F_O+F_X) of the results by two keying methods. However, the result contains the environmental noise. Therefore, we eliminate the noise using additional algorithms such as large blob detection and morphology.

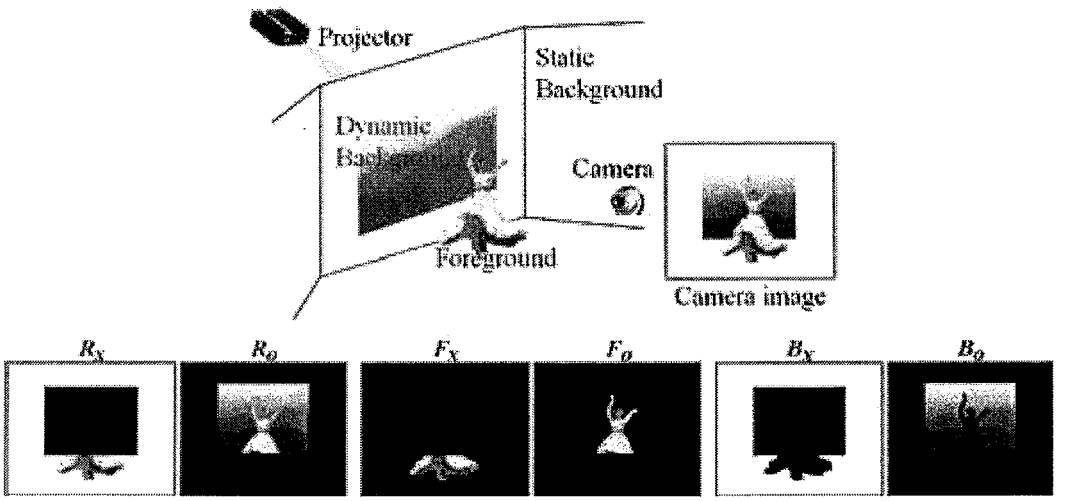


Figure 1. Two different user region in camera image.

5. EXPERIMENTAL RESULTS

A virtual environment system (see Fig. 2) was created with the purpose of analyzing the dancers' interaction with virtual environment, the effect of the virtual environment on dance performance, etc. The system has three sides (rear screens) which are projected by three projectors. A single camera is used for calibrating the system and capturing the scene. The projectors and camera were synchronized with the v-sync signal of video output from a laptop. The geometric calibration and radiometric calibration were done once in an initial step.

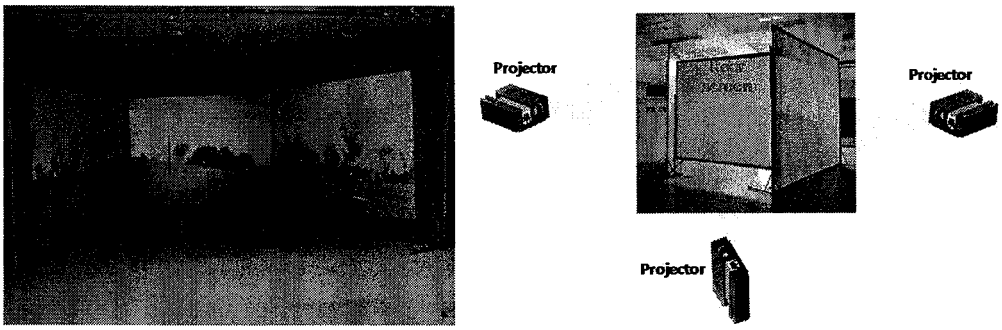


Figure 2. Our virtual environment system.

Figure 3 shows the step-by-step results for eliminating projected background in a side of our virtual environmental system. The upper left image (1) is original projector input image and the upper middle image (2) is its projection image, i.e. dynamic projected background. The upper right image (3) is the difference image between the background image predicted by geometric calibration and the projection image (2). The lower middle image (4) is the difference image between the background image predicted by both geometric and radiometric calibration and the projection image (2). The lower right image (5) is the final result after noise elimination. Consequently, by accurately predicting the projected background image from the projector input image we could simply yet precisely eliminate it from the camera image.

Figure 4 shows the results of extracting the real dancer's region in our virtual environment system. First, we extracted a part (F_X) of dancer's region using difference keying with shadow elimination. Then, we extracted the rest (F_O) of dancer's region using the proposed projected background subtraction method. By these steps, we could extract the whole dancer's region precisely and in real-time. As a practical problem, camera images may include interreflection between adjacent screens [10] and shadow cast by the dancer onto projected background. They cannot be predicted or eliminated by the proposed method. However, in our experiments, they were not found in camera images and their effect was ignorable.

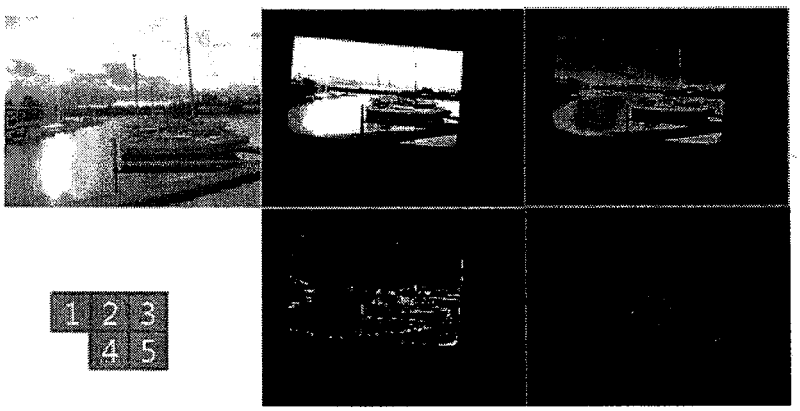
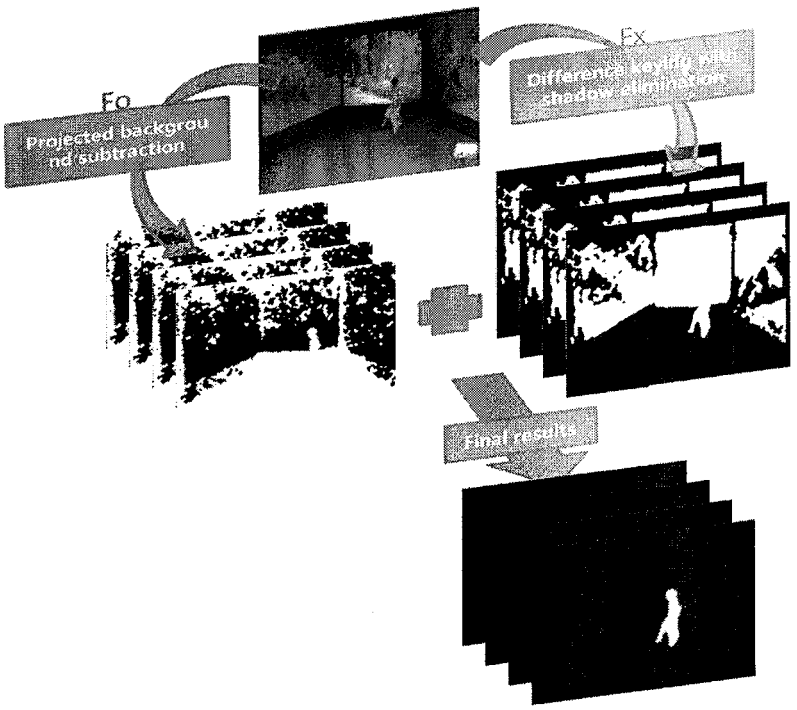
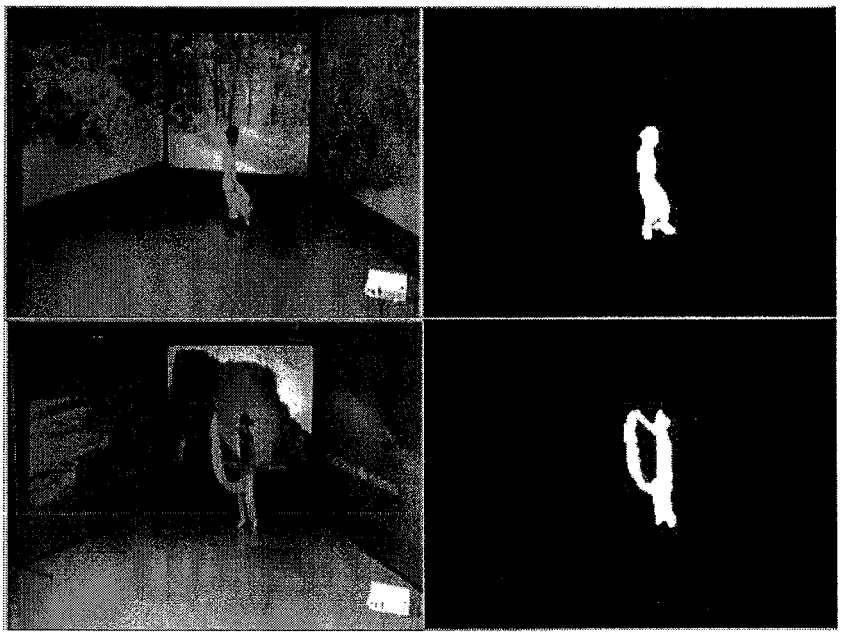


Figure 3. Result of dynamic background subtraction.



(a) Procedure for extracting dancer's region.



(b) Other input and output image

Figure 4. Results of dancer's region extraction

6. CONCLUSION AND FUTHER WORK

In this paper, we proposed a vision-based method for extracting user region from dynamic projected background based on the geometric calibration and the radiometric calibration. The proposed method could extract the user region precisely and in real-time from the monocular camera image without additional devices.

In this paper, we did not consider the more complicated light components such as inner reflection between each screen. The related techniques have been already developed [10] and thus can be applied easily. Currently, we compute the expected background image from the related projector input image. However, use of the images before and after the projector input image would be effective for reducing the computation time or improving the extraction result.Tables are handled identically to figures, except that their captions appear above the table.

ACKNOWLEDGMENTS

We would like to thank Un-Mi Kim and Mi-Hee Lee in the Department of Dance, Hanyang University for their full support in system organization and dance performance.

REFERENCES

[1] R. Sukthankar et al. "Smarter presentations: exploiting homography in camera-projector systems," *Proceedings of International Conference on Computer Vision*, 2001.
[2] A. Grundhöfer and O. Bimber, "VirtualStudio2Go: digital video composition for real environments," *Proceedings of ACM SIGGRAPH Asia*, Singapore, Article 151, December 2008.
[3] T. Uhm, H. Park, J. Park, and U. Kim, "Quantitative analysis of Korean dance performance based on a vision-based emotion recognition system," *Proceedings of International Symposium on Ubiquitous VR*, pp.113-114, Yanji City, China, July 2006.
[4] T. Horprasert, D. Harwood, and L. Davis, "A statistical approach for real-time robust background subtraction and shadow detection," Proceed*ings of IEEE Frame-Rate Workshop*, KerKyra, Greece, 1999.
[5] K. Toyama, J. Krumm, B. Brumitt, and B. Meyers, "Wallflower: principles and practice of background maintenance," *Proceedings of International Conference on Computer Vision*, 1999.
[6] G. Pingali and A. Hampapur, "Geometric image masking for segmenting dynamic projected imagery," *Proceedings of International Conference on Image Processing*, 2003.
[7] K. Yasuda, T. Naemura, and H. Harashima, "Thermo-Key: human region segmentation from video," *IEEE Computer Graphics and Applications*, 24(1): 26-30, 2004.
[8] Y. Yamanouchi, T. Fukaya, H. Mitsumine, and H. Okubo, "Three Little Pigs in the CG Theater," *Proceedings of ACM SIGGRAPH Asia*, Singapore, 2008.
[9] H. Park, M.-H. Lee, S.-J. Kim, and J.-I. Park "Surface-independent direct-projected augmented reality," *Lecture Notes in Computer Science*, vol. 3852, pp.892-901, 2006.
[10] O. Bimber, A. Grundhöfer, T. Zeidler, D. Danch, and P. Kapakos, "Compensating indirect scattering for immersive and semi-immersive projection displays," *Proc. of IEEE Virtual Reality*, pp. 151-158, 2006.

The Reality Model of Plum Tree based on SpeedTree

Zhi-yong Bai, Xin-yuan Huang
School of Information Science and Technology, Beijing Forestry University, Beijing, China.
hxy@bjfu.edu.cn

ABSTRACT

Plum Blossom as the Chinese traditional flowers may be unique all over the world and has the first right of access to international registry of flower. In this paper, the SpeedTree software is used to quickly build reality model of the plum tree. The graphics texture mapping techniques is used, and the plum tree image maps express the geometric model of the surface material, which constitutes a visual image of the graphic objects. It is significant for non-destructive study of plum and virtual garden.

SpeedTree; Plum Tree; Reality Model

1. Introduction

Plum blossom is the spiritual symbol of the Chinese nation, with a strong and universal appeal and drive[1]. Plum blossom represents the spirit of quality: fortitude and perseverance. Other flowers are open only in spring, and it is not the same. The more it is cold and the wind bully snow pressure, the more delicate and elegant it is. It is the backbone of the Chinese nation's most flowers! For thousands of years, it have encouraged the Chinese people generation after generation, keeping forging ahead and creating a good life and civilizations. In the literature and art history, there are a lot of plum poetry and painting. Plum Blossom as the Chinese traditional flower, may be unique all over the world and has the first right of access to international registry of flowers[2].

Landscape visual modeling is one of the hot areas of the research in the current 3DGIS, virtual reality technology , and its core technology is the computer graphics processing. Graphics texture-mapping technology is the computer graphics used to build a true graphical environment, and it is the most common way to build the vegetation , trees and flowers, organic life, body image maps and express the geometric model of the surface material, which constitutes a visual image of the graphic object. A flower is undoubtedly an important component of factors in a natural scene, and the complex structure of flowers is considerable difficulties.

The main topic is using interactive method to establish plum model according to the ecological characteristics of plum tree model based on the SpeedTree library and the real plum graphics. The Reality model of plum tree based on SpeedTree is different from the previous model using 3DS MAX by building a single complex model but by building a parameter method[3]. It has a bit easy and convenient operation in three-dimensional viewing and facilitates results and so on. The Reality model of plum tree based on SpeedTree which significance lies in a random change per plant is thus suitable for the production of a wide range of plum garden for games, virtual landscapes and virtual scenes. At the same time the final research results accurately and quickly build models of plum that can be used in the simulation related to non-destructive experiments, and it is thus important for the protection of plum varieties and the bio-diversity.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462J · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853212

2. SpeedTree

SpeedTree is a powerful modeling and rendering software quickly to help programmers and animation designers to move pretty dynamic three-dimensional leaf patterns to scenes. It is able to produce large areas of forest fine patterns.

SpeedTree has many special effects as well as the optimization technology providing the environment in natural conditions such as wind speed and direction and allow real-time generation of trees with an absolutely realistic swing effect. In addition, SpeedTree engine also optimize the code and is specifically responsible for the scene of "planted" tree of the program. SpeedTree not only create a very real effect of trees and forests, but also as "parts" easily embedded into other rendering engine.

3. Technology

3.1 IBR (Image-Based Rendering)

In recent years, with the continuous development of computer hardware and software technology, large-scale real-time rendering complex scenes have become possible, which made the model complexity and realism put forward new demands. The performance of large-scale visualization of trees is a concentrated expression of such a requirement. The traditional geometry-based graphics rendering needs to reduce the scene to deal with the number of patches and still needs the hardware support. IBR is to achieve ordinary computer graphics. In this method, the image rather than the polygon is used to draw the main data, which advantage is calculating the amount of drawing and geometric model of independent vertices. IBR technology can be easily applied in tree modeling and so on. IBR is based on a number of pre-acquired reference images to generate an arbitrary point of view of the target image [4].

IBR has the following characteristics: its graphics do not need much hardware, and the technology can be implemented on a computer in the general real-time rendering of complex scenes; rendering speed is independent of the geometric complexity of the scene, but only with the reference image, target image size and resolution; rendering does not require complex lighting calculations and reference images taken from actual photographs or calculated through a complex image is stronger than the traditional method of drawing; drawing does not require the construction three-dimensional geometric model of the scene.

3.2 Billboarding

In the current real-time graphics rendering systems, 3D games and graphics engines, Billboarding technology has been widely used to speed up the rendering scene. Billboarding is a polygonal line of sight based on the direction of rotation, as shown. With the alpha-texture combination, it will show a variety of solid modeling which is difficult to describe [5].

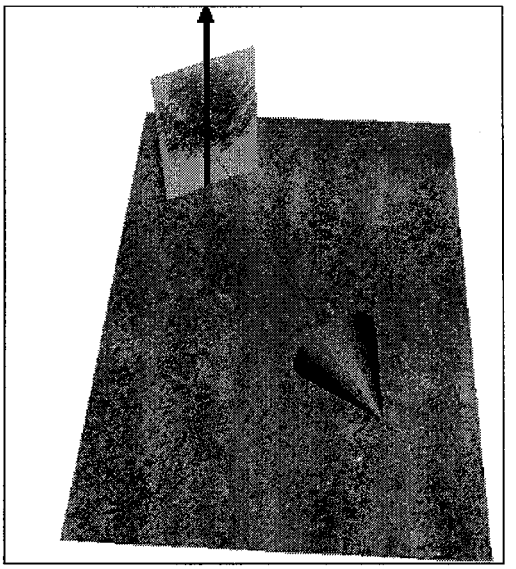


Figure 1: Billboarding

4 Production process

Use the SpeedTree to produce a model of plum tree. Steps are as follows:

1) Components of a SpeedTree Model [6].

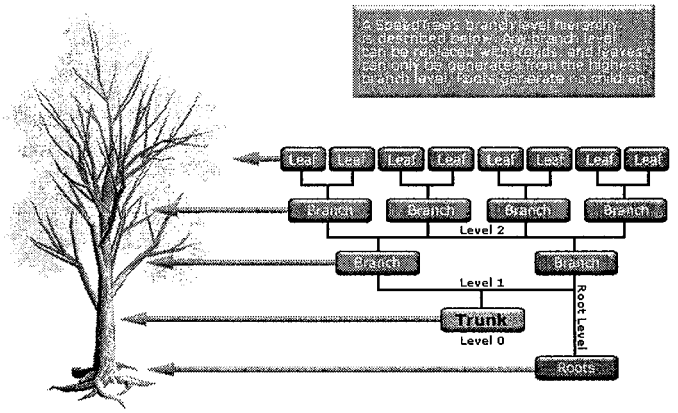


Figure 2: Components of a SpeedTree Model

2) The object of study.

Figure 3: Plum Tree

3) SpeedTreeCAD is basic SpeedTree modeling software. In order to create a simple tree to begin, there is a basic tree model and then gradually modify the parameters build the final corresponding tree.

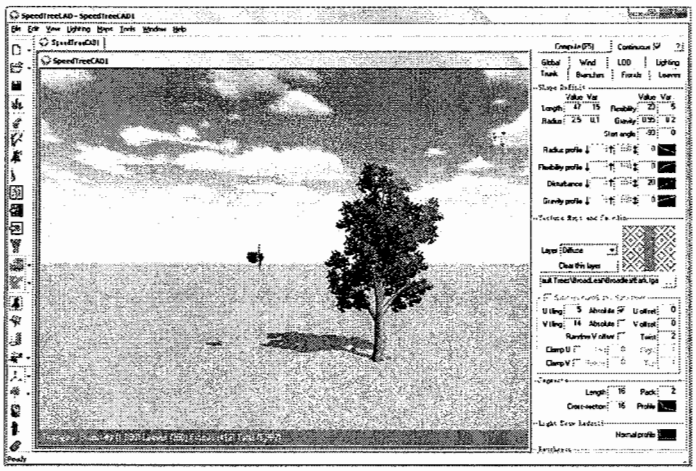


Figure 4: SpeedTreeCAD interface

4) Trunk: Replacement plum bark.

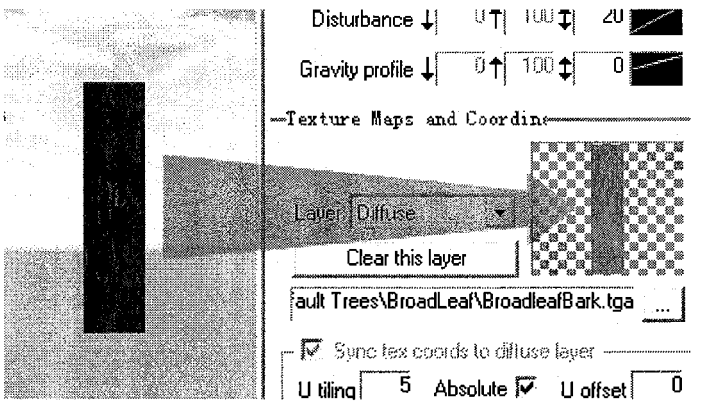


Figure 5: Bark

5) Branches.

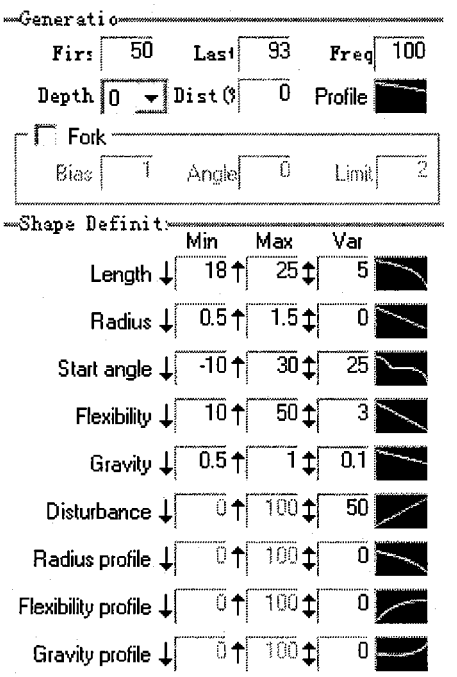


Figure 6: Branches

6) Leaves: Plum tree's leaves are its flowers.

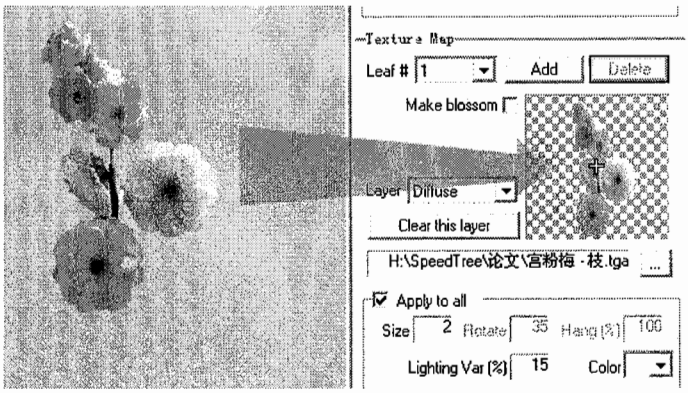


Figure 7: Branches

5 Results

Summing up the above process, finally a realistic model of plum tree is build.

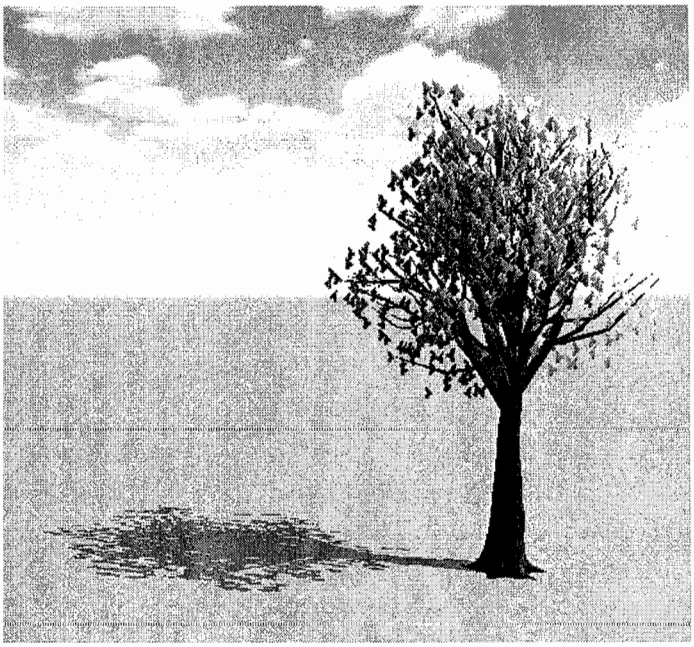


Figure 8: Results (far)

Figure 9: Results (near)

6 Summary

With the computer graphics, real-time hardware, software technology rapid development, the real-time performance of the current three-dimensional graphic image is the key research trend in the field. There is no plum visualization study. In this paper , The Reality Model of Plum Tree based on SpeedTree plays an important role, and the new technical ideas are still being continually updated and supplemented. Study of the structure plum appearance, showing a more realistic virtual plant is a very representative research on landscape design and other relevant fields.

References

[1] Dun-xian Liu, China Plum complex, Chinese Garden, 2005

[2] Jun-yu Chen, Chinese Plum Flower Species Field Guide, China Forestry Publishing House, 1989

[3] Shi-hung, Chen.3ds max 9 Command Reference, China Publishing House of Electronics Industry, 2007

[4] Kui-yu Li, Based on the depth of image-based rendering algorithm, Beijing, Chinese Academy of Sciences Master's degree thesis, 2005

[5] António. Ramires. Fernandes, Billboarding Tutorials, lighthouse3d, 2004

[6] IDV, SpeedTreeCAD User Reference, IDV Company, 2008

Face Recognition by Hopfield Neural Network and No-balance Binary Tree Support Vector Machine

Ke WANG Haitao JIA

Research Institute of U.E.S.T.C.
University of Electronic Science and Technology of China
Chengdu, China, 610054

ABSTRACT

In the biometric recognition, face recognition is the most natural, direct method. Research on face recognition has a high theoretical significance and practical value. In this paper, firstly we use the Gabor filter to extract face image features, and then denote to further dimensionality reduction by Hopfield Neural Network. At last, for face classification, a new method based on support vector machine— No-balance Binary Tree Support Vector Machine (NBBTSVM) is proposed to decide a label in this face recognition task. SVM has excellent performance to solve binary classification but for multi-classification, it's an ongoing research. According to our experiment results, NBBTSVM could do a good performance.

Keywords- Face recognition; Gabor filter; Hopfield Neural Network; Support Vector Machine (SVM); No-balance Binary Tree Support Vector Machine (NBBTSVM).

1 INTRODUCTION

Face recognition is an important research field of image processing and pattern recognition. With the development of computer science and technology, face recognition has wide application prospects in public security, customs, banking and other industries. In the last several years, there have been significant developments in the theoretical understanding on Support Vector Machine (SVM).

The main body of this paper is organized in three parts. In the first part we introduce Gabor filter. Because of the multi-convolution property, the Fourier transform of a Gabor filter's impulse response is the convolution of the Fourier transform of harmonic function and the Fourier transform of Gaussian function[1] [2]. Due to the complexity of multi-scale Gabor filter, we specially set different numbers of frequency level; and then our work is denoted to further dimensionality reduction by Hopfield Neural Network. Hopfield structures have enjoyed mired reviews over the last 20 years. In so doing, we pave the way for next step, the last part of this paper deals with classification problems. The aim of classification using SVM is to construct a hydroplane in a high dimensional feature space as the decision surface in such a way that the margin of separation between positive and negative examples is maximized [3]. Polynomial inner-product kernel exploits information about the inner products between

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462K · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855076

data items. Then, give the details of two approaches for multi-class problems. Finally, Conclusions will be presented, the SVM has a high generalization ability to solve binary classification problem, but its extension to multi-class problems is still an ongoing research issue. To deal with the multi-classification, NBBTSVM which we proposed has improved classification accuracy to SVM and works fairly well in efficiency. Experiment results are presented in this paper as well.

2 GABOR FILTER

In the last three decades, the multi-orientation histogram for feature extraction is widely reported [4] [5]. Gabor filter is given by

$$\psi_k(z)=\frac{k^2}{\sigma}\exp\left(\frac{-k^2z^2}{2\sigma^2}\right).\left(\exp(ikz)-\exp\left(-\frac{\sigma^2}{2}\right)\right) \tag{1}$$

In this equation, $z=(x,y)^T$, $k=k_v\exp(i\phi)$, $\phi=\mu\cdot\frac{\pi}{8}$, $k_v=\frac{k_{\max}}{f^v}$, $f=\sqrt{2}$ and $\sigma=\pi$ .We set $\mu=\{0,1,2,3,4,5,6,7\}$ for 8 different orientations, $v=\{0,1,2\}$ for 3 frequency levels.

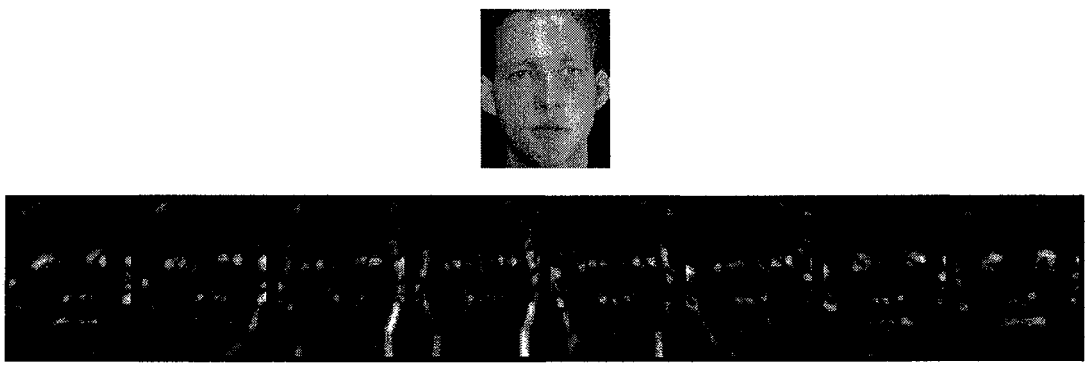


Figure 1: Images of 8 different orientations using Gabor filter

The sizes of Gabor filters are set to 9×9 , 13×13 and 17×17 respectively. Figure 1 shows the results of 8 different orientations for one face using Gabor filter.

3 HOPFIELD NERUAL NETWORK

Recently, various models of nervous system and brain which are called artificial neural networks have been extensively studied and developed. From both theoretical and applied points of view, among the most popular models in literature are Hopfield Neural Networks, which have been studied [6] [7].

In the training stage, we use Hopfield Neural Network which has only one layer of units. These units play a triple role as input, output, and processing units. The units are connected to every other unit.

The total input U_i of the i-th neuron is given by the linear combination:

$$U_i = \sum_{j=1}^{N} W_{ij} V_j + I_i \tag{2}$$

Where, N is the number of units, V_j is the output of the j-th unit, W_{ij} is a component of the connection matrix W and I_i is the bias of input to the i-th neuron. The output V_i of the i-th unit is given as:

$$V_i = sigmoid(U_i) \tag{3}$$

It is known that when the matrix W_{ij} is symmetric, we define the energy function of the Hopfield Neural Network as:

$$E = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} W_{ij} V_i V_j - \sum_{i=1}^{N} I_i V_i$$

$$+\sum_{i=1}^{N} \frac{1}{R_i} \int_0^{V_i} sigmoid^{-1}(V)\, dV \tag{4}$$

Hopfield proved that the energy function is a Liapnov function. Thus the energy function converges to a local minimum as the states of neurons converge to stable states [8].

In order to give the network the capability of the associative memory, the weights are determined as below:

$$W_{(i,j)(i',j')} = \frac{1}{N}\sum_{m=1}^{M} s^m_{(i,j)} s^m_{(i',j')} \tag{5}$$

Where, (i, j) and (i', j') are the units which are linked to each other, $s^m_{(i,j)}$ is the value of the (i, j)-th pixel of the m-th pattern. M indicates the number of patterns to be memorized.

It is clear that the weights satisfy the conditions of a Hopfield net, because the weights are symmetrical and the connection to oneself is 0. In our paper, we set N equal to 400. It attached our purpose of dimensionality reduction.

4 MUTI-CLASSIFICATION USING SVM

The SVM model is the maximal margin classifier, which works only for linearly separable data in the feature space. First, we define the linear classifier which associated with the hyperplane $(\mathbf{w}, b)$. The vector $\mathbf{W}$ is a normal vector: it is perpendicular to the hyperplane, which could achieve a functional margin of 1 on the positive

point $\mathbf{x}^+$ and the negative point $\mathbf{x}^-$. These hyperplanes can be described by the equations,

$$\begin{aligned} \mathbf{w}\cdot\mathbf{x}+b&=1 \\ \mathbf{w}\cdot\mathbf{x}+b&=-1 \end{aligned} \tag{6}$$

The geometric margin γ could be calculated as follows,

$$\begin{aligned} \gamma &= \frac{1}{2}\left(\left\langle \frac{\mathbf{w}}{\|\mathbf{w}\|_2}\cdot\mathbf{x}^+\right\rangle - \left\langle \frac{\mathbf{w}}{\|\mathbf{w}\|_2}\cdot\mathbf{x}^-\right\rangle\right) \\ &= \frac{1}{2\|\mathbf{w}\|_2}\left(\langle \mathbf{w}\cdot\mathbf{x}^+\rangle - \langle \mathbf{w}\cdot\mathbf{x}^-\rangle\right) \\ &= \frac{1}{\|\mathbf{w}\|_2} \end{aligned} \tag{7}$$

For finding the distance between these two hyperplanes, we have to minimize $\|\mathbf{w}\|$. Consider a training set of data $(x_i, y_i)_{i=1}^{\eta}$ where $x_i \in \Re^n$ are input feature vectors and $y_i \in \{-1,+1\}$ are the set of corresponding labels. The optimal hyperplane could be defined as,

$$minimise_{w,b}\langle \mathbf{w}\cdot\mathbf{w}\rangle$$

$$subject\ \text{to}\ y_i\left(\langle \mathbf{w}\cdot\mathbf{x}_i\rangle + b\right) \geq 1, \quad i=1,...,\eta \tag{8}$$

Furthermore, the optimal hyperplane can be expressed on dual representation as follows,

$$f(\mathbf{x},\alpha,b) = \sum_{i=1}^{\eta} y_i\alpha_i\langle \mathbf{x}\cdot\mathbf{x}_i\rangle + b \tag{9}$$

Where α is the solution of quadratic optimization problem.

The original optimal hyperplane algorithm proposed by Vladimir Vapnik in 1963 was a linear classifier. Considering the non-linear case might exist, the linear transform $\phi(\mathbf{x})$ could be used.

$$f(\mathbf{x},\alpha,b) = \sum_{i=1}^{\eta} y_i\alpha_i\phi(\mathbf{x}_i)^T\phi(\mathbf{x}) + b \tag{10}$$

However, in 1992, Bernhard Boser, Isabelle Guyon and Vapnik suggested a way to create non-linear classifiers by applying the kernel trick to maximum-margin hyperplanes [9] [10]. To avoid samples mapping into high dimensional space by $\phi(\mathbf{x})$.we define the kernel function as follows,

$$\mathbf{k}(\mathbf{x},\mathbf{z}) = \langle \phi(\mathbf{x}) \cdot \phi(z) \rangle \tag{11}$$

Recently, it is reported that the normalized Polynomial kernel does good performance. In this paper, we select this kernel as the kernel function. It is defined as,

$$\mathbf{k}(\mathbf{x},\mathbf{z}) = \frac{\phi(\mathbf{x})^T \phi(\mathbf{z})}{\|\phi(\mathbf{x})\| \cdot \|\phi(\mathbf{z})\|}$$

$$= \frac{(\mathbf{x}^T \cdot \mathbf{z} + c)^d}{\sqrt{(\mathbf{x}^T \cdot \mathbf{x} + c)^d (\mathbf{z}^T \cdot \mathbf{z} + c)^d}} \tag{12}$$

The classification is a multi-class issue. Two approaches are widely talked which are One-Against-All method and One-Against-One method. Because we want to compare the veracity between the binary classification and the multi-classification, One-Against-All method and NBBTSVM method will be used for the following experiments, we construct k SVM models for One-Against-All method. Figure 2 shows One-Against-All method phase. The final output is the category which corresponds with the highest output value.

$$\textit{the class of } x = \textit{argmax}_{i=1,\ldots k} \left(\langle \phi(\mathbf{x}) \cdot \mathbf{w}^i \rangle + b^i \right) \tag{13}$$

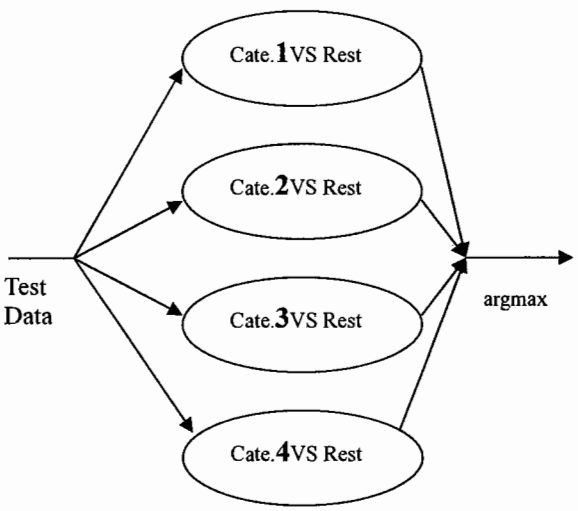


Figure 2: One-Against-All SVM classification phase

5 NBBTSVM

The NBBTSVM classification phase can be shown in Figure 3. For a training set of data $(x_i)_{i=1}^{\eta}$ where $x_i \in \Re^n$ are the input feature vectors,

1. Compute the separability measure of each class $\delta_i = \frac{\|c_j - c_i\|}{\delta_j + \delta_i}$, where c_i , c_j is the cluster center of class i , j individually , δ_j is the cluster variance of class j .
2. Select the minimum separability measure of two classes among them, the two classes can be considered as a new big class X_1 . The number of category should minus 1.
3. Repeat Step 1 until only two classes left. So these two classes contribute to a binary SVM.
4. Get the multi-class classifier by constructing modes from the root node of binary tree in sequence.

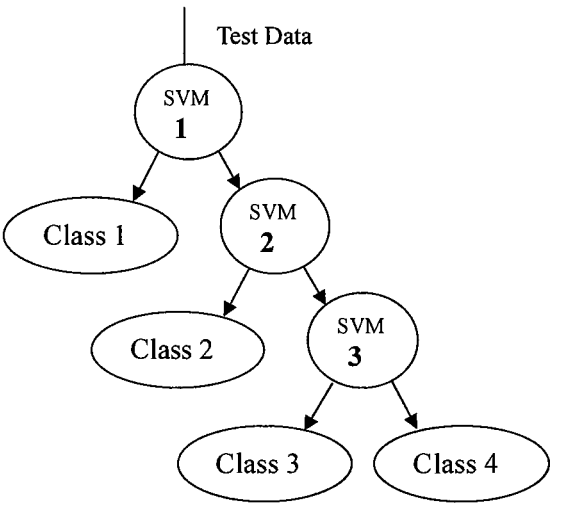


Figure 3: NBBTSVM classification phase

6 EXPERIMENTS

The goal of face recognition is to determine the identity of an individual based on a still image or video sequence of face.

In this experiment, we prepared ORL face database. The examples of face database are shown in Figure 4. For each person totally 10 images are divided into two sets averagely as training and testing images.

We use two types of classifiers, which are respectively One-Against-All SVM and NBBTSVM. Classification accuracies are shown in Table 1.

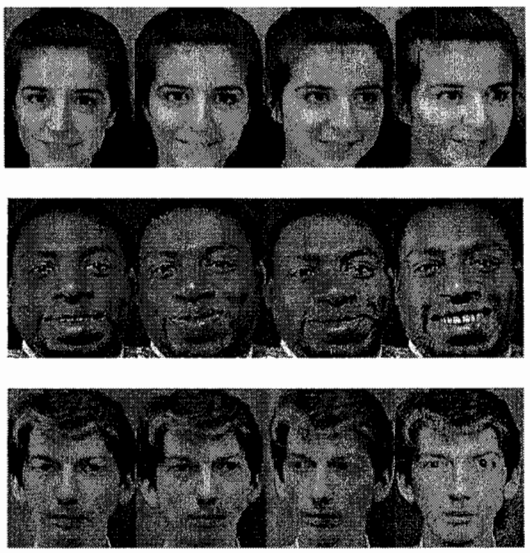


Figure 4 The examples of face database

Table 1: recognition accuracy when c=1, d=3

	Accuracy	Time(s)
OAASVM	75.0%	1.73
NBBTSVM	96.5%	1.32

In the case of One-Against-All SVM applied, the best average accuracy is 75.00% when kernel parameter c=1, d=3. The best average accuracy for NBBTSVM is improved to 96.5% when c=1, d=3.

So we could easily observe that NBBTSVM which we proposed performs much better than OAASVM. NBBTSVM's running time is much less than the OAASVM

7 CONCLUSION

In this paper, we compared the classification ability and efficiency of One-Against-All SVM, and NBBTSVM respectively based on Gabor filter and Hopfield Neural Network and obtained the conclusion, the NBBTSVM could perform much better than One-Against-All SVM that proved SVM goes well especially in binary classification.

REFERENCES

[1] Movellan, Javier R, "Tutorial on Gabor Filters", (2008).

[2] L. Vincent, "Texture Segmentation Using Gabor Filters", (2008).

[3] W. Huilan, L. Guodong, P. Zhaobang, "Multi-Class Image Recognition Based on Relevance Vector Machine", In: Intelligent Systems and Applications, 1 – 4 (2009).

[4] K. Grauman, T. Darrell, "Discriminative classification with sets of image features", In: Proc. International Conference on Computer Vision, 1458-1465 (2005).
[5] S. Lazebnik, C. Schmid, J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories", In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2169-2178 (2006).
[6] Run-Nian Ma, Peng Chu, Sheng-Rui Zhang," Stability Conditions for Discrete Delayed Hopfield Neural Networks", In: Third International Conference on Natural Computation, 56-59 (2007).
[7] Teijiro Isokawa, Haruhiko Nishimura, Naotake Kamiura, Nobuyuki Matsui, "Fundamental Properties of Quaternionic Hopfield Neural Networks", In: International Joint Conference on Neural Networks, 610-615 (2006).
[8] W. Li, M. Nasrabadi, "Object recognition based on graph matching implemented by a hopfield-style neural network." In: J. Conf. Neural Networks. 287-290 (2003).
[9] M. Aizerman, E. Braverman, L. Rozonoer, "Theoretical foundations of the potential function method in pattern recognition learning", In: Automation and Remote Control 25, 821-837 (1964).
[10] B. E. Boser, I. M. Guyon, V. N. Vapnik, "A training algorithm for optimal margin classifiers", In: D. Haussler, editor, 5th Annual ACM Workshop on COLT, Pittsburgh, PA, 144-152 (1992).

Region of interest based robust watermarking scheme for adaptation in small displays

V.Sapthagirivasan[1], Kishore Mohan K B[2], Vemula Krishna Manohar[3]
[1,2]Research Scholar, [3]Asst.Professor , Dept. of Biomedical Engineering, SRM University, India.
sapthagiri.ece@gmail.com[1],kishorekbmtech@yahoo.co.in[2], v.krishnamanohar@gmail.com[3]

ABSTRACT

Now-a-days Multimedia data can be easily replicated and the copyright is not legally protected. Cryptography does not allow the use of digital data in its original form and once the data is decrypted, it is no longer protected. Here we have proposed a new double protected digital image watermarking algorithm, which can embed the watermark image blocks into the adjacent regions of the host image itself based on their blocks similarity coefficient which is robust to various noise effects like Poisson noise, Gaussian noise, Random noise and thereby provide double security from various noises and hackers. As instrumentation application requires a much accurate data, the watermark image which is to be extracted back from the watermarked image must be immune to various noise effects. Our results provide better extracted image compared to the present/existing techniques and in addition we have done resizing the same for various displays. Adaptive resizing for various size displays is being experimented wherein we crop the required information in a frame, zoom it for a large display or resize for a small display using a threshold value and in either cases background is not given much importance but it is only the fore-sight object which gains importance which will surely be helpful in performing surgeries.

Keywords: ROI, PSNR, MSE, Adaptation, re-sizing

1. INTRODUCTION

Imaging technology and applications have advanced greatly as multimedia information is acquired, represented, stored, and distributed in digital format. Digital processing and transmission has major advantages, but problems can arise due to the ease with which digital information is reproduced and distributed. Traditional methods do not provide an effective way of securing intellectual property rights, as the original data remains unprotected once the copyright control mechanisms are surpassed.

Watermarking basic principle is to embed a weak signal in the host data without altering it significantly, this signal (watermark) is imperceptible usually, but it can be detected or extracted by specific algorithm, even after some manipulations to the watermarked data. Security applications often use the hiding of information in digital data in order to protect it or to monitor its usage. The watermark is a digital signal, usually containing information about the data origin, destination and owner, as well as additional information concerning transaction dates, serial numbers etc., which can be useful when tracking illegal use of the data. Although the modification of the original content is strictly controlled to be so slight that it is commonly imperceptible to human eyes, most of the watermarking algorithms will cause a certain amount of permanent loss of content fidelity during the embedding process. The quality loss is usually proportional to the amount of the embedded watermark. The reliability of the data is an important issue for medical images, because any manipulation or quality compromise could result in serious misdiagnosis of the patient's disease [1-2]. Thereby, in the most important part of the medical images, any slight modification is not allowed. First we identify the required information of interest in the image. We crop it out and at last we watermark the cropped out image back in the same image but in the background. The advantage thus obtained is even if hawkers cause any damage to wanted information in watermarked image by retrieving the embedded image and by superimposing the same we would be able to get back the same image without much loss. The use of our robust enhanced digital watermarking algorithm the effects due to noise are handled well than the previous algorithms. Now let us see them in detail in further sessions.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie.
Proc. of SPIE Vol. 7546, 75462L · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856295

2. PREVIOUS WORK

The advent of digital age has destroyed the security and protection of digital multimedia information. To protect this numerous schemes have been proposed in the line of Data Hiding Techniques. Some of them on what we have concentrated are discussed hereunder. Computer scientists Samir Kumar Bandyopadhyay, Debnath Bhattacharyya and Anindya Jyoti Pal examine [6] digital watermarking, the process that embeds data called a watermark into an object such that the watermark can be detected and extracted later to make an assertion about the object. Watermarking is either "visible" or "invisible". They said that the key techniques involve using secure functions to generate and embed an image mark that is more detectable, verifiable, and secure than existing protection and detection techniques. Min Wu and Bede Liu, June, 2003, proposed [7] a new method to embed data in binary images, including scanned text, figures, and signatures. The method manipulates "flippable" pixels to enforce specific blockbased relationship in order to embed a significant amount of data without causing noticeable artifacts. Shuffling is applied before embedding to equalize the uneven embedding capacity from region to region. The hidden data can be extracted without using the original image, and can also be accurately extracted after high quality printing and scanning with the help of a few registration marks. F. Bartolini, A. Tefas, M. Barni and I. Pitas discussed the problem of authenticating video surveillance image. After an introduction motivating the need for a watermarking-based authentication of VS (video surveillance) sequences, a brief survey of the main watermarking-based authentication techniques is presented and the requirements that an authentication algorithm should satisfy for VS applications are discussed. A novel algorithm, which is suitable for VS visual data authentication, has proposed [8]. Huajian Liu, Martin Steinebach 2006 proposed [11] the other alternative strategy is watermarking with region of interest (ROI). Some watermarking algorithms have been proposed with the concept of region of interest (ROI). But these algorithms are either limited to specific image format, e.g. JPEG2000 [10], or they need the precise location information of the ROI in order to extract the watermark. These requirements decrease the practicability and portability of the watermarking schemes. In the practical applications, the ROI information may be unavailable at the side of the watermark detector.

3. PROPOSED WORK

In this paper a new image watermarking blind algorithm aiming at ROI information hiding of the important images is proposed. Firstly, we segment and extract the ROI information block in original image, and then make gray value supplement in that region to generate the disguised cover image. The ROI watermarking information encrypted by chaos is embedded [9] into the disguised cover image in lifting wavelet domain, and then the watermarked image is obtained. Experiments show that the algorithm is of huge hiding information capacity, high imperceptibility and strong robustness, and blind detecting is available. Furthermore, it is valuable in practice. The figure (Fig 3.1) shows the block diagram of the proposed method.

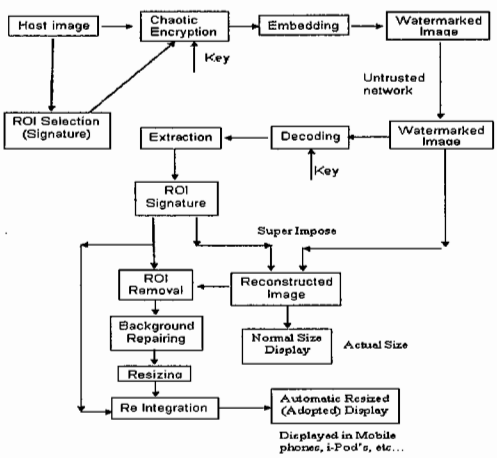


Figure 3.2: Block Diagram of ROI based robust watermarking and application for adaptation

3.1 ROI Selection

Before performing the watermark embedding process, one or multiple preferred regions of interest are specified on the image. A ROI can be either a predefined kind of object or an area with arbitrary shape that is interactively specified by the user. All the pixels or the corresponding transform coefficients inside the selected region can not be modified during the watermark embedding process. The locations of the ROI(s) are input to the watermark embedder as parameters. [3]

As the ROI is defined unchangeable, no watermark can be embedded in the ROI. The authentication of the ROI(s) must be achieved by the embedded watermarks in the other parts of the image. Therefore, how the ROI(s) and other image parts are linked together becomes an essential issue for the authentication process.

Before the watermark embedding in the corresponding domain where the watermark will be embedded, by a random permutation process we obtain a random mapping of all the image locations. The units that are used to embed the watermark, e.g. single pixel or transform coefficient or image block, are randomly permuted under the control of a secret key. All the locations inside ROI(s) are also randomly distributed into different groups, mutually referred to other locations in the same group. During the watermark detection and image authentication procedure, the reference relationship is recovered by applying the same random permutation using the same secret key. Therefore, no location information of the ROI(s) is needed to store and transfer.

3.2 Chaotic Encryption

Because the chaotic system is non-periodic as well as sensitive to initial conditions, it could be used as an encryption [9] method. In this paper Logistic mapping is adopted to generate chaotic sequence, which could be obtained by changing the system parameters and initial values. Logistic mapping express as:

$$X_{k+1}=\mu X_k(1-X_k) \qquad (1)$$

Where $X_k \in (0,1)$, $\mu \in [0,4]$, when $\mu=4$ the system is characteristic as extremely chaotic. Chaotic encryption could eliminate the data correlation, it could enhance the security of watermarking signal and decrease the influence caused by error rate. If the key is unknown, it is unavailable to recover the original image even.

3.3 Watermark Embedding

Primitive version of our algorithm is a simple digital image watermarking algorithm, which can embed the watermark image blocks into the host image based on their blocks similarity coefficient [4]. Both of watermark image and host image are divided into image blocks of size 8-by-8, and similarity coefficients between watermark image blocks and host image blocks are computed. Each watermark image block is embedded into the host image location with best similarity, namely, with the largest block similarity coefficient. Since the embedded watermark image blocks and their corresponding host image blocks are similar to some extent, the whole embedded watermark image is well imperceptible. The only change is that both host and watermark image are from the same image and this ensures a very high block similarity coefficients.
Analyze the results from the following formulas:

$$MSE=\frac{1}{N}\sum_{i=1}^{N}\left(I_i-\overline{I}_i\right)^2 \qquad (2)$$

$$MAE=\frac{1}{N}\sum_{i=1}^{N}\left|I_i-\overline{I}_i\right| \qquad (3)$$

$$PSNR=10\log_{10}\frac{P^2}{MSE} \qquad (4)$$

Where, I – Original Image Signal, $\bar{I}$ – Modified Image Signal, N – Total number of pixels in Original image, i – Pixel index value, P – Maximum gray level (peak) value Generally P will be calculated by $P = 2^n - 1$, n – Number of bits per pixel per plane

3.3.1. Image Embedding Section

The following is the algorithm for watermark embedding:

1. First the watermark and host images are divided into blocks of required size
2. Obtain the following matrices from the blocks [4]
   Row matrix, Column matrix and Interrelation matrix
3. Record the number of the watermark image embedded into the host image from top-left to bottom right.
4. Compute similarity coefficients between the watermark image block and the host image block
5. Embed the watermark image block into host image block such that host image block has the best similarity with watermark image block
6. Repeat the steps 4 and 5 until the entire image has been embedded.

3.3.2. Image Extraction Section

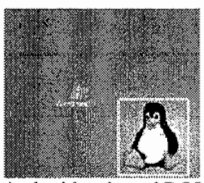
The extraction algorithm is the reverse of embedding algorithm. The watermark image can be recovered from watermarked image since watermark image blocks can be gained according to the row and column matrices. With the knowledge of the number of the watermark image embedded into the host image and using the row and column matrices recover the block of original watermark image block.

3.3.3. Super Impose and Display Adaptation

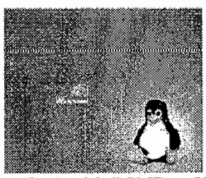
The extracted ROI information's are superimposed on the watermarked image to get the original recovered image from any ROI attacks (if any). Then the ROI regions are removed from super imposed image to get a hallow backgrounded image. From the resultant image background repairing [5] has to be made. Then resize the background image based on the display (devices) chosen by the user. Then reintegrate the resized background image with ROI region.

4. EXPERIMENTAL RESULTS

The various images, the watermarked images and extracted images with corresponding PSNR values are displayed in the following figures.

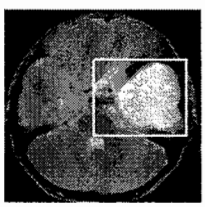


(a) Original with selected ROI (108 X 85)



(b) Watermarked with PSNR = 52.82 dB

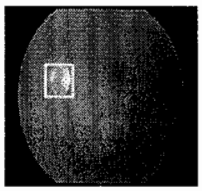
Figure 4.1. Bird Image (256 X 256)

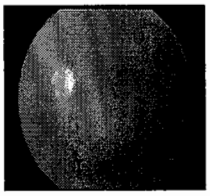


(c) Original with selected ROI (95 X 118)

(d) Watermarked with PSNR = 51.46 dB

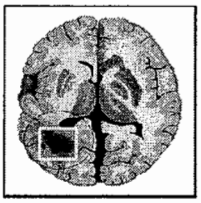
Figure 4.2. Brain MRI (256 X 256)

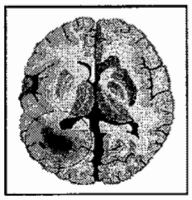




(e) Original with selected ROI (51 X 32) (f) Watermarked with PSNR = 57.49 dB

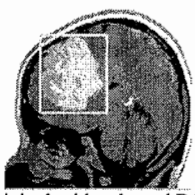
Figure 4.3. Eye (256 X 256)

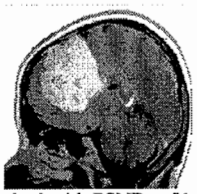




(g) Original with selected ROI (43 X 62) (h) Watermarked with PSNR = 57.38 dB

Figure 4.4. Brain CT (256 X 256)





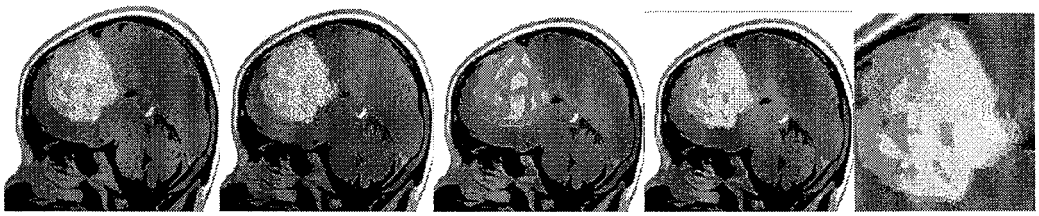
(g) Original with selected ROI (107 X 91) (h) Watermarked with PSNR = 51.09 dB

Figure 4.5. Head scan (256 X 256)

Figures 4.1, 4.2, 4.3, 4.4 and 4.5 show the original image with selected ROI and the watermarked images with corresponding PSNR values. It is significant to note that in all the cases we are able to obtain a PSNR value greater than 50 dB. The summary of the results we obtained are depicted in the tabulation below

Table 4.1 PSNR of different images

S.No	Image (256 X 256)	Image Size	Region of Interest	Mean Absolute Error (MAE)	Mean Square Error (MSE)	PSNR
	(bmp)	(kB)	(kB)			(dB)
1	Bird	192	108 X 85	0.17	0.34	52.82
2	Brain MRI	192	95 X 118	0.23	0.46	51.46
3	Eye	192	51 X 32	0.05	0.12	57.49
4	Brain CT	192	43 X 62	0.05	0.12	57.38
5	Head Scan	192	107 X 91	0.22	0.51	51.09
Average		**192**	**6.72**	**0.14**	**0.31**	**54.05**

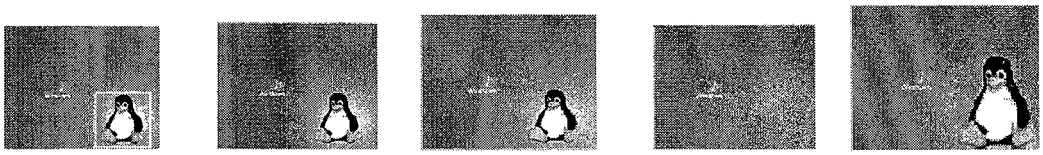


(i) (j) (k) (l) (m)

Figure 4.6. ROI Attacks

(i) ROI region attacked by Poisson noise, (j) ROI region attacked by Gaussian noise
(k) ROI region attacked by Random noise and (l) Recovered Original Image from varies ROI Attacks (super imposed (m) on (i, j, k, l)), (m) Extracted ROI from attacked images

In the Figure 4.6 various ROI attacks like Poisson noise, Gaussian noise, Random noise have been introduced to check with our algorithm's robustness and the recovered images are also shown which have proven that our algorithm though needs improvement is still better than the present algorithms.



(n) (o) (p) (q) (r)

Figure 4.7.Adapted Reconstructed image

(n) Original (256 X 256) with selected ROI (108 X 85), **(o)** Directly Recovered Original Image (256 X 256) from varies ROI Attacks, **(p)** Recovered Original Image from varies ROI Attacks and Normal Resized adapted image (160 X 180),**(q)** Resized Background repaired image , **(r)** Adapted (small) image (q+m) (160 X 180)

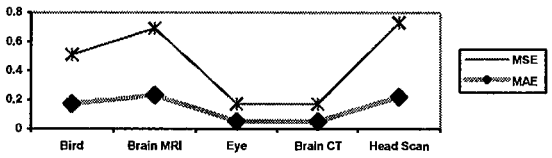


Figure 4.8 Images Vs MSE and MAE Values

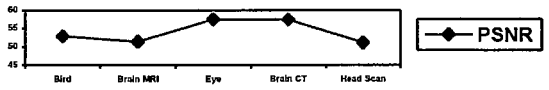


Figure 4.9 Images Vs PSNR Values

Table 4.2 Average Execution Time

Stage	Embedding	Actual Image Reconstruction	Resized Adapted Image Reconstruction
Execution Time	32 sec	18 sec	57 sec

Figure 4.8, Figure 4.9 and Table 4.1 shows the resultant parameters such as MSE, MAE and PSNR for various images. Table 4.2 shows the time consumption for reconstruction by actual and adopted methods. The Resized adapted constructed image depends on the system configuration too.

5. CONCLUSION

Thus the proposed method is robust against various ROI attacks and also avoids the need for a separate image for watermarking. Thus better block similarity coefficient is obtained in this case than using any other image. Also the results show the robustness of our algorithm to various noise attacks and we are sure that it would definitely be of great help in bio-medical instrumentation and imaging applications. Finally the automatic adaptation method also given good results and proved the efficiency of our method.

REFERENCES

[1] B. Planitz, A. Maeder, "Medical Image Watermarking: A Study on Image Degradation", in Proc. Australian Pattern Recognition Society Workshop on Digital Image Computing, WDIC 2005, Brisbane, Australia, 2005.

[2] A. Giakoumaki, S. Pavlopoulos, D.Koutsouris, "Multiple Image Watermarking Applied to Health Information Management", IEEE Transactions on Information Technology in Biomedicine, vol. 10, Issue 4, pp. 722 – 732, Oct. 2006.

[3] Huajian liu, Martin steinebach "Non-ubiquitous watermarking for image authentication by region of interest masking ",2007

[4] Haiyong Liao, Ruisong Ye`` A Novel Digital Image Watermarking Approach Based on Image Blocks Similarity", 2008 Congress on Image and Signal Processing, pp.626-630, 2008

[5] Wen-Huang Cheng, Chia-Wei Wang, Ja-Ling Wu "Video Adaptation for Small Display Based on Content Recomposition ",2007 Circuits and Systems for Video Technology, IEEE Transactions on volume 17, Issue 1, Jan. 2007 pp:43 – 58

[6] Samir Kumar Bandyopadhyay, Debnath Bhattacharyya and Anindya Jyoti Pal, Secure Delivery of Handwritten Signature", ACM Ubiquity, Vol. 7 Issue. 40 October 16, 2006.

[7] Min Wu, Member, IEEE, and Bede Liu, Fellow, IEEE, "Data Hiding in Binary Image for Authentication and Annotation", IEEE Trans. Image Processing, vol. 12, pp. 696–705, June 2003.

[8] F. Bartolini, A. Tefas, M. Barni and I. Pitas, "Image Authentication Techniques for Surveillance Applications", IEEE Proceedings, Vol. 89, No. 10, October 2001.

[9] Dashun Que, Li Zhang, Ling Lu, Liucheng Shi , " A ROI Image Watermarking Algorithm Based on Lifting Wavelet Transform ", ICSP Proceedings, 2006.

[10] W. Lie, T. Hsu, G. Lin, W. Ho, "Fragile Watermarking for JPEG-2000 Images", 16th IPPR Conf. on Computer Vision, Graphics and Image Processing (CVGIP 2003), pp. 823 – 826, Aug, 2003.

[11] Huajian Liu, Martin Steinebach "Non-Ubiquitous Watermarking For Image Authentication By Region Of Interest Masking" National Conference, 2006.

Glomeruli Extraction by Canny Operator with a Feedback Strategy

Jun Zhang[a*, b], Jinglu Hu[a] and Hong Zhu[c]

[a] Graduate School of Information, Production and Systems, Waseda University, Japan;
[b] Research Fellow of the Japan Society for the Promotion of Science, Japan;
[c] Xi'an University of Technology, China.

ABSTRACT

This paper proposes an edge detection method by Canny operator with a feedback strategy for glomeruli extraction. As we know, the effect of the Canny operator is determined by three parameters: high threshold, low threshold and standard deviation. To obtain the appropriate parameters for each image, Otsu method is used to set high and low thresholds of Canny operator firstly. And then, to select the optimal standard deviation, a feedback strategy is developed. After parameter selection, Canny operator is applied to our renal biopsy images and experimental results show that some samples can achieve successful extraction and the others result in the discontinuous edges of glomeruli. In the case of the latter, the endpoints in an image should be located and connected to form a whole edge of glomerulus. The experiments have produced the promising results for our samples.

Keywords: Glomeruli extraction, edge detection, Canny operator, parameter selection

1. INTRODUCTION

A renal biopsy is a procedure in which a sample of kidney (also called renal) tissue is obtained. Microscopic examination of the tissue can provide information needed to diagnose, monitor or treat a renal disorder. With the development of computer technology, it is possible to develop a computer aided diagnosis system for analyzing microscopic renal biopsy images automatically. In the system, glomeruli extraction is a key step for diagnosis. However, because the microscopic images always have complex shapes, poor image contrast and human tissue images always have lots of noises, it is not easy to extract glomuruli.

To solve this problem, many methods have been proposed. Unlike the round or elliptic cells, the glomerulus has irregular shapes. Ellipse fitting [1] which is commonly used in medical image processing cannot deal with all situations. Hirai *et al.* proposed to extract glomeruli using a local and an elliptical global method [2] and they improved their method in [3]. However it is still not good enough to deal with various situations. Because of the coloring uniformity of slice images, extraction by color feature [4] is difficult. Zhu *et al.* proposed to use a non-linear difference filter and region estimation model to extract glomuruli [5], [6]. But the estimation heavily depends on the low-noise. Another available method to solve this problem is using genetic algorithm (GA) to do edge fitting [7]. The heuristic nature of GA makes it capable to search for an optimal B-spline curve to represent the shape of the glomerulus with high accuracy. However, this method mainly flaws on that the searching space of GA is too large. Because there is not only glomerulus but also numerous noises in the image, edge detection method [8] cannot exclude the effect of noises, as well as active contour method [9].

Taking account the disadvantages mentioned above, this paper proposes an extraction method by Canny edge detection with a feedback strategy. Firstly, Otsu method is used to set the high and low threshold for Canny operator. Then a feedback strategy is applied to obtain the suitable standard deviation. As a result, we can get the appropriate parameters of Canny operator for each image. After above processing, some samples can achieve successful extraction results. However the others result in the discontinuous edges of glomeruli. In this case, the endpoints in an image should be located and connected to form a whole edge of glomerulus.

* Jun Zhang: E-mail: j.zhang@akane.waseda.jp
Jinglu Hu: E-mail: jinglu@waseda.jp
Hong Zhu: E-mail: zhuhong@xaut.edu.cn

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462M · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853347

The rest of this paper is organized as follows. Section 2 describes how to select the three parameters for Canny operator and gives the steps of edge detection. Section 3 gives some experimental results to demonstrate the effectiveness of the proposed method. Finally section 4 summarizes the presented research work.

2. EDGE DETECTION BY CANNY OPERATOR

The Canny operator [10] was designed to be an optimal edge detector. It works in a multi-stage process. First of all the image is smoothed by Gaussian convolution. Then a simple 2-D first derivative operator is applied to the smoothed image to highlight regions of the image with high first spatial derivatives. Edges give rise to ridges in the gradient magnitude image. The algorithm then tracks along the top of these ridges and sets to zero all pixels that are not actually on the ridge top so as to give a thin line in the output, a process known as *non-maximal suppression*.

Canny edge detector has two kinds of parameters. First is σ, the size of the Gaussian filter directly affects the results of the Canny algorithm. Smaller σ causes less blurring, and allow detection of small, sharp lines. On the contrary, a larger σ causes more blurring, smearing out the value of a given pixel over a larger area of the image. The second kind of parameters is two thresholds: the use of two thresholds with hysteresis allows more flexibility than in a single-threshold approach, but general problems of thresholding approaches still apply. A threshold set too high can miss important information. On the other hand, a threshold set too low will falsely identify irrelevant information (such as noise) as important. So how to select the three parameters is an issue.

2.1. High and Low Threshold Selection by Otsu Method

To adaptively determine the suitable thresholds, some methods have already been proposed. In [11], authors proposed to find H_{max}, the gradient value which has the largest number of pixels in the gradient histogram, and calculate the variance of whole gradients related to H_{max}. Based on the variance, they can set the high and low threshold for canny operator. However, for our renal biopsy images, the gradient histogram rolls gently and has no obvious peak. It means that the method proposed in [11] is not advisable to solve our problem. Similar issue is also in other existing methods [12], [13]. So in this paper, an Otsu method is used to set the high and low threshold. Firstly, the gradient histogram is calculated. Then Otsu method is applied to find a threshold which can distinguish edges from background by *maximizing inter-class variance*. And we set the threshold as the high threshold T_{high} for Canny operator. After this, the histogram is divided into two parts, edge and background. Now the Otsu method is applied again to the part histogram of the background and the low threshold T_{low} can be obtained.

Although the adaptive thresholds can improve the edge detection result, lots of noises are still happening when detecting edges of glomeruli. So adaptive selection of standard deviation of Canny operator becomes important.

2.2. Standard Deviation Selection by a Feedback Strategy

For our renal biopsy images, there are not only glomeruli but also other tissues in an image. It is impossible to only detect the edges of targets when we use any edge detection method. Fortunately, the targets we want are always larger than other tissues. It means that their corresponding extracted edges will be longer than other ones. Benefiting from this, we can define a feedback strategy to select the standard deviation. For the purpose of glomeruli extraction, the ideal edge detection result after denoising should be a closed curve with some length (contour of glomerulus). So we can draw the following rules:

a. The number of detected edges after denoising should be less.
b. The edge of glomeruli should be the longest one.

Let N be the number of detected edges after denoising and L be the length of edges. $\rho = L/N$ is defined as a feedback value. As we know, different standard deviation will result in the different edge detection result and we must set σ to a reasonable value. ρ is used to adjust σ until it reaches its maximum.

To realize the feedback strategy, the influence of unexpected noises must be eliminated. In this paper, we propose the following procedure to denoise. Firstly, because the glomeruli in images are always larger than other tissues, the extracted edges of them will be longer than other ones. Labeling method can be used to preserve some longest edges in the images. However, this operation may preserve some large noises. And the feedback strategy may suffer from these noises. Through mass observation, we found that these large noises are always formed by interlaced and sinuous edges. To erase interlaced edges, cross points need to be found out and deleted from images. Then the long noises will be

separated into several short segments. After this, labeling is used again to erase them all. Dilation and thinning are used to remove the sinuous edges. These operations provide a powerful support for calculating the feedback value correctly.

2.3. Steps of edge detection

According to the analysis mentioned above, steps of edge detection for our samples are listed as follows:

Step 1: Select high and low thresholds for canny operator by applying Otsu method to the gradient histogram after non-maximal suppression and set the initial standard deviation $\sigma = 0.5$.

Step 2: Detect edges by Canny operator with selected parameters.

Step 3: Get the edges of glomerulus by a series of denoising operations and calculate ρ_0.

Step 4: Try to increase σ and repeat Step 2 and 3. And then the new ρ_1 will be obtained. If $\rho_1 > \rho_0$, $\sigma = \sigma + 0.3$, continue to increase σ till $\sigma = \sigma_{max} = 5$; if $\rho_1 < \rho_0$, $\sigma = \sigma - 0.3$, end.

Step 5: When ρ reaches its maximum ρ^*, the corresponding images is regarded as final edge detection output.

3. EXPERIMENTAL RESULTS

To examine the performances of our proposed method, about 80 samples were used. We will give some of experimental results in this section. All experiments were implemented on a personal computer with CPU of 1.83GHz using MATLAB7.0 programming language.

Experiment 1: As previously mentioned, to guarantee correctness of the feedback value calculation, noises must be erased clearly. The interlaced edges will be the large noises after labeling. To erase the noises, the cross points need to be located and deleted from edge images. Then the large noises are separated into several short segments and labeling method is applied again to remove them. Fig. 1 shows this procedure.

Experiment 2: To obtain the edges of target and achieve noise suppression at the same time, we propose to use Otsu method to set the high and low thresholds, and use a feedback strategy to determine the standard deviation for Canny operator. In this experiment, we give an example for standard deviation selection on the premise that the high and low threshold of Canny operator are already determined by Otsu method. The range of σ is [0.5, 5.0]. Fig. 2 shows different edge detection results under different standard deviation from 1.1 to 3.3 with step 0.3. To select the best one, ρ is used to optimize σ. We can see that when $\sigma = 2.3$, ρ reaches its maximum $\rho^* = 750$. shown in Fig. 2(e). So for this sample, standard deviation is determined as 2.3 and the whole edge of target can be detected successfully.

Experiment 3: In this experiment, the proposed method is applied to different samples, shown in Fig. 3(a), to prove its effectiveness. Because the three parameters of Canny operator are determined adaptively, the images with different brightness and contrast can achieve successful extraction, shown as Fig. 3(b). This experiment proved that the proposed method can solve the problem caused by constant parameters of Canny operator.

Experiment 4: Although we proposed to select parameters for Canny operator adaptively and we also proved its performance, some samples such as images in Fig. 4(a) still result in the discontinuous edges, shown in Fig. 4(b). In this case, the endpoints, marked by '✳' in Fig. 4(c), should be found out and connected with each other to form a whole edge. In this paper, we simply connect the endpoints directly as a line. The connected edges are shown in Fig. 4(d). The results are acceptable.

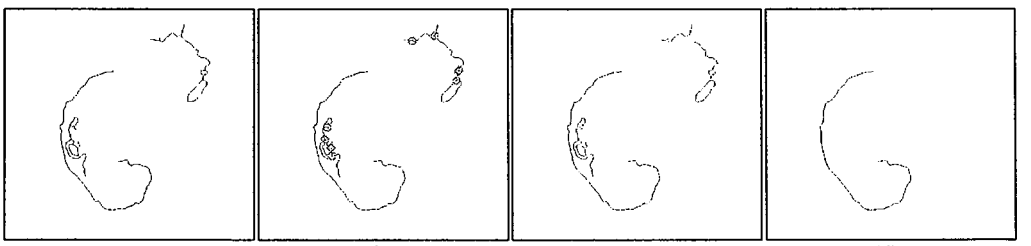
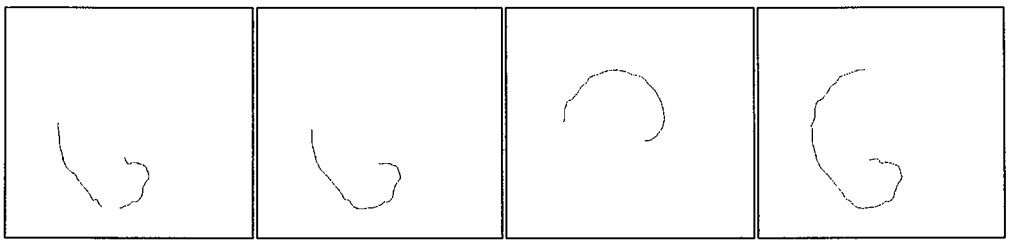
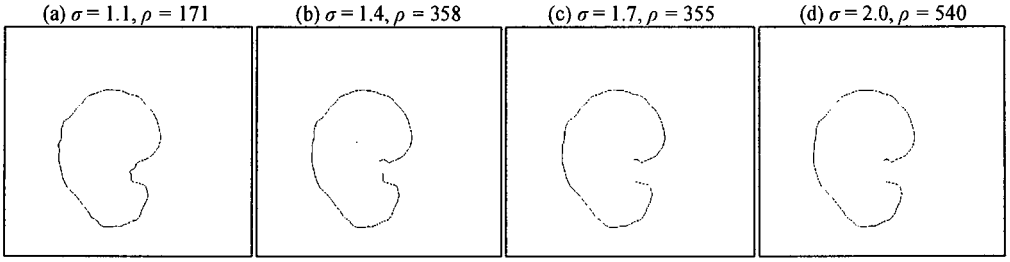


Figure 1. Large noises removal. (a) is the image after shorter edges removal by labeling; the cross points are marked in (b); (c) shows the image after cross points removal and the large noise is separated into several segments; (d) is the clear image by labeling to remove noises again.

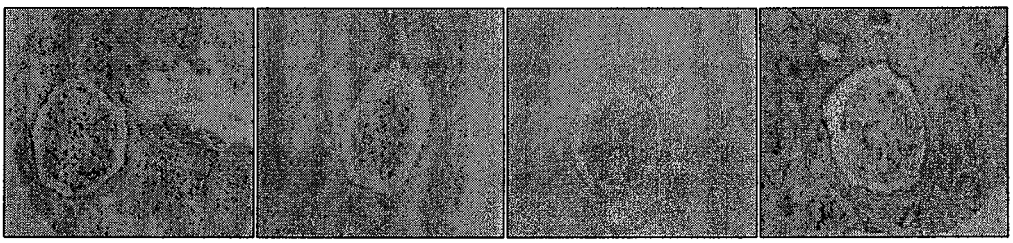


(a) $\sigma = 1.1, \rho = 171$ (b) $\sigma = 1.4, \rho = 358$ (c) $\sigma = 1.7, \rho = 355$ (d) $\sigma = 2.0, \rho = 540$

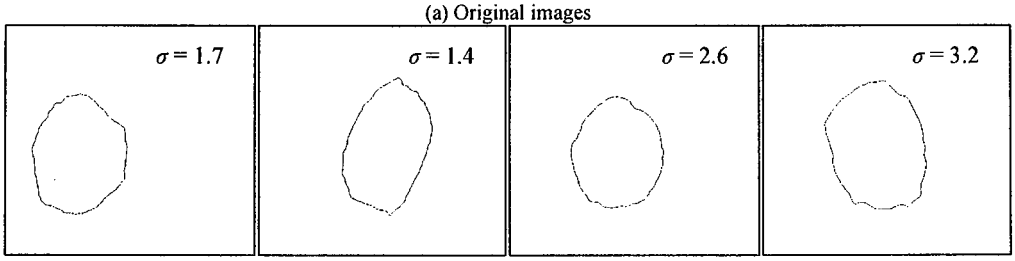


(e) $\sigma^* = 2.3, \rho^* = 750$ (f) $\sigma = 2.6, \rho = 745$ (g) $\sigma = 2.9, \rho = 721$ (h) $\sigma = 3.2, \rho = 719$

Figure 2. Edge detection by proposed method.

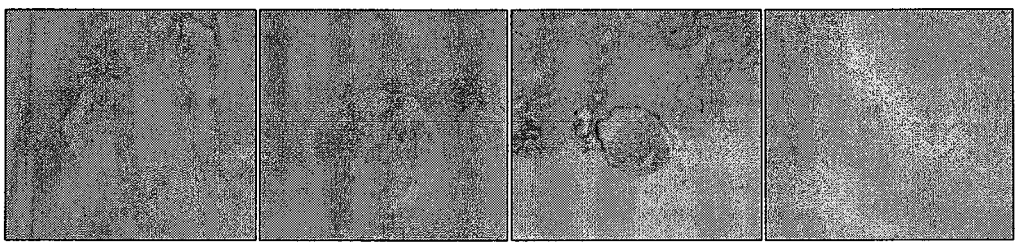


(a) Original images



(b) Edge detection results

Figure 3. Edge detection examples.



(a) Original images

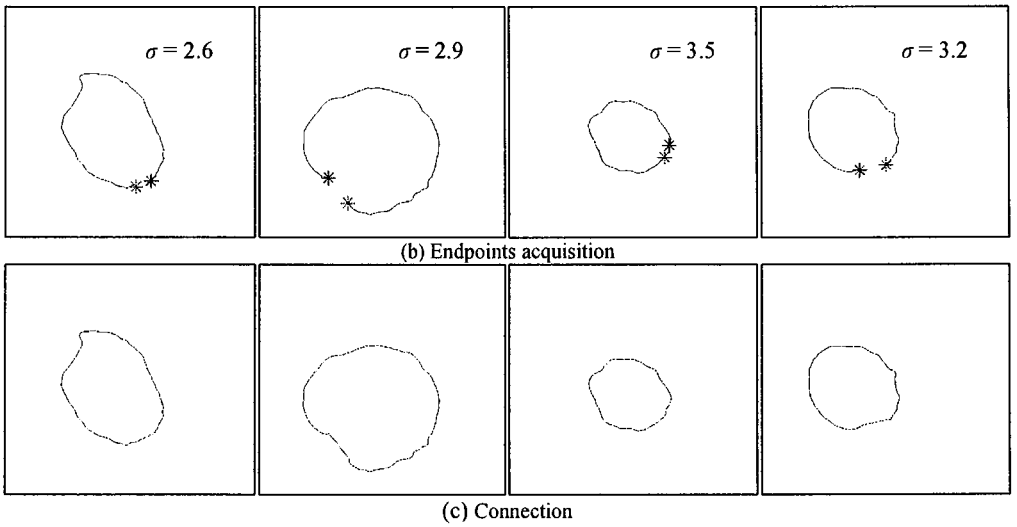


Figure 4. Endpoints connection for whole edge acquisition.

Table 1. Comparative result.

Sample No.	N_{org}	Proposed method		Conventional method [7]	
		N_{pro}	ME (%)	N_{con}	ME (%)
1	197867	198211	0.33	193945	1.78
2	204646	204498	0.14	203267	0.62
3	205409	205237	0.18	204620	0.63
4	208726	208617	0.51	206928	0.86
5	198193	199012	0.47	196267	1.25

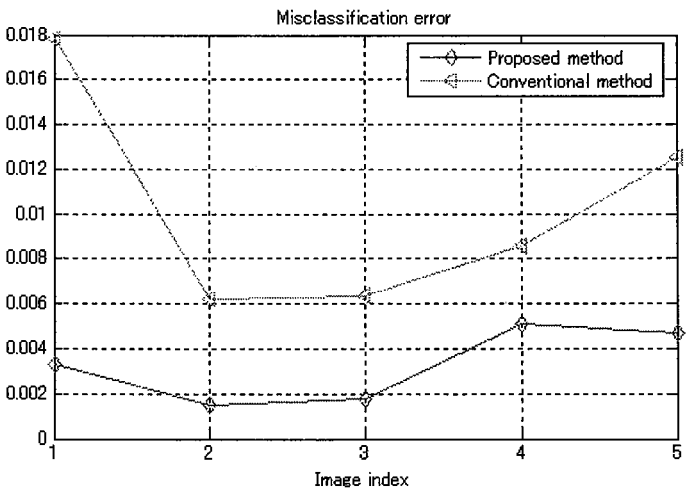


Figure 5. Quantization error by different methods.

Experiment 5: In this experiment, we compared the proposed method with the method presented in [7]. To evaluate both methods precisely, we present comparative results for them based on Misclassification error (ME) [14]. The ME varies from 0 for a perfectly classified image to 1 for a totally wrongly binarized image. Table 1 gives the comparative values. N_{org} is the real area of target, N_{pro} is the extracted area by our method and N_{con} is the extracted area by conventional method. To show the different performance of the two methods, we plot the error in Fig. 5. It is obvious that the proposed method performs better than [7].

4. CONCLUSION

This paper presents an edge detection method by Canny operator with a feedback strategy. Firstly, Otsu method is applied to the gradient histogram to determine the high and low thresholds for Canny operator. Then a feedback strategy is used to select the suitable standard deviation. During this procedure, the noises are erased clearly to guarantee the correctness of the feedback value calculation. As a result, the edges of target without noises can be obtained. Experiments showed that some samples resulted in a whole edge but others produced the discontinuous edges. In the case of the latter, endpoints should be located and the endpoint pairs which should be connected are also found out and connected to form a whole edge. Experimental results indicate that the proposed method has produced the promising results for our renal biopsy images. However, for the connection of endpoint pairs, we only use a line to connect them, a more precise method need to be studied for the future work.

REFERENCES

[1] Kharma, N., Moghnieh, H., Yao, J. et al., "Automatic Segmentation of Cells from Microscopic Imagery Using Ellipse Detection," IET Image Processing, 1 (1), 39-47 (2007).

[2] Hirai, K., Watanabe, S., Taniguchi, K., et al., "Extraction of the Glomerular Region in Kidney Images Using a Local and an Elliptical Global Method," Med. Imag. Tech., 20(6), 685-693 (2002).

[3] Hirai, K., Watanabe, S., Taniguchi, K., et al., "Improved Method for Extracting the Glomerular Region in Kidney Images," Med. Imag. Tech., 21(5), 377-381 (2003).

[4] Fan, J., Yau, D., Elmagarmid, A., Aref, W., "Automatic Image Segmentation by Integrating Color-edge Extraction and Seeded Region Growing," IEEE Trans. on Image Processing, 10(10), 1454-1466 (2001).

[5] Zhu, H., Taniguchi, K., Watanabe, S., et al., "Extraction of Renal Glomerular Regions Using a Non-linear Difference Filter and Region-estimation Model," Med. Imag. Tech., 17(6), 669-680 (1999).

[6] Zhu, H., Taniguchi, K., Zhang, X., et al., "A Method of Binalization for Renal Images using the Self-organized Neural Network," IEEJ-C, 118(9), 1397-1398 (1998).

[7] Nian, W., Okazaki, K., Taniguchi, K., et al., "Contour Extraction of Renal Glomeruli by GA," Med. Imag. Tech., 15(6), 725-733 (1997).

[8] Bruni, C., Santis De, A., Iacoviello, D., et al., "Modeling for Edge Detection Problems in Blurred Noisy Images," IEEE Trans. on Image Processing, 10(10), 1447-1453 (2001).

[9] Chan, T. and Vese, L., "Active Contours without Edges," IEEE Trans. on Image Processing, 10(2), 266-277 (2001).

[10] Canny, J., "A Computational Approach to Edge Detection," IEEE Trans. on Pattern Analysis and Machine Intelligence, 8(6), 679-714 (1986).

[11] Yan, F., Shao, X., Li, G., et al., "Edge Detection of Tank Level IR Imaging Based on the Auto-adaptive Double-threshold Canny Operator," Second Int' l Symp. on Intelligent Information Technology Application, 3, 366-370 (2008).

[12] He, W. and Yuan, K., "An Improved Canny Edge Detector and its Realization on FPGA," Proc. of the 7th World Congress on Intelligent and Automation, 6561-6564 (2008).

[13] Zhou, Z. and Zhang, J., "Object Detection and Tracking Based on Adaptive Canny Operator and GM(1,1) Model," IEEE Int'l Conf. on Grey Systems and Intelligent Services, 434-439 (2007).

[14] Zhang, J. and Hu, J., "Renal Biopsy Image Segmentation Based on 2-D Otsu Method with Histogram Analysis," Med. Imag. Tech., 27(3), 185-192 (2009).

Incorporating Multiple SVMs for Active Feedback in Image Retrieval Using Unlabeled Data

Zongmin Li[1], Yang Liu[1],Hua Li[2]

[1]Department of Computer Science, University of Petroleum, Dongying, P.R China
[2]Institute of Computing Technology, Chinese Academy of Sciences, Beijing, P.R China
Email: lizongmin@upc.edu.cn; liuyangsunnygirl@126.com; lihua@ict.ac.cn

ABSTRACT

Active learning with support vector machine(SVM) selects most informative unlabeled images for user labeling, however small training samples affect its performance. To improve active learning and use more unlabeled data, we propose a new algorithm training three SVMs separately on the color, texture and shape features of labeled images with three different kernel functions according to the features' distinct statistical properties. Different algorithms are used in the selection of disagreement and agreement samples from unlabeled data and also in the calculation of their confidence degrees. The lowest confident disagreement samples are returned to user to label and added to the training data set with the highest confident agreement samples. Experimental results verify the high effectiveness of our method in image retrieval.

Multiple SVMs; Active Learning; Relevance Feedback; Image Retrieval

1. INTRODUCTION

The semantic gap between high-level semantics and low-level image features is a key problem of Content-based Image Retrieval [1,2]. Relevance feedback studies from the interactive process between the system and the user, establishes and revises the relation of high-level semantics and low-level image features, bridges the semantic gap to augment the performance of the retrieval system [3,4]. However, for large scale image database retrieval problem, labeled images are always rare compared with unlabeled images. The key problem of relevance feedback is how to learn the semantics from limited labeled data in high dimensional feature space [5]. In order to enhance relevance feedback by using the unlabeled data of the database [6,7] and avoid giving extra burden to users, active learning with SVM selects the most informative samples which can provide the classifier with more information to generate decision boundary when labeled by user [8].

Tong and Chang proposed the SVM_{active} algorithm [9] and selected the image near the SVM boundary for user labeling which can approximately maximally reduce the size of the version space. J.He et al extended SVM_{active} algorithm theoretically and proposed the concept named Mean version space [10] which tailored for each specific learning task. They take both the size of the version space and the posterior probabilities into consideration. The above are strategies for single sample and apparently it is more efficient to select and label a set of samples before repeatedly running the training algorithm. Therefore K.Brinker proposed angle-diversity algorithm [11] trying to select batches of new training examples by balancing the distance from samples to the classification hyper-plane and the diversity among these samples which is verified a faster method to attain a level of generalization accuracy in terms of the number of labeled samples. Considering of the redundancy problem, R.Liu [13] used a different simple classifier to classify the unlabeled data and added the most irrelevant images into the labeled images. Then the images near the SVM boundary were clustered and the most informative one was chosen from each cluster for user labeling. Most methods combine all the different features of the image into one feature view without considering the contribution of different features is significantly different. J.Cheng and K.Wang proposed a new algorithm named Co-SVM [12] to improve the active learning algorithm. In Co-SVM , color features and texture features are taken as two sufficient and uncorrelated views of an image. SVM is used to separately learn two classifiers on the labeled samples only with color view and texture view [12]. A set of unlabeled samples that disagree between the two classifiers are chosen as the contention samples, and the contention

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462N · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853383

samples with the lowest confidence degree are recommended to user to label.

However Co-SVM only uses one kind of kernel function to train SVMs for color and texture features without considering their different statistical properties. J.Fan [14] mentioned that the performance of image classifiers largely depends on three inter-related issues: representative visual features, suitable functions for image similarity characterization and effective algorithms for image classifier training. It is very hard to use one single type of similarity function to characterize the diverse visual similarities between the images effectively [14]. And only using the color and texture feature to represent the visual properties of the image is not sufficient. Another problem in relevance feedback is that the number of samples for user labeling each time must be small preventing user losing patience. So even after a few rounds of feedbacks the labeled images are still rare compared with unlabeled images. But the performance of SVM classifier largely depends on the number of the labeled data.

Based on the above analysis and be inspired by Co-SVM, we propose a new strategy to incorporate multiple SVMs in relevance feedback to improve the performance of image retrieval. Firstly we extract the color, texture and shape features from the labeled images. Secondly, we separately train three SVMs on the three features with different kernel functions according to the different statistical properties of the features. Then the unlabeled samples that disagree between three classifiers are named as the "disagreement samples" and that classified consistently to be negative samples by three classifiers are named as the "agreement samples". Different algorithms are used to calculate the confidence degrees of the two kinds of samples. The disagreement samples with lowest confident scores are returned to user to label and added to the training data with the agreement samples that have the highest confident scores. Therefore, we extend the size of labeled data by adding a bunch of the most irrelevant images with the highest confident score into the labeled images while choosing a bunch of images for user labeling. Experiments show our algorithm can reach a good result in a much less rounds of feedback.

2.TWO-CLASS SVM

Support Vector Machine (SVM) has strong theoretical foundation and excellent empirical success as a core machine learning technology. The main idea is learning a boundary (i.e., hyperplane) separating the training data by a maximal margin. There are many types of SVM including one-class, two-class and multi-class SVM. In this paper we only consider two-class SVM.

Given training data and their labels: $(x_i, y_i)_{1 \le i \le N}$, $x_i \in R^d$, $y_i \in \{-1, 1\}$, the general form of decision function is $g(x) = w \cdot x + b$. Normalize $g(x)$, make all x_i satisfy $|g(x)| \ge 1$, then the margin between two boundaries is equal to $2/\|w\|$. Solving the quadratic optimization problem equals to minimize $\|w\|$, therefore the objective function is:

$$\min \phi(w) = \frac{1}{2}\|w\|^2 = \frac{1}{2}(w \cdot w) \qquad (1)$$

$$Subject\ to: y_i(w \cdot x_i + b) \ge 1, i = 1....N$$

Use the Lagrange's method of multipliers introducing $\alpha = (\alpha_1, ..., \alpha_N)$ to solve this quadratic optimization problem and get the optimal separating hyperplane. All vectors lying on one side of the hyperplane are labeled as 1, and all vectors lying on the other side are labeled as -1. The training instances that lie closest to the hyperplane are called support vectors. The final form of the function is:

$$f(x) = sign(\sum_i \alpha_i y_i x_i \cdot x + b). \qquad (2)$$

For the nonlinear case, SVM allows one to project the original training data in space X to a higher dimensional feature space F via a Mercer kernel operator K. Firstly it introduces the slack variable and penalty factor making the objective function become:

$$\varphi(\omega, \xi) = \frac{1}{2}(\omega \cdot \omega) + C\left[\sum_{i=1}^{N} \xi_i\right]. \qquad (3)$$

Then SVM projects the original training data to a higher dimensional feature space with the nonlinear transformation, and obtains the optimal separating hyperplane in the new space. The inner product would be $k(x, y) = \phi(x) \cdot \phi(y)$, then

the final function is :

$$f(x) = sign(\sum_i \alpha_i y_i k(x_i \cdot x) + b). \quad (4)$$

By choosing different kernel functions we can implement the nonlinear transformation. The commonly used kernels are polynomial kernel, radial basis function kernel, and sigmoid kernel. In this paper, we will choose the kernel functions according to the statistical properties of different features.

3.FEATURES AND KERNEL FUNCTIONS

3.1. Color

The color feature we use is HSV color histogram uniformly into 64 bins, and the kernel function we adopt is the χ^2 kernel function which is a Mercer kernel [14]. Given two color histograms u and v, their χ^2 statistical similarity is defined as:

$$\chi^2(u,v) = \frac{1}{2}\sum_{i=1}^{64} \frac{(u_i - v_i)^2}{u_i + v_i} \quad . \quad (5)$$

where u_i and v_i are the corresponding color bins from two color histograms u and v.

The kernel function $K(u,v)$, which is used to characterize the similarity between these two color histograms, is defined as:

$$K(u,v) = e^{-\chi^2(x,y)/\sigma} . \quad (6)$$

where σ is set to be the mean value of the χ^2 distances between all the image pairs in our experiments.

3.2. Texture

The texture feature we adopt is the 51-dimensional texture spectrum descriptor vector [15].Given two texture vectors x and y, their Euclidean distance is defined as:

$$d^2(x,y) = \sum_{i=1}^{51} (x_i - y_i)^2 . \quad (7)$$

To incorporate the Euclidean distance into our kernel-based image similarity characterization framework, we can extend Gaussian kernel as:

$$K(x,y) = e^{-d^2(x,y)/\gamma} . \quad (8)$$

where γ is set as the mean value of $d^2(x,y)$ of the image pairs in our experiments.

3.3.Shape

We use 23-dimensional moment vector including 7-dimensional Hu moment vector and 16-dimensional Zernike moment vector as shape features. Our experiment verifies Mahalanobis distance is more suitable to calculate the similarity between two compositional shape feature vectors than Euclidean distance here. The formula is:

$$m^2(x,y) = (x-y)^T S(x-y) . \quad (9)$$

where S is the covariance matrix of the two shape feature vectors. Then by incorporating the Mahalanobis distance into Gaussian kernel, the final formula is defined as:

$$K(x,y) = e^{-m^2(x,y)/\tau} . \quad (10)$$

where τ is settled as the mean value of $m^2(x,y)$ of the image pairs in our experiments.

4.THE PROPOSED ALGORITHM

To provide the training data to SVMs initially, three features are combined as one feature view and images in the database are ranked in descending order of similarity to the query by calculating the euclidean distance of their features during the first round of querying. The first m images are returned to user to label and meanwhile the last n images are added into the labeled data as irrelevant images.

After training three SVM classifiers on the labeled samples with three different kernel functions according to their different statistical properties, the unlabeled samples that disagree between three classifiers are selected as the "disagreement samples" and that classified consistently to be negative samples by three classifiers are selected as the "agreement samples".

Then the confidence degrees of the disagreement samples are calculated based on angle-diversity algorithm [11] by the formula defined as :

$$C(x_i) = \lambda\left|g(x_i)\right| + (1-\lambda)\max_{j\in S} k^*(x_i, x_j).$$

$$0 \le \lambda \le 1. \tag{11}$$

where S is the candidate data set for user to label. Set $S = \phi$ at first. $g(x_i)$ is the real-valued output of the classifier f. $k^*(x_i, x_j)$ denotes the (undirected) angle between two hyperplanes h_i and h_j which correspond to examples x_i and x_j. Its formula is defined as:

$$k^*(x_i, x_j) = \left|\cos(\angle(h_i, h_j))\right| = \frac{\left|\langle\phi(x_i), \phi(x_j)\rangle\right|}{\left\|\phi(x_i)\right\|\left\|\phi(x_j)\right\|} = \frac{\left|k(x_i, x_j)\right|}{\sqrt{k(x_i, x_i)k(x_j, x_j)}} \tag{12}$$

The individual influence of each requirement can be adjusted by the trade-off parameter λ. Setting $\lambda = 1$ restores the distance strategy same as J.Cheng's algorithm [12], whereas for $\lambda = 0$ the algorithm focuses exclusively on maximizing the angle diversity. Setting $0 < \lambda < 1$ considers both factors. In this paper, we set $\lambda = 0.5$.

We incrementally construct a new training batch S:

$1: S = \varnothing$

$2: p = \arg\min_{i\in(I^*-S)}(\lambda\left|g(x_i)\right| + (1-\lambda)\max_{j\in S} k^*(x_i, x_j))$

$3: S = S \cup \{x_p\}$

$4:$ *repeat step* 2 *and* 3 *until* $\left|S\right| = m$.

where I^* denotes the disagreement samples.

As there are 3 classifiers, so we can construct S_1, S_2, S_3 respectively for three SVMs using this algorithm. The final candidate set $S = S_1 \cap S_2 \cap S_3$. If $\left|S\right| < m$, add the samples with lowest $C(x_i)$ in S_1, S_2, S_3 into S until $\left|S\right| = m$. Return samples in S to user to label. Meanwhile the confidence degrees of the agreement samples are calculated by the formula defined as:

$$C'(x_i) = \max\left|\sum_{j=1}^{3} g_j(x_i)\right|. \tag{13}$$

where $g_j(x)$ is the real-valued output of the classifier f_j.

Then the highest n agreement samples and the samples in S are added into the training data and three SVMs are retrained on the labeled data set until the user is satisfied with the results.

5.EXPERIMENTS

To validate our algorithm's performance, we compare it with K.Brinker's angle-diversity algorithm, J.Cheng's Co-SVM algorithm. Experiments are performed on a subset selected from the Corel image CDs. There are 50 categories in our subset. Each category contains 100 images, 5000 images in all. The categories have different semantic meanings such as dinosaur, sunrise, bus, butterfly, waterfall, cloud, elephant, etc. The parameter m is set to 10 and n is set to 5 in the experiments.

We choose 100 images as query images randomly to probe the retrieval performance. We use the precision of top k result images as the evaluation of three algorithms. The final accuracy of our experiments is the averaging accuracy of all test images. We compare three algorithms with our algorithm under the same conditions. The color and texture features used in Co-SVM are same as our algorithm and the feature used in angle-diversity algorithm is the combine of three features of our algorithm which is 138-dimensional feature vector in total. Figs.1 and 2 are the accuracy vs. scope curve of the three algorithms after the third and fifth rounds of relevance feedback, respectively. We can see that Co-SVM is better than angle-diversity algorithm and our new strategy to use the unlabeled data shows a best performance among the three algorithms. The detailed results are summarized in Table 1. The results depicted in Table 1 show that our new strategy using the unlabeled data achieves the higher performance than others.

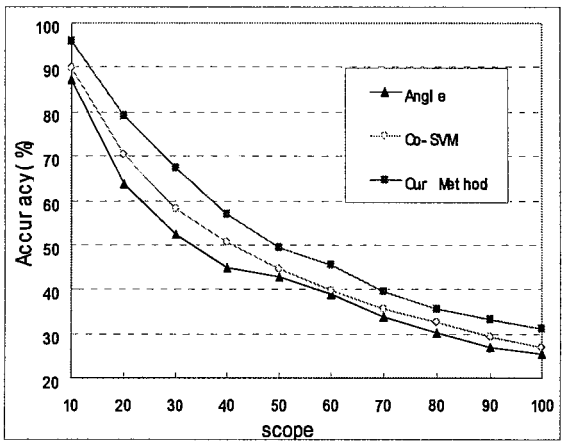


Figure 1.The average top-k accuracy of image retrieval with three feedbacks.

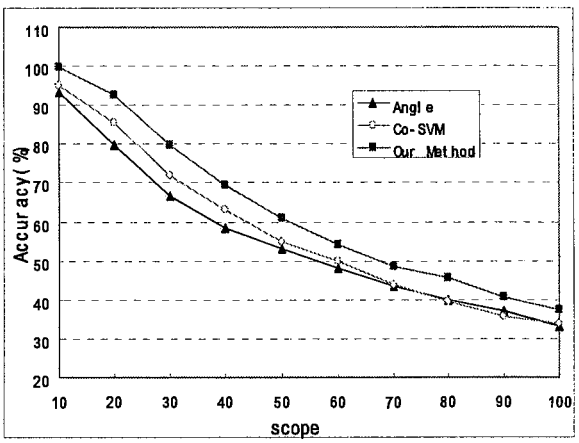


Figure 2. The average top-k accuracy of image retrieval with five feedbacks.

6.CONCLUSION

This paper trains three SVMs respectively with three different kernel functions according to the different statistical properties of the color, texture and shape features extracted from images. We use angle-diversity algorithm to choose the unlabeled data from the disagreement samples for user labeling and in order to augment the training data of SVMs and decrease the feedback counts, we add the most irrelevant images into the training data in every loop. Experiments show that our algorithm improves the retrieval performance obviously.

ACKNOWLEDGMENTS

This work was supported by National Natural Science Foundation of China (grant No: 60533090), National Key Basi c Research Plan (grant No: 2004CB318000) , PetroChina Innovation Foundation(grant No: 2009D-5006-01-07) and Natural Science Foundation of Shandong(grant No: 2009ZRA05136).

REFERENCES

[1] Y. Liu, D. S. Zhang, G. J. Lu, and W. Y. Ma. A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 2007, 40 (1): 262-282.

[2] X. S. Zhou, and T. S. Huang. CBIR: from low-level features to high-level semantics. *Proceedings of SPIE Image and Video Communication and Processing*, San Jose, CA, January 2000, pp. 426-431.
[3] Y. Rui, T. S. Huang, and M. Ortega. Relevance feedback: A power tool in interactive content-based image retrieval. *IEEE Tran.Circuits and Systems for Video Tech.*1998, 8(5): 644-65.
[4] Y. Rui, T. S. Huang, and S. Mehrotra. Content-based image retrieval with relevance feedback in MARS. *Proc.IEEE Int.Conf.on Image Proc.*1997.
[5] P. Hong, Q. Tian, and T. S. Huang. Incorporate Support Vector Machines to Content-Based Image Retrieval with Relevant Feedback. *Proc.IEEE Int'l Conf Image Processing*, 2000: 750-753.
[6] Z. H. Zhou, K. J. Chen, and Y. Jiang. Exploiting unlabeled data in content-based image retrieval. *Proceedings of the 15th European Conference on Machine Learning*, Pisa, Italy, 2004.
[7] Z. H. Zhou, K. J. Chen and H. B. Dai. Enhancing relevance feedback in image retrieval using unlabeled data. *ACM Trans. on Information Systems 2006*, 24(2): 219–244.
[8] L. Wang, K. L. Chan, and Z. Zhang. Bootstrapping SVM active learning by incorporating unlabelled images for image retrieval. *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, 2003, pp. 629–634.
[9] S. Tong, and E. Chang. Support vector machine active learning for image retrieval. *Proceedings of ACM Multimedia*, Ottawa, Canada, 2001.
[10] J. He, M. Li, H. J. Zhang, et al. Mean version space: a new active learning method for content-based image retrieval. *Proceedings of the 6th ACM SIGMM International Workshop on Multimedia Information Retrieval (MIR)*, 2004: 15–22.
[11] K. Brinker. Incorporating diversity in active learning with support vector machines. *Proceedings of the 20th International Conference on Machine Learning (ICML)*, 2003: 59–66.
[12] J. Cheng, and K. Wang. Active learning for image retrieval with Co-SVM. *Pattern Recognition*, 2007, 40(1): 330-334.
[13] R. Liu, Y. Wang, T. Baba, et al. SVM-based active feedback in image retrieval using clustering and unlabeled data. *Pattern Recognition*, 2008, 41(8): 2645-2655.
[14] J. Fan, Y. Gao, H. Luo, et al. New Approach for Hierarchical Classifier Training and Multi-level Image Annotation. *MMM 2008*: 45-57.
[15] Z. P. Shi, H. He, Y. Liz, Z. Sic, and L. Duan. Texture spectrum descriptor based image retrieval. *Journal of Software*,2005,16(6):1039−1045.DOI: 10. 1360/jos161039.

Table 1. The comparison of three algorithms.

	Top10			Top20			Top30		
	Angle	Co-SVM	Our Method	Angle	Co-SVM	Our Method	Angle	Co-SVM	Our Method
FB1	71.5	71.5	73.8	51.9	51.9	55.4	42.4	42.4	49.3
FB2	78.2	83.0	85.2	57.8	61.6	65.3	45.6	51.1	55.9
FB3	87.0	89.9	95.7	63.7	70.2	79.1	52.4	58.2	67.2
FB4	90.9	94.4	98.9	75.2	82.8	89.8	63.4	69.6	78.0
FB5	93.4	95.1	99.7	79.8	85.4	92.4	66.7	71.8	79.8

	Top40			Top50			Top60		
	Angle	Co-SVM	Our Method	Angle	Co-SVM	Our Method	Angle	Co-SVM	Our Method
FB1	36.3	36.3	40.1	32.7	32.7	35.8	29.2	29.2	32.2
FB2	38.9	43.5	47.7	35.9	38.5	42.5	33.0	34.9	38.3
FB3	45.0	50.6	56.9	42.8	44.7	49.3	38.9	39.8	45.4
FB4	55.8	60.2	68.2	52.6	53.6	59.5	44.3	47.9	52.6
FB5	58.3	62.9	69.4	53.0	54.8	60.8	48.1	49.8	54.2

	Top70			Top80			Top90		
	Angle	Co-SVM	Our Method	Angle	Co-SVM	Our Method	Angle	Co-SVM	Our Method
FB1	26.6	26.6	29.4	24.5	24.5	27.1	23.1	23.1	25.2
FB2	29.0	31.5	34.8	25.8	28.9	31.9	25.6	26.8	29.6
FB3	33.7	35.7	39.6	30.3	32.5	35.7	27.0	29.4	33.1
FB4	39.8	43.0	47.4	35.1	39.0	42.6	33.3	33.6	38.0
FB5	43.5	43.8	48.3	40.0	39.7	45.6	37.1	35.5	40.6

	Top100		
	Angle	Co-SVM	Our Method
FB1	21.6	21.6	23.6
FB2	22.3	25.1	27.6
FB3	25.4	27.0	31.0
FB4	31.9	32.5	35.9
FB5	33.1	33.9	37.5

Content Based Image and Video Retrieval

Ms. Shubhangi H. Patil H.O.D of Electronics Dept, Asst.prof. & Co-ordinator, FD cell, Dr. J.J.M.C.O.E., Jaysingpur,Dist. Kolhapur, Maharashtra, India, Mrs P.P.Belegali,Asst Prof. E& C dept, Dr. J.J.M.C.O.E., Jaysingpur,Dist. Kolhapur, Maharashtra, India ,Mr. Patil B.S., H.O.D., IT dept, P.V.P.I.T.,Budhgaon,India, Ms Mohite T.H.,Lecturer Electronics Dept, Dr. J.J.M.C.O.E., Jaysingpur., Ms. Dhobale Dhanashri D, Lecturer, IT Dept, P.V.P.I.T., Budhgaon,Dist Sangli,Maharashtra,India

Email: sbp_jjm2004@yahoo.co.in,

ABSTRACT

The growing capacity of computers, the abundance of digital cameras and the increased connectivity of the world all point to large digital multimedia archives. They include images and videos from the World Wide Web, museum objects, flowers, trademarks, and views from everyday life. The faster they grow, the more prominently needed is the efficient access to the content of the images and videos. In this paper we have given important step of feature extraction, will be discussed in detail such as color, shape and texture information, particularly paying attention to discriminatory power and invariance. Then, we focus on the concepts of indexing and genre classification as intermediate step to sort the data. We pay attention to (interactive) ways to perform browsing and retrieval by means of information visualization and relevance feedback. Methods are being discussed to localize the retrieved objects in images.

We adopt a hybrid approach for such text extraction by exploiting a number of characteristics of text blocks in color images and video frames. Our system detects both caption text as well as scene text of different font, size, color and intensity. Such texts are used for retrieval of video clips based on any given keyword. Content-Based Image And Video Retrieval addresses the basic concepts and techniques for designing content-based image and video retrieval systems.

Keyword list- Video retrieval, image retrieval, image extraction, *Containment Tree,* Hidden Markov Model.

I. INTRODUCTION

In the image retrieval task a user must search a database of many thousands, or millions, of images. User goals vary; in some cases the task is to find a particular image, in other cases any image from a class will do. The optimal interface would provide a very flexible query mechanism, perhaps through a natural language interface. In fact, many "stock photo houses" currently provide such an interface to their collections. Advertisers and publishers present a description of their requirements: "an image of the beach with athletic people playing volleyball". Human clerks then scan many images by hand using keywords. Recently a large number of automated image retrieval systems have appeared [7, 11, 12, and 15]. Rather than describe an image using text, in these systems an image query is described using a set of example images. In some of these systems a user's only interaction with the retrieval engine is through example images, in others the user is also asked to weight a set of "intuitive" features, such as color, texture and shape. Image retrieval differs from the more common task of classification which includes tasks such as face detection and character recognition. In retrieval the number of potential image classes is extremely large and the number of example images is very small. For example, a user may wish to retrieve example images of "cars on the road" using perhaps three example images. Conventional machine learning methods, such as neural networks or support vector machines, are not well suited to this task because they often require a small number of classes and a large set of labeled data (see [1, 2] for example).

An effective solution to this problem hinges on the discovery of a simplifying structure in the distribution of images. A learning algorithm can then take advantage of this structure to learn an image class from a small number of examples.

When a human clerk is shown three example images containing "a car on the road", he concludes that other images must contain both "car" and "road". A photograph chosen at random from the Web might contain a "car", a "road", the "Eiffel Tower", the "Taj Mahal", or any one of a thousand other objects. But, while there are a very large number of objects which might be present in any one image, any particular image will contain at most a few of these

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462O · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856296

objects. This is not unlike the structure of English text: there are over 100,000 possible English words, but any given sentence will contain roughly six words. The clerk's actions in the above example are justified because the probability of a car and a road appearing by random chance in all three images is quite low.

Content-based image and video retrieval is concerned with retrieving images and video sequences on the basis of automatically derived features such as color, texture, and shape information that is present in the images. The need for efficient storage and retrieval of images is not new, but the increase in the number and size of digital image collections made the problems of image retrieval widely recognized. The shortcomings with traditional methods of indexing, often based on keywords, have led to the interest in retrieval on the basis of features that are automatically derived from the image content.

The emergence of content-based image and video retrieval has brought together and broadened the extend of a number of disciplines, such as image understanding, database indexing, object recognition, visual data modeling, feature extraction, visual querying, perception and cognition, and information visualization.

Fig. 1 shows one sample for image extraction.

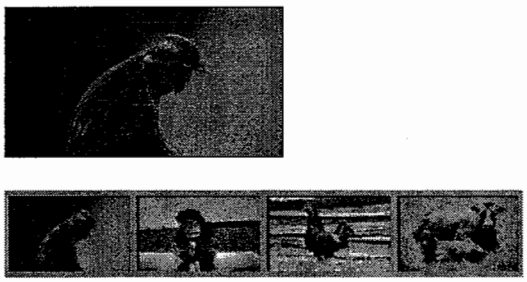


Fig.1 Sample of Image extraction

II RELATED WORK FOR IMAGE EXTRACTION

Numerous approaches for content-based image retrieval have been proposed in the literature. They are based on features like color , shape or texture . In a graph-based approach similar to the one described in section III (Point B) is used, while in

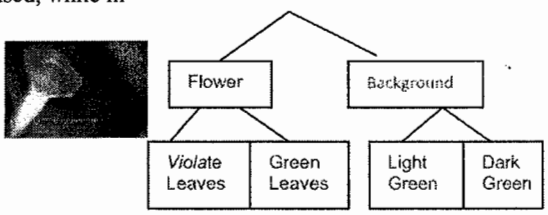


Fig.2 . An image and its inherent structure.

an cdit distancc mcasurc is uscd to mcasurc similarity of topological arrangements. gives a nice overview of the different approaches. However, all those approaches concentrate on single image features like color or texture. Our method differs from those proposed in the literature as we combine different approaches in order to achieve more meaningful results.

III CONTENT-BASED IMAGE RETRIEVAL

In content-based image retrieval the use of simple features like color, shape or texture is not sufficient. Instead, the ultimate goal is to capture the content of an image via extracting the objects of the image. Usually images contain an inherent structure which may be hierarchical. An example can be seen in figure 2. In the following, we describe two

models for image representation and similarity measurement, which take structural as well as content features like color into account.

A. IMAGE REPRESENTATION AS CONTAINMENT TREES

One way to model images for content-based retrieval is the use of trees representing the structural and content information of the images. In this section, we describe, how the structure of images can be extracted automatically based on the color of its segments. Additionally we show how the similarity between two such trees can be measured.

a. Transforming an Image into a Containment Tree : To utilize the inherent structure of images for content-based retrieval, we model them as so called containment trees. Containment trees model the hierarchical containment of image regions within others. To extract the containment tree of an image we first segment the image based on the colors of the regions using a region growing algorithm. The resulting segments are attributed with their color and size relative to the complete image. In a second step, the containment hierarchy is extracted from the set of segments by determining which regions are completely contained in other regions. In this context, a region R_{in} is said to be contained in a region R_{cont} if for every point $p \in R_{in}$ and every straight line $L \in p$ there exist two points o_1, $o_2 \in R_{cont}$ with $o_1, o_2 \in L$ and o_1, o_1 are on opposite sides of p.

b. Measuring the distance between two Containment Trees: To measure the similarity of containment trees, special similarity measures for attributed trees are necessary. A successful similarity measure for attributed trees is the edit distance. Well known from string matching the edit distance is the minimal number of edit operations necessary to transform one tree into the other. The basic form allows two edit operations, i.e. the insertion and the deletion of a node. In the case of attributed nodes the change of a node label is introduced as a third basic operation. A great advantage of using the edit distance as a similarity measure is that along with the distance value, a mapping between the nodes in the two trees is provided in terms of the edit sequence. The mapping can be visualized and can serve as an explanation of the similarity distance to the user. However, as the computation of the edit-distance is NP-complete, constrained edit distances like the degree-2 edit distance have been introduced. They were successfully applied to trees for web site analysis, structural similarity of XML documents, shape recognition or chemical substructure search.

Definition1-(degree-2 edit distance). *The edit distance between two trees* t1 *and* t2, ED2 (t1, t2), *is the minimum cost of all degree-2 edit sequences that transform* t1 *into* t2 *or vice versa. A degree-2 edit sequence consists only of insertions or deletions of nodes* n *with* degree(n) <= 2, *or of relabeling:-*

ED2 (t1, t2)= min{c(S) | S is a degree-2 edit sequence transforming t1 into t2}

The main idea behind this distance measure is that only insertions or deletions of nodes with a maximum number of two neighbors are allowed.

c. Efficient Similarity Search for Containment Trees: While yielding good results, the degree-2 edit distance is still computationally complex and, therefore, of limited benefit for searching or clustering in large databases. A filter and refinement architecture for the degree-2 edit distance is presented to overcome this problem. A set of new filter methods for structural and for content-based information as well as ways to flexibly combine differ -ent filter criteria are presented.

B. IMAGE REPRESENTATION AS SEGMENTATION GRAPHS.

Graphs are another way to model images for content-based similarity search. They were successfully used for shape retrieval, object recognition or face recognition. In this section, we describe a content-based image retrieval system based on graphs which are extracted from images in a similar way as the trees in the preceding section.

***a. Transforming an Image into a Segmentation Graph*:** To extract graphs from the images, they are segmented with a region growing technique and neighboring segments are connected by edges to represent the neighboring relationship. Each segment is assigned four attribute values, which are the size, the height and width of the bounding box and the color of the segment. The values of the first three attributes are expressed as a percentage relative to the image size, height and width in order to make the measure invariant to scaling.

b. Measuring the distance between two Segmentation Graphs: Most known similarity measures for attributed graphs are either limited to a special type of graph or are computationally extremely complex, i.e. NP-complete. Therefore they are unsuitable for searching or clustering large collections.

Definition2 -(edge matching distance).

Let G1 (V1, E1) *and* G2 (V2, E2) *be two attributed graphs. Without loss of generality, we assume that* $|E1| >= |E2|$. *The complete bipartite graph* G_{em} (V_{em} = E1 U E2 U Δ, E1 × (E2 U Δ)), *where* Δ *represents an empty dummy edge, is called the edge matching graph of* G1 *and* G2. *An edge matching between* G1 *and* G2 *is defined as a maximal matching in* G_{em}. *Let there be a non-negative metric cost function* c: E1 × (E2 U Δ) → IR^{+}_{0}. *The edge matching distance between* G1 *and* G2, *denoted by* d_{match} (G1, G2), *is defined as the cost of the minimum-weight edge matching between* G1 *and* G2 *with respect to the cost function* c. We can demonstrate that the edge matching distance is a meaningful similarity measure for attributed graphs and that it enables efficient clustering of structured data.

c. Efficient Similarity Search for Segmentation Graphs: There is also filter refinement architecture and an accompanying set of filter methods presented to reduce the number of necessary distance calculations during similarity search. We can employ the same approach to ensure efficient query processing in many experiments.

d. Concept-Specific Image Models

Image retrieval for a text query c is essentially the task of deciding, for each image I, whether or not it contains the concept c. This motivates the following concept-specific model, in which one pair of
HMM is estimated for each concept: an HMM H^{+}_{c} from all images that contain the concept c, and an H^{-}_{c} from all images that do not contain the concept. Both H^{+}_{c} and H^{-}_{c} have simple left-to-right topologies, with as many states as there are rectangular image regions, as shown in Figure 1. Formally, the state space of both H^{+}_{c} and H^{-}_{c} is S = {1, . . . , T}, and unlike the preceding joint model, these states do not have any interpretation in terms of the concept vocabulary V and instead simply model spatial properties of the images.

Fig.3 shows the Hidden Markov Model Topology for the Concept-Specific Models.

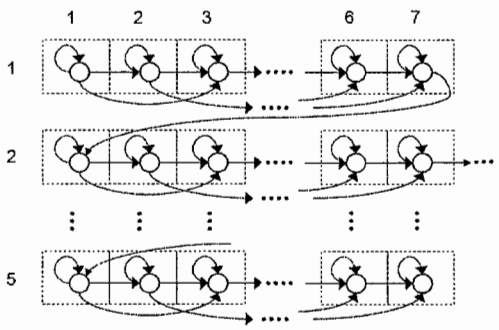


Fig. 3. Hidden Markov model Topology for the Concept-Specific Models

IV CONCLUSION

Content-based image& Video retrieval addresses the problem of finding images relevant to the users' information needs, based principally on low-level visual features for which automatic extraction methods are available. For the development of CBIR applications, an important issue is to have efficient and objective performance assessment methods for different features and techniques.

REFERENCES

[1] H. A. Rowley, S. Baluja, and T. Kanade. Neural networkbased face detection. *IEEE Patt. Anal. Mach. Intell.*, 20(1):23–28, 1998.

[2] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *ICCV98*, pages 555–562, 1998

[7] M. Kelly, T. M. Cannon, and D. R. Hush. Query by image example: the candid approach. In *SPIE Stor. and Ret. Image & Video Databases III*, volume 2420, pages 238–248, 1995.

[11] A. Pentland, R. W. Picard, and S. Sclaroff. Photobook:
content-based manipulation of image databases. *Int. J. Comp. Vis.*, 18(3):233–254, 1996.

[12] J. Huang, S. Kumar, M. Mitra,W. Zhu, and R. Zabih. Image indexing using color correlograms. In *CVPR97*, pages 762– 768, 1997.

[15] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. 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.

[16] Flickner, M., Swahney, H., Niblack,W., Ashley, J., Huang, Q., Dom, B., Gorkani, M., Hafner, J., Lee, D., Petkovic, D., Steele, D., Yanker, P.: Query by image and video content: The QBIC system. IEEE Computer (1995)

[17]. Mehtre, B., Kankanhalli, M., Lee, W.: Shape measures for content based image retrieval: A comparison. Information Processing Management **33** (1997) 319–337

[18]. Cullen, J., Hull, J., Hart, P.: Document image database retrieval and browsing using texture analysis. In: Proc. 4th Int. Conf. Document Analysis and Recognition. (1997) 718–721

[19]. Fuh, C.S., Cho, S.W., Essig, K.: Hierarchical color image region segmentation and shape extraction. IEEE Transactions on Image Processing **9** (200) 156–163

[20]. Tagare, H., Vos, F., Jaffe, C., Duncan, J.: Arrangement - a spatial realtion between parts for evaluating similarity of tomographic section. IEEE Trans. PAMI **17** (1995) 880–893

[21] Huet, B., Cross, A., Hancock, E.: Shape retrieval by inexact graph matching. In: Proc. IEEE Int. Conf. on Multimedia Computing Systems. Volume 2. (1999) 40–44

[22]. Kubicka, E., Kubicki, G., Vakalis, I.: Using graph distance in object recognition. In: Proc. ACM Computer Science Conference. (1990) 43–48

[23] Wiskott, L., Fellous, J.M., Kr¨uger, N., von der Malsburg, C.: Face recognition by elastic bunch graph matching. IEEE Trans. PAMI **19** (1997) 775–779

Effective Phonocardiogram Segmentation Using Time Statistics and Nonlinear Prediction

S.RAJESWARI*[1], Dr.J.JANET[2]
[1] Project Manager, Conrad Technologies Pvt. Ltd, Chennai, India
[2] Professor,IT Department Veltech Dr.RR. & Dr.SR Technical University, Chennai, India
Email: sridharanrajeswari@yahoo.com, janetjude@rediffmail.com

ABSTRACT

In the fields of image processing, signal processing and recognition, image Segmentation is an efficient method for segmenting the phonocardiograph signals (PCG) is offered. Primarily, inter-beat segmentation is approved and carried out by means of DII lead of the ECG recording for identifying the happenings of the very first heart sound (S1). Then, the intra-beat segmentation is attained by the use of recurrence time statistics (RTS), and that is very sensitive to variations of the renovated attractor in a state space derived from nonlinear dynamic analysis. Apart from this if the segmentation with RTS is unsuccessful, a special segmentation is proposed using threshold that is extracted from the high frequency rate decomposition and the feature extraction of the disorder is classified based on the murmur sounds. In the Inter-beat segmentation process the accuracy was 100% of the over all PCG recording. Taking into account a different level of PCG beats were strongly concerned by different types of cardiac murmurs and intra-beat segmentation are give up for an accurate result.

Keyword Terms—Segmentation, intra-beat, Phonocardiogram, Cardiac murmurs

1. INTRODUCTION

The phonocardiogram includes information by previously established value in the diagnosis of heart disease. The predictable methods are previously used for the processing of information's. Apart from this in which the doctor listens to the signal directly and makes an interpretation of the heart's condition based on the sound of the signals [2].From this signals the graphs are plotted and the segmentation process is been carried out. Segmentation is the process of extracting the PCG sound signals (Graphs) as a significant in order to get the murmur sounds and to detect them, and also to diagnosis them in order to find out the cardiac disorders using computer analysis. Thus, it is necessary that different mechanism of heart cycle can be timed and divided [1]. Segmentation is carried out in two major steps: inter-beat segmentation and intra-beat segmentation. Here initially the first mechanism which refers to the identification and separation of the heart beats or the heart sounds with the aim of performing an individual diagnostic mechanism in each one of them. The next step contains the separation of the heart beat sound into its different components: the first cardiac sound is (S1), systole, and the second cardiac sound is (S2) and diastole [2]. Here the Cardiac murmurs are been present in systole, diastole or both. By this mechanism, it is critical to precisely identify the boundaries in order to aptly locate the components of heart sounds and to perform an accurate diagnosis and classification of the murmur which are presented along with the heat sounds. In the existing system the main problem that occurs in the Heart sound segmentation is the murmur sounds which are present and produces high distortion in the existing temporal tracing mechanism [13]. Previously a large range of algorithms are used to perform PCG segmentation that has been proposed in the literature Survey[1].Here PCG signals are being segmented by the synchronized electrocardiograph (ECG) signal that are being acquired with PCG. On the further method, the envelope of the PCG can be used to perform the segmentation [3] .Here the envelope is being estimated by means of homomorphic filtering methodology, whereas it is been estimated by means of Shannon normalized average energy[4]. Moreover, Shannon energy can be used with more compound methodologies: here the wavelet transforms is in order to enhance the spectral components of S1 and S2; or with Mel-scaled filter-banks and also with Hidden Markov Models. Apart from this another mechanism to segment the PCG signals is based on time-frequency

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462P · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856301

analysis, which is to implement the wavelet transformations or with that of the distributions which are belonging to the Cohen's quadratic class. Here the PCG signals, that exhibits a nonlinear dynamics [5] which motivates the main purpose of complexity measures, and RTS [6]. In this paper we are discussing about the merging of these different methodologies with the aim of developing a trustworthy algorithm that is capable of segmenting the PCG signal with high accuracy. Apart from this here we also propose a new methodology for the boundary identifications of S1 and S2. By using these PCG signal that are being the references in order to split each of the Heart beat Sounds according to the new methodology which is been proposed in [7]; later on, the end of S1, the beginning of S2 and end of S2 are detected by using the RTS and the threshold values. The results that are obtained in the earlier stages are authenticated by using the biomedical features. Here if we consider the various points are out of different boundaries then these values will be given by physiological considerations or of with an alternative segmentation methodology which is being used that is based on wavelet decomposition, and that of the threshold values. Furthermore the proposed algorithms are experienced in a PCG database and are marked by the experts of the cardiology [5].

2. METHODOLOGY

The main objective of this study is to obtain a threshold and segmentation based heart beat results. However, it is necessary to obtain a solution that is suitable for all kind of circumstances. Here we are going to describe about the various techniques that are being used for the extraction of the PCG signals.

2.1. R-peak detections in ECG Signals

The database used in this work, contains an ECG signal which are synchronized with PCG, thus detecting correctly the R-peaks in ECG, makes it possible to locate the beginning of S1. The R-peaks in ECG can be detected employing algorithms for QRS complex detection, which have been widely reported in the literature. In this case, it is used a method proposed in [7] which consists basically in three steps: Linear filtering, nonlinear transformations and decision rules.

2.2. Recurrence of time statistics (RTS)

RTS are used in order to detect abrupt changes in the signal dynamics, corresponding to S1 and S2. An arbitrary state is chosen on the trajectory where upon all recurrences within a hypersphere of radius that are selected, and is then defined as the total amount of states in the set, and are related to the information dimension, which motivates its ability to detect weak signal transitions based on the amplitude, period, dimension and complexity of the signal [14]. A sliding window is used to partition the recorded PCG signal into overlapping segments, and is calculated for each segment. The r value is a very important parameter in the detection algorithm. If it is chosen too low, the hypersphere would be low on data, and if r is chosen too high, the hypersphere will contain misleading information from erroneous parts of the reconstructed state space. In this work, it becomes lower if there is not lobe detection corresponding to S1 andS2 [6].

2.3. Wavelet decomposition

These methods are used in the detection of S2 when it is considered that RTS did not give a good estimation of the boundaries. Basically, three steps are implemented: high pass filtering, wavelet decomposition and Shannon energy operator applied to the detail coefficients. From the knowledge of cardiac functionality and genesis of S1 and S2 sounds, it is known that aortic valves close with relatively large pressure difference across the valve, which originates the high frequency content in S2 sound [2]. This is the motivation to use the approximation coefficients to perform the detection of S2. To extract the signal envelope from the detail coefficients, the Shannon energy operator is applied [1]. Here the different levels are detailed with higher coefficients of the wavelet transformed heart sound Signal and N is the number of samples in the selected window. This technique emphasizes the medium intensity signal components and attenuates the effect of low intensity components [12]. Using the signal envelopes provided by the Shannon energy, sound lobe boundaries are identified applying decision rules based on

thresholds. By using these methodologies the accuracy of the proposed algorithm which is been increased with the higher propobability of prediction of different disorders.

2.4. Proposed algorithm for PCG segmentation

According to the methods mentioned in sections 2.1 to 2.3, a segmentation algorithm was designed in order to find the points corresponding to the boundaries of S1 and S2 in PCG. This algorithm is described below:

Step 1. Wave Acquisition and preprocessing

The sounds were recorded with a higher accuracy using different sampling frequency, in two channels: DII derivation of ECG and PCG recording. The database used in this work is made up of 25 adults are subjects, who gave their informed consent, and underwent a medical examination. That, 10 different recordings, of 7 seconds each one, corresponding to the four traditional focuses of auscultation were taken for each patient in the phase of post expiratory and post inspiratory [10]. Finally, the most representative beats (180 normal and 180 pathologic) were selected and labeled according to the criterion of cardiologists. In this stage, PCG signals are normalized to be in the interval [-1, 1], and filtered with wavelet denoising up to 8th level of decomposition using the wavelet. Later, the PCG recording is high-pass filtered in order to enhance the frequency components of HS.

Step 2. QRS and R-peak detection of PCG Signals

The beats are segmented using the method explained in section 2.1. Henceforth, the processing is applied on the individual beats.

Step 3. Locate at the end of S1

RTS are obtained using a sliding window of 2000 points over the sampled signal with 90% overlap for taking well resolution in time-domain. The dynamic attractor is reconstructed by using time delay embedding, where the embedding parameters (i.e., embedding dimension and time lag) have to be estimated. The embedding dimension d = 5, is obtained using the false neighbor method and the time lag is calculated minimizing the mutual information between each state variable and maximizing the distance from each point to the main diagonal of the state space. The end of S1 is located in a window between 70 and 160ms after R-peak; hence the search is limited to this time Interval. Furthermore, we have local minima in S1 and S2, in this way the sounds are segmented and are computed in order to detect positive lobes corresponding to HS. Is computed with an initial radius of r0 = 0.35, but sometimes S1 has low intensity and there is no lobe detection using this value. Therefore r is decreased until it exceeds a predefined threshold of 0.7 [9]. Some systolic murmurs give rise to extra lobes at the end of the analysis window, to solve this problem, if a lobe is detected in this zone, and if it exceeds certain threshold (80 % of the maximum value), it is removed. Finally, a forward search is performed from the maximum of the lobe corresponding to S1, until the signal has an amplitude or slope below established thresholds.

Step 4. Locate at the beginning and end of S2

The estimation is computed in the same way aforementioned for S1, but in a different analysis window (250 -610ms).Moreover, systolic murmurs can produce lobes in the first part of the window, so if they are present, they are taken away. After this, it is performed a lobe search based on an ascending amplitude threshold (50 % of the maximum value), and a descending threshold (10% of the maximum). Once the lobes are detected, they are validated, removing the ones that start beyond 430ms, which is physiologically impossible. Finally, S2 is assumed to be in the remaining lobe with major amplitude and its boundaries are located computing backward and forward search from the maximum of the lobe [8].

Step 5. Alternative segmentation of S2 Values

The boundaries estimated in Step 4 are compared with some criteria in order to decide whether they are correct or wrong:

* S2 length must be more than 30ms.
* S2 length must be less than 230ms.
* S2 beginning must be at least 225ms after R-peak.
* S2 beginning cannot be more than 400ms after R-peak.
* There must be at least one lobe detection in Step 4.

If one of the aforesaid criteria is not meet, an alternative segmentation method is used, based on wavelet decomposition and Shannon energy. First, the signal is down sampled to 4000Hz, later; the wavelet decomposition is calculated up to 6th level using the wavelet Daubechies 6 [9]. Then, the energy is computed in the detail coefficients, using the expression (11), with N = 100 and the envelope Ed of the PCG signal is obtained. Ed is analyzed within the same window used in Step 4, and the lobes

corresponding to possible murmurs are removed if they are present. Afterwards, a lobe search is performed in order to find the lobes that exceed certain amplitude threshold (30% of the maximum of Ed) [2]. Then, the lobes beginning before than 265ms after R-peak or beyond 400ms before R-peak are deleted. Finally, if there is no lobe that meets the conditions, it is assigned a lobe with physiological mean values: beginning at 265ms, end at 610ms and maximum point located 440ms after R-peak. The last step is to detect the boundaries of the major lobe as explained in Step 4. Figure1 shows a PCG signal with severe systolic murmur effectively segmented by the system [3].

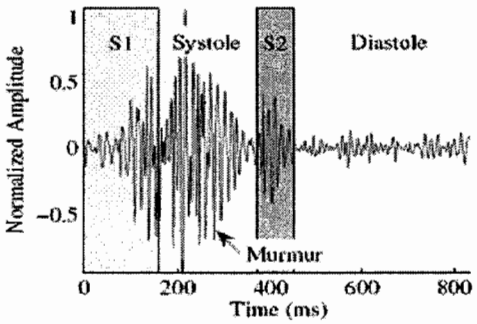


Figure1. Correctly segmented PCG signal with systolic murmur.

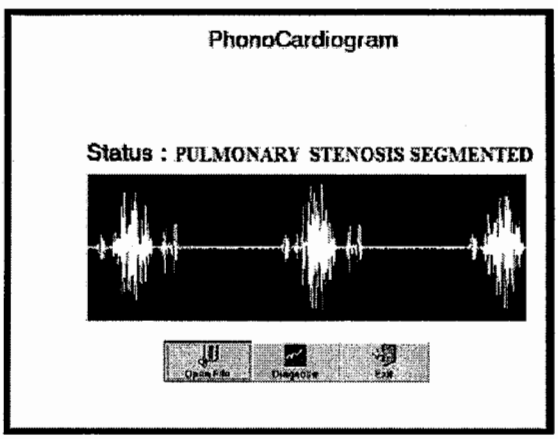


Figure 2. Disordered Patients PCG Signals

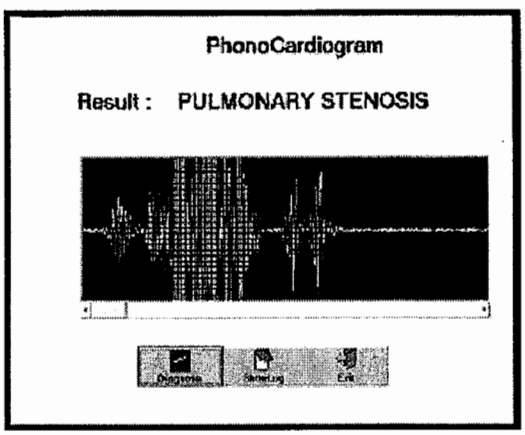


Figure 3. Analyzing the Disordered Portions

3. RESULTS

In this section, we apply the proposed method of PCG Sound signal segmentation, and confirm with a variety of heart sound results that are obtained from the different patients. In addition, the algorithms are tested for different people's data in a PCG database that are labeled by different expert cardiologists. Here the Heart beat separations that are being used by the ECG channel that is of inter-beat segmentation that will give a higher performance rate of 100% in the entire database. In order to evaluate the entire performance of intra-beat segmentation, there are three points that are to be considered and which are being labeled by expert cardiologists: 1. end of S1, 2.beginning of S2 and 3.end of S2. Figure2 shows the location of the three intra-beat segmentation of the heart beat sound of different disordered Patients PCG signals that are being labeled by the expert cardiologists. The end of S1 is situated just about 120ms after R-peak with low distribution. Apart from this the average location that is the beginning of S2 is 350ms after Rpeak and the end of S2 at 430ms. Figure3 show the absolute differentiation in time (ms) of the points that are being calculated by the proposed segmentation algorithm with respect to that of the trained labels. The

performance results for each segmentation event are been shown by the Mean Squared Error (MSE) with respect to the points that are being marked by experts and the number of points whose difference with the label is more than 30ms.Here the total MSE in all intra-beat segmentation points and the total accuracy of the algorithm is 97.7%.

4. CONCLUSIONS

In this paper, we proposed an accurate PCG signal segmentation based on threshold values and the results that are obtained are with a high accuracy rates. Apart from this the proposed methodology has outcome with that of an application and the test results which includes both the threshold based segmentation technique along with the feature estimation methods in order to identify cardiac diseases that includes with the valvular problems. Apart from this the cardiac murmurs whose detection has been identified and the classification of the disorder is done. This method has a good trustworthiness and the results are shown in Figure2 & Figure3 with high precision rate.

REFERENCES

[1] El-Segaier M, Lilja O, Lukkarinen S, Srnmo L, Sepponen R, Pesonen E. Computer-based detection and analysis of heart sound and murmur. Ann Biomed Eng 2005;33(7):937–942.

[2] Liang H, Lukkarinen S, Hartimo I. A heart sound segmentation algorithm using wavelet decomposition and reconstruction. In Proc. 19th IEEE-EMBS. 1997; 1630–1633.

[3] Gupta CN, Palaniappan R,, Swaminathan S, Krishnan SM. Neural network classification of homomorphic segmented heart sounds. Applied Soft Computing 2007;(1):286–297.

[4] Liang H, Lukkarinen S, Hartimo I. Heart sound segmentation algorithm based on heart sound envelogram. In Proc. Computers in Cardiology 1997; 105–108.

[5] Nigam V, Priemer R. Accessing heartdynamics to estimate durations of heartsounds. Physiol Meas 2005;26(6):1005–1018.

[6] Ahlstrom C, Liljefeldt O, Hult P, Ask P. Heart sound cancellation from lung sound recordings using recurrence time statistics and nonlinear prediction. IEEEJSPL 2005; 12(12):812–815.

[7] Laguna P, Srnmo L. Bioelectrical signal processing in cardiac and neurological applications. Elsevier,2005.

[8] Floyd CE, Baker JA, Chatos HG, Delong DM, Ravin CE. Selenium-based digital radiography of the chest: radiologists preference compared with film-screen radiographs. AJR Am J Roentgenol 1995; 165: 1353–1358.

[9] C. Ahlstrom, P. Ask, P. Rask, J-E Karlsson, E. Nylander, U. Dahlstr¨om, and P. Hult. Assessment of suspected aortic stenosis by auto mutual information analysis of murmurs. In 29th Annual International Conference of the Engineering in Medicine and Biology Society (EMBC 2007), pages 1945–1948,2007.

[10] K. H¨oglund, C. Ahlstrom, J. H¨aggstr¨om, P. Ask, P. Hult, and C. Kvart.Time-frequency and complexity analysis – a new method for differentiation of innocent murmurs from heart murmurs caused by aortic stenosis in boxer dogs. A J Vet Res, 68:962–969, 2007.

[11] A. N. Pelech. The physiology of cardiac auscultation. Pediatr Clin North Am, 51:1515–1535, 2004.

[12] D. K. Dey, V. Sundh, and B. Steen. Do systolic murmurs predict mortality in the elderly? a 15-year longitudinal population study of 70-year-olds. Arch Gerontol Geriatr, 38:191–200, 2004.

[13] H. Olsen, E. Vernersson, and T. L¨anne. Cardiovascular response to acutehypovolemia in relation to age. Implications for orthostasis and hemorrhage. Am J Physiol Heart Circ Physiol, 278(1):H222–

[14] A. Johansson, C. Ahlstrom, T. Lanne, and P. Ask. Pulse wave transit time for monitoring respiration rate.Med Biol Eng Comput, 44(6):471–478, 2006.

[15] C. Ahlstrom, A. Johansson, T. L¨anne, and P. Ask. Noninvasive investigation of blood pressure changes using the pulse wave transit time: a novel approach in the monitoring of hemodialysis patients. J Artif Organs, 8:192–197, 2005.

Advances in the biometric recognition methods: A survey on iris and fingerprint recognition

Naser Zaeri* and Fuad Alkoot**

*Faculty of Computer Studies, Arab Open University – Kuwait, P O Box 3322, Safat 13033, Kuwait, E-mail: n.zaeri@aou.edu.kw, **TNI-PAAET, P O Box 4575, Alsalmia, Kuwait, 22046, Email: f_alkoot@yahoo.com*

ABSTRACT

Human recognition based on biometrics finds many important applications in many life sectors and in particular in commercial and law enforcement. This paper aims to give a general overview of the advances in the biometric recognition methods. We concentrate on main methods and accessible ideas presented for human recognition systems based on two types of biometrics: iris and fingerprint. We present a quick overview of the landmark papers that laid the foundation in each track then we present the latest updates and important turns and solutions that developed in each track in the last few years.

Keyword list: iris, fingerprint, biometrics recognition, classification

1. INTRODUCTION

An important and vital method for identification and verification is using the person's physiological or behavioral characteristics, known as biometrics. Some of these characteristics are the person's fingerprints, hand geometry, face, iris, and voice. In the biometrics context, identification or recognition is the task in which, given a biometric signal, the identity of the person related to the signal is retrieved from a database of N subjects. Here we try to match the biometric signal under inspection to a subject in the database. Identification in this case is a one-to-N matching process. Another related issue is the verification or authentication, where a biometric signal and a hypothetical identity are given. The task is to verify or ensure that the signal is related to a given identity. Thus biometric data is to be checked against a claimed identity and in this case we can consider the verification as a one-to-one matching process.

In this paper, we present a survey on two types of biometrics: iris and fingerprint, with a special focus on the recent advances presented in last few years. In the next section we present the pioneering work on iris recognition followed by a survey on the latest advancements on iris recognition in section 3. In section 4 we present the foundation of research on fingerprint followed by the latest advances on fingerprint recognition in section 5. The paper is brought to conclusion in section 6.

2. FOUNDATION OF RESEARCH ON IRIS

The iris is formed in early life in a process called morphogenesis. Once fully formed, the texture is stable throughout life. It is an elastic, pigmented, connective tissue that controls the pupil. It is the only internal human organ visible from the outside and is protected by the cornea. It has a unique pattern, from eye to eye and person to person, making it suitable for automated and highly reliable personal identification. This means that the probability of finding two people with identical iris patterns is almost zero [1]. An automated iris scaner will analyze over 200 points of the iris, such as rings, furrows, freckles, the corona and will compare it to a previously recorded template. In an attempt to

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462Q · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853523

automate identity recognition and verification, Daugman presented a prototype system for iris recognition [2] and reported that it has excellent performance on a diverse database of many images. Also, Wildes *et al.* [3] described a prototype system for personal verification based on automated iris recognition. These recent prototype systems considered a number of implementation issues from the practical point of view. Both the systems of Daugman and Wildes *et al.* concentrated on ensuring that repeated image captures produced irises in the same location within the image, had the same resolution, and were glare-free under fixed illumination. These requirements were essential for the accurate extraction of iris features in order for processing to be successful. The prototype of [3] relied on image registration, which is very computationally demanding. Daugman's system filters transformed images with oriented, quadrature pair, bandpass filters and coarsely quantizes the resulting representation for byte-wise matching. Both systems in [2] and [3] have been much more extensively tested on databases of hundreds of images and have been shown to produce remarkable results. However, many have further attempted to improve the robustness, speed and accuracy of iris recognition methods.

3. NEW DEVELOPMENTS ON IRIS

Boles and B. Boashash [4] proposed a new algorithm for extracting unique features from images of the iris of the human eye and representing these features using the wavelet transform (WT) zero crossings. The advantage of the method presented in [4] is processing 1-D iris signatures rather than the 2-D images as used in both [2] and [3]. However, their technique has been tested on a small number of real images (with and without noise). Further, Daugman *et al.* [1] presented a study where they have assessed the randomness and singularity of iris patterns, and their phenotypic distinctiveness as biometric identifiers, based on video images acquired in public trials of pattern recognition methods proposed in [5]. They have found that the probability of two different irises agreeing by chance in more than 70% of their phase sequence is about one in 7 billion. The detailed phase information was extracted from each isolated iris pattern using complex-valued two-dimensional Gabor wavelets.

Huang *et al.* [6] proposed an iris recognition method which constructs basis function for training set by Independent Component Analysis (ICA), represents iris pattern with ICA coefficients, determines the center of each class by competitive learning mechanism and finally recognizes the pattern based on Euclidian distances. They have shown that their system can work well for blured iris image, variable illumination, and interference of eyelids and eyelashes. Lim *et al.* [7] decomposed an iris image into four levels using 2-D Haar wavelet transform and quantized the fourth-level high-frequency information. Park *et al.* [8] used a directional filter bank to decompose an iris image into eight directional subband outputs and extracted the normalized directional energy as features. Kumar *et al.* [9] utilized correlation filters to measure the consistency of iris images from the same eye. The correlation filter of each class was designed using the two-dimensional (2-D) Fourier transforms of training images. If the correlation output exhibited a sharp peak, the input image was determined to be from an authorized subject, otherwise an imposter. Bae *et al.* [10] projected the iris signals onto a bank of basis vectors derived by independent component analysis and quantized the resulting projection coefficients as features. In [11], the global texture features of the iris were extracted by means of well-known Gabor filters at different scales and orientations. Based on the experimental results and analysis obtained in [11], authors in [12] constructed a bank of spatial filters, whose kernels are suitable for iris recognition, to represent the local texture features of the iris and achieved much better results. From the methods described above, it can be concluded that there are four main approaches to iris representation: phase-based methods [5,12,14], zero-crossing representation [15], [16], texture analysis [17], [18], [7], [11], [12], [8], and intensity variation analysis [10], [19].

To answer the question of which approach is most suitable for extracting iris features, Ma *et al.* [20] carried out extensive quantitative comparison among some existing methods and provided detailed discussions on the overall experimental results. They have discussed that the iris consists of many irregular small blocks, such as freckles, coronas, stripes, furrows, crypts, and so on. Furthermore, the distribution of these blocks in the iris is also random. Such randomly distributed and irregular blocks constitute the most distinguishing characteristics of the iris. As such, local sharp variations denote the most important properties of a signal. In their framework, they recorded the position of local sharp variation points as features instead of locating and recognizing those small blocks. The characteristics of the iris can be considered as a sort of transient signals. Local sharp variations are generally used to characterize the important structures of transient signals.

4. FOUNDATION OF RESEARCH ON FINGERPRINT

The second biometric signature that we survey is the fingerprint. Fingerprint matching is one of the most popular biometric techniques used in automatic personal identification. The researchers' interest in image database representation and content-based retrieval stems from the application domain of fingerprint matching. Law enforcement agencies use it routinely for criminal identification. Now it is also being used in several other applications such as access control for high security installations, credit card usage verification, and employee identification. The FBI database size has grown so large requiring terabytes of storage space.

The main reason for the popularity of fingerprints as a method of identification is that the fingerprint of a person is unique and features used in matching remain invariant with age. The law enforcement agencies have developed a standardized method for manually matching rolled fingerprints and latent or partial fingerprints (lifted from the scene of a crime). However, the manual matching of fingerprints is a highly tedious task, time-consuming and unsuitable for an on-line fingerprint verification system. A fingerprint database is characterized by a large number of records (in the order of millions). In other word, the fingerprint matching complexity is a function of the size of the image database. which can vary from a few hundred records to several million records. As such, automatic fingerprint recognition is essential. For the purpose of automation, a suitable representation (feature extraction) of fingerprints is required. Although significant progress has been made in designing automatic fingerprint identification systems over the past 30 years, a number of design factors (lack of reliable minutia extraction algorithms, difficulty in quantitatively defining a reliable match between fingerprint images, fingerprint classification, etc.) create bottlenecks in achieving the desired performance.

5. NEW DEVELOPMENTS ON FINGERPRINT

A human expert can perform the classification relatively easily. For an automatic system, the problem is much more difficult because the system has to take into account the global ridge directions and also their local connectivity. Many researchers have addressed the fingerprint classification problem and many approaches to automatic fingerprint classification have been presented in the literature and the research on this topic is still very active. The first scientific studies on fingerprint classification were made by Galton [21], who divided the fingerprints into three major classes. Later, Henry [22] refined Galton's classification by increasing the number of the classes. All the classification schemes currently used by police agencies are variants of the so-called *Henry's classification scheme*.

Since the techniques presented by Rao and Black [23], Kawagoe and Tojo [25], Srinivasan and Murthy [24], Blue *et al.* [28], and Kamijo [29]; the PCASYS approach (Pattern-level Classification Automation SYStem) proposed by Candela *et al.* at NIST [26, 27] is, in our opinion, the most promising approach for exclusive classification since it assigns

fingerprints to six non-overlapping classes. Before computing the directional images, the ridge-line area is *separated* from the background and an *enhancement* is performed in the frequency domain. The directional image is then *registered* with respect to the core position which corresponds to the fingerprint center. The dimensionality of the directional image, considered as a vector of 1,680 elements, is reduced to 64 elements by using the *principal component analysis* (KL transform). At this stage, a *PNN* (Probabilistic Neural Network) is used for assigning each 64-element vector to one class of the classification scheme. In order to improve the classification reliability, especially for whorl fingerprints, the authors also implemented an auxiliary module (called *pseudoridge tracer*), which works by analyzing the ridge-line concavity under the core position.

To handle very large databases, Ratha *et al.* [30] described an integrated multilevel approach that uses a content-based retrieval method that can tolerate noisy and distorted query images. Due to the advances in technology a need for online processing of inkless scanner fingerprint images became apparent. Jain *et al.* [31] presented a solution where they have proposed a hierarchical approach to obtain a smooth orientation field estimate of the input fingerprint image, which improves the performance of minutia extraction. For minutia matching, they have proposed an alignment-based elastic matching algorithm. This algorithm is capable of finding the correspondences between minutiae without resorting to an exhaustive search and has the ability to adaptively compensate for the nonlinear deformations and inexact pose transformations between different fingerprints. Chong *et al.* [32] based the classification on the ridge-line geometrical shape. B-spline curves were used to model fingerprint ridge lines; adjacent curves were merged to limit noise artifacts. Classification was performed by tracing the resulting curves in order to detect turns (i.e., complete direction exchanges). An alternative classifier model was proposed by Senior [33] where he used a hidden Markov model classifier whose input features are the measurements (ridge angle, separation, curvature, etc.) taken at the intersection points between some horizontal-vertical fiducial lines and the fingerprint ridge lines.

Cappelli *et al.* [34] presented a new fingerprint classification method which uses dynamic masks for directional image partitioning. They perform a "guided" segmentation of the directional image with the aim of drastically reducing the degrees of freedom during the partitioning process and, consequently, conferring stability to the solutions. Their approach is translation and rotation invariant and it does not require the singularities to be detected.

While computationally efficient, filtering based methods require careful thresholding strategies to avoid detection mistakes. They often cannot be avoided with low-quality fingerprints. Additional problems with filtering methods arise when processing fingerprints that do not contain singular points. Recently, Kryszczuk and Andrzej Drygajlo [36] presented a method in which the singular point detection is performed by analyzing the local quadrant change of the ridge gradient vectors. Singular points (SP) are defined as discontinuities in the directional field [35]. Formally this can be stated as the area where ridges oriented rightwards change to leftwards and those that were oriented upwards turn downwards, and opposite. Their algorithm performs a robust estimation of the local ridge gradient. They employed a modified version of the "*squared average gradients*" [35] to estimate the direction of the smoothed gradient vectors. As opposed to the existing method, they allowed canceling out of the opposite local gradients, achieving a more robust average local ridge gradient estimation. They have demonstrated that their algorithm handles well fingerprints of good and degraded quality alike.

6. CONCLUSION

We have surveyed the latest research on iris and finger print biometric methods. We introduced a quick overview of

pioneering work on both biometrics, and then presented the latest advancements in each field. Attempts have been made on achieving real time processing and speed to handle large databases. It is clear that methods for both biometrics have reached high levels of robustness and speed. Nevertheless, there are still some design factors that worth more investigation in order to achieve the optimal performance.

References

1. Daugman, J. and Downing C., "Epigenetic randomness, complexity and singularity of human iris patterns", The Royal Society, London 268, 1737-1740, (2001).
2. Daugman, J., "High confidence visual recognition of persons by test of statistical independence," IEEE Trans. Pattern Anal. Machine Intell., vol. 15, pp. 1148–1161, Nov. (1993).
3. Wildes, R. et al., "A system for automated iris recognition," in Proc. 2nd IEEE Workshop Applicat. Comput. Vision,. pp. 121–128, Dec. (1994).
4. Boles, W. W. and Boashash, B. "A Human Identification Technique Using Images of the Iris and Wavelet Transform", IEEE TRANSACTIONS ON SIGNAL PROCESSING, VOL. 46, NO. 4, APRIL (1998).
5. Daugman, J., "High confidence visual recognition of persons by a test of statistical independence," IEEE Trans. Pattern Analy. Machine Intell., vol. 15, pp. 1148–1161, Nov. (1993).
6. Huang, Y-P., Luo, S-W., and Chen, E-Y., "An Efficient Iris Recognition System", Proc. 1st Int'l. Conf. Machine Learning and Cybernetics, China, (2002).
7. Lim, S., Lee, K., Byeon, O., and Kim, T., "Efficient iris recognition through improvement of feature vector and classifier," ETRI J., vol. 23, no. 2, pp. 1–70, (2001).
8. Park, C., Lee, J., Smith, M. and Park, K., "Iris-based personal authentication using a normalized directional energy feature," in Proc. 4th Int. Conf. Audio- and Video-Based Biometric Person Authentication, pp. 224–232, (2003).
9. Kumar, B., Xie, C. and Thornton, J. "Iris verification using correlation filters," in Proc. 4th Int. Conf. Audio- and Video-Based Biometric Person Authentication, pp. 697–705, (2003).
10. Bae, K., Noh, S. and Kim, J. "Iris feature extraction using independent component analysis," in Proc. 4th Int. Conf. Audio- and Video-Based Biometric Person Authentication, pp. 838–844, (2003).
11. Ma, L., Wang, Y. and Tan, T. "Iris recognition based on multichannel Gabor filtering," in Proc. 5th Asian Conf. Computer Vision, vol. I, pp. 279–283, (2002).
12. Ma, L., Wang, Y. and Tan, T. "Iris recognition using circular symmetric filters," in Proc. 16th Int. Conf. Pattern Recognition, vol. II, pp. 414–417, (2002).
13. Daugman, J., "Statistical richness of visual phase information: update on recognizing persons by iris patterns," Int. J. Comput. Vis., vol. 45, no. 1, pp. 25–38, (2001).
14. Daugman, J., "Demodulation by complex-valued wavelets for stochastic pattern recognition," Int. J.Wavelets, Multi-Res. and Info. Processing, vol. 1, no. 1, pp. 1–17, (2003).
15. Daugman, J. and Downing, C., "Demodulation, predictive coding, and spatial vision", J. Opt. Soc. Am. A 12, 641-660, (1995).
16. Sanchez-Avila, C. and Sanchez-Reillo, R. "Iris-based biometric recognition using dyadic wavelet transform," IEEE Aerosp. Electron. Sys. Mag., vol. 17, pp. 3–6, Oct. (2002).
17. Wildes, R., Asmuth, J., Green, G., Hsu, S., Kolczynski, R., Matey, J.and McBride, S. "A machine-vision

system for iris recognition," Mach. Vis. Applic., vol. 9, pp. 1–8, (1996).
18. Zhu, Y., Tan, T., and Wang, Y., "Biometric personal identification based on iris patterns," in Proc. Int. Conf. Pattern Recognition, vol. II, , pp. 805–808, (2000).
19. Ma, L. "Personal identification based on iris recognition," Ph.D dissertation, Inst. Automation, Chinese Academy of Sciences, Beijing, China, June (2003).
20. Ma, L., Tan, T., Wang, Y. and Zhang, D. "Efficient Iris Recognition by Characterizing Key Local Variations", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 6, JUNE (2004).
21. Galton, F. [Finger Prints], London: McMillan, (1892).
22. Henry, E.R. [Classification and Uses of Finger Prints], London: Routledge, (1900).
23. Rao, C.V.K. and Black, K. "Type Classification of Fingerprints", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 2, no. 223-231, pp. 31, (1980).
24. Srinivasan, V.S. and Murthy, N.N. "Detection of Singular Points in Fingerprint Images," Pattern Recognition, vol. 25, no. 2, pp. 139-153, Feb. (1992).
25. Kawagoe, M. and Tojo, A. "Fingerprint Pattern Classification, Pattern Recognition, vol. 17, no. 3, pp. 295-303, (1984).
26. Candela, G.T. and Chellappa, R. "Comparative Performance of Classification Methods for Fingerprints," NIST Technical Report NISTIR 5163, Apr. (1993).
27. Candela, G.T. et al., "PCASYS—A Pattern-Level Classification Automation System for Fingerprints." NIST Technical Report NISTIR 5647, Aug.(1995).
28. Blue, J.L., Candela, G.T., Grother, P.J., Chellappa, R. and Wilson, C.L. "Evaluation of Pattern Classifiers for Fingerprint and OCR applications," Pattern Recognition, vol. 27, no. 4, pp. 485-501, Apr. (1994).
29. Kamijo, M. "Classifying Fingerprint Images Using Neural Network: Deriving the Classification State," Proc. Third Int'l Conf. Neural Network, pp. 1,932-1,937, (1993).
30. Ratha, N. K. et al. "A Real-Time Matching System for Large Fingerprint Databases", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 18, NO. 8, AUGUST (1996).
31. Jain, A., Hong, L. and Bolle, R. "On-Line Fingerprint Verification", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 19, NO. 4, APRIL (1997).
32. Chong, M.M.S. et al., "Geometric Framework for Fingerprint Image Classification," Pattern Recognition, vol. 30, no. 9, pp. 1,475- 1,488, (1997).
33. Senior, A. "A Hidden Markov Model Fingerprint Classifier," Proc. 31st Asilomar Conf. Signals, Systems, and Computers, pp. 306-310, (1997).
34. Cappelli, R., Lumini, A., Maio, D., and Maltoni, D. "Fingerprint Classification by Directional Image Partitioning", IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 21, NO. 5, MAY (1999).
35. Liu, T., Hao, P., and Zhang, C. "Fingerprint Singular Points Detection and Direction Estimation with a "T" Shape Model", Proc. of the AVBPA 2005, Springer, Hilton Rye Town, NY, USA, pp. 201-207, July (2005).
36. Kryszczuk, K. and Drygajlo, A. "Singular point detection in fingerprints using quadrant change information", The 18th International Conference on Pattern Recognition (ICPR'06), (2006).

Document Image Database Indexing with Pictorial Dictionary

Mohammad Akbari[a] and Reza Azmi[b*]

[a] Engineering Department,I.A.U., Shahr-e-Qods Branch,Tehran, Iran
[b] Engineering Department, Alzahra University, Tehran, Iran

ABSTRACT

In this paper we introduce a new approach for information retrieval from Persian document image database without using Optical Character Recognition (OCR).At first an attribute called subword upper contour label is defined then, a pictorial dictionary is constructed based on this attribute for the subwords. By this approach we address two issues in document image retrieval: keyword spotting and retrieval according to the document similarities. The proposed methods have been evaluated on a Persian document image database. The results have proved the ability of this approach in document image information retrieval.

Keywords: Document retrieval, Document image database, Pattern recognition ,Image processing, Pictorial dictionary.

1. INTRODUCTION

Modern technology has made it simply possible to produce, process, store and transmit document images efficiently. Many organizations currently use and are depend on document image databases. However, these databases often don't have adequate and proper index information. The technology of Document Image Processing (DIP) can convert digital images of documents to machine-readable text format by using Optical Character Recognition (OCR) technology. However, typically it is not a practical and cost effective approach to process huge number of paper documents. Therefore, you may conclude that the practical solution is storing legacy and traditional documents in image format.

We can use a DIP system for Document Image Retrieval, but the performance of a DIP system relies heavily on the quality of scanned images. It is generally acknowledged that the required recognition accuracy in document image retrieval methods are considerably lower than what is required in many document processing applications [1]. Consequently, it is an alternative way to match directly word images of document image for information retrieval.

1.1 Related Previous Work

In recent years, A lot of efforts regarding document image retrieval research have been reported [2][3]. For example, Chen and Bloomberg [4] have described a method for selecting sentences and keywords to create summery from an image document without any need for recognition of characters of each word. Liu and Jain [5] addressed an approach to image-based form document retrieval. Niyogi and Serihari [6] have introduced an approach to retrieve information from document images stored in a digital library by means of knowledge-based layout analysis and logical structure derivation techniques. Tang et al [7] proposed a method that analyzes geometric and logical structure of image and acquires atomically knowledge from document images.

Spitz described character shape codes for duplicate document detection [8], information retrieval [9]. Yu and Tan [21], proposed a method for text retrieval from document images without the use of OCR. In recent years, some methods are suggested to search words in handwritten [11] and printed documents [12]. Some work have been done on the documents' pages classification and some methods have been proposed but most of these methods require domain-specific models [13]. In most cases, a general class of documents is assumed (such as letters or forms) and classification is reduced to identification of specific instances of the class. Taylor [14] has benefited from geometric-based and content-based features to process heterogeneous batches of documents including business letters, memos, newspapers and technical articles as a part of Intelligent Document Understanding System (IDUS).

* Further author information: (Send correspondence to M. Akbari)
M.Akbari: E-mail: akbarimohammad@gmail.com, Webpage: http://mohammad.akbari.asia
R.Azmi: E-mail: azmi@alzahra.ac.ir

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462R · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856302

In this paper, we introduce a method for retrieving Persian document images from image databases. This method is designed especially for poor quality images. At first words are represented by their upper contour and then are categorized based on their features. Categorized sub-words make an indexing structure called pictorial dictionary. Each sub-word of document is represented by its entry number in pictorial dictionary. Using pictorial dictionary, we make a feature vector for each image. These feature vectors are used in document image retrieval.

3. PICTORAL DICTIONARY AS INDEXING STRUCTURE

Intensive research has been done and now available on OCR and DIR systems[15]. However, Farsi/Arabic texts have main specifications which make them difficult to recognize and also retrieve. Farsi/Arabic texts are cursive and are written from right to left. A Farsi character might have several shapes –from 1 to 4 shapes- depending on its relative position in the word. In addition, some Farsi characters have the same shape and differ from each other only by existing of dots or zigzag bar. A subword is either a single character or a set of connected characters. The neighboring characters, separated or connected, may overlap vertically. These characteristics of Farsi script are shown in Fig. 1. There are many works reported on the recognition of Farsi texts e.g. [16, 17,18]. The main problem in Farsi/Arabic segmentation-based systems is character segmentation where each word or subword is first split into a set of single characters and then is recognized by its individual characters.

We used technique called pictorial dictionary for indexing document images. To produce a pictorial dictionary first we define some features for sub-words and extract the value of these features, and then we put similar sub-words in the same entry in the pictorial dictionary. In fact, each entry in pictorial dictionary is a collection of sub-words that the value of their features is equivalent. The idea of pictorial dictionary has been used in recognizing Latin and Persian words and we have used it to create a content-based feature vector for document image [19, 20].

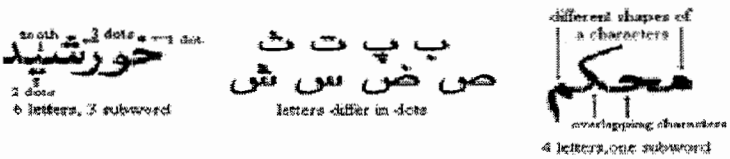


Figure 1. Some characteristics of Farsi/Arabic script

3.1 Upper Contour Labels Pictorial Dictionary

A set of labels are being considered as sub-words upper contour. Labels are specified as below: first, the surrounding contour of sub-word is computed; after that, the upper contour is specified. At the end, considering the position of dots to the base line, upper contour is divided to upper bit path, middle bit path and down bit path. In reference [17], we discussed about the details of computing bit paths and specifying labels. Determining these features includes computing base line, modifying and computing upper and down edges and labeling upper contour by applying state diagram [17] [19] .

We will describe the algorithm in this section briefly. To learn more about the algorithm see [17]. The most frequent size of the black-pixel runs in the vertical histogram of each line columns is adopted as the pen size, w. Each subword is the combination of regions, including bodies, points, zigzag bars, etc. If a region overlaps with base line in some pixels, it is a body (Fig. 2).

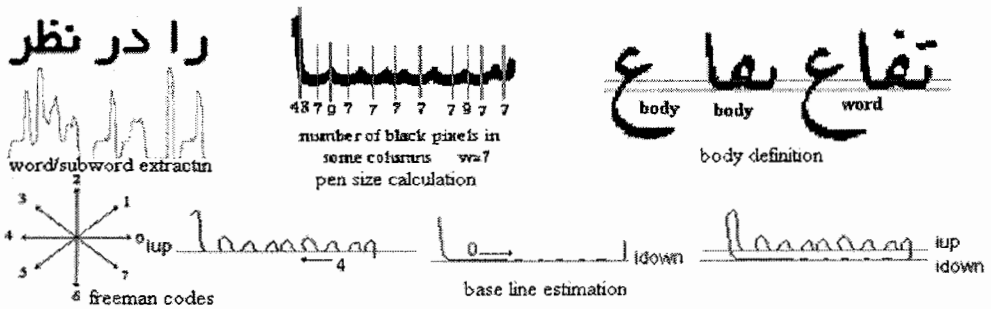


Fig. 2. Preprocessing phase

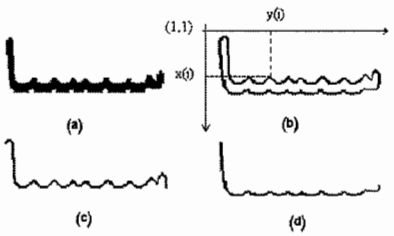


Fig. 3. (a) Body of word. (b) its contour. (c) up contour. (d) down contour.

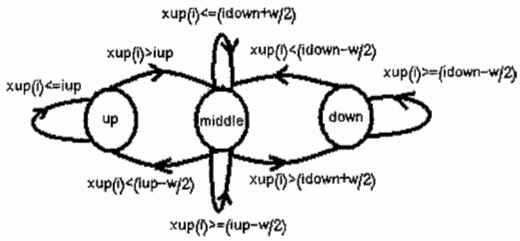


Fig. 4. State diagram of up contour labeling algorithm.

Our pictorial dictionary is based on the conditional labeling of the up contours of each subword (Fig. 3). Tracing the up and down contours from right to left in CCW, each point is labeled -1, 0 and -1 standing for up, middle and down, respectively - depending on its distance from the base line and the label of its preceding point. Figure 4 shows state diagram of up contour labeling algorithm [17, 19]. Fig. 5 shows a sample word and it's labeled up contour.

Fig. 5. A word and its labeled up contour.

4. KEYWORD BROWSING SUB-SYSTEM

The structure of system has been shown in figure 6. As we mentioned before words are basic units of meaning. As a result, producing feature vector according to the body of sub-words is not proper for document indexing. To solve this problem, we used N-Gram algorithm to produce more proper feature vectors. In fact, we applied Bi-Grams of sub-words to produce the feature vector of document. When user starts his/her query, first the query is segmented to its sub-words. Then, the upper contour of sub-word is recognized by means of labeling algorithm and peer entry of the sub-word in pictorial dictionary is recognized. After that, noting the feature vector of each document, we assign if an image contains the Bi-Gram or not. Finally, the documents that contain searched Bi-Grams are represented to the user as the results.

4.2 Performance Evaluation of Keywords Query Sub-System

To evaluate the performance of proposed system in querying keywords, we collected a set of 117 different document images from different sources like students' booklets, conference articles and journal papers from internet. Some of them were text documents so to access their image; first, we printed them and then, took their pictures with the precision of 300 dpi. To evaluate the performance of the system, we chose 150 different keywords and presented them to system as samples. The results were collected and the accuracy of the system was measured. The mean accuracy was 72.48% and the mean recall was 92.17% in our experiments.

5. RETRIEVAL SUB-SYSTEM OF CONTENT-BASED DOCUMENT IMAGES

Querying by means of samples is an applicable technique in image databases. A document image retrieval system should be able to retrieve document images by means of sample. A user may be unaware of correct keywords that should use. Consequently, querying by using a sample document will be useful. In this approach user will present a sample document image to the system and will define necessary terms. The system will retrieve document images according to user's query and defined necessary terms and represents the results. Figure 7 illustrates this retrieval sub-system.

5.1 Production of Document's Feature Vector

Document image is segmented into sub-words via a connected components labeling algorithm. By using pictorial dictionary, bodies of sub-words are grouped and their peer entries in dictionary are recognized. After that, a feature vector is produced for document image that we call it document feature vector or document vector. Each entry of this vector holds the index of one Bi-Gram and its occurrence frequency in document. At last, each occurrence frequency is normalized by dividing it by the total number of Bi-Grams in document. We can model feature vector of document m like equation 1.

In which N is the number of different Bi-Grams in document m. In other words, X_m is an n-dimensional vector and each member of this vector is computed as below:

$$X=X_{m1}\, X_{m2} X_{m3} \ldots X_{mN} \quad \text{(Eq.1)} \qquad X_{mi} = \frac{f(P_{mi})}{\sum_{i=1}^{N} f(P_{mi})} \quad \text{(Eq. 2)}$$

That f is a function that returns the frequency of P_{mi} in document image and N is the number of different Bi-Grams in document.

5.2 Similarity Measure

Feature vectors of similar documents point to the same position. Therefore, we can measure the similarity of two document images by dot product of their vectors, which is equal to the cosine of the angle between two document vectors and is a proper criterion for measuring similarity (difference) between two feature vectors. Hence, the similarity of two documents is measured as below:

$$Similarity\ (X_m, X_n) = \frac{\sum_{j=1}^{J} X_m X_n}{\sqrt{\sum_{j=1}^{J} X_{mj}^2 \sum_{j=1}^{J} X_{nj}^2}} \quad \text{(Eq. 3)}$$

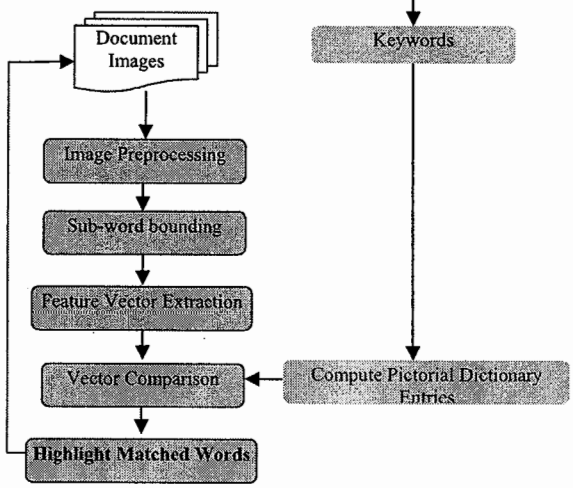


Fig. 6. System diagram for searching a user-specified word in a document image

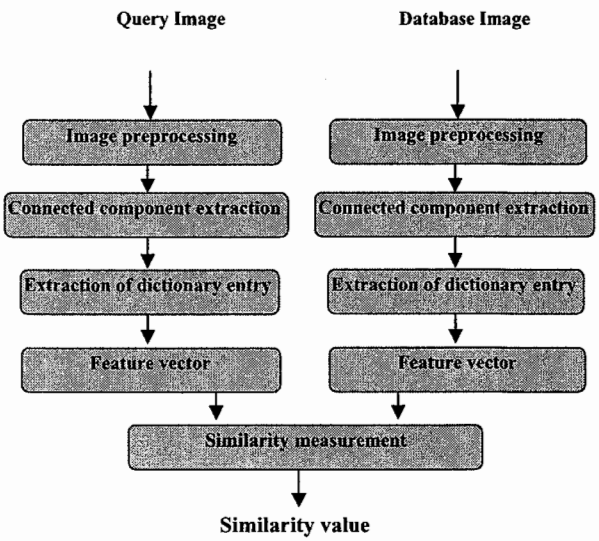


Fig. 7. System diagram for measuring the similarity between two document images.

6. EVALUATION OF RETRIEVAL SUB-SYSTEM OF CONTENT-BASED DOCUMENT IMAGES

To evaluate the efficiency of our suggested approach in retrieving related documents (similar in content), we prepared a set of 147 document images and categorized these documents to eight different groups according to their subject. Then, to run our experiments we made a text version and an image version for each of them. Figure 8 shows the similarity between C5 document in-group C and the three first documents of each group. We can retrieve documents of group C (C1, C2, C3) if we choose C5 as query document and pick up a proper similarity threshold.

To have a proper criterion for performance evaluation, we selected the first image of each group as query image and compared them with other images of database. Then, we computed accuracy and recall value for each of these queries. These values are also computed for content-based N-Gram method. The results are shown in table 1. We also computed F1 compound rate for each query. This criterion is used in many retrieval algorithms and is defined as:

$$F1 = \frac{2RP}{R+P} \qquad (4)$$

We gain F1 rate from both accuracy and recall rates, and compute it for both content-based and image-based methods. The results of retrieval performance evaluation for various similarity thresholds are shown in table 1. The mean of accuracy and recall for threshold 0.8 are 67.05% and 48.37%. By choosing 0.7 for threshold, the mean of accuracy and recall will be 47.93% and 78.29%. The standard value of F1 for threshold 0.8 is 54.58% and for threshold 0.7, is 55.22%. However, the standard value of F1 for threshold 0.75 is 55.26% that is more than its value for both 0.7 and 0.8 thresholds. This shows that 0.75 is a proper choice for threshold. Table 1 shows even content-based method has low performance for subject five.

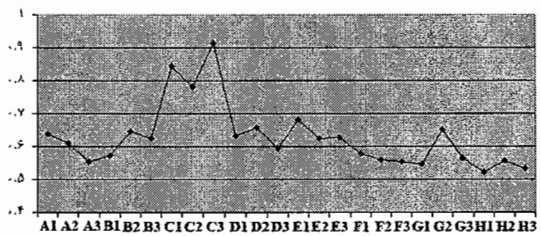


Fig. 8. The similarity with document C5.

Document Type	Topic id	NR	Threshold = 0.8			Threshold = 0.75			Threshold = 0.7		
			R%	P%	F_1%	R%	P%	F_1%	R%	P%	F_1%
Image Based	A	21	33.33	100	49.99	57.14	66.67	61.54	85.71	42.86	57.1
	B	19	21.05	33.33	25.80	36.84	25.92	30.43	68.42	26	37.6
	C	17	64.70	61.11	62.85	82.35	45.16	58.33	94.12	23.88	38.0
	D	17	64.70	78.57	70.96	70.59	70.59	70.59	76.47	56.52	64.9
	E	19	10.53	28.57	15.39	15.79	12	13.64	47.37	14.51	22.2
	F	20	75	100	85.71	80	100	88.89	95	100	97.4
	G	17	52.94	42.86	47.37	58.82	35.71	44.44	70.59	19.67	30.7
	H	17	64.71	100	78.57	82.35	100	90.32	88.23	100	93.7
	Average		48.37	68.05	54.58	60.48	57.01	57.27	78.29	47.93	55.2
Text Based	A	21	47.62	100	64.52	80.95	89.47	84.99	90.48	70.37	79.1
	B	19	84.21	100	91.43	94.74	100	97.30	100	41.30	58.4
	C	17	94.12	88.89	91.43	100	41.46	58.62	100	16.04	27.6
	D	17	88.23	88.23	88.23	100	48.57	65.38	100	14.17	24.8
	E	19	5.26	33.33	9.09	10.52	18.18	13.33	94.74	25.71	40.4
	F	20	75	100	85.71	100	100	100	100	100	10
	G	17	70.59	92.31	80.00	82.35	93.33	87.49	94.12	53.33	68.0
	H	17	94.12	100	96.97	100	100	100	100	27.41	43.0
	Average		69.89	87.84	75.92	83.57	73.88	75.89	97.42	43.54	55.2

7. CONCLUSION AND FUTURE WORKS

Document images has become a very important resource in today modern society and information retrieval from document image databases has become an important subject in knowledge and data engineering. Information retrieval without using OCR has its practical value but it also has its challenging problems. In this paper, we proposed new methods for keyword search problem and measuring document similarities problem by using pictorial dictionary.

REFERENCES

[1] K. Tagvam, J. Borsack, A. Condir, and S. Erva, "The Effects of Noisy Data on Text Retrieval", J. Am. Soc. For Information Science, vol. 45, no. 1, pp. 50-58, 1994.

[2] D. Doermann, "The Indexing and Retrieval of Document Images: A Survey", Computer Vision and Image Understanding, vol. 70, no. 30, pp. 287-298, 1998.

[3] M. Mitra and B.B. Chaudhuri, "Information Retrieval from Documents: A Survey", Information Retrieval, vol. 2, nos 2/3, pp. 141-163, 2000.

[4] F.R. Chen and D.S. Bloomberg, "Summarization of Imaged Documents without OCR", Computer Vision and Image Understanding, vol. 70, no. 3, pp. 307-319, 1998.

[5] J. Liu and A.K. Jain, "Image-Based Form Document Retrieval", Pattern Recognition, vol. 33, no. 3, 2000.

[6] D. Niyogi and S. Srihari, "The Use of Document Structure Analysis to Retrieve Information from Documents in Digital Libraries", Proc. SPIE, Document Recognition IV, vol. 3027, pp. 207-218, 1997.

[7] Y.Y. Tang, C.D. Yan, and C.Y. Suen, "Document Processing for Automatic Knowledge Acquisition", IEEE Trans. Knowledge and Data Eng., vol. 6, no. 1, pp. 3-21, Feb. 1994

[8] A.L. Spitz, "Duplicate Document Detection", Proc. SPIE, Document Recognition IV, vol. 3027, pp. 88-94, 1997.

[9] A.F. Smeaton and A.L. Spitz, "Using Character Shape Coding for Information retrieval", Proc. Fourth Int'l Conf. Document Analysis and Recognition, pp. 974-978, 1997.

[10] Z. Yu and C.L. Tan, "Image-Based Document Vectors for Text Retrieval", Proc. 15th Int'l Conf. Pattern Recognition, vol. 4, pp. 393-396, 2000.

[11] A. Kolcz, J. Alspector, M. Augusteijn, R. Carlson, and G.V. Popescu, "A Line-Oriented Approach to Word Spotting in Handwritten Documents", Pattern Analysis and Applications, vol. 3, no.2, pp. 153-168, 2000.

[12] Kuo and O.F. Agazzi, "Keyword Spotting in Poorly Printed Documents Using Pseudo 2-D Hidden Markov Models", IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 8, pp. 842-848, Aug. 1994.

[13] G. Maderlechner, P. Suda, T.Bruckner, "Classification of documents by form and content", Pattern Recognition Letters, 18: 1225-1231, 1997

[14] S.L. Taylor, M. Lipshutz, R.W. Nilson, "Classification and functional decomposition of business documents", In Proceedings of International Conference on Document Analysis and Recognition, pp. 563-566, 1995

[15] J.J Hull, S.N. Srihari, "A computational approach to visual word recognition: hypothesis generation and testing", Proceedings of IEEE Conference on Computer Vision and PatternRecognition, CVPR'86, Washington, DC, (1986), pp. 156-161, 1986.

[16] A. Amin, "Off-line Arabic character recognition: the state of the art". Pattern Recognitiou 31, (1998), pp. 517-530.

[17] R. Azmi, "Recognition of omnifont printed Farsi text". Ph.D. Thesis, Tarbiat Modarres University, Tehran, (1999).

[18] R. Azmi, E. Kabir, "A recognition algorithm for hand printed Farsi characters".Proceedings of the International Conference on Telecommunication, ICT '96, Istanbul, (1996), pp. 852-855.

[19] R. Azmi, E. Kabir, "Creation of pictorial dictionary for Persian sub-words with visual feature and fuzzy labell".Proceedings of the 6th Iranian Conference on Electerical Engineering, Iran, 2001, pp 147-1

[20] M. Akbari , R. Azmi , 'Information Indexing and Retrieval from Document Image Database', 14th Iranian Conference on Electrical Engineering, May 2006, Tehran, Iran.

Classification of Cast Iron Based on Graphite Grain Morphology using Neural Network Approach

Pattan Prakash C, V.D. Mytri, P.S. Hiremath

System Analyst, PDA College of Engineering,Gulbarga, India
prakashpattan@gmail.com
GND College of Engineering, Bidar, India
vdmytri@yahoo.com
Professor, Department of Computer Science,
Gulbarga University, Gulbarga, India
hiremath53@yahoo.com

ABSTRACT

The ISO-945[2] committee has defined six classes of grain morphology through reference drawings for cast iron graphite grain classification. These reference drawings are universally accepted for classification of graphite grains. The main aim of this work is to propose a neural network approach for cast iron classification based on graphite grain morphology by processing microstructure images. The two sets of shape features investigated are, Simple Shape Descriptors (SSDs) and Moment Invariants(MIs). The classifiers like, feed forward neural network with back propagation and radial basis functions are also investigated. The experimentation is carried out using the metallographic images from the well known microstructures library[4]. For training and testing the networks, the grain shapes identified in ISO-945 reference drawings and the grain classification by the experts are used. The moment invariant shape features and neural network classifier with radial basis function yield better classification results for graphite grains.

Feed-forward neural network, radial basis function, morphology, shape descriptors, moment invariants, ISO-945.

1. INTRODUCTION

The classification of objects into groups based on their close proximity of features is an ever challenging task for human being. The accuracy is a common suffering factor in human recognition system. It is due to physiological limits and variations from one person to other person. Such an automated classification of the objects is one of the common tasks in computer vision applications. For the recognition of the objects, the objects are represented using shape descriptors. An efficient set of shape descriptors and a suitable classifier are of paramount importance in object classification process. Visual inspection by human expert method always a challenging job and results are not consistent due to many physiological limits. Many efforts have been made in this area to achieve the optimum recognition rate. The simple shape descriptors (SSDs) are the general choice of many for extracting the features. The comparison of shape descriptors is given in[5] and it is observed that SSDs with Euclidian distance classifier are efficient features. Out of many features discussed, not all the features are suitable for recognition purpose. Choosing the most suitable features for the case in hand is a challenge. The Fourier descriptors[6] are popularly used as object features. The curvature scale space representation[7] is also are the popular options for shape representation. The moment invariants (MIs) are also of high interest because of their invariance property[1]. M.K. Hu[10] has defined seven moment invariants which are applied for object recognition such as content-based image retrieval[4]. Another method, known as local moment invariant analysis[11], is also of interest of many authors. With this motivation and background we propose a most suitable method for classification of grains in to six classes from a given microstructure image, based on, the features extracted from reference drawings proposed by ISO-945[2].

In the present study, we have employed two shape features, namely, simple shape descriptors(SSDs) and moment invariants (MIs) for classification of graphite grains in microstructure images of cast iron using neural network approach. The experimentation has been carried out using features extracted from ISO-945 reference drawings[2] for training neural

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462S · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853286

networks with back propagation and radial basis functions, and digital metallographic images from the well known microstructures libraries[4,9] for classification (testing). The experimental results demonstrate the efficiency of the proposed method.

1.1 The Domain: Classes of grains in cast iron

When the carbon is diluted in iron at molten state with an amount of more than 2.1%, it precipitates to form graphite particles. The shape of these particles can vary widely depending on the presence of other impurities and on the cooling rate. The thermo-mechanical properties of cast-iron are strongly dependent on these shapes. For example, nodular cast iron, in which the particles are merely spherical in shape, is normally less brittle than gray iron, where sharp graphite flakes contribute to stress concentration and crack initiation. Thus, it is relevant to classify cast-iron according to graphite grain shape. The ISO-945 standard defines six classes for cast iron, based on graphite grain shape. These classes are shown in Fig.1. In the manual method, an expert chooses the correct class through visual comparison with the reference drawings. Manual method produces varying results in many cases. Instead, the digital image processing and analysis method produces consistent and accurate results[2].

Fig. 1. Reference drawings for the ISO-945 defined six classes of grains in cast iron

1.2 Shape features

We choose two types of shape descriptors for grain classification, namely, simple shape descriptors and moment invariants which are described below.
The Simple Shape Descriptors(SSDs): The definitions of the simple shape descriptors[3] are,
Roundness:

$$Roundness = \frac{4Area}{\pi Length^2_{\max}} \quad (1)$$

where $Length_{\max}$ is the maximum length between two boundary points of the region.

Eccentricity: It is the ratio of the length of the longest chord of the shape to the longest chord perpendicular to it.

$$Eccentricity = \frac{Length_{major\_axis}}{Length_{\min or\_axis}} \quad (2)$$

Elongation: It is the ratio of the height and width of a rotated minimal bounding box.

$$Elongation = \frac{height_{bounding\_box}}{width_{bounding\_box}} \quad (3)$$

Rectangularity: It is the ratio of area of the object to the area of minimum bounding rectangle around the object.

$$Rectangularity = \frac{Area_{Object}}{Area_{bounding\_rectangle}} \quad (4)$$

Tortousity: It is the ratio of contour length to the maximum linear length between any two points on the contour.

$$Tortousity = \frac{Length_{\max}}{Length_{contour}} \quad (5)$$

where $Length_{\max}$ is the maximum length between two boundary points of the region and $Length_{contour}$ is the length of the contour of the region boundary.

Moment Invariants (MIs):

The seven moment invariants I_1 to I_7 proposed by Hu[10], which are invariant under translation, rotation and scaling, are defined as:

$$I_1 = \mu_{20} + \mu_{02} \quad (6)$$

$$I_2 = (\mu_{20} + \mu_{02})^2 + 4\mu_{11}^2 \quad (7)$$

$$I_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \quad (8)$$

$$I_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \quad (9)$$

$$\begin{aligned} I_5 = {} & (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})\left[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2\right] \\ & + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})\left[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2\right] \end{aligned} \quad (10)$$

$$\begin{aligned} I_6 = {} & (\mu_{20} - \mu_{02})\left[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2\right] + 4\mu_{11} \\ & (\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \ . \end{aligned} \quad (11)$$

$$\begin{aligned} I_7 = {} & (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})\left[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2\right] \\ & - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})\left[3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2\right] \end{aligned} \quad (12)$$

where,

$$\mu_{pq} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(x - \bar{x}\right)^p\left(y - \bar{y}\right)^q f(x,y)dxdy \quad (13)$$

here, $\bar{x} = \frac{\mu_{10}}{\mu_{00}}, \bar{y} = \frac{\mu_{01}}{\mu_{00}}$ are the components of the region's centroid and

$f(x,y)$ is the intensity function.

Moments η_{ij}, where $i + j \geq 2$, can be constructed to be invariant to both translation and scale by dividing the corresponding central moment by the property scaled $(00)^{th}$ moment, using the relation,

$$\eta_{ij} = \frac{\mu_{ij}}{\mu_{00}^{(1+\frac{i+j}{2})}} \quad (14)$$

1.3 Materials used

We have used the microstructure images of cast iron, that have been taken from steel bars, that have all types of graphite grain structures. The etching medium used for preparing specimen is 3% alcoholic nitric acid. The images are drawn from microstructure libraries[4] for experimentation. The microstructure images used for testing are of cast iron with

varying compositions and magnifications. The ISO-945 reference drawings of grain morphology[2] are used for training neural network.

2. PROPOSED METHOD

The proposed method is based on neural network approach and has two phases: training and testing. Each phase consists of preprocessing and feature extraction stages as described below.

2.1 Preprocessing

Generally, the microstructure images suffer from noise and artifacts developed at the time of specimen preparation. This stage is of high importance in achieving good results in segmentation and further process. We have used circular averaging filter with radius 2, followed by the morphological closing with fixed structuring element, 'disk' of radius 2, in order to de-noise the image.

We have used the watershed segmentation method for segmenting the graphite grains from the background ferrite matrix[12]. The segmentation of the image starts from the local minima of the grey tone image. In some cases, images may be over segmented due to improper preprocessing method which might not have removed even insignificant noise. Watershed is noise sensitive. Tiny and insignificant noise also forms its own catchments basin. Hence, noise free image is an important precaution to be taken when watershed segment method is used for segmentation. This problem is overcome by denoising the image by average filtering and morphological operators. The segmented image is labeled. Each labeled segment is a grain structure which is subjected to the shape feature extraction.

2.2 Neural network classifier

We use the neural network classifier for grains classification in the segmented image. The training and classification phases of neural network are given in the following algorithms.

Phase 1: Neural network training

Step 1: Input grayscale microstructure image (training image).
Step 2: Perform preprocessing (averaging, morphological operations and watershed segmentation) and obtain segmented image. The segmented regions are known grains.
Step 3: Compute the shape features for each labeled region (grain), which are of known grain-class.
Step 4: Input the shape features computed in step 3 as inputs to the neural network.
Step 5: Repeat the steps 1 to 4 for all training images.

Phase 2: Neural network classifier

Step 1: Input grayscale microstructure image (test image).
Step 2: Perform preprocessing (averaging, morphological operations and watershed segmentation) and obtain segmented image.
Step 3: Compute the shape features for each labeled region.
Step 4: (Classification)

Input the shape features computed in step 3 for a region to the neural network which is trained using the Algorithm 1 for grains classification.

Step 5: The output of neural network indicates the grain class to which the region belongs.
Step 6: Repeat steps 4 and 5 for all regions of the segmented image.

3. EXPERIMENTAL RESULTS AND DISCUSSION

For the experimentation, the neural network is trained using six ISO 945 reference drawings with six grain classes[2]. The grain class I,II,III,IV,V and VI contain 35,18,53,24,19 and 19 grains, respectively. The images are of size 525x525 and the sample images are shown in the Fig.1. During the testing phase, 50 microstructure images, each with grains of

different types containing 25 grains on an average, are used. These images are drawn from the microstructure library[4] and the sample microstructure images are shown in the Fig. 2.

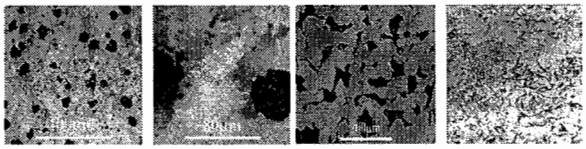


Fig. 2. Sample Microstructure images of Cast iron showing various types of grain structures.

Out of many SSDs defined in the literature, the five SSDs and the first four MIs defined in the section 1.2, employed by the neural network classifier with back propagation or radial basis function, have yielded better classification results during our experimentation. For experimentation, the following neural networks are configured and their performance is analyzed.

Experiment #NN1: We have trained the back propagation neural network with simple shape descriptors (SSDs) as shape features. The classification is tested with microstructure test images which are drawn from microstructure library[4]. The average classification rate of six grain classes is about 73.83%.

Experiment #NN2: We have trained the neural network with radial basis function using simple shape descriptors (SSDs) and tested using test microstructure images. The average classification rate of six grain classes is about 79.66%.

Experiment #NN3: First four moment invariants (I_1 to I_4) are used as inputs for training the back propagation neural network. The classification is tested with microstructure test images. It has yielded a classification rate of 76.66%.

Experiment #NN4: The radial basis neural network is trained with first four moment invariant features. Then the classification is tested using test microstructure images. It is observed that the average classification rate of all grain classes is about 87%, which is a better result as compared to all the previous three experiments.

Also, it is observed a low rate of classification in the case of grain class IV and class V. It is due to their very similar shape features. Such a low rate of distinction of classes IV and V is observed with experts also. The training performance of such a neural network is shown in the Fig.3. Thus, a radial basis neural network with first four moment invariants as shape features has yielded optimal classification of graphite grains. The comparison of classification results are given in Table 3 and Fig.4. The proposed grain classification method is automatic, efficient and fast compared to manual system. The expert's intervention is limited to only building the neural network and the knowledge-base of grain classes.

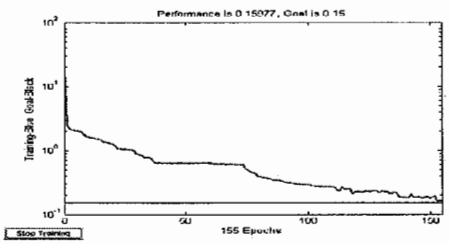


Fig. 3. Training performance of neural network with moment invariants and radial basis

Table 1. Comparison of classification rates for different neural networks and feature sets

Shape Features	Neural Network	
	Back propagation	Radial basis
SSDs	73.83%	79.66%
MIs	76.66%	87%

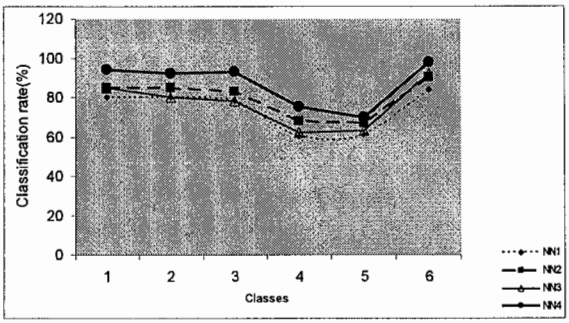


Fig. 4. Performance of the neural network classifier with four combinations of shape features and neural network functions.

The architecture of neural network is described below:
In the case of back propagation neural networks, the input layer has 5 neurons and 5 shape features as inputs, and output layer has one output (grain class number). The transfer function used is 'tan sigmoidal', training function used is Levenberg-Marquardt back propagation, the weight/bias learning function is 'gradient descent' function and the performance function is 'mean square error(mse)' which is set to 0.01. In the case of radial basis neural network, the shape features are used as inputs. The error function is 'mean square error(mse)' which is set to 0.15. The spread for radial basis function is 1.0 and the maximum number of neurons allowed to add during training is 300. The Fig. 4 shows the performance of the optimal classification achieved by the combination of moment invariants and radial basis neural network for each grain class.

4. CONCLUSION

A novel efficient automatic method for classification of cast iron based on graphite grains morphology using neural network approach is proposed. The method is robust and invariant under geometry, lighting transforms, quality and variety of magnifications of microstructure images. The experimental results show that the moment invariant features and radial basis neural network classifier yield better classification performance. The results can be further improved by using suitable preprocessing methods, feature sets and classifiers. This will be considered in our feature work

.

5. REFERENCES

[1] Jain A.K., "Fundamentals of Digital Image Processing". Prentice-Hall, Englewood Cliffs, NJ (1989).
[2] Handbook Committee, Hand book of ASM International, Vol 9, Metallography and Microstructures, ISBN:0-87170-706-3.
[3] Jamil N, Abu Bakar Z, & Sembok T.M.T., "Image Retrieval of Songket Motifs using Simple Shape Descriptors", in Proceedings of the Geometric Modeling and Imaging – New Trends, 171-176 (2006).
[4] Microstructures library: http://www.metalograf.de/start-eng.htm
[5] Sarfraz M and Ridha A, "Content-based Image Retrieval using Multiple Shape Descriptors", IEEE 1-4244-1031-2/07 (2007).
[6] Persoon and Fu ," Shape discrimination using Fourier descriptors, IEEE transactions on Systems, Man and Cybernetics", 7:170-179 (1977).
[7] Abbasi S. and Mokhtarian F., "Robustness of Shape Similarity Retrieval under Affine Transformation", in Challenge of Image Retrieval, Newcastle upon Type, UK, Feb. 25-26 (1999).
[8] Maitra S., "Moment invariants", Proceedings IEEE, 67(4):697-699 (1979).
[9] University of Cambridge , Contributed Microstructures library: http://www.doitpoms.ac.uk
[10] Hu M.K , "Visual pattern recognition by moment invariants". IRE Transactions Information Theory, 8(2):179-197 (1962).
[11] Yufeng Chen, Mandun Zhang, Peng Lu, Yangsheng Wang, "Local Moment Invariant Analysis", IEEE Proceedings of the "Computer Graphics, Imaging and Vision : New Trends(CGIV'05), 0-7695-2392-7/05 (2005).
[12] Xaviour Arnould, Michel Coster, Jean-Louis Chermant, Liliane Chermant, Thierry Chartier and Abder Elmoataz, "Segmentation and Grain Size of Ceramics", Image Anal Stereol, France ,131-135 (2001).

MULTI-HOP PATH TRACING OF MOBILE ROBOT WITH MULTIRANGE IMAGE

RamakantaChoudhury(Sr.Lecturer)M..A.I.T,Delhirkchoudhury1@gmail.com
Chandrakanta Samal (Asst.Professor) A.N.Dev College,Delhi Universitycksamal@gmail.com
DrUmakanta Choudhury(Professor)IdealCollege Ghaziabad(U.P)umakanta.choudhury@gmail.com

ABSTRACT

It is well known that image processing depends heavily upon image representation technique . This paper intends to find out the optimal path of mobile robots for a specified area where obstacles are predefined as well as modified. Here the optimal path is represented by using the Quad tree method. Since there has been rising interest in the use of quad tree, we have tried to use the successive subdivision of images into quadrants from which the quad tree is developed. In the quad tree, obstacles-free area and the partial filled area are represented with different notations. After development of quad tree the algorithm is used to find the optimal path by employing neighbor finding technique, with a view to move the robot from the source to destination. The algorithm, here , permeates through the entire tree, and tries to locate the common ancestor for computation. The computation and the algorithm, aim at easing the ability of the robot to trace the optimal path with the help of adjacencies between the neighboring nodes as well as determining such adjacencies in the horizontal, vertical and diagonal directions. In this paper efforts have been made to determine the movement of the adjacent block in the quad tree and to detect the transition between the blocks equal size and finally generate the result.

GENITOR, QUADTREE, SHAREDSIDE, MIRROR, ABUT, PROGENYTYPE, NEIGHBOUR FINDING

1. INTRODUCTION

At the core of mobile robot the problem of automatic collision free path tracing representation for manipulators [11] is a more well-known case, but path tracing for mobile robot is not exactly the same as this. Some of the draw backs are : A mobile robot should keep as far away from obstacles as possible. It has to avoid collision and get an optimal path. Mobile robot will ordinarily negotiate any given path only once, as opposed to a manipulator, which might perform the same task thousands of times. This implies that it is more important to develop a negotiable path quickly than to develop an "optimal path". This is usually a costly operation. The mobile robot cannot make a complete picture of the surroundings and hence it cannot distinctly identify spatial spread of the object. The conventional path tracing can be divided broadly into two categories. In the first category are the methods which make trivial changes to the representation of the image map before planning a path. The regular grid search and vertex graph methods [7] [17][1] fall in this category. The second method makes elaborate representational changes to convert to a representation, which is easier to analyze before planning the path. A potential and practical shortcoming of such method for mobile robot navigation is that the path tracing cost is still very high as it involved representation conversion process. [1-8, 11].

A quad-tree-based hierarchical representation method of path tracing for mobile robots has been elaborated in this paper,Further; a staged search based method for initializing the hierarchical method of representation has been described. The first part is a Quad tree approach based on A* search and the second part, a staged path tracing algorithm, which has computational advantage as compared to the pure A* search on Quad tree. Yellow blocks are introduced in the staged search method.

Let us focus on the basics of quad tree. A quad tree is a two dimensional picture and uniformly colored on $2^i \times 2^i$ blocks. A block of a quad tree has been represented on $2^i \times 2^i$ square region of the picture. A free block of a Quad tree is a block representing a region of free space. An obstacle block is a block representing a region having a mixture of free space and obstacles. A leaf block of a Quad tree is the tip block of the tree. In an ordinary Quad tree, leaf blocks are free blocks, but in pruned Quad tree, they may also be Yellow blocks. For any block G and S (G) denotes the sub tree. For the block G and L (G) denotes the number of leaf blocks in S (G). The Yellow content of a Yellow block G is defined as the number of obstacle pixels in the region represented by G and the Yellowness of G is the percentage of obstacles in that region. When some of the Yellow blocks of a Quad tree are made leaf blocks, the resulting structure is called pruned Quad tree . It represents same space as the original Quad tree but with reduced resolution.

1.1 Verbiage

A* can be regarded as minimum cost graph algorithm with optimality property .In this approach an open block is not expanded. This improves the efficiency of the algorithm. In the algorithm search process, the block that expands correctly is considered as the block.During evaluation such a block is able to trace the shortest path. The predecessor of a block N in the search graph is the block preceding N on the current best path to N (for the start block).

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462T · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856309

2 . *QUADTREE ASSOCIATED PATH TRACING*

We have developed an algorithm for mobile robot path tracing based on a quad tree representation of the robot's immediate environment. If there are large areas of free space obstacle, then those can be represented by a few large blocks in the quad tree and can be dealt by the algorithm. Given a binary array or raster representation of robot's immediate environment obstacles grow by the radius of robot's cross section by converting the raster into Quad tree conversion algorithm. This algorithm is of complexity O (n), where *n* is the number of pixels in the image being converted. In the resulting Quad tree , blocks of zero represent free space blocks.

In the second spacer we have computed the distance transform of the set of 0's i.e., the free space blocks. This determines, for each block of free space, the minimal distance between the center of that block and the boundary of a block of obstacles. An algorithm is introduced for computing the distance transform for Quad tree which is of complexity O (n), where n is now the number of blocks in the quad tree.

2.1 PATH PLANNING ALGORITHM

Given the start and goal points. We first determine the quad tree leaf blocks 'S' (start) and 'G' (goal blocks), representing the image containing these points. Next we plan a minimal cost path between S (start) and G (goal) in the graph formed by the non obstacle leaf blocks of the quad tree, using the A* search algorithm with the evaluation function f of a block c defined as f (c)=g(c)+ h(c).Here g (c) represents the cost of the path from s to c and h(c) represents the heuristic estimate of the cost of remaining path from C and G. Since the cost of the path depends both on the actual distance travelled along the path and clearance of the path from the obstacle, we define g (c) as g(c) = g(p) + g'(p, c).Where g (p) is the cost of the path from S to C predecessor P on the path line and g' (p, c) is the cost of the path segment between p and c. The letter faction is defined gap'(p ,c)= D(p, c)+ α d(c) With D (p, c) representing the actual distance between the blocks p and c, given as half of the sum of block sizes, and d (c) representing the cost incurred by including block c on the path, d (c) depends upon the clearance of the block c from the near by obstacle shape for the cost function d, defining d(c) as . d(c) = O_{max} - O(c)where O(c) is the distance of the block C from the nearest obstacle given by the quad tree distance transform and O_{max} is the maximum such distance for any block in the quad tree so that d(c) is always positive. α is the equation for g' (p, c) positive constant which determines how the resultant path will avoid obstacles. Finally h(c) calculated as the Euclidian distance between the midpoints of the region represented by C and G. Clearly this measure is a lower bound on the actual minimum cost path between C and G ; thus on A* search with this measure as its heuristic estimate is admissible. The power of this heuristic depends on the average deviation of the minimum deviation cost path from the straight line path. It is highest when cost α is zero and decreases as α Increases. It is of course possible to use more informed, but inadmissible heuristic to speed up this search.Example: Number of obstacles intersecting the straight line path between C and G the total area of the obstacle intersecting the straight line path C and G are more powerful heuristics then the one we are using, but they are not admissible. The block expansion processes involving finding the non obstacle leaf blocks adjacent to the block being expanded .We accomplished this by using neighbor finding strategy. First, only the neighbors in the horizontal and vertical direction are considered .Diagonal neighbors which share only single points with the current block would result in inflexible corners.

Second, when one of the neighbors given by the quad tree neighbor finding algorithm is a grey code, we find the mono-stable leaf blocks if any ,of the quad tree rooted at that grey block are adjacent to the block being expanded and considered them as neighbors. The result applying the above A* algorithm to the quad tree is a list of blocks from the quad tree which defines a set of path between the short and goal blocks. If desired an optimal path , these blocks can be computed or the center points of consecutive blocks on the list can be connected to compute a negotiable path.

2.1.1 REPRESENTATION

The quad tree is an approach to image representation based on the successive subdivision of image into quadrants. It is represented by a tree of out degree 4 in which the root represents a block and four PROGENIES represent in order the NW ,NE,SW,SE quadrants . We assume that each block is stored as a record containing six fields. The first field s containing pointer to the block's GENITOR and it's four PROGENIES. which corresponds to the four quadrants. If P is a block and I is a quadrant then these fields are referenced as GENITOR (P, I) respectively. We can determine specific quadrant in which a block say p, lies relative to its GENITOR by use of the function PROGENYTYPE (P) ,which has a value of I if PROGENY(GENITOR(P),I = P. The six field, BLOCKTYPE, describes the contents of block of the image which

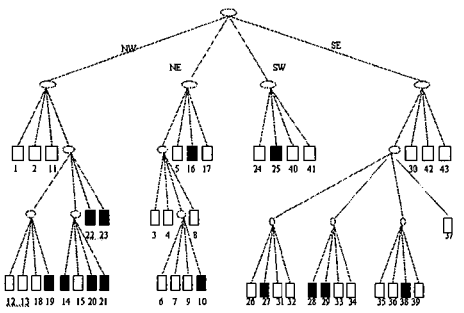


FIG1(b) Quad tree representation of the blocks

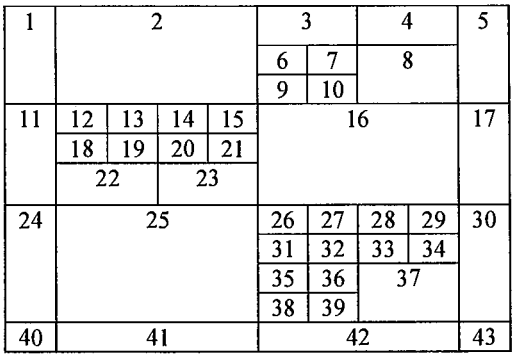


FIG1(a) Dividing Block into equal size quadrants

The block represents i.e., RED if the block contains number of 1's , VIOLET if the block contain only 1's and Yellow if it contain 0's and 1's.alternatively,VIOLET and RED blocks are terminal blocks ,while YELLOW are non terminal blocks. For example 1(b) is a block decomposition of region in fig 1(a) while fig 1(b) is the corresponding quad tree. Let the four sides of the blocks be as N, E, S and W sides. They are also termed boundaries and at time we speak of them as if they are directions. We define following predicates and functions to aid in the expression of operations involving blocks quadrants and its boundaries. ABUT (B,I) is true if and only if quadrant I is adjacent to boundary B of the blocks block's e.g. ABUT(W,SW) is true. MIRROR (B,I) yields the PROGENYTYPE value of the block of equal size that is adjacent to side B of a block having PROGENYTYPE value of the block of equal size that is adjacent to side B of PROGENYTYPE Value I MIRROR(N,SW)=NW,SHAREDSIDE(Q1,Q2) indicates the boundary of the block containing quadrants Q1 and Q2 i.e common to them e.g. SIIAREDSIDE(SW,NW) = W. If Q1 and Q2 are not adjacent brother quadrants (e.g. NE, SW) or if Q1andQ2 are the same, then the value of common side is undefined .FCQUAD(Q) is the quadrant which does not share a block boundary with quadrants of a node and its bounders while TABLES1-4 below show the definition of ABUT, MIRROR,FCQUAD,SHAREDSIDE relationship respectively. Corresponds to an undefined value, for a quadtree corresponding to a 2×2 array we say that the root is level n, and that block at level i is at distance of n-i from the root of the tree .In other wards, for a node at level i we must ascend n-i GENITOR links to reach the root of the tree is at level ≥ 0,a node at level corresponding to a single pixel in the image. Also, we say that a block is of size 2^s if it is found at level s in the tree.

ABUT

S \ Q	NW	NE	SW	SE	
	Yes	Yes	No	No	N
	No	Yes	No	Yes	E
	No	No	Yes	Yes	S
	Yes	No	Yes	No	W

TABLE 1 .ABUT(S,Q)

MIRROR

S \ Q	NW	NE	SW	SE	
	SW	SE	NW	NE	N
	NE	NW	SE	SW	E
	SW	SE	NW	NE	S
	NE	NW	SE	SW	W

TABLE 2.MIRROR(S,Q)

Q2 \ Q1	NW	NE	SW	SE	
	Θ	N	W	Θ	N
	N	Θ	Θ	E	E
		Θ	Θ	S	S
	Θ	E	S	Θ	W

TABLE 4. SHAREDSIDE (Q1,Q2)

NW	SE
NE	SW
SW	NE
SE	NW

TABLE 3. FCQUAD (Q)

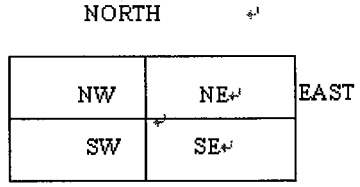


FIG2.RELATION BETWEEN BLOCKS, QUADRANTS AND ITS BOUNDRIES

3. NEIGHBOUR FINDING ALGORITHM

Given a block corresponding to a specific block in the image, its neighbor of equal size in the horizontal or vertical direction is determined by locating a common ancestor .Next , we retrace the path while mirror image moves about an axis formed by the common boundary between the block associated with the two blocks. The common ancestor is simple to determine – e. g to find the eastern neighbor, the common ancestor is the first ancestor block which is reached via its NW or SW progeny. For example, the eastern neighbor of a block A in fig 3(a) is G. It is located by ascending the tree until the D, is found. This requires going through a NE link to reach B, a NE link to reach C ,and a NW link to reach D. Block G is now reached by backtracking along the previous path with the appropriate mirror image move. This requires descending a NE link to reach E, a NW link to reach F, and a NW link to reach G, FIG 3(a) and 3(b), and FIG 4 (a) and 4(b) show how the eastern neighbor of a block A is located. The algorithm for locating an equal sized neighbor in a given horizontal or vertical direction is given below using a variant as

```
Block procedure EQUAL_ABUT_NEIGHBOR(P,D);/* locate an equal sized neighbor of block P
   in horizontal or vertical direction D.*/
begin
value direction D ;
return(PROGENY(If (D,PROGENYTYPE(P)) then
EQUAL_SIZE_NEIGHBOR(GENITOR(P),D)
Else MIRROR(D,PROGENYTYPE(P)))); end;
```

Finding a block in diagonal direction (i.e., corresponding block touches the given block at a corner) is more complex. Given a node corresponding to a specific block in the image, its neighbor of equal size in a diagonal direction is determined by a three step process.

First, we locate the given block's nearest ancestor who is adjacent (horizontal or vertically) to an ancestor of the sought neighbor. Next, we make use of EQUAL_ABUT_NEIGHBOUR to access the ancestor of the sought neighbor in the direction of the adjacent, Finally, we retrace the remainder of the path while making directly opposite moves (i.e. 180 ° Opposite so that a NW move become SE move).The nearest ancestor of the first step is the first ancestor which is not reached by a link equal to the direction of the desired neighbor -e.g. to find a SE neighbor, the nearest such ancestor block that is not reached via its SE PROGENY. For example, The SE neighbor of block A .

is G in fig4 (a). It is located by ascending the tree until the nearest ancestor ,B which is also adjacent horizontally(in this case)to an ancestor of G ,i.e. F is found .This require going through a NE link to reach B .Block F is now reached by applying EQUAL_ABUT_NEIGHBOUR in the direction of the adjacency(i.e. east). This forces us to go through a NE link to reach C and a NW link to reach F .Finally, we backtrack along the remainder of the path making 180 ° moves – i.e. we descend a SW link to reach G. Figure a and b show how SE neighbor of block A is located. Note that , at times , EQUAL_ABUT_NEIGHBOUR may not need to apply. This is the case when nearest ancestor of the first step is reached by equal to a direction opposite that of desired neighbor (e.g. in fig 1, the SW neighbor 16 is 25 with the nearest ancestor of step1 being block A). The algorithm for locating an equal size neighbor in a given diagonal direction is given below we assume that the neighbor in the specified direction does indeed exit (i.e. we are not in the border of the image)

```
begin
value block P ; value quadrant C ;
return(PROGENY(if PROGENYTYPE(P))=FCQUAD(C)
 then GENITOR(P)
```

else if PROGENYTYPE (P) = C then
EQUAL_CORNER_NEIGHBOR(GENITOR(P),C)
Else EQUAL_ABUTTING_NEIGHBOR(GENITOR(P),
SHAREDSIDE(PROGENYTYPE(P),C);
FCQUAD(PROGENYTYPE(P)))); end;

It is the case that neighbors are of different sizes. In such case, we say that the neighboring terminal blocks having equal or greater size e.g., the eastern neighbor terminal block 23 in fig. is 16) If such a block does not exit, then we return a YELLOW block of equal size if possible (e.g. the northern neighbor of block in fig1is J). Otherwise the block is adjacent to the border of the image(not the region) and NULL is return since there is no neighbor in the specified direction (e.g. the northern neighbored block 2 in fig 1 is NULL).When anode does not have a neighbor terminal block of equal or greater size, returning a YELLOW block of equal size is reasonable because the given block is the given direction .Then algorithm for locating neighbors of equal or greater size in horizontal or vertical directions as well as diagonal directions are given below using procedures GTEQUAL_ABUT_NEIGHBOUR and GTEQUAL_CORNER_NEIGHBOUR respectively.

Note that a diagonal direction, say C, will not always about against corner C of the block whose neighbor is sought (e.g. block 16 is a non abutting NE neighbor of block 23 in fig 1).

Block procedure
Neighbor GTEQUAL_ABUT_NEIGHBOR(P,D);
/*locate a neighbor of block P in horizontal or vertical direction D. if such a block does not exist, then return NULL*/
Begin
Value block P; Value direction D; Block Q;
If not NULL (GENITOR (P)) and ABUT(D,PROGENYTYPE(P)) then
/* find a common ancestor */
Q← GTEQUAL_ABUT_NEIGHBOR (GENITOR (P),D)
Else Q← GENITOR (P);
/* follow the reflected path to locate the neighbor */
Return (! NULL (Q) and YELLOW(Q) then
PROGENY (Q, MIRRORED (D, PROGENYTYPE (P)))
Else Q); End;
Block procedure GTEQUALCORNER_NEIGHBOR(P,C);
/* locate a neighbor of block P in the direction of quadrant. If such a block does not exist, then return NULL*/
Begin
Value block P; Value quadrant C;
Block Q;
If not NULL(GENITOR(P)) and PROGENYTYPE(P)!=FCQUAD(C) then
/* find a common ancestor */
If PROGENYTYPE(P)=C then
Q← GTEQUAL_CORNER_NEIGHBOR(GENITOR(P),C)
Else Q←GTEQUAL_ABUT_NEIGHBOR(GENITOR(P),SHAREDSIDE(PROGENYTYPE(P),C))
Else Q← GENITOR(P);
/* follow opposite path to locate the neighbor*/
Return(!NULL(Q)andYELLOW(Q)then PROGENY(Q,FCQUAD(PROGENYTYPE(P),C))
Else Q); End;

If neighbors are of different sizes, we may wish to know the size of the adjacent or abutting neighbor. In such a case we want our neighbor finding algorithm to return both a pointer to the neighbor block a value from which the block's size can be easily computed.

This is relatively straightforward when we know the level in the tree at which a block is found whose neighbor is being sought. In fact, such an algorithm need only increment of the level counter by 1 for each link that 1 ascended while locating common ancestor, and then decrement of the level counter by 1 for each link that is descended while locating the appropriate neighbor. The algorithm for locating neighbor of equal or greater size, with their corresponding level positions in horizontal and vertical directions as well as diagonal directions are given below using procedures GTEQUAL_ABUT_NIGHBOR@ and GTEQUAL_CORNER_NIEGHBOR respectively.

Note thee use of reference parameters to transmit and return results. An alternative is to define a record of type block having two fields of type *block* and *integer*, whose values are a pointer to the neighboring block and its level, respectively.

Procedure GTEQUAL_ABUT_NEIGHBOR(P<D<Q<L);
/*return in Q the neighbor of block P in horizontal or vertical direction D/ L. Denotes the level of the tree at which block P is initially found and the level of tree at which block Q is ultimately found. If such a block does not exists, then return NULL*/
Begin
Value block P; Value direction D; Reference block Q;
Reference integer L; L←L+1;
If (!GENITOR(P) and ABUT(PROGENYTYPE(P))) then
/* find a common ancestor*/
GTEQUA_ABUT_NEIGHBOR2(GENITOR(P),D,Q,L)
Else Q←GENITOR(P);
/* follow the reflected path to locate the neighbor*/
If (!NULL(Q) and YELLOW(Q)) then
Begin
Q←PROGENY(Q,MIRROR(D,PROGENYTYPE(P)));
L←L-1; End; End;

Procedure GTEQUAL_CORNER_NEIGHBOR2(P,C,Q,L);
/* return in Q the neighbor of block P in the direction of quadrant C. L denotes the level of the tree at which block P is initially found and the level of the tree at which block Q is ultimately found. If such a block does not exist, then return NULL*/
Begin
Value block P; Value quadrant C; Reference block Q; Reference integer L;
L←L+1;
If(!NULL (GENITOR(P)) and PROGENYTYPE(P)!=FCQUAD(C)) then
/* find a common ancestor*/
If PROGENYTYPE (P)==C then
GTEQUAL_CORNER_NEIGHBOR2(GENITOR(P),C,Q,L)
ELSE GTEQUAL_ABUT_NEIGHBOR2(GENITOR(P),SHAREDSIDE(PROGENYTYPE(P),C),Q,L)
Else Q←GENITOR(P);
/* follow the opposite path to locate the neighbor*/
If nor NULL(Q) and YELLOW(Q) then
Begin
Q←PROGENY(Q,FCQUAD(PROGENYTYPE(P)));
L←L-1;
End; End;

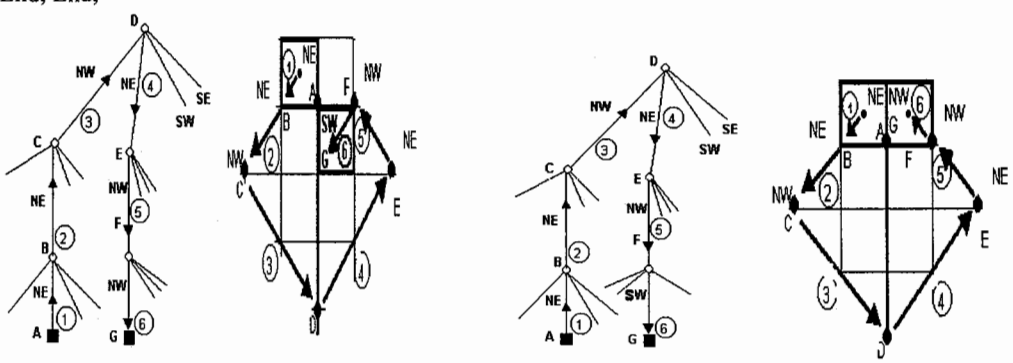


FIG:3(a) process of representing eastern neighbor of node A(i. e G)FIG:4(a) process of locating SE neighbor of node A

FIG: 3 (b) Tree representations

FIG: 4 (b) Tree representations

At times we may wish to locate an adjacent horizontal or vertical neighbor regardless of size. In such a case, we also specify a corner of the block corresponding to the block whose neighbor is being sought. Neighbor block must be adjacent to the corner (e.g. block 21 is the northern neighbor of block 23 which is adjacent to the NE corner of block 23). The algorithm for computing such a neighbor is given below by procedure CORNER_ABUT_NEIGHBOR, which make use of GTEQUAL_ABUT_NEIGHBOR.

Block procedure CORNER_ABUT_NEIGHBOR(P,D,C);

/* locate a neighbor of block P in horizontal or vertical direction D which is adjacent to corner C of block P. If such a block does not exist, then return NULL*/

Begin

Value block P; Value direction D; Value quadrant C;

P<-GTEQUAL_ABUT_NEIGHBOR(P<D);

While YELLOW(P) do P<-PROGENY(P,MIRROR(D,C)); /*Descend to the desired corner*/

Return (P); End;

Similarly, in the case of a diagonal neighbor, we may also wish to locate the neighbor in the given direction regardless of its size (e.g., block 20 is a NE neighbor of block 22 in Fig. that is smaller in size). The algorithm for locating an arbitrary-sized diagonal.

neighbor is given below by procedure CORNER_CORNER_NEIGHBOR which use of GTEQUAL_CORNER_NEIGHBOR

Block procedure CORNER_CORNER_NEIGHBOR (P<C);

/*locate a neighbor of block P in the direction of quadrant C that abuts against corner C of block P. If each does not exists, then return NULL*/

Begin

Value block P; Value quadrant C; Block Q;

Q<-GTEQUAL_CORNER_NEIGHBOR(P,C);

While YELLOW(Q)do(Q<-PROGENY(Q,MIRROR(D,C));/*Descend to the desired corner*/

Return (Q); End;

It should be clear that procedure similar to CORNER_ABUT_NEIGHBOR and CORNER_CORNER_NEIGHBOR can be constructed that also return the level at which the desired neighboring block is found. This will not be done here.

The procedure outlined above always return NULL when a neighbor in a specified direction does not exists. This situation arises whenever the block whose neighbor is sought is adjacent to the border of the image along the specified direction. At times the NULL pointer is not convenient. Instead, we could assume that the image is surrounded by RED or VIOLET for the surrounding blocks depends on the particular applications. For example, we use RED in the case of the Quadtree to boundary code conversion algorithm[2] while VIOLET is more useful in the case of the computation of distance[15] and the construction of a Quadtree Medial Axis Transform[16].

At times it is useful to determine if certain edges of the blocks corresponding to two neighboring blocks extend past each other or aligned. For example, in fig 1(a), block 16 extends past block 10 with respect to their western boundaries, , while the western boundaries of blocks 9 and 16 are aligned. We assume that the level of the tree at which each of the two blocks, say P and Q at levels LP and LQ blocks must be visited. This can be seen by observing that the LQ, respectively. It should be clear that at most LP- smaller of the two blocks cannot extend GENITOR than the other because this would imply that the blocks properly overlap, which is impossible. at best, the smaller block can be aligned with the other block, and this occur if and only if the smaller block is adjacent to the extreme side in the designated direction of the nearest common ancestor of the two blocks. The algorithm for computing the aligned relationship is given below by procedure ALIGNED.

Boolean procedure ALIGNED (P, LP, LQ, D);

/* given two blocks P and Q at levels LP and LQ respectively, which are adjacent along side CCSIDE(D) of block P, determine whether either of P or Q extends further in direction D than the other(return FALSE), or their two sides in direction D are aligned(return TRUE)*/

Begin

Value block P, Q; Value integer LP, LQ; Value direction D;

Block R; Integer I;If LP+LQ then return (TRUE)

Else if LP>LQ then R←Q Else R←P;

For I←1 step 1 until ABS (LP-LQ) do

Begin

If (!ABUT(D,PROGENYTYPE(R)) then return (FALSE)

Else R←GENITOR(R);
End; Return (TRUE); End;
The above techniques should be contrasted with other methods of locating neighbors. In this method it is described for moving between adjacent blocks of equal size that are brothers (i.e., have the same GENITOR block). This method does not make use of the tree structures; instead, coordinate information and knowledge of the size of the image are used to locate a neighboring brother in a given horizontal or vertical direction. This is accomplished by a number of primitives termed MOVE UP, MOVE DOWN, MOVE RIGHT and MOVE LEFT. Transitions to non-brother neighboring blocks require the use of approximation through the use of primitives named MORE, LESS and GAMMA. The disadvantages of these methods are that they require computation (rather than chasing links) and are clumsy when adjacent blocks are not brothers as well as when they are of different size than the block whose neighbor are sought.
A number of algorithms are described for operating on images using Quadtree. Transitions between neighboring blocks are made by use of explicit links from a block to its adjacent neighbors in the horizontal and vertical directions. This is achieved through the use of adjacency trees, "ropes", and "nets". An adjacency tree exists whenever a leaf block, say X, has a YELLOW neighbor, say Y, of equal size.
In such a case, the adjacency tree of X is a binary tree rooted at Y whose blocks consist of all Progenies of Y(VIOLET, RED and YELLOW) that are adjacent to X, for example, for block 16 in FIG.1, the western neighbor is YELLOW block F with an adjacency tree as shown in fig6 , a rope is a link
between adjacent blocks of equal size at least one of which is a leaf block. For example, in fig.1, there exists a rope between block 16 and blocks G, 17, H and F. similarly, there exists a rope between block 37 and blocks M and N; However, there does not exist a rope between block L and blocks M and N.
The algorithm for finding a neighbor using a roped Quadtree is quite simple. We want a neighbor, say Y, on a given side, say D, of a block, say X. If there is a rope from X on side D, then it leads to the desired neighbor
If no such rope exists, then the desired neighbor must be larger. In such a case, we ascend the tree until encountering a block having a rope on side D, that leads to the desired neighbor in effect, we have ascended the adjacency tree of Y. For example, to find the eastern neighbor of block 21 in FIG., we ascend through block J to block F, which has a rope along its eastern side leading to block 1. At times it is not convenient to ascend blocks searching for ropes. A data structure named a net is used to obviate this step by linking all leaf blocks to their neighbor regardless of their relative size. Thus in the previous example there would be a direct link between blocks 21 and 16 along the eastern side of block 21. The advantage of ropes and nets is that the number of links that must be traversed is reduced. However, the disadvantage is that the storage requirements are considered increased since many additional links are necessary. In contrast, our methods are implemented by algorithms that use of the existing structure of the tree- i.e., four links from a non leaf block to its PROGENIES, and a link from a non root block to its GENITOR. The execution time of the neighbor finding algorithms presented in section3 depends on the relative positions of the blocks in question. Clearly, the execution time depends on the number of blocks that must be traversed in locating the desired neighbor. In the following we analyze the average execution time of
EQUAL_ABUT_NEIGHBOR,GTEQUAL_ABUT_NEIGHBOR,
CORNER_ABUT_NIEGHBOR,EQUALCORNER_NIEGHBOR,
GTEQUAL_CORNER_NEIGHBOR,CORNER_CORNER_NEIGHBOR,andALIGNED.
GTEQUAL_ABUT_NEIGHBOR2 and GTEQUAL_CORNER_NEIGHBOR2 have the same execution time as the GTEQUAL_ABUT_NEIGHBOR and GTEQUAL_CORNER_NEIGHBOR, respectively, since they visit the same number of blocks.
At this point it is appropriate to elaborate on our notion of average. We assume a random image in the sense that a block is equally likely to appear in any position and level in the Quadtree. This means that we assume that all neighbor pairs (i.e., configurations of adjacent blocks of varying sizes) have equal probability. This is different from the more conventional notion of a random image which implies that every block at level 0(i.e., pixel) has an equal probability of being VIOLET or RED. Such an assumption would lead to a very low probability of any blocks corresponding to blocks of size larger than 1. Clearly, for such an image, the Quad tree is the wrong representation.

FLOW CHART

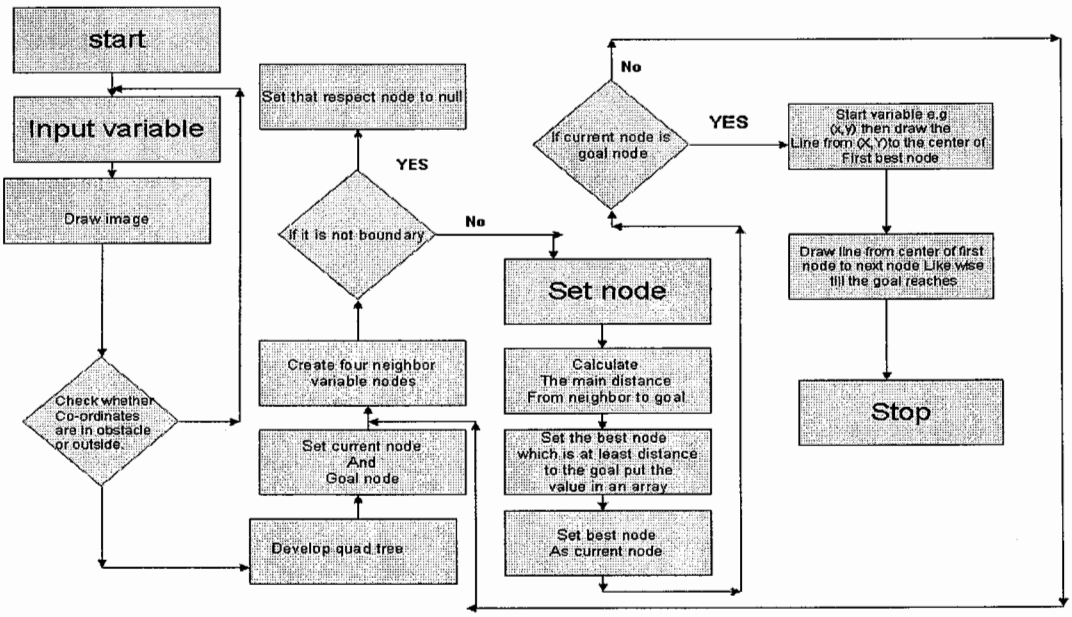


FIG. 5 , Flow chart

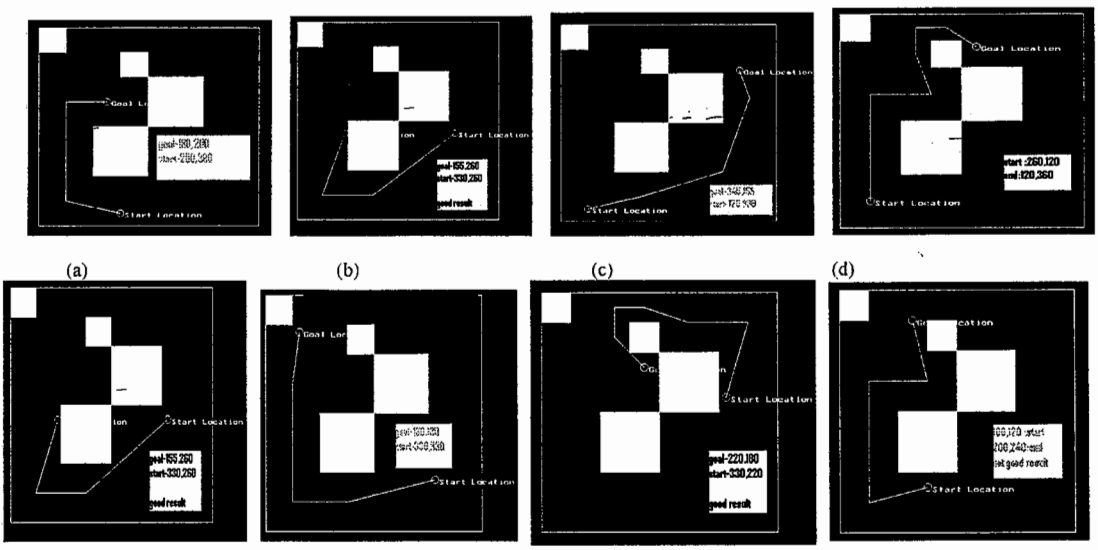


FIG: 6(a to h) Path tracing with different start and goal points

CONCLUSION

It has been shown the neighbor finding technique for Quadtree in detail. The analysis of various algorithms demonstrate that the operation is quite different. We have used different kind of model of randomly picked image in our analysis. It is possible to use some optimal algorithm for path tracing based on micro mouse and some practical heuristic rules. This paper deals with the optimal collision free detection and path planning for mobile robot.

REFERENCES

[1] G.M.Hunter, *Efficient Computation and Data Structures for Graphics*, Ph.D. dissertation, Department of Electrical Engineering and Computer Science, Princeton University,Princeton. N.J. 1978.

[2] G.M.Hunter and K.Steiglitz, Operations on images using Quadtree, *IEEE Trans. Pattern Anal. Mach. Intell.* 1. 1979,145-153.

[3] G.M.Hunter and K.Steiglitz, Linear transformation of pictures represented by Quadtrees. *Computer graphics and Image Processing* 10, 1979,289-296.

[4] A.Klinger,Patterns and search statistics, in *Optimizing Methods in Statistics*(J.S.Rustagi, Ed.),Academic Press, New York, 1971.

[5] A.Klinger and C.R.Dyer, Experiments in picture representation using regular decomposition *ComputerGraphics and Image Processing* 5,1976,68-105.

[6] A.Klinger and M.L.Rhodes, Organization and access of image data by areas, *IEEE Trans, Pattern Anal, Mach. Intell.* I.1979,50-60.

[7] H.Moravec, "Rover visual obstacle avoidance, "in *Proc. Seventh Int. Joint Conf. Artificial Intell.*,1981.

[8] N.J.Nilsson, "Amobile automation: an application of artificeal intelligence techniques," in *Proc. First Int. Joint Conf, Artificial Intell.*,1969.

[9] H.Samet, Region representation:Quadtrees from binary arrays, Computer Graphics and Image Processing 13,1980,88-93.

[10] B.A. Brooks, "Solving the findpath problem by good representation of free space," *IEEE Trans. Syst. Man, Cybern.*,Vol.10,pp.190-197,1983.

[11] C.Thorpe, "Path relaxation: path planning for a mobile robot," in *Proc. Nat. Conf. Artificial Intell.*,1984.

[12] Charles R. Dyer, "The Space Efficiency of Quadtrees", Computer Graphics and Image Processing, vol. 19, p.p. 335-348, 1982.

[13] David Nitzan, "Development of Intelligent Robots" IEEE journal of Robotics and Automation, vol. RA-1, No. 1, p.p. 3-12, 1985.

[14] H.Samet, A distance transform for images represented by quadrees for images represented by Quadtrees, *IEEE Trans. Pattern And Mach. Intell.*,in press.

[15] H.Samet, *A Quadtree Medial Axis Transform*, Computer Science IR-803, University of Maryland College Park, Md, August 1979.

[16] James L. Crowley, "Navigation for an intelligent Mobile Robot", IEEE journal of Robotics and Automation, vol. RA-1, No. 1, p.p. 31-40, 1985.

[17] James L. Crowley, "Navigation for an intelligent Mobile Robot", IEEE journal of Robotics and Automation, vol. RA-1, No. 1, p.p. 31-40, 1985.

[18] A.M.Thompson, "The navigation system of the JPL robot" in *Proc. Fifth Int. Joint Conf. Artificial Intelligence*,1977.

[19] C.Thorpe, "Path relaxation: path planning for a mobile robot," in *Proc. Nat. Conf. Artificial Intell.*,1984.

[20] R.Wallace, "Two-dimensional path planning and collision avoidance for three-dimensional robot manipulators," in *Representation and Processing of spatial Knowledge*, tech. rep.1275, Department of Computer Science, University of Maryland, May 1983.

[21] H. Samet, "An Algorithm for converting rasters to Quadtrees," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 3, pp. 93-95. 1981.

[22] H. Samet, "An overview of Quadtrees, oct-trees, and related Hierarchical Data structures", NATO ASI Series, vol. F40, 1988.

[23] S.Ghoshray, K.K.Yen, "A comprehensive robot collision avoidance scheme by two dimensional geometric modeling" in proc. Of the 1996. IEEE Int. conf. On Robotics and Automation, minne polis, USA, 1996.

[24] Danny Z Chen, Robert J. Szczerba and John J. Uhran. Jr, "A Framed-Quadtree approach for determining Euclidean shortest paths in a 2-D environment". IEEE Transactions on Robotics and Automation, vol. 13. No. 5. P.p. 668-680, 1997.

Research and Realization of Signal Simulation on Virtual Instrument

ZHAO Qi, HE Wenting, GUAN Xiumei
School of Electronics and Information Engineering
Beijing University of Aeronautics and Astronautics
Beijing, China
E-mail: zhaoqi@buaa.edu.cn

ABSTRACT

In the engineering project, arbitrary waveform generator controlled by software interface is needed by simulation and test. This article discussed the program using the SCPI (Standard Commands For Programmable Instruments) protocol and the VISA (Virtual Instrument System Architecture) library to control the Agilent signal generator (Agilent N5182A) by instrument communication over the LAN interface. The program can conduct several signal generations such as CW (continuous wave)、AM (amplitude modulation)、FM (frequency modulation)、ΦM (phase modulation)、Sweep. As the result, the program system has good operability and portability.

Signal generator; SCPI protocol; VISA library; Signal modulation

1 Introduction

In the modern engineering projects, arbitrary waveform generator controlled by software interface is needed by simulation and test. This article discussed program which can generate several signals and control the signal generator by instrument communication over the LAN interface. The signal simulation on virtual instrument not only can fit the need of the engineering projects, but also has well maneuverability and transplant.

With the rapid development of the computer, the virtual instrument has been widely used. Virtual instrument established a powerful, flexible, volatile computer-based test and measurement and control system, which combined with the current computer mainstream technology and flexible software and high capability modular hardware[1]. The combination of the modular hardware and powerful software made monitoring and controlling more convenient.

2 Hardware environment

Agilent N5182A vector signal generator output frequency range is 250kHz to 3 or 6 GHz, including step & list sweep of frequency, power, or frequency and power, user flatness correction, automatic leveling control, power calibration. Options can be achieved by adding the AM (amplitude modulation), FM (frequency modulation) and ΦM (phase modulation) modulation capabilities, and external input modulation can also be achieved. Besides, N5182A also has pulse modulation function. The instrument has GPIB, USB2.0, and 100Base-T LAN interface for remote controlling, and SCPI and IVI-COM driver, 8648/ESG code compatible, and fits the LXI Class C standards. For the need of the external analog I/Q inputs, the instrument also can satisfy users' requirements.

In this article, the main achievement was the overall control of the vector signal source based on PC. And Agilent N5182A worked in the LAN communication interface mode.

The SCPI is supported by Agilent N5182A. By the SCPI-compatible protocol the users can control and operate the instrument directly. This paper used VXI-11 bus technology and the VISA library to control the signal generator.

3 VISA and SCPI

VXI (VMEbus Extension for Instrumentation) system is a modular instrumentation platform, and it is the extension of VME bus. It introduced an open new generation of automated test industry bus standard in 1987. VXI bus specification

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462U · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.852960

achieved the standardization of VXI hardware system, but truly "plug and play" also need to achieve standardization of the software.

In 1993, five companies, including NI, Tektronix etc., founded the VPP (VXI plug & play) Union. They advanced a series of VPP software standards such as system frame, drivers, softpanel, VISA (Virtual Instrument System Architecture), parts repository and so on, which promoted the process of the software standardization.

3.1 VISA

VISA (Virtual Instrument Software Architecture) defined a new generation of I/O interface software specifications, not only for VXI interface, but also for GPIB, serial ports (such as the RS-232) and other interfaces. VISA library functions were used to write drivers for devices to complete commands and data transmissions between computer and instrument, in order to achieve the program-controlled. The location of VISA in the whole automatic test system was shown in Fig. 1.

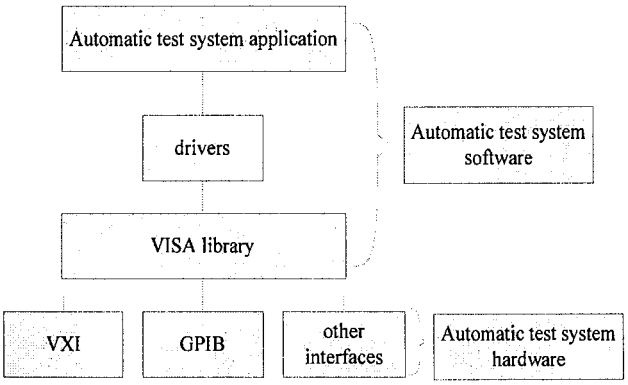


Figure 1. Automatic test system diagram.

As the underlying I/O interface software, VISA library resided in the manager system (computer system), and achieved the commands and data transmissions between computer systems and instruments.

For the users, they only need to know the format and parameters of the VISA I/O function, and they can write instrument drivers without caring about the details of how the VISA library communicates with the instruments.

3.2 SCPI

SCPI (Standard Commands for Programmable Instruments) was proposed in 1990. SCPI Union was committed to develop the common interface language between the computer and the testing systems. SCPI standard established above the standard codes and formats proposed by IEEE488.2 (SCPI standard contained the command of IEEE488.2). The instruments commands were required by SCPI to follow the standard of IEEE488.2. However, SCPI was a pure software standard. SCPI was hardware independent. SCPI string can be sent by any instrument interface, whether it was GPIB, RS-232, VXI bus, or local area network, it can work just as well.

SCPI statement existed as ASCII text, so it can be added to any computer testing program language, such as BASIC, C or C++; it also can be used in the testing program development environment, such as Lab Windows/CVI, Lab VIEW or HP VEE.

4 Agilent N5182A communications setting and implement of SCPI

4.1 Control Setting

In engineering projects, testing and controlling often required various forms of signals, including continuous wave, swept wave, amplitude modulation wave, frequency modulation wave, phase modulation wave, pulse wave and other signals; for the modulation signal, the signal waveform also included sine, square and triangular wave. The controlling of signal types, the controllable output power, the controllable center frequency and so on, were also the main functions of the system. For the different signal types, the specific parameters have different settings; such as the swept wave had the scanning modes of frequency, power, or frequency and power, each mode had the scan parameters, such as points and step

dwell. This system can control generation of all kinds of signals such as the continuous wave, swept wave, AM wave, FM waves, ΦM wave, pulse wave and other forms of signals.

In this project, PC directly connected the signal generator (Agilent N5182A) by a crossover cable, and the standard language was SCPI. In order to produce the required function signal, the appropriate SCPI program was only needed. And Agilent N5182A which fully supported the SCPI commands can be easily carried out the setting of the signal frequency, amplitude, voltage bias and other parameters through the SCPI commands [2].

For the modulation waveform, there were mainly three categories: sine, square, triangle wave. Sine wave generation can be used by the established Agilent N5182A waveform output. The square and triangle wave signals needed locally generated waveform data uploaded to the instrument as modulated signal source. System was mainly to achieve frequency modulation and phase modulation of the sine, square, triangle wave. The system used SCPI commands, embedded in the communication codes, to control the various types of waveforms. The system used Visual C++ 6.0 to design the control panel, users entered the required waveform and parameters into the computer software, the computer calculated the values in the background, embedded the values into the SCPI commands, the communication part of system will write the commands into the instrument, and read the necessary information from the instrument.

No matter what the waveform will be generated, the parameters, such as frequency, amplitude and so on, need to be set up. For some modulation, also need to enter some specific parameters, such as bias, rate etc. For the sine wave modulation, the input parameters can be directly embedded in the SCPI commands, while the parameters of the square and triangle wave such as bias, rate and other parameters needed to make the appropriate background data processing, and processed the data embedded in SCPI commands. The system communicated the instrument through the LAN interface, so the system can be connected to more than one signal generators, and at the same time, produced variety types of signals.

4.2 Design and Realization of the LAN communication and parameter account

The system used VXI-11 bus technology and the VISA library to control the signal generator; the users selected the signal type and set the corresponding parameters by controlling panel, the system got the communication with the signal generator through the LAN port to make the control. The signal generator control program was divided into four parts, which are declaration, opening, the device I/O and closing. The flow chart was shown in Fig. 2.

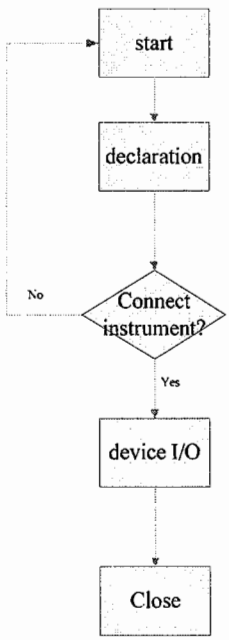


Figure 2. Signal generator control flow chart.

1) Declaration area. Declare all the program variables data types. The types of the function used in the declaration are the VISA data types, which was independent from the programming language. VISA data types, programming language data types and the corresponding instructions contained in a specific header file. The header file in the C++ language is visa.h. Open the visa.h you can see, ViSession and ViUInt32 correspond to the unsigned long.

2) Opening area. First call viOpenDefaultRM (ViPSession vi), open the connection with the default resource manager, get the connecting. Default Resource Manager is the agency used to manage all the resources in VISA. Calling the function returns a value as an input parameter of the viOpen (). Second, call the function viOpen (ViSession sesn, ViRsrc name, ViAccessMode mode, ViUInt32 timeout, ViPSession vi). VISA can be used by many instruments irrespective of the interface types. In the viOpen() function, the second parameter means the mode of the interface type. Because this system used the LAN interface, so the second parameter of the viOpen function is set to the IP address of Agilent N5182A. The system, which controlled multiple signal generators at the same time, identified them by the different IP addresses.

3) Device I/O area. In this process, the device I/O area primarily completed and sent the SCPI commands, read back the response string. Writing commands called the function viWrite (ViSession vi, ViBuf buf, ViUInt32 cnt, ViPUInt32 retCnt), and the read-back function is viRead (ViSession vi, ViPBuf buf, ViUInt32 cnt, ViPUInt32 retCnt).

4) Closing area. By the end of the operation, the program must call the viClose () for closing the communication with the device and the default resource manager respectively.

In the opening area with the devices, the system set the connection status detection. When the connection with the signal generator can not be established, the system will give error messages; prompt the users to check the equipment and connection.

Parameters calculation module got the response of the setting parameter values on the control panel, and translated the parameter values into the SCPI language, sent the commands through the communication module in the device I/O area[3].

4.3 The Design of Control Panel

Control panel was mainly user-oriented, designed with the object-oriented MFC, had strong integration and easy operation. Fig. 3 showed the setting panel of the FM modulation, the users can directly set FM waveform parameters in the control panel. When click OK button, the selected output of the Agilent N5182A vector signal generator can generate a corresponding signal waveform. Fig. 4 was the Pulse modulation setting panel; Fig. 5 was the Sweep setting panel.

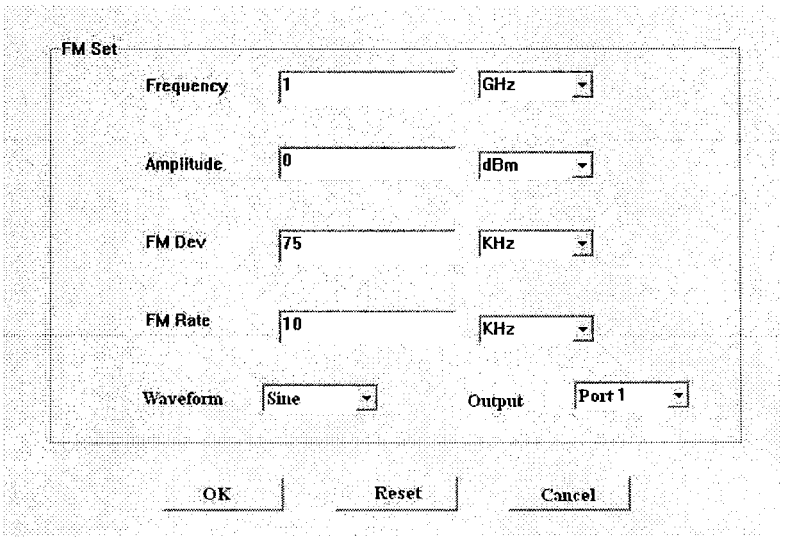


Figure 3. FM setting panel.

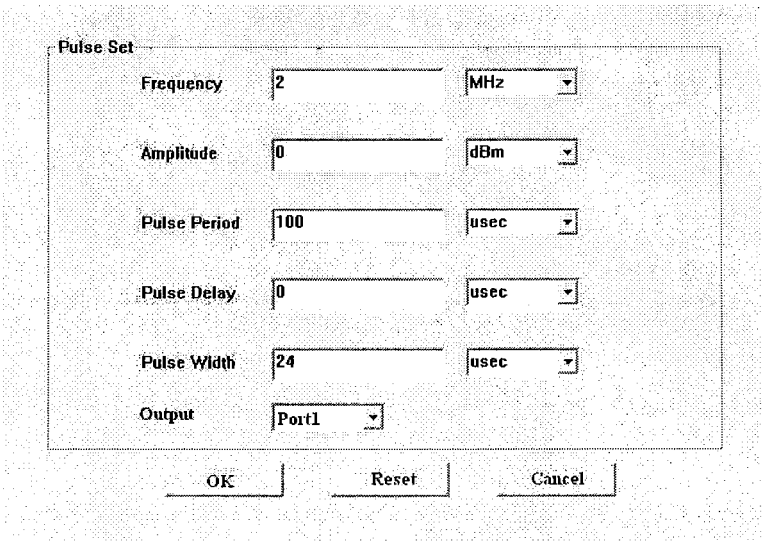


Figure 4. Pulse setting panel.

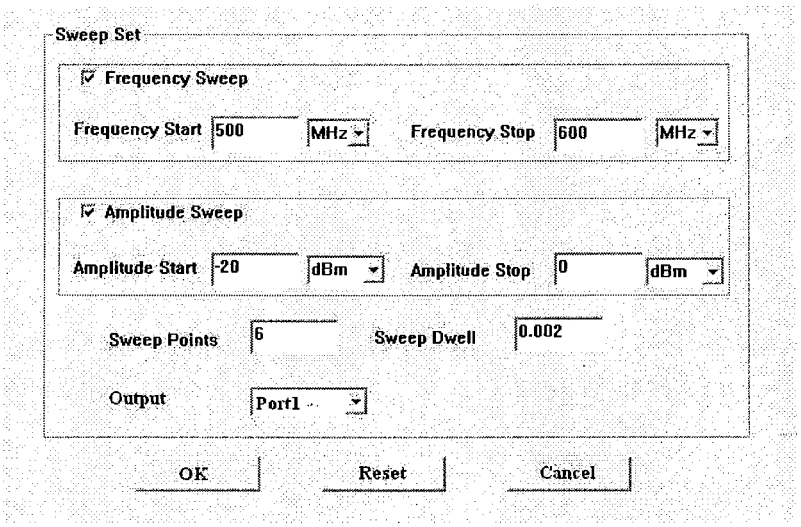


Figure 5. Sweep setting panel.

4.4 Results of the signal simulation

Through the debugging process, this system can be applied to engineering projects, be able to control any signal waveform generation through the software interface. The system not only can produce the frequency and amplitude adjustable continuous-wave, AM, FM, ΦM, pulse, sweep signals, but also can satisfy the other square, triangle wave signal modulation requirements.

5 Conclusions

Compared with the traditional instruments, virtual instrument obviously had much more technology advantages in the intelligence process, processing capacity, cost performance, operability, etc. And the greatest feature of this article was the instrument control in the virtual instrument environment. Using virtual instrument designed the different operation display interface for different applications, used the computer data processing, made the operation of the instrument become more

directly perceived, simple, easy to understand. The data control can be directly translated through the network without operating Agilent personally. Because of the virtual instrument technology advantages, the standardization of software interfaces, and the standardization of programmable instrument object become the mainstream of the virtual instrument development. As discussed in this article, the development of virtual instrument based on the standards VISA and SCPI was versatile, and has better reusability.

References

[1] Bo Lin. "Network Virtual Instrument Core Technology and Typical Cases Research," ChongQi: Chongqing University, (2004). (in Chinese)

[2] Li Qiong, Fu Ping and Zhang Zhongting. "The Research of Virtual Instrument System Architecture (VISA)," Measurement and Control Technology. Papers 18(8), 22-24(1999). (in Chinese)

[3] "Agilent Technologies," SCPI Command Reference, (2008).

Application of photogrammetry technology to industrial inspection

Zhang De-hai,Liang Jin, Guo Cheng,Chen Zhi-xin,

(School of Mechanical Engineering, Xi'an Jiaotong University, Xi'an 710049, China

Email:zhangdehai0318@163.com , gch@mail.xjtu.edu.cn)

ABSTRACT

In order to meet the requirement of obtaining the object figure quickly and accurately, XJTUDP software has been developed successfully by oneself based on photogrammetry theory. The contents of composing and explored of this system are introduced in this paper. The VDI/VDE2634 testing program is taken as referenced project, the self-designed framework of cube is taken as tested target, and then the conclusion that precision of XJTUDP may meet the standard of industrial measurement has been reached out. Finally, large-scale waterwheel leave is taken as example to conduct measuring, it is proved that photogrammetry system explored by ourselves may be imposed applying on the field of industrial measuring successfully.

Keywords: Photogrammetry ; Bundle adjustment; Large-scale waterwheel leave; XJTUDP

1 INTRODUCTION

As early as the 1960s, photogrammetry principles developed for shape and motion measurements were used to estimate plate deflections [1]. Morimoto and Fujigaki [2] discussed the use of multiple cameras with images of a deforming rectangular grid and FFT methods for image analysis and surface motion estimates. Andresen [3] used photogrammetry principles and a stereo system to record and analyze images of a grating undergoing large deformation. Basic studies of typical high speed imaging systems were performed by Tiwari et al. [4]. Bay et al. [5,6] proposed that the 2D-DIC concepts used successfully for 2D and 3D computer vision be extended and used to match small sub-volumes before and after undergoing loading to obtain a full volumetric field of three-dimensional motions.

In recently, with the development of computer, numerical control and laser measurement technology, the products of modern industrial manufacturing such as automobiles, airplanes, household appliances require a large number of A-class surfaces. It is urgent to need 3D full-size measuring technology. The module method is used to be applied under workshop conditions, Contact measurement methods such as CMM possesses higher precision, but its disadvantages is easier to scratch the surfaces, harder to acquire continuous surface data. Non-contact optical methods such as laser triangulate and Moiré projection method, which measurement accuracy are relatively lower due to restrictions on the working principle [7].In order to overcome existing shortcoming of measurement methods, a new measuring method based on close range photogrammetry [8, 9] principle is put forward, and then the related software XJTUDP has been developed by Xi'an Jiaotong University. The software applying to industrial inspection is possessed of advantages such as higher precision, higher speed and not scratches parts.

In this paper, digital close-range industrial photogrammetry software XJTUDP may compute un-coded points and coded-points fixed on the surface of objects, it takes a solid basement for the fusion of much pieces points-clouds. Surface scanner software XJTUOM can capture 3D data of the object surface. According to the whole coded-points derived from XJTUDP, every piece of point-clouds may be assembled automatically. On the basis of this points-clouds , 3D geometric features of the parts are reconstructed using post-processing software. Experiment implies that this reliable method may be suitable to industrial measurement.

2 WORKING PROCEDURE OF PHOTOGRAMMETRY TECHNOLOGY

The working procedure is divided into several phase based on photogrammetry technology. First , paste un-coded points and coded-points on the surface of the object, and then, take a group of pictures from different view as 2D original data.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462V · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.852490

Once these pictures are loaded on the measurement software XJTUDP, it begins to inspect each picture, detect and locate un-coded points and coded-points. Second, according to their 2D coordinates on different pictures, 3D coordinates of the un-coded points and coded-points are reconstructed, which in fact are the different projection of the same un-coded and coded-points. Third, the non-coded marked-points matrix of the whole spatial coordinates is derived from XJTUDP, and then it is imported into surface scan software XJTUOM. Finally, the final points-clouds are arrived out using post-processing software, and then the related CAD model is constructed. Consequently, geometric features of part are reconstructed.

3 EXPLOITATION OF PHOTOGRAMMETRY MEASURING SYSTEM

Digital close-range industrial photogrammetry measuring system XJTUDP is mainly place coded-point and un-coded points on the surface of an object and around. A certain number of photos are obtained through taking photo from different angles and locations. After image processing, marked-points position determination and code identification, 3D point cloud of objects are matched and reconstructed depended on these marked-points finally. On this basis of these points, the distance of key points on the object and other data could be measured precisely.

3.1 Composition of photogrammetry system XJTUDP is shown in Fig.1 and as follows :

1) System measurement software It installs in the high-performance desktop or portable computers.
2) Reference coded-points They are constituted of a central point and the ring code around the central point, each code-point has its own number.
3) Reference un-coded points They are used to measure 3D coordinates of pasted position on the surface of objects.
4) Professional digital cameras They are digital cameras with fixed focal length and high-resolution, but lens are interchangeable.
5) The high-precision gauge Gauge is taken as a proportion of measurement results, which is precisely to measure reference points and determine their distance between points.

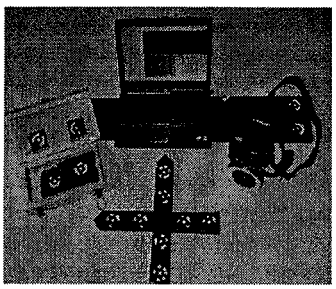


Fig.1 Hardware composition of XJTUDP

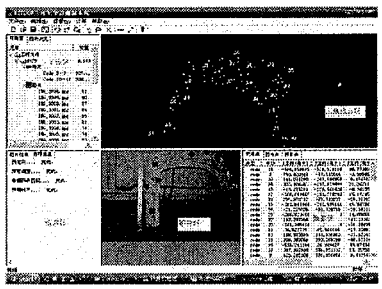


Fig.2 Main program interface of XJTUDP

3.2 The main contexts of photogrammetry system

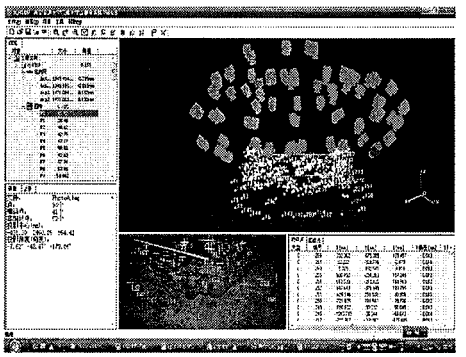
The main contents explored of this system are list as follows: the high accuracy center measuring method of marked points technique based on the fitting sub-pixel edge, the methods of coded-points designing and coded-points auto-detecting, the camera calibration techniques based on photogrammety and 3D reconstruction, photographic measurement system accuracy testing and evaluation methods.

3.3 Developed interface of program

1) The developed interface of the main program is shown in Fig.2. It may implement the functions including multi-window interface, multi-mode switching and measuring and compared mode etc.

2) The interface of OPENGL software is developed and shown in Fig.2, which may implement displaying, rotating, panning, zooming and selecting of 3D graphics, predetermining of pictures and displaying of camera position.

3) Fig.3 shows image displaying plug-ins of .bmp format, it is constituted of displaying of 2D image, zooming, panning, and selecting of arbitrary center point, color displaying of elliptic, number display functions and distance measurement module.
4) Fig. 4 shows that color map displaying, which is the comparison results between measured data and number model. Fig.5 shows "321" alignment between digital model and global marked-point cloud, which can achieve arbitrary conversion of coordinate system.
5) Fig.6 shows project tree, the list display and export data of the global marked-points.



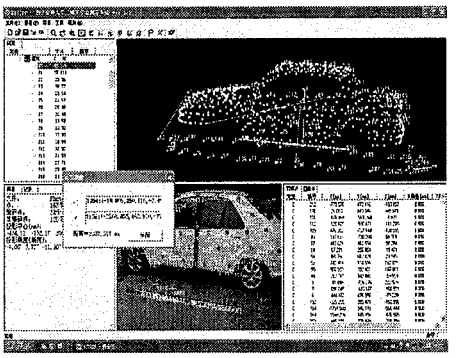
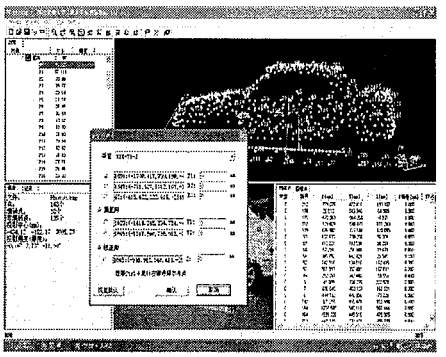


Fig.3 Bundle and predetermed of picture and displaying of camera position Fig.4 Measured distance between marked-points



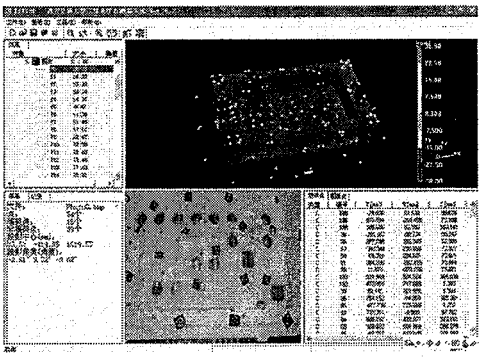
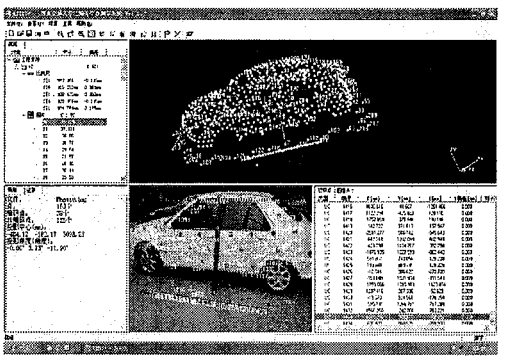


Fig.5 The 321 alignment of point-clouds and digital model Fig.6 Comparison between measurement data and digital modeling



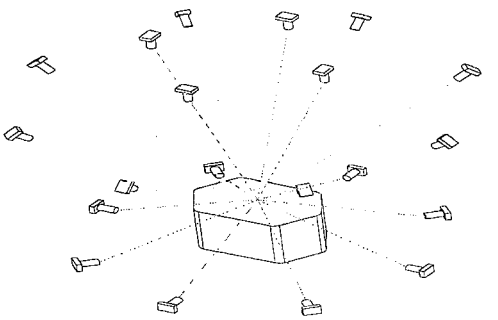


Fig.7 Obtaining global marked-points using image alignment and adjustment Fig.8 Multi-position photogrammetry

3.4 The key technologies of XJTUDP

In order to reconstruct the matrix of discrete points of 3D coordinates from the structural features of parts, close-range industrial photogrammetry system XJTUDP adopts digital image processing and photogrammetric technology. The digital cameras take photos to obtain part images from different angles and positions shown in Fig.8, the part is equipped with coded and un-coded points. The images are processed with sub-pixel extraction technology based on least-squares fitting method. At first, exact geometric centers of coded-points and un-coded points are obtained. The center coordinate and coded information of the above coded points may be solved based on collinear equation. And then the 3D coordinates of object points are reconstructed after coplanarity equation, direct linear transformation solution, the outer polar geometric constraints, 3D reconstruction and bundle adjustment solution. The results of the former four algorithms provide initial values to bundle adjustment algorithm. The bundle adjustment algorithm is the final step to optimize all parameters including internal and external orientation parameters of the camera and 3D coordinates of object points using a non-linear method. While adjusting the 3D coordinates of object points, the camera is calibrated with greater precision so as to control the error of object features points and obtain the greater precision matrix of points-cloud of 3D coordinates.

The problem of camera calibration has found great interest in the computer vision community and a large number of calibration methods have been proposed over the years. Among the linear calibration methods, the algorithm proposed by Ravn et al. [10] is attractive since it is simple to implement and only requires a planar calibration target that is easily manufactured. The linear calibration method does not permit the determination of the non-linear distortion parameters, and it requires that the calibration points on the target are known with a high degree of accuracy. These limitations can be overcome using the bundle adjustment technique developed for photogrammetry applications. The bundle adjustment method permits the calibration of non-linear camera models while simultaneously determining the shape of the calibration object up to a scale factor. This significantly reduces the requirements on the calibration target, since only the distance between two points has to be accurately known to resolve the scale factor.

So bundle adjustment method can greatly improve the accuracy of photogrammetric system. The error equation of bundle adjustment based on collinearity equation is written as follows [11] :

$$v = At + BX_1 + CX_2 + D_{ad}X_{ad} - L \qquad (1)$$

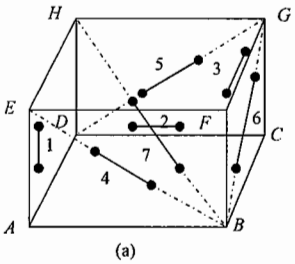
Where t, X_1, X_2 and X_{ad} are the vector of correction number of external parameters , intrinsic parameters , coordinate of object and distortion parameters, respectively. A, B, C, D_{ad} matrix of the corresponding coefficients. L vector of observer and test value.

4 EXPERIMENT BASED ON VDI2634

As can be seen from Fig.9 (a), its design is according to Germany VDI2634 technique standard. The distance errors between spheres are determined and tested in the whole measurement body, testing framework (2000mm × 2000mm × 1500mm) made by ourselves is adopted to conduct measurement (Fig.9 b). The reference point and the gauge are arranged as Fig.9 (a).It is a key factor that taken photos from different location , 36 pieces of pictures are taken according to camera station shown in Fig.8. The other 3 pieces of pictures which are elevations of gauge must be considered. Hence, total of about 39 photographs are arranged as a group. XJTUDP and TRITOP are tested, respectively. The reference point coordinate is reconstructed, measured and compared. (Fig.10). This experiment shows that the accuracy and robustness of XJTUDP is good for measurement [12].

5 APPLICATION OF XJTUDP TO INDUSTRIAL INSPECTION

That close-range industrial photogrammetry software XJTUDP and surface scanning software XJTUOM are combined[13], which can improve the measurement accuracy and avoid the accumulated error of point-cloud matching using surface-scanning software single. Moreover, it can implement automatic matching of point-clouds, reducing workers labor



(a)

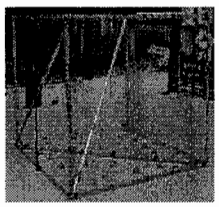

(b)

Fig.9 Sketch and framework of measuring and testing

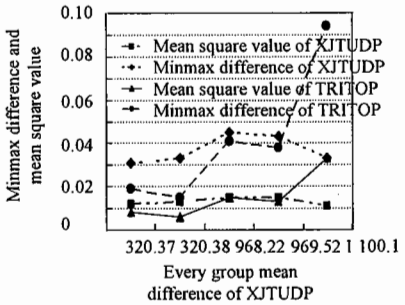
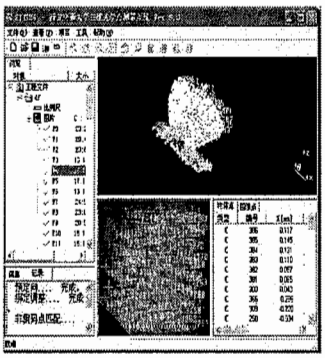


Fig.10 Comparison between XJTUDP precision and TRITOP precision

intensity. The coded-points are placed on the surface of object waiting for measuring, and then the global coordinates of an object are obtained through multi-stations taking. On the basis of imported un-coded point's matrix, surface scanning software is accessed to scan each pieces of point cloud through dynamic coordinate transformation, and then the final points-cloud has been merged and arrived out using Geomagic Studio.Fig.11 shows the program of obtaining large-scale waterwheel leaves shape, which is adopted by XJTUDP and surface scanning software XJTUOM. The actual results imply that photogrammetry technology may be applied in industrial measurement field.


(a)


(b)

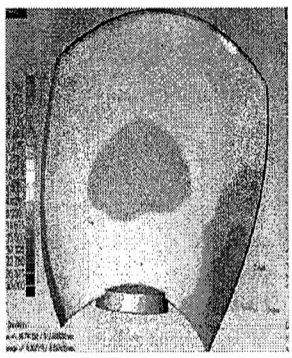

(c)

Fig.11 Measuring program of large-scale waterwheel leave using photogrammery software XJTUDP (a)fixed the marked points on the surface of part (b) compute coordinate of marked points (c) the final shape of parts in Geomagic Studio

6 CONCLUSIONS

XJTUDP software has been developed successfully based on photogrammetry theory. The VDI/VDE2634 testing program is taken as referenced project, the self-designed framework of cube is taken as tested target, large-scale waterwheel leave is taken as experiment example, all of them are proved that photogrammetry system explored by ourselves may be imposed applying to industrial inspection.

Acknowledgement

This work is supported by the National Natural Science Foundation of China for financial support (No. 50975219).

References

[1] Wasil, A., and Merchant, D., C., "Plate-deflection measurement by photogrammetric methods,"Exp.Mech. Papers 4(3), 77–83(1964).

[2] Morimoto, Y., and Fujigaki, M., "Automated analysis of 3-D shape and surface strain distributions of a moving object using stereo vision," Opt. Lasers Eng.Papers 18(3), 195–212(1993).

[3] Andresen, K., "Strain tensor for large three-dimensional surface deformation of sheet metal from an object grating," Exp.Mech.Papers 39(1), 30–35(1999).

[4] Tiwari, V., Sutton, M., A., and McNeill, S., R.,"Assessment of high speed imaging systems for 2D and 3D deformation measurements: Methodology development and validation,"Exp.Mech.Papers 47(4), 561–79(2007).

[5] Bay, B., K., Smith, T., S., Fyhrie, D., P., and Saad, M.,"Digital volume correlation: Three-dimensional strain mapping using X-ray tomography, "Exp.Mech.Papers 39(3), 217–226(1999).

[6] McKinley, T., O., and Bay, B., K., "Trabecular bone strain changes associated with subchondral stiffening of the proximal tibia," J.Biomech.Papers 36(2), 155–63(2003).

[7] Gao, W., Dejima, S., Yanai, H., Katakura, K., Kiyono, S., and Tomita, Y.," A surface motor-driven planar motion stage integrated with an $XY\theta_Z$ surface encoder for precision positioning,"Precis.Eng.Papers 28(3), 329-37(2004).

[8] Malgorzata, K., "Architecture of a multipurpose fringe pattern analysis system," Opt. Lasers Eng.Papers 19(4-5), 261-268(1993).

[9] Bobick, A., and Davis, J., "The recognition of human movement using temporal templates," IEEE Trans. Pattern. Analysis and Machine Intelligence. Papers 23(3), 257-267(2001).

[10] Ravn, O., Andersen, N., A., and Sorensen, A., T., "Auto-calibration in Automation Systems using Vision," Proc.ISER'93,206–218 (1993).

[11] Feng, W., H., [Industrial Measuring], Wuhan University Publishing, Wuhan, 113-253(2004).

[12] Zhang,D.,H., Liang,J., Tang,Z.,Z., Guo,C.,Cao,J.,M., and Wang,H., "Exploitation and study of close range industry photogrammetry system based on large-scale complex surface products," Opto-Electronic Eng.Papers 36(5),122-28 (2009).

[13] Zhang,D.,H., Liang,J., Tang,Z.,Z., and Guo,C., "New measuring method of large size measurement based on close range photogrammetry and 3D optical measurement," China Mech.Eng.Papers 20(7), 817-22 (2009).

Wavelet Based Technique for Texture Classification

Yogita K.Dubey*, Amoli D.Belsare, Milind M.Mushrif
Department of Electronics & Telecommunication;
Yeshwantrao Chavan College of Engineering, India

ABSTRACT

This paper presents a technique for texture feature extraction and classification using wavelet transform. A image is decomposed into no. of sub-bands after applying Wavelet transform to it. A three level decomposition is carried out. A number of sub-bands are generated after wavelet decomposition. An energy signature is computed for each sub-band of these wavelet coefficients. A k-nearest neighbor's classifier is then employed to classify texture patterns. To test and evaluate the method, several sets of textures along with different wavelet bases are employed. Experimental results show superiority of the proposed method.

Keywords: Texture classification, Wavelet transform, Feature Extraction, Classification

1 INTRODUCTION

Today, texture analysis plays an important role in many tasks, ranging from remote sensing to medical imaging. Image analysis is mainly based on three main image features: color, shape and texture. Texture plays an important role in human vision. The past few decades have witnessed a substantial interest in the analysis of textured images. The interest has been motivated in large part by the huge number of applications in such diverse areas as remote sensing, medical image analysis, and classification. Image texture, defined as a function of the spatial variation in pixel intensities (gray values), is useful in a variety of applications and has been a subject of intense study by many researchers. One immediate application of image texture is the recognition of image regions using texture properties. Texture is the most important visual cue in identifying the types of homogeneous regions. This is called *texture classification*.

The goal of texture classification is to produce a classification map of the input image where each uniform textured region is identified with the texture class it belongs to. Texture analysis is a very vast area and it has been widely studied by many researchers. The reason for this is the complexity and lack of ability to clearly define the significant features of textures. So a number of challenging problems still need to be addressed. The main difficulty of texture analysis in the past was the lack of adequate tools to characterize different scales of textures effectively. Many texture classification techniques have been proposed in the literature.

Among the many texture analysis algorithms proposed in the literature, statistical approaches [1] and model-based approaches [2] are some of the more commonly used methods. A review of invariant texture analysis method is given in [3].Wavelet based multi scale analysis is an important way in model based texture analysis. Traditional multi scale decomposition of the 2-D image is to perform wavelet transform with 2-D image data directly. It provides spatial/frequency information of texture, which is very useful for classification and segmentation. Wavelet transform captures the variation along specific directions (namely, horizontal, vertical and diagonal) due to severability of the wavelet basis functions. Chang and kuo [4] use the tree-structured wavelet transform for texture classification

In this paper, the DWT is applied on a set of texture images and energy features are extracted from the approximation and detail regions of DWT decomposed images, at different scales. In order to improve the success rate of classification, the feature values are normalized by the maximum value in feature set.

This paper is organized as follow: In section 2, a brief review of Wavelet Transform is given. In section 3, a method for texture classification is given. Experimental results are described in section 4 and conclusions are presented in section 5.

* Email Id. yogeetakdubey@yahoo.co.in , Mb. No. 09326351135

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462W · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856315

2 BASIC DETAILS OF WAVELET TRANSFORM AND THEIR PRPERTIES

2.1 Wavelet Transform

The beginning of wavelet transform as a specialized field that can be trace to the work of Grossman and Morlet in 1984. For a real and complex value continuous time function ψ(t) with the following properties,

1) The function integrates to zero

$$\int_{-\infty}^{\infty} \psi(t)dt = 0 \tag{1}$$

2) It is square integral or equivalently has finite energy.

$$\int_{-\infty}^{\infty} |\psi(t)|^2 dt < 0 \tag{2}$$

the function ψ(t) is a mother wavelet or wavelet, if it satisfies these two properties. Property 1&2 are sufficient to define the continuous wavelet transform (CWT), and they acquire essentially the reasons for calling the function a wavelet. Property 2 implies that most of the energy in ψ(t) is confined to finite duration. Property 1 is suggestive of a function that is oscillatory or that has a wavy appearance. Thus in contrast to sinusoidal function, it is a "small wave" or a wavelet. Wavelet transform provides a time-frequency representation of a signal. Wavelet coefficients of signal f(x) are the projection of signal onto the multi-resolution subspaces $V_j = \overline{span_k\{\varphi_{j,k}(x), k \in j\}}$ and $W_j = \overline{span_k\{\Psi_{j,k}(x) k \in j\}}$, k ∈ Z(set of integer) where the basis functions $\varphi_{j,k}(x)$ and $\psi_{j,k}(x)$ given by equation (3) and (4) are constructed by dilation and translation of the scaling and wavelet function *Φ(x)* and *ψ(x)*

$$\varphi_{j,k(x)} = 2^{j/2}\varphi(2^j x - k) \tag{3}$$

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \tag{4}$$

The scaling and wavelet function satisfy the dilation and wavelet equation

$$\varphi(x) = \sum_n h_\varphi(n)\sqrt{2}\varphi(2x - n) \tag{5}$$

$$\psi(x) = \sum_n h_\psi(n)\sqrt{2}\psi(2x - n) \tag{6}$$

Where $h_\varphi(n)$ and $h_\psi(n)$ are called as scaling function coefficient and wavelet function coefficient respectively and n ∈ Z.

2.2 Wavelet Series Expansion

For any function f(x) ∈ L^2(R), (the set of all measurable, square-integral function), relative to wavelet $\psi(x)$ and scaling function $\varphi(x)$ we have

$$f(x) = \sum_k c_{j_0}(k)\varphi_{j_0,k}(x) + \sum_{j=j_0}^{\infty}\sum_k d_j\psi_{j,k}(x) \tag{7}$$

Where j_0 is any arbitrary starting scale and $c_{j_0}(k)$ and $d_j(k)$ are called as approximation or scaling coefficient and detail or wavelet coefficients respectively are given by equation (8) and (9).

$$c_{j_0}(k) = \int f(x)\varphi_{j_0,k}(x)dx \tag{8}$$

$$d_j(k) = \int f(x)\psi_{j,k}(x)dx \tag{9}$$

Using the coefficients $c_{j,k}$ at specific level j, we can calculate the coefficients at level j-1 using a filter bank as shown fig.1. H and G are low pass and high pass filters corresponding to $h_\varphi(n)$ and $h_\psi(n)$.

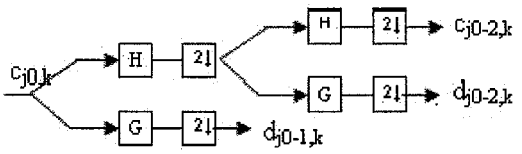


Figure 1 Filter bank used to calculate the wavelet coefficients

2.3 Discrete Wavelet Transform

If the function being expanded is a sequence of numbers like samples of continuous function f(t) the resulting coefficients are called discrete wavelet (DWT) of f(t). so far in this case the equation (8), (9) and (7) becomes as follows

$$W_\varphi(j,k) = \frac{1}{\sqrt{M}}\sum_x f(x)\varphi_{j_0,k} \tag{10}$$

$$W_\psi(j,k) = \frac{1}{\sqrt{M}}\sum_x f(x)\psi_{j_0,k} \tag{11}$$

$$f(x) = \frac{1}{\sqrt{M}}\sum_k W_\varphi(j_0,k)\varphi_{(j_0,k)}(x) + \frac{1}{\sqrt{M}}\sum_{J=J_0}\sum_k W_\psi(j,k)\psi_{(j,k)}(x) \tag{12}$$

Here f(x), $\varphi_{j_0,k}(x)$ and $\psi_{j,k}(x)$ are the function of discrete variable x=0,1,2,....M-1. The normalizing factor $1/\sqrt{M}$ has been added to both forward and inverse impressions. This factor alternately be incorporated as 1/M in the inverse alone or forward alone. The wavelet decomposition of 2-D signal can be achieved by applying the 1-D wavelet decomposition along the rows and columns of the image separately.

3 PROPOSED METHOD FOR TEXTURE ANALYSIS

In this section, the texture analysis technique using wavelet transform is introduced. The block diagram is shown in Fig. 2.

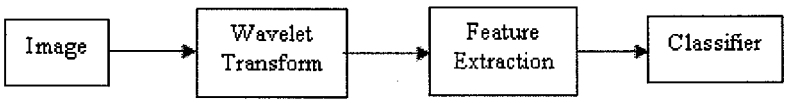
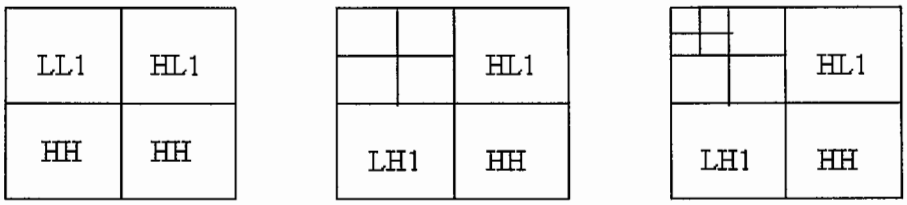


Figure 2 Block diagram representation of proposed method for texture feature classification.

3.1 Discrete Wavelet Transform

Using Wavelet transform, the image is decomposed into four sub-bands and critically sub-sampled by applying DWT as shown in Fig. 3(a). These sub-bands (LH1, HL1 and HH1) represent the finest scale wavelet coefficients i.e., detail images, while the sub-band LL1 corresponds to coarse level coefficients i.e., approximation image. To obtain the next coarse level of wavelet coefficients, the sub-band LL1 alone is further decomposed. This results in two-level wavelet

decomposition as shown in Fig. 3(b). Similarly, to obtain further decomposition in third level, LL2 will be used to decompose Fig.3(c). Thus the images are decomposed by three level of wavelet decomposition (three different scales).



(a) (b) (c)

Figure 3 Decomposition of Image. (a) One-level, (b) Two-level, (c) Three-level

3.2 Wavelet Feature Extraction

By applying the wavelet transform to the image, a number of sub-bands are produced. For each sub-band the following features are calculated

$$e_1 = \frac{1}{MxN}\sum_{i=1}^{M}\sum_{j=1}^{N}|C_{i,j}| \tag{13}$$

$$e_2 = \frac{1}{MxN}\sum_{i=1}^{M}\sum_{j=1}^{N}\sqrt{|C_{i,j}|} \tag{14}$$

where M and N are the dimension of each sub-band. Feature e_1 shows amount of signal energy, e_2 is also the amount of signal energy with scaling provided by taking square root of absolute values.After three level decomposition of image, 12 sub-bands are produced. For each sub bands, feature values e_1 and e_2 are calculated. Feature vector for each feature can be formed as

$F_1=[f_{11},f_{12},\ldots\ldots\ldots\ldots f_{112}]$, for energy feature e_1 and

$F_2=[f_{21},f_{22},\ldots\ldots\ldots\ldots f_{212}]$, for energy feature e_2.

Since different features have different range of possible values and the entire feature may not have the same level of significance because after decomposition of image the sub bands with the highest resolution corresponds to noise and are not valuable for classification. So all these feature values are normalized in the range 0 and 1 by the maximum value in the feature space before classification.

3.3 Classification

The K-NN classifier[5] with Euclidian distance, equation (15) is used to classify each image into appropriate class. In this algorithm the unknown sample is assigned to the class most commonly represented in the collection of its k-NN. The k-NN are the samples of the training set which have minimum distance to the unknown sample in the feature set. The minimum distance is calculated as

$$di_{EUC} = \sqrt{\sum_{j=1}^{m}(x_i - p_{i,j})^2} \tag{15}$$

where m represent the number of features. X_j represent the j^{th} feature of unknown texture, while $P_{i,j}$ represents the j^{th} features of known i^{th} class in the feature set. Where i=1,2,.....n (n represent number of classes). In this paper 30 texture images are considered, so number of classes are 30.

4 EXPERIMENTAL RESULTS

For the method, 30 texture images of size 640x640 from Brodatz album [6] were used. 30 images of size 512x512 are created from the center of each of the images. The images are shown in fig 4.

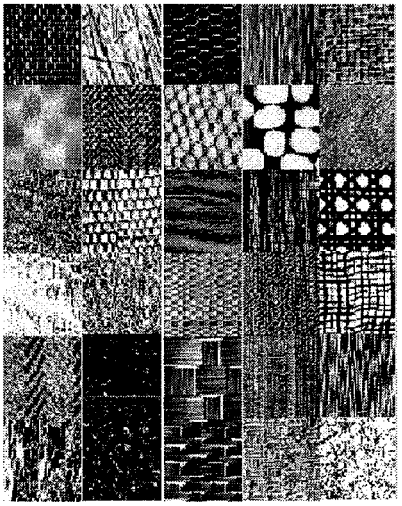


Figure 4 Row 1: d3, d11, d24, d68, d84; Row 2: d2, d12, d25, d75, d93; Row3: d6, d15, d26, d76, d102; Row4: d7, d16, d29, d77, d104; Row5: d8, d17, d64, d78, d105; Row6: d9, d23, d65, d82, d110.

We demonstrate the efficiency of our approach using two different data sets. Data set I consist of texture images of size 128x128 and the other set is of size 64x64 (data set II).The data sets were obtained by dividing each texture image into 16, 128x128 non overlapping region and 64, 64x64 non overlapping regions. For data set I the 4 subsample images out of 16 and for data set II the 16 subsample images out of 64 were used to create the training set of 120(30x4) images (data set I) and 480(30x16) (data set II). The remaining 360(30x12) from data set I and 1440(30x48) from data set II were used for testing. The proposed algorithm is applied to each texture image in the database. Three level of wavelet decomposition with Daubechies wavelet of length 4 and 6 (db4 and db6) is done. After three level decomposition of image, 12 sub-bands are produced. For each sub bands, feature values e1 and e2 are calculated. Experimental results with correct classification percentages are shown in Table 1.With the proposed method average classification of 96.1 % with e_1 and 96.9 % with e2 for data set I and 95.89% with e_1 and 96.3 % e_2 features is obtained by using Daubechies wavelet of length 6 (db6), and 4 (db4) respectively. For data set II the correct classification percentages of 91.94 % with e_1 and 92.98 % with e2 and 92.36 % with e_1 and 93.6 % with e_2 features is obtained by using Daubechies wavelet of length 6 (db6), and 4 (db4) respectively as shown in Table 1.

In this way, a feature database of all feature vectors of all images in the dataset is created. For the similarity matching, the k-nearest neighbor search is used to classify the similar images in the dataset. When the system receives a query image, it computes the feature vector of the query image by the proposed algorithm and compares with all feature vectors in the feature database. The similarity between the query image x and the i^{th} image in the dataset is defined by the Euclidean distance and the query image is classified to its proper class on the basis of minimum distance to its nearest neighbor by the K-NN classifier, where K represent the number of neighbors.

5 CONCLUSIONS

In this paper a technique for texture classification using Wavelet transform is presented. The discrete wavelet transform was demonstrated to be an excellent feature extraction tool. In this work, it is shown that the classification performance varies for different wavelet bases. The choice of wavelet basis plays a crucial role in wavelet-based analysis. Large databases of 1800(360 + 1440) samples of images are used to verify the result.

Table 1 Correct Classification Percentage

		Data Set I				Data Set II			
Wavelet		**db6**		**db4**		**db6**		**db4**	
Sr.No.	Image	e1	e2	e1	e2	e1	e2	e1	e2
1	d3	100	100	100	100	100	100	100	100
2	d5	91.667	100	91.667	100	91.667	100	89.583	100
3	d6	100	100	100	100	85.417	75	87.5	87.5
4	d7	100	91.667	100	100	87.5	79.167	89.583	81.25
5	d8	100	100	100	100	100	100	100	100
6	d9	91.667	91.667	83.333	83.333	68.75	70.833	70.833	70.833
7	d11	91.667	91.667	91.667	91.667	85.417	89.583	89.583	87.5
8	d12	100	100	100	100	100	100	100	100
9	d15	75	83.333	75	75	64.583	72.917	66.667	72.917
10	d16	75	75	75	75	79.167	79.167	75	77.083
11	d17	100	100	100	91.667	87.5	85.417	89.583	87.5
12	d23	100	100	100	100	93.75	95.833	89.583	97.917
13	d24	100	100	100	100	87.5	93.75	85.417	91.667
14	d25	100	100	100	100	100	100	97.917	97.917
15	d26	100	100	100	100	100	100	100	100
16	d29	100	100	100	100	100	100	100	100
17	d64	100	100	100	100	93.75	97.917	97.917	100
18	d65	100	100	100	100	95.833	95.833	100	97.917
19	d68	100	100	100	100	100	100	100	100
20	d75	100	100	100	100	95.833	100	97.917	100
21	d76	100	100	100	100	91.667	93.75	93.75	97.917
22	d77	100	100	100	100	100	100	100	100
23	d78	91.667	100	91.667	100	95.833	93.75	97.917	97.917
24	d82	100	100	100	100	100	100	100	100
25	d84	100	100	100	100	100	100	100	100
26	d93	66.667	75	66.667	75	60.417	72.917	62.5	68.75
27	d102	100	100	100	100	100	100	100	100
28	d104	100	100	100	100	100	100	97.917	100
29	d105	100	100	100	100	100	100	100	100
30	d110	100	100	100	100	93.75	93.75	91.667	93.75
Avg. classification percentage		**96.111**	**96.944**	**95.833**	**96.389**	**91.944**	**92.986**	**92.361**	**93.611**

REFERENCES

1. R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," IEEE Trans. Syst., Man, Cybern., vol. SMC-3, no. 6, pp. 610–621, 1973.
2. R. C. Dubes and A. K. Jain, "Random field models in image analysis," J. Appl. Stat., vol. 6, pp. 131–164, 1989.
3. J. Zhang and T. Tan, "Brief review of invariant texture analysis methods." Pattern Recognition vol. 35 pp. 735–747, 2002
4. T. Chang and C.J. Kuo, "Texture analysis and classification with tree structured wavelet transform." IEEE Trans. Image Processing, vol. 2 no. 10. pp. 429-441 Oct. 2003.
5. E. Gose, R. Johnsonbaugh, and S. Jost, [Pattern Recognition and Image Analysis], NJ: Prentice-Hall, 1996.
6. P. Brodatz, Texture: A Photographic Album for Artist and Designers. New York: Dover, 1996.

Image processing techniques applied to the detection of Optic disk-a comparison

V.VIJAYA KUMARI
Department of ECE, V.L.B. Janakiammal College of Engineering and Technology
Coimbatore 641 042, India.
email: ebinviji@rediffmail.com

N.SURIYANARAYANAN
Department of Physics, Government College of Technology
Coimbatore 641 013, India.
email: esnsuri@yahoo.co.in

ABSTRACT

In retinal image analysis, the detection of optic disk is of paramount importance. It facilitates the tracking of various anatomical features and also in the extraction of exudates, drusens etc., present in the retina of human eye. The health of retina crumbles with age in some people during the presence of exudates causing Diabetic Retinopathy. The existence of exudates increases the risk for age related macular Degeneration (AMRD) and it is the leading cause for blindness in people above the age of 50.A prompt diagnosis when the disease is at the early stage can help to prevent irreversible damages to the diabetic eye. Screening to detect diabetic retinopathy helps to prevent the visual loss. The optic disk detection is the rudimentary requirement for the screening. In this paper few methods for optic disk detection were compared which uses both the properties of optic disk and model based approaches. They are uniquely used to give accurate results in the retinal images.

Key words: Diabetic Retinopathy, optic disk, exudates, retina, macular degeneration.

1 INTRODUCTION

Diabetic Retinopathy (DR) is a result of long term diabetes mellitus. It has been noted as a significant growing public health problem. It is one of the predominant causes of blindness. It causes pathological changes of the retina such as microaneurysms, intraretinal microvascular abnormalities, venous bleeding and neovascularities as well as haemorrhages, exudates and retinal oedema. Regular screening of Diabetic Retinopathy is indispensable so that appropriate and timely treatment can be given which thereby reduces the incidence of impaired vision and blindness from this condition. Current methods of detection and assessment of diabetic retinopathy are manual, expensive and require highly trained personnel to read large number of fundus images. The efficiency can be improved by automating the initial task of analyzing the huge amount of retinal fundus images.

The optic disk is the brightest part in fundus images that can be seen as a pale, round or slightly oval disk. It is the entrance region of blood vessels and also acts as a landmark and reference for the other features in the fundus image. There are several methods for optic disk detection. In the Principal Component analysis method (PCA), the minimum distance between the original retinal image and its projection onto disk space is located as the center of Optic disk[1]. This detection is accurate but more time consuming. In the next method PCA and active shape model [2] is used, where the shape of the OD is obtained by an active shape method. Here the affine transformation is used to transform the shape from shape space to the image space. This algorithm takes the advantage of top down processing that increases the robustness yet it is time consuming. In lab color morphology[3],the location of optic disk was by both the automatic initialization of snake and the application of morphology in color space[4].In Sobel edge detection and least square regression[5], the detection is performed in the red component in three steps-candidate area identification, sobel edge detection and estimation step[6]. Lalonde et.al used a Hausdorff-based template matching technique on edge map, guided by a pyramidal decomposition [7]for large scale object tracking .The cholesky algorithm[8] was used but the snake had failed to locate the boundary in the upper right quadrant. In this paper, the detection of optic disk is done using combination of various concepts. Initially the candidate regions are selected and by using different methods such as

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462X · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.852380

iteration, binary imaging, clustering and PCA, propagation through radii, the location of optic disk is detected and compared.

2 METHODS

2.1 Binary imaging method

The image acquired from the camera is converted to gray scale containing 256 levels. Then the threshold is calculated by finding the mean intensity of the image and multiplying it by suitable scaling factor. This factor is decided by user intervention which is usually done on a trial and error basis. Then the gray image is scanned pixel by pixel. For every pixel, if it exceeds the threshold level it is made white else the pixel is made black. Thus the gray scale image is converted into a binary image containing only two levels. The white pixels in the binary image constitute the optic disc and these pixels can be mapped on to the RGB image and the optic disc can be identified.

The snag in this method is that in few images exudates becomes brighter than the optic disc. The processing steps require different threshold parameters for different types of retinal images and need users intervention on a case by case basis. As a result this thresholding-based algorithm is not scalable for analyzing large number of retinal images.

2.2 Iteration method

This method is based on density. Initially the RGB image obtained from the fundus camera is converted to Gray-scale and then a window of arbitrary size is chosen and it is made to scan the entire retinal image. This works in such a way that, the number of pixels exceeding a intensity value in the window is compared with a certain threshold value, when the density of pixels is greater than the threshold it makes the entire area under the window to be white, instead if the density of pixels is less than the threshold, it makes the entire area under the window to be black. This is the first iteration. In the subsequent iterations, the window size is increased from the previous one and the output of first iteration is fed as input and the above steps are repeated again. The whole process is repeated for a finite number of times. Finally we get a binary image with a white rectangle representing the location of optic disc. This can be mapped on to the RGB image. Thus in this iterative manner the approximate localization of Optic Disc is done.

In this method the window moves sequentially. So a tiny part of the optic disc perhaps the edge may fall into a window such that the major part is the surrounding pixels of the optic disc. Since the entire area under the window is made black the edges may get eroded and this affects the final result.

2.3 Clustering method

This is a distance based method. The calculation of threshold is done by simple mean estimation method and multiplying the mean image by a scaling factor. A scaling factor of 0.85 is appropriate. The gray scale image is initially scanned pixel by pixel. For every pixel exceeding the threshold a box of size (5 x 5) is constructed. This box is called a cluster. The x and y co-ordinates of the pixel is updated into two vectors row and column. Starting from the first entry in the vectors a box with coordinate lying in an acceptable level of distance is searched. To compute the distance basic mathematical formula

$$\sqrt{(x2-x1)^2+(y2-y1)^2} \quad \text{----------(1)}$$

is used. Once found the boxes are joined. The boxes are joined based on their location.

Say for example, (x1, y1) be the centroid of first cluster and (x2, y2) be the centroid of second cluster. After this the basic mathematical formula to evaluate the distance between two points is applied to find out the distance between the clusters. Keeping (x1, y1) as the reference point the distance between this reference and the center point of every other cluster is determined. The clusters having centroids within a specified distance from this reference centroid point are combined. The north-west point of the reference cluster and the South-east point of the cluster which is closer to it are merged to form a single cluster. Then subsequently these centroids are removed from the list of centroids to be evaluated. This phenomenon is repeated for every other centroid and the clusters are regrouped. This method is repeated above for certain time which gives certain candidate regions. For every candidate, number of pixels exceeding the

threshold is found. If it is below a certain level determined by the size of the candidate region the candidates are discarded. Finally we have an image in which the candidate region completely encloses the optic disc though it may not exactly square it.

This is suitable for normal healthy fundus images where in optic disc is alone the brightest region of the image. But our images contain exudates and lesions. The exudates' intensity value is similar to that of optic disc. If it is small in size then it can be characterized by the intense pixels density. But in some images the exudates are as large and dense as exudates. In such cases it is difficult to find which of the candidate contains the optic disc.

2.4 Principal component analysis between clusters

PCA is a statistical method. For every pixel a window is constructed around it with the pixel as the centre .For every box PCA is applied and the one having minimum Euclidian distance contains optic disc. Our test image size is 240 x 180.So the PCA process has to be applied 43200 times. So it is very time consuming and the method is not accurate enough to wait so long. Clustering does not produces correct output for images with exudates. The advantage of the PCA and clustering can be combined. The main extract of this method is that, the candidate regions are first determined by clustering the brightest pixels in intensity image. Principal Component Analysis is then applied to these candidate regions. The candidate region having the least Euclidian distance contains the optic disc. Though the centre of the candidate region and centre of the optic disc does not coincide, the region encloses the optic disc.

Candidate region is determined by clustering method. The clustering method produces candidate regions of varying dimensions. But for applying PCA we need the candidate regions to be of fixed predetermined size. (40 x 40) in our case, so the candidates are resized. For every fixed size cluster we find the brightest pixel and construct a 40 x 40 box. The fixed size clusters and these boxes are the entries for PCA. Hence if there are n candidates we will have 2n PCA entries. Now we apply PCA only to these entries rather than to every box cantered around every pixels of the image. The training set can be extracted individually by moving a window of size 40 x40 over this single image[fig 2.1]. These extracted sub images are stored in a three dimensional matrix. So the third dimension of this matrix represents a page and each page contains a individual sub image. By doing this the access of the individual sub images becomes easy as they have the same name and only differs in the index of the third dimension which is a numerical value. By incrementing this numerical value in a for loop, the consecutive sub images which form the training dataset can be accessed.

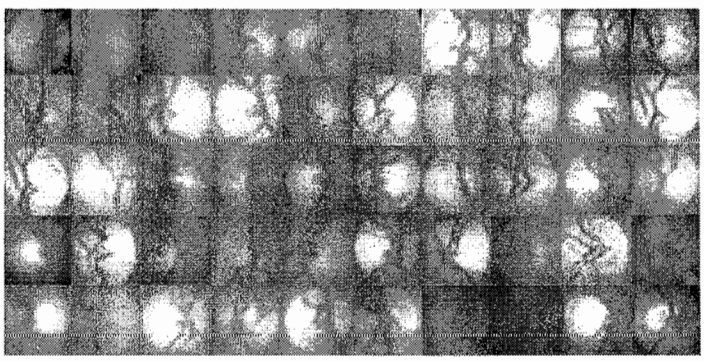


Figure 2.1 Training image

2.5 Propagation through radii

PCA between clusters produces output such that the optic disc is enclosed in a square box of size (40 x 40). For fixed size boxes 1600 pixels are extracted considering it to be containing the optic disc. But in many images the optic disc occupies only around 1000 pixels. If any exudates lie within proximity of the optic disc then it may also be

extracted. This may increase the risk of abnormal retina being reported normal. In few images the optic disc is too large to completely fit within the (40 x 40) box. So few of the optic disc pixels may remain unextracted. This may intervene in exudate detection. So in either case extracting pixels under a fixed size area is not efficient. To avoid this we tried to define a circle (circle of best fit) that encloses the optic disc with minimum number of pixels that does not belong to the optic disc. The original fundus image is subjected to contrast limited Adaptive Histogram Equalization.

The centre of the box containing optic disc is defined by coordinates (x, y). In general to represent a circle in discrete space we use $x+r\cos\theta$, $y+r\sin\theta$. θ lies from 0 to 360. The arbitrary initial value of r is 10. Thus we have ({x+10cos0, y+10sin0}, {x+10c0s1, y+10sin1}... {x+10cos360,y+10sin360}). Each of them representing a point in the circumference of the circle. Now the first point (corresponding to $\theta=0$) on the circumference is taken. Keeping it as the centre point and a pixel is chosen above $(x+(r+1)\cos\theta, y+(r+1)\sin\theta)$ and below$(x+(r-1)\cos\theta, y+(r-1)\sin\theta)$ that. The difference of pixel intensity value between the upper point and center point and also between lower point and center point is computed. A threshold is set (5 in our case) and if the difference value is greater than the threshold make the center point is made black. This procedure is repeated until the difference value becomes greater than the threshold, each time shifting the centre point upward along the radius by incrementing the initial r value .The final r value is the radius corresponding to that θ. This is done for all the points present in the circumference of the circle i.e., for θ ranging from 0 to 360. Hence we get radius for all 360 degrees, and then the mean radius is found. Our next step is to find mean centre. This is done by finding midpoint of line joining circum points of supplementary angles for $\theta=0$ to 180.The mean of these x and y co-ordinates of the midpoints gives us the mean centre. The circle of best fit is drawn with mean centre as the centre and mean radius as the radius. As error margin can be added to the mean radius, this ensures that the optic disc is completely enclosed.

3 RESULTS

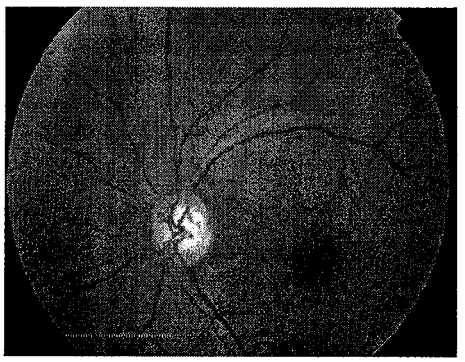


Figure 3.1 Input RGB image

Figure 3.2 Output binary image

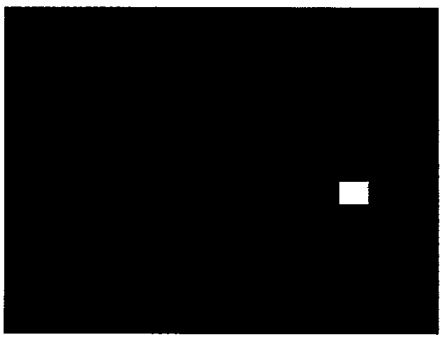


Figure 3.3 Iteration output

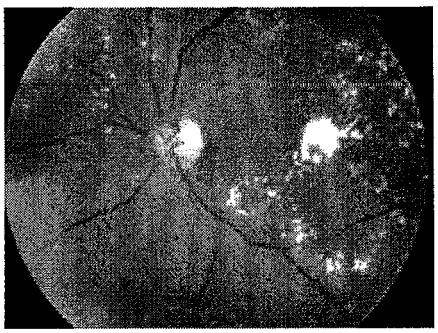


Figure 3.4 Abnormal input image

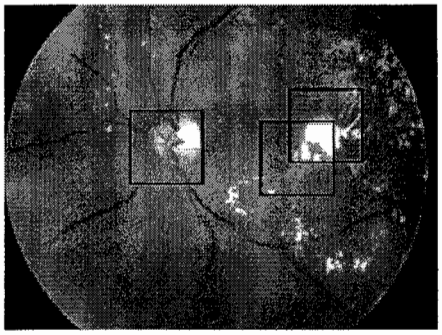


Figure 3.5 Clustering output

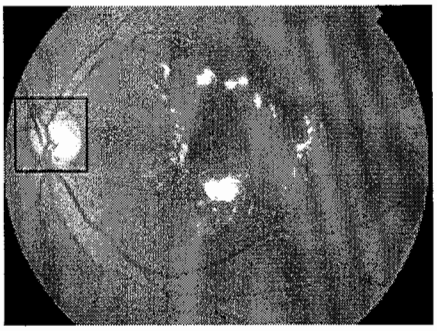


Figure 3.6 PCA Output

Figure 3.7 Input image with Optic disk squared

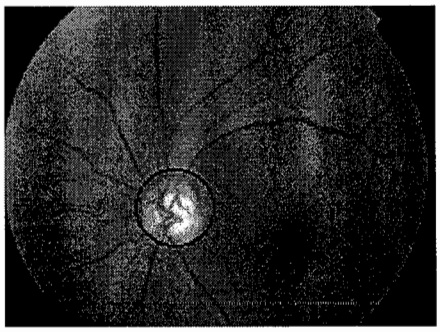


Figure 3.8 Propagation through radii output

4 DISCUSSION AND CONCLUSION

Binary imaging method can be used as a trial method. It is not suitable if there are lesions present and are brighter than the optic disk. Iteration method is better than binary imaging but gives an approximate location .Scaling factor of 0.85 is appropriate in clustering method and is suitable for normal healthy images where optic disk alone is the brightest region in the image. Using PCA between clusters, out of 50 images, accurate results are obtained for 44 images. The detection of exudates involves the removal of optic disk. Under these circumstances propagation through radii method can be employed. The various methods for optic disk detection is given in which the time consumption of PCA is overcome by applying PCA only to the clusters. The propagation through radii method is suitable for removing optic disk in the detection of exudates since it does not require the exact center of optic disk.

ACKNOWLEDGEMENT

Our heartfelt thanks to THE EYE FOUNDATION , Coimbatore for providing the images. Our thanks to V.H.Raghuvarran, R.Sujitha for their valuable support.

REFERENCES

1. Li. H. and Chutatape. O, "Automatic location of Optic Disc in Retinal Images", IEEE ICIP, pp: 837-840 (2001).
2. Li. H. and Chutatape. O. "A Model - based approach for Automated feature extraction in fundus Images", Proceedings of the 9th IEEE International Conference on Computer Vision, (2003).

3. Osareh. A, Mirmehdi. M, Thomas. B. and Markham. R, "Color Morphology and snakes for Optic Disc
4. Localisation", The 6th Medical Image Understanding and analysis Conference (2002).
5. Osareh. A, Mirmehdi. M, Thomas. B and Markham. R, "Comparison of Colour spaces for Optic Disc localization in Retinal images" Proceedings of 16th International Conference on Pattern Recognition, volume 1, pp:743-746(2002).
6. Li, H. and Chutatape, O, " Fundus image features extraction", Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Volume.4, pp:3071-3073(2000) .
7. Lupascu. C.A, Tegolo. D and Di Rosa. L, "Automated Detection of Optic Disk localization in Retinal images", 21st IEEE Int.Symp. on Computer-based medical System, pp :17-22(2008).
8. Lalond. M, Beaulieu, M. and Gagnon. L. "Fast and Robust Optic Disc Detection using Pyramidal decomposition", IEEE Transactions on Medical Imaging, Vol.20, No.11, pp:1193-1200 (2001).
9. Morris. D. T. and Donnison, C. "Identifying Neuro Retinal Rim Boundary Using Dynamic contours", Image Vision Computing, volume:17, pp:169-174(1999).

Association rule based tuberculosis disease diagnosis

ASHA.T*[a], S NATARAJAN[b], K.N.B. Murthy[a]
[a] Principal and Director, PESIT, Bangalore, India
[b] Professor, Department of Information Science and Engineering, PESIT, Bangalore 560085, India

ABSTRACT

Tuberculosis (TB) is a disease caused by bacteria called Mycobacterium tuberculosis. It usually spreads through the air and attacks low immune bodies such as patients with Human Immunodeficiency Virus (HIV). This work focuses on finding close association rules, a promising technique in Data Mining, within TB data. The proposed method first normalizes of raw data from medical records which includes categorical, nominal and continuous attributes and then determines Association Rules from the normalized data with different support and confidence. Association rules are applied on a real data set containing medical records of patients with TB obtained from a state hospital. The rules determined describes close association between one symptom to another; as an example, likelihood that an occurrence of sputum is closely associated with blood cough and HIV.

Keywords: Association Rule, Data Mining, Medical Diagnosis, Tuberculosis

1. INTRODUCTION

Data Mining is an active research area. One of the most popular approaches to do data mining is discovering association rules [7][8]. Association rules are generally used with basket, census or financial data. If used in the medical domain it has a great potential to improve disease prediction [10][17]. In this research work, association rules are mined on a medical data set which includes categorical, nominal and continuous attributes to improve TB disease diagnosis. Association rules have important advantages over traditional supervised machine learning or statistical algorithms. Tuberculosis is a common and often deadly infectious disease caused by mycobacterium; in humans it is mainly *Mycobacterium tuberculosis*. It is a great problem for most developing countries because of the low diagnosis and treatment opportunities. Tuberculosis has the highest mortality level among the diseases caused by a single type of microorganism. Thus, tuberculosis is a great health concern all over the world, and in India as well [wikipedia.org]. Work has already been carried out on TB prediction using artificial neural networks[1][2].

In this paper, the study involves finding close association rules within TB dataset using apriori algorithm. The focus is on two concepts described as follows. The first aspect is normalization of raw data from medical records which includes categorical, nominal and continuous attributes. The next aspect is discovering association rules from the normalized data with different support and confidence. Association rules are applied on a real data set containing medical records of patients with TB obtained from a state hospital. Rest of the paper is organized as follows. Section II states definitions of association rules while section III provides brief overview of literature survey. Section IV describes tuberculosis medical data and Section V addresses normalization process where different types of attributes are mapped to integer numbers. Section VI introduces experimental results of association rules obtained after applying apriori with different support and confidence. Finally Section VII contains conclusions.

2. DEFINITIONS

Here we give the classical definition of association rules. Let $\{ t_1, t_2, \ldots t_n \}$ be a set of transactions and let I be a set of items, $I=\{ I_1, I_2, \ldots I_m \}$. An association rule is an implication of the form $X \rightarrow Y$, where X,Y are disjoint subsets of item I and $X \cap Y = \phi$. X is called the *antecedent* and Y is called the *consequent* of the rule. In general, a set of items such as

*ashamasthi@rediffmail.com; phone 91-80-9880511358

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462Y · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853291

the antecedent or consequent of a rule is called an *Itemset*. Each *itemset* has an associated measure of statistical significance called *support*. *support(x)=s* is the fraction of the transactions in the database containing X. The rule has a measure of strength called *confidence* defined as the ratio *support(X ∪ Y) / support(X)*.

3. LITERATURE SURVEY

Association rules were introduced in a well-known article [7],[8] . More general research using association rules in the medical community includes the following. Researchers in other fields may not be aware of the particular constraints and difficulties of the privacy-sensitive, heterogeneous, but voluminous data of medicine. Ethical and legal aspects of medical data mining are discussed, including data ownership, fear of lawsuits, expected benefits, and special administrative issues. The mathematical understanding of estimation and hypothesis formation in medical data may be fundamentally different than those from other data collection activities discussed in [19]. Predicting heart disease has been automated by association rules. An algorithm that uses search constraints to reduce the number of rules was proposed which searches for association rules on a training set, and finally validates them on an independent test set. The medical significance of discovered rules is evaluated with support, confidence, and lift[3], [5] and [6] . Association rules describe what drugs are frequently coprescribed with antacids [18] . Association rules are applied for the analysis of human sleep data using polysomnographic readings. They introduced a specialized association rule mining technique that can extract patterns from complex sleep data comprising polysomnographic recordings, clinical summaries, and sleep questionnaire responses. The rules mined can describe associations among temporally annotated events and questionnaire or summary data; e.g., the likelihood that an occurrence of a rapid eye movement (REM) sleep stage during the second 100 sleep epochs of the night is associated with moderate caffeine intake[4] .

Boolean analyzer (BA), which finds dependency/association rules was used for indications of computed tomography in a head trauma dataset . The BA algorithm is given one or more metrics to partition the entire dataset. The weighted rules are in the form of Boolean expressions. To augment the analysis of the rules produced, they applied a probabilistic interestingness measure (PIM) to order the generated rules based on event dependency, where events are combinations of primed and unprimed variables. Interpretation of the dependency rules generated on the clinical head trauma data resulted in a set of criteria that identified minor head trauma patients needing computed tomography (CT) scans [12] . Frequent patterns in gene data have been discovered with association rules [11], [13] . Tuberculosis diagnosis is done using multilayer neural networks with two hidden layers and a genetic algorithm for training algorithm has been used[1] and prediction of active pulmonary tuberculosis[2] using general regression neural network has been carried out . Data mining approach was adopted to classify genotype of mycobacterium tuberculosis using c4.5 algorithm[9] . Important issues when using machine learning or data mining techniques in the medical domain [17] include fragmented data collection, strict privacy regulations, rich data types (image, numeric, categorical, missing information), complex taxonomies classifying attributes, and an already rich and complex knowledge base. Analysis on the database of diabetes patients with their death certificates was carried out to find association rules that describe associations made of patients at their first visit to the hospital and early mortality[10]. A new feature selection method based on association rules(AR) and neural network(NN) is presented for the diagnosis of erythemato-squamous diseases where AR is used for reducing the dimensions and NN is used for efficient classification[14].

4. MEDICAL DATA SET DESCRIPTION

The medical dataset we are mining includes real records of patients suffering from TB obtained from a state hospital. The entire dataset is put in one file having many records. Each record corresponds to most relevant information of one patient. The symptoms of each patient such as age, chroniccough(weeks), weightloss, intermittentfever(days), nightsweats , Sputum, wheezing and TBtype etc.,are considered as attributes

Table 1 shows names of 13 attributes considered along with their datatype(DT). Type N-indicates numerical and C-categorical.

Table 1. List of Attributes and their Datatypes

No	Name	DT
1	Age	N
2	chroniccough(weeks)	N
3	weightloss	C
4	intermittentfever(days)	N
5	nightsweats	C
6	Bloodcough	C
7	chestpain	C
8	Diabetesmellitus	C
9	HIV	C
10	Radiographicfindings	C
11	Sputum	C
12	wheezing	C
13	TBtype	C

5. NORMALIZATION

The above medical attributes are normalized to a suitable binary format. A categorical data field has a value selected from an available list of values. Such data items can be normalized by allocating a unique column number to each possible value. Numerical data fields take values that are within some range defined by minimum and maximum limits. In such cases we divided the given range into a number of sub-ranges and allocate a unique column number to each sub-range respectively.

Here we give a small example of five patients medical records with five attributes. Table 2 shows original data. Table 3 contains schema of how the attributes are mapped to individual column numbers. Table 4 is the final translated or normalized data.

Table 2. Original (raw) Data

Age	Chronic cough(weeks)	Weight loss	HIV	Sputum
17	3	Yes	Negative	Yes
13	6	Yes	Negative	Yes
45	6	Null	Negative	Yes
32	Null	Yes	Positive	Null
22	Null	Yes	Positive	Yes

Table 3. Schema Table

Age < 25	Age >= 25	Chronic cough (weeks)< 4	4 <= Chronic cough (weeks) < 8	Chronic cough (weeks) = Null	Weight loss = Yes
1	2	3	4	5	6

Weight loss = Null	HIV = Positive	HIV = Negative	Sputum = Yes	Sputum = Null
7	8	9	10	11

In the above tables, note that Age is a numerical attribute and its cut off point is <25 & >=25. Similarly HIV is a categorical attribute where positive value is assigned one number and negative another. The value Null for categorical attribute weightloss is equivalent to No and is assigned a unique number. By using the schema table above we map each tuple in the original data of table 2 to a resulting normalized table shown in table 4. Resulting table has the same number of columns as the original table but filled with unique integer values.

Table 4. Normalized Table

Age	Chronic cough(weeks)	Weight loss	HIV	Sputum
1	3	6	9	10
1	4	6	9	10
2	4	7	9	10
2	5	6	8	11
1	5	6	8	10

6. DISCOVERED ASSOCIATION RULES

Several medically important association rules are obtained after applying apriori algorithm to normalized table. Each association rule shows the relation between one symptom with the other. Data set was first tested by fixing support=70% and confidence=90%.We could get very few association rules, some shown in Listing 1.Relationship between all the attributes were not shown here. Next with 50% support and 90% confidence we could get large number of association rules, few listed in Listing 2 that provides more relationship with many frequent attributes.

Association Rules Using Output Schema with support=70% and confidence=90%

1. {chroniccough(weeks)<13.0} -> {Bloodcough=null} 100.0
2. {weightloss=yes} -> {Bloodcough=null} 100.0
3. {HIV=negative} -> {Bloodcough=null} 100.0
4. {Sputum=yes} -> {Bloodcough=null} 100.0
5. {HIV=negative} -> {Sputum=yes} 100.0
6. {Sputum=yes} -> {HIV=negative} 100.0
7. {Bloodcough=null HIV=negative} -> {Sputum=yes} 100.0
8. {wheezing=null} -> {Bloodcough=null} 100.0
9. {Bloodcough=null Sputum=yes} -> {HIV=negative} 100.0
10. {HIV=negative} -> {Bloodcough=null Sputum=yes} 100.0
11. {HIV=negative Sputum=yes} -> {Bloodcough=null} 100.0
12. {Sputum=yes} -> {Bloodcough=null HIV=negative} 100.0

Listing 1

Association Rules Using Output Schema with support=50% and confidence=70%

1. {chroniccough(weeks)<13.0 Bloodcough=null chestpain=null} -> {Sputum=yes TBtype=PTB} 75.0
2. {chroniccough(weeks)<13.0 Bloodcough=null Sputum=yes} -> {chestpain=null TBtype=PTB} 75.0
3. {weightloss=yes intermittentfever(days)<15.0} -> {HIV=negative} 100.0
4. {Diabetesmellitus=null} -> {weightloss=yes Bloodcough=null} 100.0
5. {Bloodcough=null HIV=negative Sputum=yes} -> {nightsweats=null wheezing=null} 80.0
6. {intermittentfever(days)<15.0} -> {HIV=negative} 100.0

7. {Sputum=yes TBtype=PTB} -> {HIV=negative} 100.0
8. {Radiographicfindings=yes} -> {chroniccough(weeks)<13.0 Bloodcough=null} 100.0
9. {Radiographicfindings=yes} -> {chestpain=null} 100.0
10. {intermittentfever(days)<15.0 Sputum=yes} -> {weightloss=yes Bloodcough=null} 100.0
11 {chestpain=null HIV=negative} -> {TBtype=PTB} 100.0
12. {chroniccough(weeks)<13.0 chestpain=null HIV=negative} -> {TBtype=PTB} 100.0
13. {chestpain=null Sputum=yes} -> {TBtype=PTB} 100.0
14. {HIV=negative Sputum=yes} -> {nightsweats=null wheezing=null} 80.0

Listing 2

Under Listing 1, rule 1 implies that those who have cough for less than 13 weeks may not cough blood. Rule 5 says that people without HIV may have their sputum tested positive which is the main test for TB. Rule 9 implies that those with sputum tested positive and without blood cough may not have HIV. Thus the rules above are close symptoms for diagnosing TB. People with sputum=yes or / and people with HIV= positive are more vulnerable to TB attack. Under Listing 2, rule 6 implies that if fever is less than for 15 days, their HIV is negative. Rule 13 implies that for those with no chest pain and sputum tested positive their TB type is pulmonary tuberculosis(PTB).

7. CONCLUSIONS

The proposed work helps in discovering association rules on tuberculosis data which associates closely related symptoms of TB. Apriori algorithm has been applied on real data of TB patients. This work proceeds in two steps. At first the original data is normalized into transaction format suitable for mining process. In this step unique integer numbers are assigned to each attribute based on whether the attribute is categorical or numerical. Next, association rules are discovered from the normalized data with different support and confidence values. Results of association have been shown in listings 1 and 2. The rules mined describes close association between one symptom to another; eg., likelihood that an occurrence of sputum is closely associated with blood cough and HIV. The same problem can be examined with missing values not considered in our problem. Tuberculosis clinical data set can also be tried with the same algorithm.

REFERENCES

[1] Orhan Er., Feyzullah Temurtas, Tantrikulu, A.C., " Tuberculosis disease diagnosis using Artificial Neural networks, " Journal of Medical Systems, category: online submission, Springer, DOI 10.1007/s10916-008-9241-x (2008)

[2] Ali. A. El-Solh, M.D., Chiu-Bin Hsiao, M.D., Susan Goodnough RN, et al " Predicting Active pulmonary Tuberculosis using an Artificial Neural network ," CHEST journal 116(4), 968-973 (1999)

[3] Carlos Ordonez, "Association Rule Discovery With the Train and Test Approach for Heart Disease prediction," IEEE Transactions on Information Technology in Biomedicine ,10(2), 334-343 (2006)

[4] Parameshvyas Laxminarayan, Sergio A. Alvarez, Carolina Ruiz, and Majaz Moonis, " Mining Statistically Significant Associations for Exploratory Analysis of Human Sleep Data," IEEE Transactions on Information Technology in Biomedicine, 10(3) ,440-450 (2006)

[5] Carlos Ordonez, Edward omiecinski, Cesar A. Santana, et al "Mining Constrained association Rules to Predict Heart Diseases," Proc. ICDM Nov., 433-440 (2001)

[6] Carlos Ordonez, Cesar A. Santana , Levien de Braal, "Discovering interesting association rules in medical data," Proc. ACM DMKD 2000 , 78-85 (2000)

[7] Rakesh Agrawal and Ramakrishnan Srikant, " Fast algorithms for mining association rules in large databases ," Proc. VLDB conference Sept 12-15, 487-499 (1994)

[8] Rakesh Agrawal, Tomasz Imielinski, and Arun Swamy, " Mining association rules between sets of items in large databases, " Proc. ACM SIGMOD international conference on management of data, 22(2), 207-216 (1993)

[9] Sebban, M. , Mokrousov, I., Rastogi, N. and Sola, C. " A data-mining approach to spacer oligo nucleotide typing of Mycobacterium tuberculosis," Bioinformatics, 18(2), 235-243 (2002)

[10] Richards, G., Rayward-smith, V.J., Sonksen, P. H., et al "Data Mining for indicators of early mortality in a database of clinical records," Artificial Intelligence in Medicine,Elsevier, 22(3), 215-231 (2001)
[11] Elisabeth Georgii, Sabine Dietmann, Takeaki Uno, Philipp Pagel and Koji Tsuda "Mining expression-dependent modules in the human interaction network " BMC *Bioinformatics*, 1 8(8), 1471-2105 (2007)
[12] Imberman, S., Domanski, B., Thompson, H. "Using dependency/association rules to find indications for computed tomography in a head trauma dataset, " *Artificial Intelligence in Medicine*, Elsevier, 26(1), 55-68 (2002)
[13] Keivan Kianmehr, Reda Alhajj "A class association rule-based classification framework and its application to gene expression data Export," *Artificial Intelligence in Medicine*, Elsevier, 44(1), 7-25 (2008)
[14] Murat Karabatak, Cevdet Ince, M. "A new feature selection method based on association rules for diagnosis of erythemato-squamous diseases," Expert Systems with Applications:, Elsevier, 36(10), 12500-12505 (2009)

[15] Tamura, Makio , D'haeseleer, Patrik "Microbial genotype-phenotype mapping by class association rule mining," Bioinformatics, 24(13) , 1523-1529 (2008)

[16] Elisabeth Georgii, Lothar Richter, Ulrich Ruckert and Stefan Kramer "Analyzing Microarray data using quantitative association rules, " Bioinformatics , 21(2) , 123-129 (2005)

[17] Roddick, J. F., Fule, P. and. Graco, W. J "Exploratory medical knowledge discovery: Experiences and issues," *SIGKDD Explorations*, 5(1), 94–99 (2003)
[18] Chen, T. J., Chou, L. F. and. Hwang, S. J "Application of a data mining technique to analyze coprescription patterns for antacids in Taiwan," *Clin.Ther.*, 25(9), 2453–2463 (2003)
[19] Krzysztof J. Cios, William Moore,G "Uniqueness of medical data mining" *Artificial Intelligence in Medicine*, Elsevier, 26(1), 1–24 (2002)

Feature based registration of thorax x-ray images for lung disease diagnosis

Rininta Putri Nugroho, Astri Handayani, Tati Latifah Rajab Mengko
Biomedical Engineering Research Division, School of Electrical Engineering and Informatics,
Bandung Institute of Technology, Indonesia
Email: rininta_nugroho@yahoo.com

ABSTRACT

In diagnosing lung diseases using x-ray images of a human thorax, there is a huge risk of error in detecting abnormalities of the lung. This may be caused by geometric differences in the images that are being compared. To minimize the possible errors, a system is proposed to assist in the diagnosis process. In implementing this system, a registration process of the images is required as the first step in minimizing the human errors. A feature based method is used to solve the registration of images by using a scale invariant feature transform (SIFT) as the method of feature extraction. Using this feature based method is hoped to result in a better registration than the area based method that was previously used.

Registration, Feature based, Scale invariant feature transform

1. INTRODUCTION

There are several methods for medical image acquisition that are non-invasive, but the most commonly used method is a medical x-ray machine. These images are important in the diagnosis process of diseases that are not visible to the human eye. Interpreting x-ray images can sometimes be difficult because of their two-dimensional form. This difficulty can increase the possibility of human errors in interpreting or detecting abnormalities in the images affecting the diagnosis process of the patient. To reduce the human errors, a diagnosis assisting system is proposed to detect the abnormalities. This system is categorized as a Computer Aided Diagnosis (CAD) that is defined as a diagnosis made by a radiologist who takes into account the result of a computer analysis[1]. The implementation of CAD can be used to diagnose Tuberculosis (TBC) using x-ray images. The detection of abnormalities caused by tuberculosis is done by comparing images that were taken at different times to observe the differences of certain visible features. The result of this detection is then used to isolate a diagnosis and decide a further course of action. Translation and rotation of the object in the image are just some of the difficulties found when comparing the images. The solution for this problem is by using an image registration process.

Previous research was done by using an intensity based registration method with mutual information criterion[2]. That particular method was used because of its global characteristic that is not affected by the local variances of the images used, but it had a disadvantage of having a long computing time. This research was done to improve the previous research by exploring the implementation of a feature based method instead. The calculations needed for a feature based registration method only depends on the extracted features and not the entire image, making it more efficient in solving the computing time issue. This research also explores the automatic detection of features by the developed system, making it independent from user-based feature selection. Hence, the detected features must be rotation, scale, and translation invariant.

2. FEATURE BASED REGISTRATION

Image registration can be defined as the process of overlaying two or more images of the same scene that were taken at a different time, from a different viewpoint and/or using a different sensor[3]. The registration process, for this research, is done towards two x-ray images of a human thorax from a single patient taken at two different times. One image is the reference image and the other the sensed image that needs to be overlaid with the reference image.

2.1 Feature detection

For the purpose of this research, the algorithm used for the feature detection is a scale invariant feature transform (SIFT) that was previously developed by David Lowe, 2004[4]. This algorithm extracts distinctive features that are scale and rotation invariant from the two images needed in the registration process. The name SIFT comes from its characteristic that transforms image data into coordinates that are scale invariant and is relative to the local features of the image. In

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75462Z · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853407

this particular research, the SIFT algorithm used is the open implementation developed by Andrea Vedaldi[5]. The algorithm itself can be divided into four main processes that is the detection of key point candidates using the Difference of Gaussian of the image, then the localization of the detected points, the assignment of orientation and lastly the calculation of the proper descriptor for each key point. These processes can be seen in the block diagram in Figure 1. Images that are used as inputs for the feature detection process are changed into grayscale single precision type data from the original forms. Both the reference image and the sensed image of the registration process are detected. Then the scale space extremes of the images are found by calculating the peaks of the difference of Gaussians that is saved in the variables x_1, x_2 and s. Where x_1 and x_2 represent the coordinates of the detected points and s is the scale (σ).

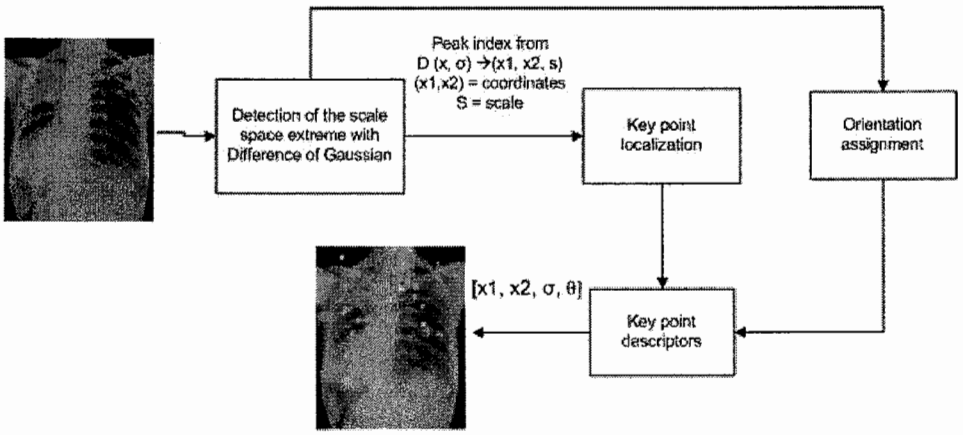


Figure 1. Block diagram of the feature detection process using the SIFT method.

The next step is the localization which is done by fitting the indexes (x_1, x_2, s) to a quadratic interpolation while eliminating key points with low contrast and high edge responses. This step results in key points with high stability making it better image features. After the localization, each key point is assigned a specific orientation by making a window around every detected key point and calculating the histogram of the orientation gradient of that point. A smoothing process is then done to the histogram as a filter and then the global maximum is calculated and assigned as the orientation. Besides the global maximum, every local maximum with the value of 0.8% above the global maximum is also separated as the orientation of each window.

2.2 Feature matching

Feature matching is done by the correlation similarity measure[6] of each key point detected in the previous process of feature detection to find the best match between the reference image and the sensed image. This measure is done by calculating the correlation value between two 11x11 windows that is made around each key point that has been detected. The basic concept of the correlation similarity measure is to find a match between a subimage $w(x, y)$ of size $J \times K$ within an image $f(x, y)$ of size $M \times N$, where it is assumed that $J \leq M$ and $K \leq N$. Mathematically, calculating the correlation value is done according to this definition:

$$r(x,y) = \frac{\sum_s \sum_t [f(s,t) - \bar{f}(s,t)][w(x+s, y+t) - \bar{w}]}{\left\{\sum_s \sum_t [f(s,t) - \bar{f}(s,t)]^2 \sum_s \sum_t [w(x+s, y+t) - \bar{w}]^2\right\}^{1/2}} \quad (1)$$

where $x = 0, 1, 2, \ldots, M-1$, $y = 0, 1, 2, \ldots, N-1$, $\bar{w}$ is the average value of the pixels in w (computed once), $\bar{f}$ is the average value of f in the region coincident with the current location of w, and the summations are taken over the coordinates common to both f and w.

The data processed is the key points from each set of images (reference and sensed). The first step in this measure is to eliminate key points that are distanced more than r pixels apart. For this research r is defined as 300 pixels which is a quarter of the average image size used. After that, an 11x11 sized window is made around each detected key point in both images and the correlations between the windows in the two images are calculated. The calculated

correlation is then normalized to achieve a more optimal correlation value. The maximum value is then searched in both the row and column direction and the indexes that contain this maximum is set aside as indexes that refer to coordinates of the matching features between the reference and sensed image. A flowchart is used to illustrate the whole process of the feature matching using correlation similarity measure that is seen in figure 2.

2.3 Transformation

A mapping function from the matched features can now be constructed to transform the sensed image onto the reference image. The key points that have been matched must be as close as possible after the transformation has been applied. For the purpose of this research, a global mapping model is used or more specifically a similarity transform that only consists of rotations, translations, and scaling parameters. An affine transform has six parameters that is scaling, rotation, horizontal shearing, vertical shearing, and translation where the lines and angles are not kept but parallelism of the image is still maintained. To accommodate x-ray images that consist of many geometrical differences and different contours of the lung and ribcages, an affine transformation is chosen as the transformation model in this research.

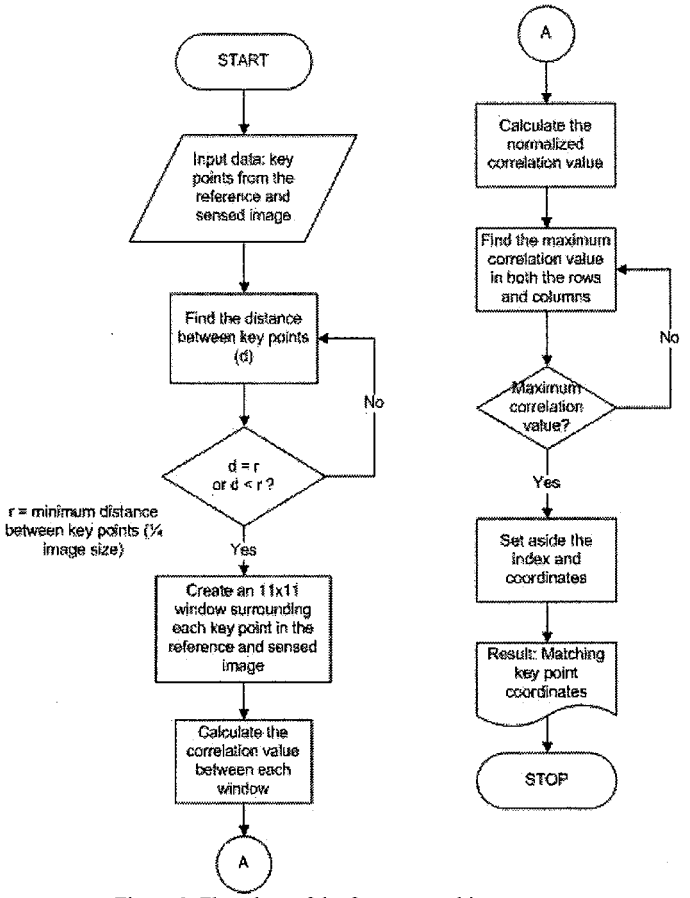


Figure 2. Flowchart of the feature matching process.

3. Experimentation

The implementation of the proposed algorithm was done using Matlab version 7.4.0 R2008a with the hardware specifications of a Intel Core Duo 1.67 GHz processor, 1 GB RAM, and a Windows XP Professional operating system. The input images used were 13 pairs (26 images) of digitized thorax x-ray images from 13 different patients that were obtained from Rumah Sakit Paru Rotinsulu Bandung (Rotinsulu Lung Hospital in Bandung). Digitizing the images was done using a Umax Powerlook 2100 XL scanner that resulted in 24-bit bitmap RGB images with a resolution of 100 dpi.

3.1 Procedure

Figure 3 illustrates the registration process that is executed using this algorithm. In deciding the performance of this algorithm, a few parameters are calculated as a measure such as, the standard deviation of the joint histogram[7] (between the reference image and the registered image in comparison to between the reference image and the sensed image), normalized mutual information or the NMI[8] (between the images before and after registration), the time needed to run the process and also the energy of the pixel difference between the images (before and after registration). A well registered image will result in a joint histogram that is well clustered towards a single diagonal line compared to before registration where usually it is more spread out. Sometimes a visual observation is difficult to read, hence the usage of the standard deviation value of the joint histogram. The smaller the standard deviation, the better the registration performed because a small standard deviation value is a result of clustered values in the histogram.

Normalized mutual information (NMI) of the images is calculated from the entropy of the images. This value represents the mutual information in each pixel of two images. If the NMI is high, this means that the information each pixel shares is also high which is a result of a well performed registration process. NMI before and after registration is compared to see whether there is an increase in value. Another evaluation parameter used is the energy of the pixel difference between the reference and registered image. A good registration will leave a small pixel difference making the energy also small. The energy is calculated by the summation of the squared difference of each pixel. If there is a decrease of energy from before and after registration, that is the result of a well performed registration as well. As was mentioned in the previous section, the previous research done using the area based method of registration had a problem with the time elapsed for each computation of the registration process. Time constraints were also evaluated in this research using a general stopwatch to measure the time elapsed from loading the images until the end of the transformation.

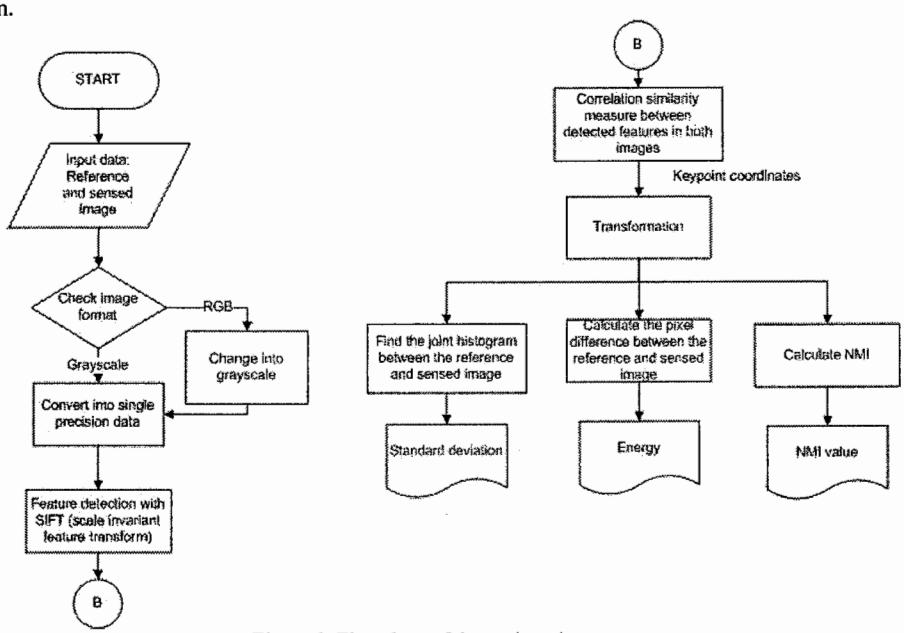


Figure 3. Flowchart of the registration process

3.2 Result analysis

For all the images processed using the proposed method of registration, the amount of features extracted in general are 4319 features from the reference image, 4376 features from the sensed image and around 866 features as a result of matching the two image features. This amount is quite large, considering that the affine transformation only needs six sets of key points to perform the necessary transformation. After running the process on all 13 image sets, it is found that all resulted in a decrease in standard deviation of the histogram. The decrease has the average of 14.28% for each set of images. This steady decrease suggests that the feature based registration process results in a well registered image.

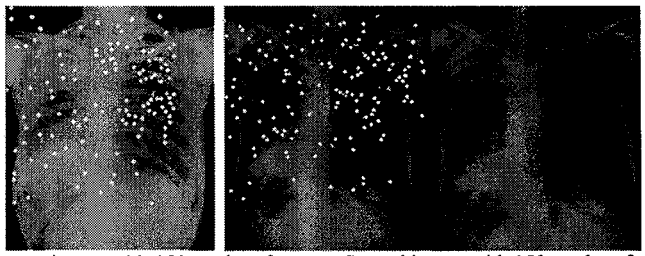


Figure 4. (Left to right) Reference image with 150 random features; Sensed image with 150 random features; Registered image

Table 1. Evaluation parameters from the images processed in figure 4.

Evaluation parameters	Result
Reference image features	4714
Sensed image features	4970
Amount of matched features	694
Standard deviation decrease percentage	32.813%
Difference image energy decrease percentage	1.666%
Computing time	1 minute 40 second
NMI increase percentage	1.372%
Transformation scale	1.1046
Transformation orientation	5.0421

Another parameter that is observed is the value of the normalized mutual information between the images in a set. As was mentioned before, if there is an increase in NMI the registration process can be suggested as well performed. From the example in table 1, there is a 1.372% increase in NMI and all 13 images resulted in an average increase of 1.032%. Some sets of images did not comply to this rule but with consideration that the images had different intensities which would affect the information within each pixel. The difference of intensities in the image may also affect the calculation of the difference image energy as another evaluation parameter. Two pixels with different intensities will always have a high difference. Overall, the images proved to have a noticeable decrease in energy from before to after the registration process. An average of 7.68% decrease was found on all 13 sets of images.

One of the most noticeable improvements from the previous research using the area based method of registration is the computing time used. Each image set needs only an average of 1 minute and 25 seconds to perform the entire process of registering the image. The area based method needed an average of 20 minutes to register a set of images. This shorter computing time is caused by the calculations involved in the process that are done locally on the features without involving the entire image globally.

4. Conclusion

This research implements a feature based registration method with an automatic feature detection algorithm. The feature detection method used is the scale invariant feature transform (SIFT) that extracts rotation, scale, and translation invariant features (RST invariant) making the method applicable for the medical images used. The algorithm was tested on 13 sets of medical x-ray images of the human thorax in support of lung disease diagnosis. To match the extracted features, a correlation similarity measure was implemented and then transformed using an affine transformation model. A few evaluation parameters were used a guidelines in determining the performance of the proposed algorithm for registration such as the standard deviation of the joint histogram (14.28% increase), normalized mutual information (1.032% increase), difference image energy (7.68% decrease), and the computing time (1 minute 25 seconds per set).

It is determined that the usage of the SIFT method of feature detection and correlation similarity measure for feature matching has an optimal performance in terms of computing time and the other evaluation parameters considered for the necessary registration. It also proves to be more efficient in comparison the previous research. A note that must be taken into consideration is the pre-processing of the input images to eliminate the affect of the difference of intensities between the reference and sensed images. This difference may have an effect on the feature detection process and also the calculation of necessary parameters of evaluation.

5. References

[1] Ginneken, B. v, [Computer-aided Diagnosis: A Survey], Ponsen and Looijen, Netherlands, 19-35 (2001)

[2] BNPH, L., [Registrasi Citra Sinar-X Thorax Berbasis Area Dengan Kriteria Mutual Information Untuk Aplikasi Analisis Penyakit Paru-paru], Unpublished Final Project, Bandung, Electrical Engineering Study Program, Bandung Institute of Technology (2009)

[3] Zitova, B., & Flusser, J., "Image Registration Methods: A Survey", Image and Vision Computing Elsevier. (2003)

[4] Lowe, D., "Distinctive Image Features from Scale Invariant Keypoints", Int. Journal of Computer Vision (2004)

[5] Vedaldi, A., An Open Implementation and Portable Library of Computer Vision Algorithms, From VLFeat: http://www.vlfeat.org (2008)

[6] Gonzalez, R., & Woods, R. E., [Digital Image Processing. 2nd ed], Prentice Hall, New Jersey, 701-703 (2002).

[7] Pass, G., & Zabih, R., "Comparing Images Using Joint Histogram", Computer Science Department Cornell University (1997)

[8] Pluim, J., Maintz, J. A., & Viergever, M. A., "Mutual Information based registration of medical images: a survey. IEEE Transactions on Medical Imaging, 22, 986-1004 (2003)

[9] Aston, R., [Principles of Biomedical Instrumentation], Macmillan, New York (1990)

[10] Bohm, J., & Becker, S., "Automatic Marker-Free Registration of Terrestrial Laser Scans Using Reflectance Features", Institute for Photogrammetry, Universitaet Stuttgart, From: http://ifp.uni-stuttgart.de (2007)

[11] Brown, L. G., "A Survey of Image Registration Techniques", *ACM Computer Survey,* vol. 24, no. 4 (1992)

[12] Dai, X., & Khorram, S., "A Feature-based Image Registration Algorithm Using Improved Chain-code Representation Combined with Invariant Moments", IEEE Transactions of Geoscience and Remote Sensing, vol. 37, no. 5 (1999)

[13] Gonzalez, R., & Woods, R. E., [Digital Image Processing. 2nd ed], Prentice Hall, New Jersey, 34-75 (2002).

[14] Mikolajczyk, K., & Schmid, C., "A Performance Evaluation of Local Descriptors", Proc. CVPR (2003)

[15] Schmid, C., Mohr, R., & Bauckhage, C., "Evaluation of Interest Point Detectors", International Journal of Computer Vision, vol. 37, no. 12 (2000)

[16] Vedaldi, A., "Local Features, All Grown Up", University of California, From: www.cs.ucla.edu (2006)

A Fuzzy Expert System Design for Diagnosis of Cancer

Milindkumar V. Sarode, Dr. Prashant R. Deshmukh
(Department Of Computer Science & Engineering, Jawaharlal Darda Institute of Engineering & Technology, Yavatmal, India , College of Engineering Amravati, India)

ABSTRACT

Here a fuzzy expert system design for diagnosing, analyzing and learning purpose of the cancer diseases is described. For this process prostate specific antigen (PSA), age and prostate volume (PV) has been used as an input parameters and prostate cancer risk (PCR) as an output. This system allows determining if there is a need for the biopsy and it gives to user a range of the risk of the cancer diseases. It is observed that this system is rapid, economical, without risk than traditional diagnostic systems, has also a high reliability and can be used as learning system for medicine students.

Key words: prostate specific antigen, prostate volume, prostate cancer

1. INTRODUCTION

In recent years, the methods of Artificial Intelligence have largely been used in the different areas including the medical applications. In the medicine area, many expert systems (ESs) were designed. ONCOCIN and ONCO-HELP are the ESs for diagnosis of the general cancer diseases. For example, ONCO-HELP is a multimedia knowledge based decision support system for individual tumor entities. It makes individual and prognosis-oriented treatment of patient's tumor possible (if corresponding predictor's respective prognostic factors are known). Trough registration of individual patient data over tumor type, histology, metastatic type, metathesis localization and amount, as well as corresponding laboratory parameters together with a corresponding knowledge based on a patient individual prognosis-score can be determined. Using this score, a therapy concept is drafted. ONCO-Help evaluates this concept by using therapy controls with regards to tumor progression/regression and side effects of the therapy. Soft computing technology is an interdisciplinary research field in computational science. Various techniques in soft computing such as ESs, neural networks, fuzzy logic, genetic algorithms, Bayesian statistics, khaos theory, etc, have been developed and applied to solve many challenging tasks in medicine and engineering design. There are some publications in the area prostate cancer prognosis or diagnosis by aid of soft computing methods. A fuzzy logic based method for prognostic decision making in breast and prostate cancers is developed. In the study five different trainable neuro-fuzzy classification algorithms based on different approaches to organize and classify biological data sets by the construction of a fuzzy interference system were investigated. The best classifier based on a mountain clustering algorithm reached recognition rates high in comparison to the Bayes classifier and the KNN classifier. These results suggest that neuro-fuzzy algorithms have the potential to improve common classification methods significantly for the use in ultrasonic tissue characterization. As seen in these studies, it is not quite possible to diagnose of prostate cancer fully based on only ultrasonography and image processing. We have developed a rule-based fuzzy expert system (FES) that uses the laboratory and other data and simulates an expert-doctor's behavior. As known when the prostate cancer can be diagnosed earlier, the patient can be completely treated. If there is a biopsy for diagnosing, the cancer may spread to the other vital organs. For this reason the biopsy method is undesirable. As laboratory data, prostate specific antigen (PSA) and prostate volume (PV) and age of the patient are used. Using this data and help from an expert-doctor, the fuzzy rules to determine the necessity of biopsy and

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754630 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853225

the risk factor was developed. The developed system gives to the user the patient possibility ratio of the prostate cancer. The system was developed by aid of the Mat lab.

2. FUZZY EXPERT SYSTEM

The units of the used factors are: PSA (ng/ml), age (year), PV (ml) and PCR (%). Parts of the developed fuzzy rules are shown in the Table 1. Total of 90 rules are formed.

Rule Number	PSA	AGE	PV	PCR
Rule 1	VL	Very Young	VS	VL
.....				
.....				
Rule 43	VL	MA	H	VL
.....				
.....				
Rule 77	VH	Old	VS	H
.....				
.....				

Table 1. Fuzzy rules

For example, Rule 1, Rule 43 and Rule 77 can be interpreted as follows:

Rule 1: If PSA=VL and Age=Very Young and PV=VS, then PCR=very low, i.e. if the patient's PSA is very small and patient is very young and patient's PV is very small, then patient's prostate cancer risk is very low.

Rule 43: If PSA=VL and Age=Middle Age and PV=H, then PCR=VL, i.e. if the patient's PSA is very low and patient has middle age and patient's PV is high, then patient's prostate cancer risk is very low.

Rule 77: If PSA=VH and Age=Old and PV=VS, then PCR=VH, i.e. if the patient's PSA is very high and patient is old and patient's PV is very small, then patient's prostate cancer risk is high. Fuzzfication of the used factors are made by aid of the follows functions. These formulas are determined by aid both of the expert-doctor

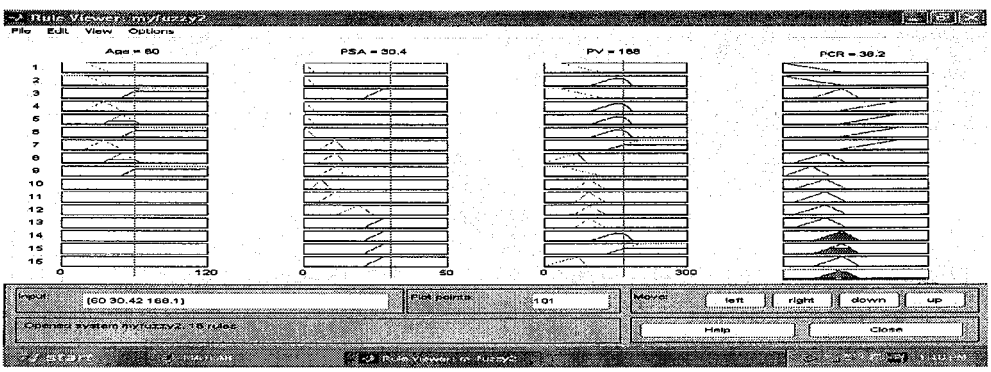


Figure 1. Calculation of the value PCR for the values PSA=30.4ng/ml, Age=60, PV=168 ml

2.1 Fuzzy inference process:

There are five parts of the fuzzy inference process: Fuzzification of the input variables,
Application of the fuzzy operator (AND or OR) in the antecedent, Implication from the antecedent to the consequent, Aggregation of the consequents across the rules, Defuzzification.

Step 1. Fuzzify Inputs: The first step is to take the inputs and determine the degree to which they belong to each of the appropriate fuzzy sets via membership functions. In the Fuzzy Logic Toolbox, the input is always a crisp numerical value limited to the universe of discourse of the input variable (in this case the interval between 0 and 10) and the output is a fuzzy degree of membership in the qualifying linguistic set (always the interval between 0 and 1). Fuzzification of the input amounts to either a table lookup or a function evaluation.

Step 2. Apply Fuzzy Operator: Once the inputs have been fuzzified, we know the degree to which each part of the antecedent has been satisfied for each rule. If the antecedent of a given rule has more than one part, the fuzzy operator is applied to obtain one number that represents the result of the antecedent for that rule. This number will then be applied to the output function. The input to the fuzzy operator is two or more membership values from fuzzified input variables. The output is a single truth value. As is described in the section on fuzzy logical operations, any number of well-defined methods can fill in for the AND operation or the OR operation. In the Fuzzy Logic Toolbox, two built-in AND methods are supported: *min* (minimum) and *prod* (product). Two built-in OR methods are also supported: *max* (maximum), and the probabilistic OR method *probor*.

Step 3. Apply Implication Method: Before applying the implication method, we must take care of the rule's weight. Every rule has a *weight* (a number between 0 and 1), which is applied to the number given by the antecedent. Generally this weight is 1 (as it is for this example) and so it has no effect at all on the implication process. From time to time you may want to weight one rule relative to the others by changing its weight value to something other than 1. Once proper weighting has been assigned to each rule, the implication method is implemented.

Step 4. Aggregate All Outputs: Since decisions are based on the testing of all of the rules in an FIS, the rules must be combined in some manner in order to make a decision. Aggregation is the process by which the fuzzy sets that represent the outputs of each rule are combined into a single fuzzy set. Aggregation only occurs once for each output variable, just prior to the fifth and final step, defuzzification. The input of the aggregation process is the list of truncated output functions returned by the implication process for each rule. The output of the aggregation process is one fuzzy set for each output variable.

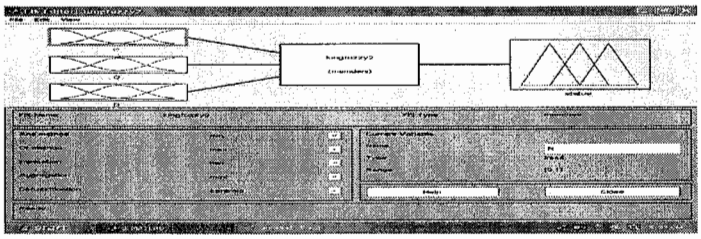
Step 5. Defuzzify: The input for the defuzzification process is a fuzzy set (the aggregate output fuzzy set) and the output is a single number. As much as fuzziness helps the rule evaluation during the intermediate steps, the final desired output for each variable is generally a single number. However, the aggregate of a fuzzy set encompasses a range of output values, and so must be defuzzified in order to resolve a single output value from the set.

3. IMPLEMENTATION

3.1 Lung cancer

3.1.1. The FIS Editor:

Above diagram shows the names of each input variable on the left, and those of each output variable on the right. The FIS Editor displays general information about a fuzzy inference system. The sample membership functions shown in the boxes are just icons and do not depict the actual shapes of the membership functions



3.1.2. The Membership Function Editor

The Membership Function Editor shares some features with the FIS Editor. In fact, all of the five basic GUI tools have similar menu options, status lines, and **Help** and **Close** buttons. The Membership Function Editor is the tool that lets you display and edit all of the membership functions associated with all of the input and output variables for the entire fuzzy inference system. When you open the Membership Function Editor to work on a fuzzy inference system that does not already exist in the workspace, there are not yet any membership functions associated with the variables that you have just defined with the FIS Editor.

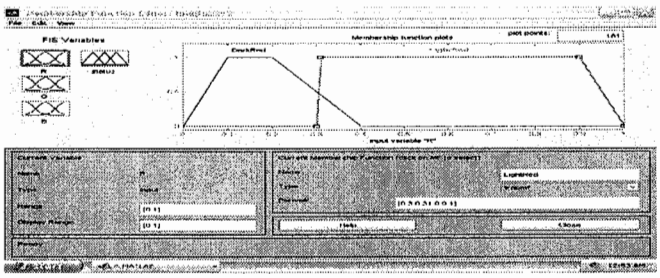


Figure 2 FIS Editor

3.1.3 The Rule Editor

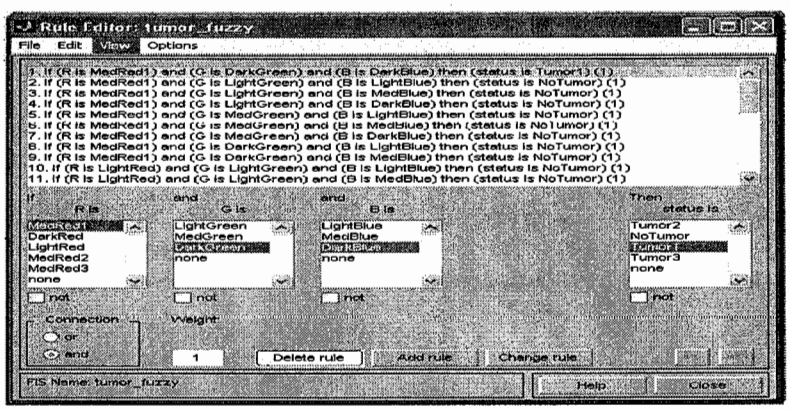
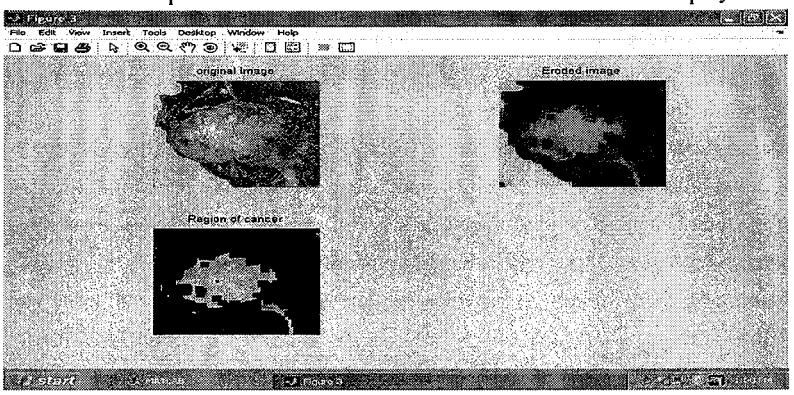


Figure 3. Rule Editor

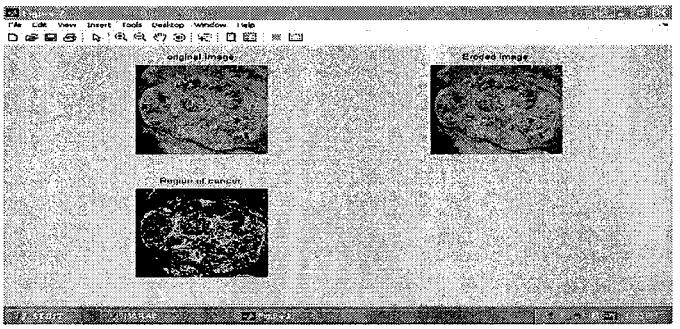
Constructing rules using the graphical Rule Editor interface is fairly self-evident. Based on the descriptions of the input and output variables defined with the FIS Editor, the Rule Editor allows you to construct the rule statements automatically, by clicking on and selecting one item in each input variable box, one item in each output box, and one connection item. Choosing **none** as one of the variable qualities will exclude that variable from a given rule. Choosing

not under any variable name will negate the associated quality. Rules may be changed, deleted, or added, by clicking on the appropriate button. The Rule Editor also has some familiar landmarks, similar to those in the FIS Editor and the Membership Function Editor, including the menu bar and the status line. The **Format** pop-up menu is available from the **Options** pull-down menu from the top menu bar — this is used to set the format for the display.



3.1.5 Renal Cancer

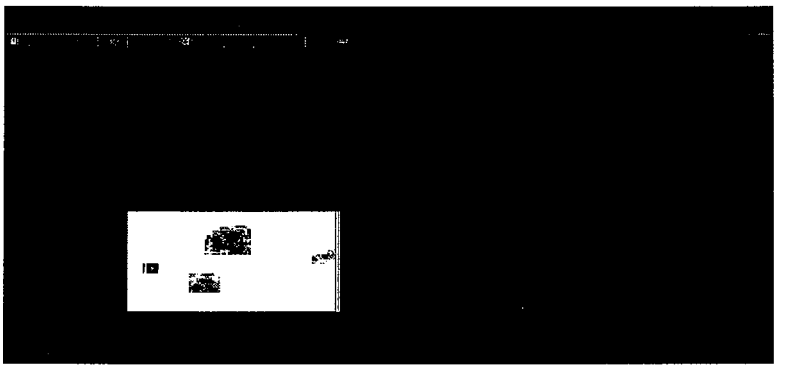
Renal cell cancer is a disease in which malignant (cancer) cells form in tubules of the kidney. Renal cell cancer (also called kidney cancer or renal Aden carcinoma) is a disease in which malignant (cancer) cells are found in the lining of tubules (very small tubes) in the kidney. There are 2 kidneys, one on each side of the backbone, above the waist. The tiny tubules in the kidneys filter and clean the blood, taking out waste products and making urine. The urine passes from each kidney into the bladder through a long tube called a ureter. The bladder stores the urine until it is passed from the body. Thus in this , first we are taking original image as input image which is then eroded by using the normalization technique in which the color with shade yellow is considered as normal while the other color are the region affected by cancer following figure shows original image then eroded image and finally image showing region of cancer.



3.1.6 Colon Cancer

An automated algorithmic approach, based on quantitative measurements, is a valuable tool to a Pathologist for fast verification of colon cancer image abnormalities for effective treatment. In this project a novel method which automatically locates differences in colon cell images and classifies the colon cells into normal and malignant cells is presented. The system fuzzifies image feature descriptors and incorporates a clustering paradigm with neural network to classify images. The novelty of the algorithm is that it is independent of the feature extraction procedure adopted and overcomes the sharpness of class characteristics associated with other classifiers. It incorporates feature analysis and selection and differs markedly from other approaches which either ignore them or perform them as separate tasks prior to

classification. The innovative method has been evaluated using 116 cancerous and 88 normal colon cell images and resulted in a very high classification rate of 96.435%. The percentage error rate of 2.6% is primarily due to preprocessing anomalies. The proposed system was evaluated using 116 cancers and 88 normal colon cell images and shown to be more efficient, simple to implement and yields better accuracy than other methods.



4. CONCLUSION

Here we have used the fuzzy expert systems that are capable of mimicking the behavior of a human expert. Fuzzy approach is useful to detect the region of various type of cancer from the original image by performing erosion operation. It is also helpful to determine need for the biopsy and it gives to user a clear idea of spread and severity level of cancer. Fuzzy logic addresses such applications perfectly as it resembles human decision making with an ability to generate precise solutions from certain or approximate information. It fills an important gap in engineering design methods left vacant by purely mathematical approaches (e.g. linear control design), and purely logic-based approaches (e.g. expert systems) in system design

5. REFERENCES

[1] Abbod M. F., von Keyserlingk D. G., Linkens D. A., and Mahfouf M., "Survey of Utilization of Fuzzy Technology in Medicine and Healthcare", Fuzzy Sets and Systems, 331–349 (2001).

[2] Minsong Cao, Yun Liang, Changyu Shen, Kathy D. Miller, and Keith M. Stantz, "Developing DCE-CT to Quantify Intra-Tumor Heterogeneity in Breast Tumors With Differing Angiogenic Phenotype", IEEE Transactions On Medical Imaging, Vol. 28, No. 6, 861-871, (2009).

[3] Allah Verdi N., Expert Systems. An Artificial Intelligence Application, Istanbul: Atlas, 248 (2002).

[4] Allah Verdi N., and Yaldiz S., "Expert System Applications in Medicine and Example of Design of a Pre-Diagnosis Expert System", Proc. Second Turkish-German Joint Computer Application Days, 15-16 Oct.1998, Konya, 175-192 (1998).

[5] Boegla K., Adlassniga K.-P., Hayashic Y., Rothenfluhd T. E., and Leiticha H., "Knowledge Acquisition in the Fuzzy Knowledge Representation Framework of a Medical Consultation System", Artificial Intelligence in Medicine, 676, 1–26 (2002).

[6] Ivana I˘sgum, Marius Staring, Annemarieke Rutten, Mathias Prokop, Max A. Viergever, and Bram van Ginneken, "Multi-Atlas-Based SegmentationWith Local Decision Fusion—Application to Cardiac and Aortic Segmentation in CT Scans" IEEE Transactions On Medical Imaging, Vol. 28, No. 7, 1000-1009, (2009).

Artifact reduction using two-mode filters for compressed images

Ying-Wen Chang and Yen-Yu Chen
Department of Information Management, ChungChou Institute of Technology, Taiwan, R.O.C.
ywchang@dragon.ccut.edu.tw

ABSTRACT

The blocking effect is a major drawback of the DCT-based compression scheme at low bit rates. Significantly decreasing blocking effects can raise compression ratios for a particular image quality, or improve the quality with regard to the specific bit rate of compression. This work presents a scheme based on two mode filters in terms of the activity across block boundaries. For smooth regions, the strong smooth filter exploits the correlation between the neighboring blocks to reduce the discontinuity of the pixels across the boundaries. The weak smooth filter employs an edge-preserving smooth filter for texture and edge-based regions. Simulation results reveal that the proposed algorithm significantly lowers the blocking artifact, as judged by both objective and subjective measures.

KEYWORDS: blocking artifact, DCT, enhancement.

1. INTRODUCTION

Block-based DCT coding has been successfully adopted in several applications owing to its energy compacting property and relative ease of implementation. For high compression ratio, the blocking artifact caused by coarse quantization of the coefficients appears to the boundaries between adjacent image blocks. Such degradation is highly objectionable, and is likely to annoy users with low bit rates. Reducing the blocking effect has received considerable attention [1-14].

Recently, the processing domain of enhancing Schemes turns to DCT-domain. Zeng's approach [12] first recognizes that visible boundaries between two adjacent blocks in the coded image are primarily oriented along the horizontal and vertical directions. Some AC components of those adjacent blocks are zeroed out to lower the visibiltiy of blocking artifacts. Luo's algorithm [13] detects the smooth and non-smooth regions, and applies DCT and edge-preserving filters. To improve the visual quality, Zeng and Luo's methods need to use forward DCT and inverse DCT in a back-and-forth manner because of the high computation complexity. Ref. [14] enhanced Zeng and Luo's methods to improve the subjective and objective image quality. However, the false edges in smooth regions are hard to be eliminated.

This work presents an easy-to-implement and low-complexity post-processing method. The rest of this paper is organized as follows. Section 2 describes the proposed method in detail. Section 3 presents the experimental results. Simulation results demonstrate that the proposed algorithm efficiently reduces the blocking effect in both smooth and edge-texture regions. Section 4 gives conclusions and discussions.

2. PROPOSED ALGORITHM

The JPEG standard, which is the still image compression standard, recommends the use of BDCT (Blocked-DCT). For high compression ratios, the blocking artifact resulting from coarse quantization of the coefficients appears to the boundaries between adjacent image blocks. This degradation is highly objectionable, and is likely to annoy audiences on low bit rates. A JPEG-coded block of 8×8 pixels becomes smoother when transformed by DCT and quantized by the quantization table. The false steep edges might appear in the boundary of neighboring JPEG-coded blocks due to quantization. This solution can reduce steep edges, and is widely employed in image enhancing. Ref. [14] recommended a global operator to set several DCT coefficients of a shift block zero, thus improving the objective image quality. However, the operator cannot get the excellent perception quality in the smooth region. This investigation develops a two-mode filter scheme, which not only eliminates the false edges but also promotes the subjective and objective image quality metric.

In the two-mode filter scheme, the smooth mode and edge mode are applied to those regions A smooth region is applied to the strong smooth filter as well as edge or texture region take advantage of weak smooth filter. Figure 1 shows

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754631 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853213

that the proposed block diagram involves a JPEG decoder and a two-mode filter. The decoded image is reconstructed using entropy coding, de-quantization and inverse DCT transform in the JPEG decoder. The DCT classification seeks the DCT coefficients, and classification map marks the smooth DCT block. A smooth or weak smooth filter is applied according to the mark on the classification map to implement to the JPEG decoded image to lower the blocking artifact.

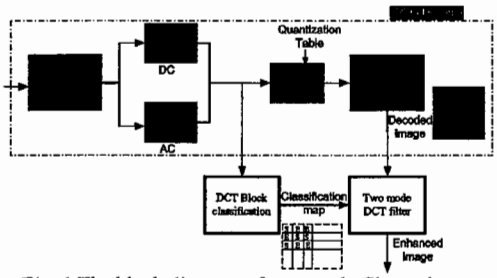


Fig. 1 The block diagram of two mode filter scheme

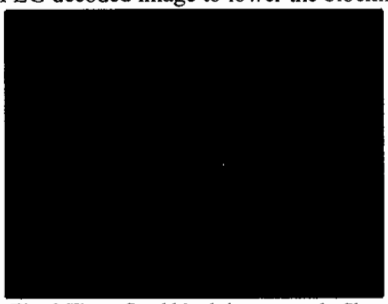


Fig. 2 The refined block in two mode filter

In DCT block classification, the DCT block containing many zero coefficients is a smooth area. The classification seeks those smooth areas, which are denoted as "S" on the classification map. In this process, the number of DCT coefficients belonging to smooth block is equal to one. An edge block has more than one DCT coefficient.

In the two-mode filter process, the decoded image and classification map resulting from the classification process are applied as the input data for the proposed deblocking approach. First, the decoded image is divided into 8×8 blocks, and the block is represented as a refined block (R) depicted in Fig. 2. A refined block is placed in the center of four neighboring DCT blocks. Several methods [12][13] select the block in the vertical and horizontal direction. Nevertheless, a vertical block results in horizontal false edges, and a horizontal block contains vertical false edges. Zeng and Luo [12][13] cannot cut down the blocking noise in a single iteration, and thus require two stages, one for each direction. Since the refined blocks illustrated in Fig. 2 carry both vertical and horizontal false edges that need to be removed from the target, the proposed filter is applied in a single iteration, thus reducing the computation complexity. Figure 3 illustrates the proposed two-mode filter scheme. The proposed approach has three stages. First the refined block is transformed with 2-DCT. Second, if the refined block transformed by DCT is in a smooth region, then the smooth weak smooth filter is applied to it. Finally, the processed block is transformed using inverse DCT transformation.

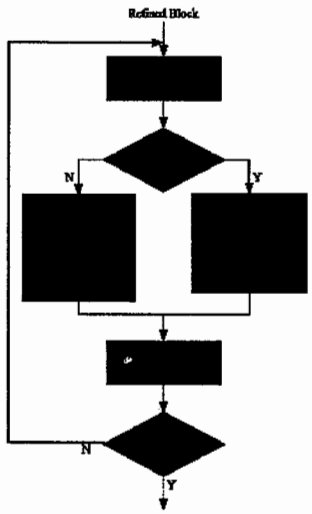


Fig. 3 The flow chart of two mode filter scheme

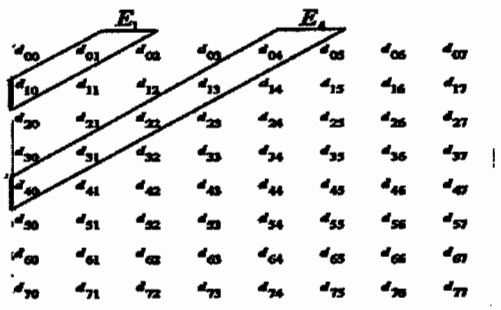


Fig. 4 15 frequency band with a block

Let $x_{i,j}$ denote the gray values of an 8×8 refined block in the decoded image, with its DCT given by $d_{k,l}$. The coefficients $d_{k,l}$ in the output refined block are arranged left to right and top to bottom in order of rising spatial frequencies in the horizontal and vertical spatial dimensions, respectively.

A refine block is located on the center of four neighboring DCT blocks shown in Fig. 2. For instance, refined block R is a neighbor to A, B, C and D. If four neighboring DCT blocks are smooth type, then R belongs to a smooth region, and the strong smooth filter is applied. In contrast, if one of neighboring DCT block (i.e. A, B, C, D) contains texture or edges, then the weak smooth filter should be applied to R. Equation 1 illustrates the decision rule of strong smooth filter and weak smooth filter. The classification map is obtained by the numbers of DCT coefficients before de-quantization in JPEG decoder. If four neighboring sub-blocks that involves only one DCT coefficients, the filter mode of this refined block is belonged to strong smooth filter.

$$filter\ \ mode = \begin{cases} strong\ \ smooth\ \ filter, if\ \ count(classification\ \ map == 's') = 4 \\ weak\ \ smooth\ \ \ filter,\ \ otherwise \end{cases} \tag{1}$$

The proposed strong smooth filter is a low-pass filter. The aim of a strong smooth filter is to reduce the false edges without producing new blocking artifacts. Accordingly, the low-pass filter should contain several DCT coefficients, rather than only a DC coefficient. Since the filter retains merely DC, a new blocking artifact is more seriously. Shen [16] proposed three DCT coefficients (i.e. $d_{0,1}$,$d_{1,0}$,$d_{1,1}$) corresponding to horizontal, vertical and diagonal edges. The strong smooth filter determines the values of these coefficients in image reconstruction.

Tang [17] presented a contrast measure that can be applied to measure the contrast of images in the DCT domain. This work employs a contrast measure given by the ratio of high-frequency and low-frequency content in the bands of the DCT matrix, as illustrated in Fig. 4. The nth band consists of the coefficients with $n=k+l$. For example, E_1 is constructed from $d_{0,1}$ and $d_{1,0}$. Due to the significant coefficient $d_{1,1}$, the coefficients of E_2 should be preserved. In strong smooth filter, the coefficients located on the frequency band above 2 are set to zero.

In the edge region, the strength of false edges is smaller than the signal. Modifying the coefficients on low and high frequency band degrades the image quality. Setting the DCT coefficients zero can significantly enhance the perceived image quality. The DCT coefficients of a refined block in row 1 and column 1 excluding DC denote the information of the vertical and horizontal edge, respectively. In order to reduce the vertical and horizontal edges, equation (2) lists behavior of weak smooth filter. The DCT coefficient on row 5 can be considered as the false edge of horizontal edge and column 5 is the vertical noise. Removing those coefficients can effectively enhance the perceptive quality. Since the human eyes are not sensitive to high frequency information, setting the coefficients in row 7 and column 7 to zero would reduce the quantization noise.

$$d_{k,l} = \begin{cases} 0, & k = 0,1,...,7, \quad l = 5,7 \\ 0, & k = 5,7, \quad l = 0,1,...7 \\ d_{k,l}, & otherwise \end{cases} \tag{2}$$

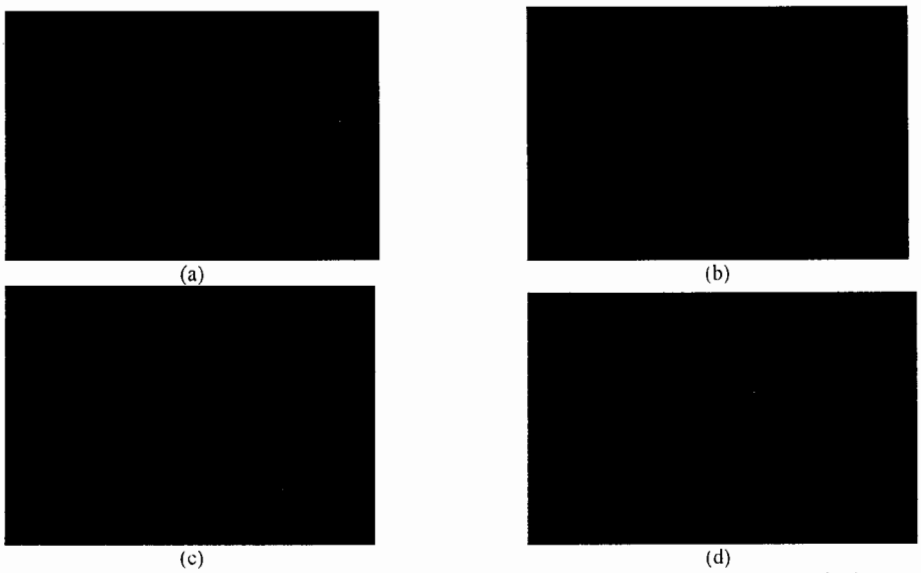
3. Simulation Results

The proposed algorithm was applied to several JPEG-coded images at low bit rates. The results for these images, namely Lena, Pepper and Boots, at size 512×512 pixels with 8 bit per pixel, are presented. For the subjective evaluation of the proposed technique, Fig. 5 (a)-(d), an enlarged top left part of the image Lena at 0.3 bit per pixel(bpp), were obtained after running the methods of Ref. [12], [13], [14] and the two-mode filter, respectively. Figures 6 and 7 illustrate enlarged parts of the images Bike and Café using the methods from Ref. [12], [13],[14] and two-mode filter. These figures demonstrate that the proposed method significantly reduces the blocking artifacts, particularly in the smooth areas. In Fig. 5-7, the results of proposed method show less blocking artifact in both smooth and edge regions.

BDCT suffers from both the burring and blocking effect. JPEG_quality_score, which is an NR perceptual quality assessment scheme [18], was adopted to assign each image a quality score between 0 and 10 (where 10 represents the best quality, and 0 the worst). Table 1 lists the comparison results of the proposed algorithm and some de-blocking approaches at 0.3 bpp. These results indicate that the proposed scheme has the best perceptual image quality.

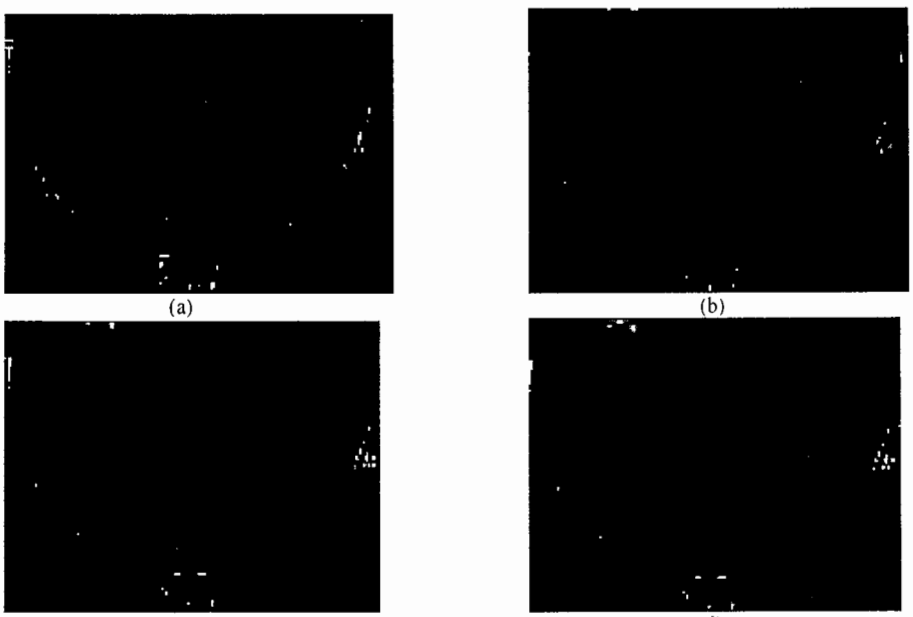
Table 1 JPEG_quality_score comparison of test images at 0.3bpp for results of Ref.[12], Ref. [13], Ref. [14], and proposed scheme.

	JPEG	Proposed scheme	Ref. [14]	Ref. [13]	Ref. [12]
Lena	4.51	9.08	9.81	5.97	8.07
boots	2.99	8.24	8.18	4.35	6.46
Pepper	4.59	9.20	9.05	6.07	8.33
Bike	4.78	9.08	8.99	6.02	8.05
Café	3.17	8.21	8.13	4.27	6.42



(a) (b) (c) (d)

Fig. 5 Lena 0.30 bpp, (a) Ref. [13] (b) Ref. [12] five iteration (c) Ref. [14] (d) proposed scheme

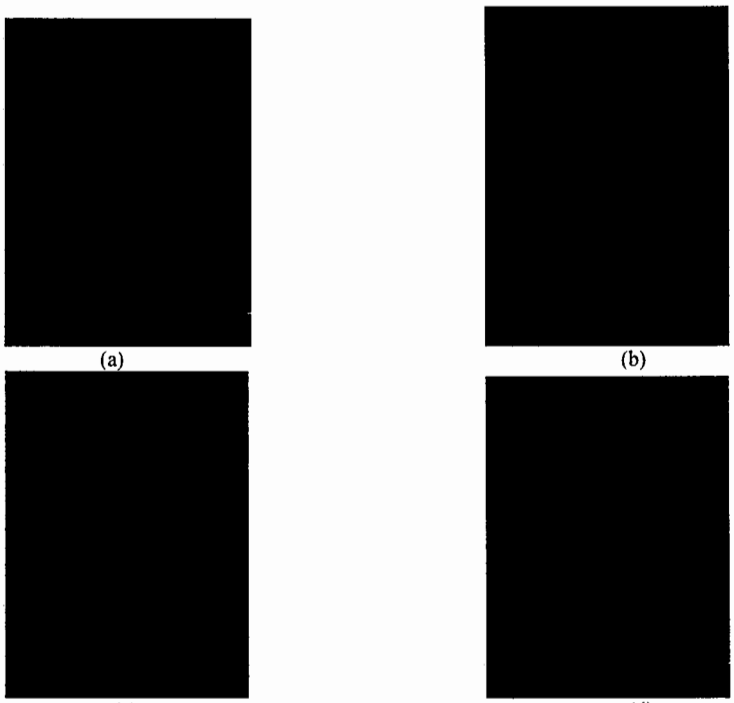


(a) (b) (c) (d)

Fig. 6 Bike 0.25 bpp, (a) Ref. [13] (b) Ref. [12] five iteration (c) Ref. [14] (d) proposed scheme

Table 2 shows the Peak Signal to Nose Ratio (PSNR) results, where ΔPSNR indicates the difference in PSNR between the processed image and the JPEG coded image. These results demonstrate that the subjective metric, PSNR, is better by 0.3dB on average at 0.1-0.3bpp in the proposed scheme. In the objective quality metric, the proposed two-mode filter scheme is superior to Ref. [12] [13]. Comparing to Ref. [14], applied proposed method is barely superior to Ref. [14] in quantitative metrics. However, Ref. [14] takes a global mechanism while more artifacts are still emerged in smooth regions. Although the result of Ref. [14] and proposed method are similar in the objective criteria (Table 1-2), the simulation results shows applied proposed methods would provide better perception quality in Fig. 5-7.

Table 3 presents the computational complexity of the proposed blocking method and those of Zeng [12] and Luo [13]. As compare with Ref. [12] and ref. [13] the computational complexity of the proposed scheme is reduced without sacrificing the PSNR.



(a) (b) (c) (d)

Fig. 7 Café 0.30 bpp, (a) Ref. [13] (b) Ref. [12] five iteration (c) Ref. [14] (d) proposed scheme

Table 2 ΔPSNR comparison of test images for results of [12], [13], [14] and proposed scheme.

image \ *bpp* / *ΔPSNR*		0.15	0.20	0.25	0.30
lena	Proposed scheme	0.57	0.55	0.47	0.37
	Ref .[14]	0.51	0.51	0.45	0.36
	Ref. [13]	0.14	-0.03	-0.30	-0.26
	Ref. [12]	0.25	-0.12	-0.37	-0.59
Pepper	Proposed scheme	0.53	0.49	0.37	0.22
	Ref. [14]	0.46	0.43	0.34	0.20
	Ref. [13]	0.33	0.21	-0.11	-0.40
	Ref. [12]	0.20	-0.04	-0.36	-0.74
Boots	Proposed scheme	0.55	0.53	0.47	0.43
	Ref. [14]	0.46	0.47	0.45	0.41
	Ref. [13]	0.29	0.24	-0.07	-0.27
	Ref. [12]	0.48	0.36	0.05	-0.20
Goldhill	Proposed scheme	0.39	0.35	0.27	0.20
	Ref. [14]	0.33	0.33	0.27	0.20
	Ref. [13]	0.43	0.15	-0.27	-0.58
	Ref. [12]	0.37	0.13	-0.18	-0.43
Baboo	Proposed scheme	0.23	0.23	0.23	0.18
	Ref. [14]	0.21	0.22	0.21	0.17
	Ref. [13]	0.20	0.17	0.01	0.00
	Ref. [12]	0.16	0.10	-0.01	-1.28
boat	Proposed scheme	0.44	0.43	0.35	0.25

	Ref. [14]	0.41	0.41	0.33	0.24
	Ref. [13]	0.32	-0.05	-0.37	-0.63
	Ref. [12]	0.33	-0.16	-0.58	-0.84
F-16	Proposed scheme	0.52	0.56	0.54	0.48
	Ref. [14]	0.45	0.52	0.51	0.46
	Ref. [13]	0.35	0.02	-0.22	-0.71
	Ref. [12]	0.34	0.04	-0.36	-0.82

Table 3. Comparing the computational complexity using Ref. [12], Ref. [13], Ref. [14] and proposed scheme

	proposed	Ref. [13]	Ref. [12] one iteration	Ref. [14]
DCT	MN-M-N-1	3NM-N-M	2NM-N-M	MN-M-N-1
Comparison	MN-M-N-1	3*(2NM-M-N)	0	0
Add/sub	3MN-3M-3N-3	7*(2NM-M-N)	0	0
Mul/div	0	5*(2NM-M-N)	0	0

image size:H=8*N,W=8*M

4. CONCLUSIONS

This study proposes a new post-processing approach for the reduction of block artifacts in a transform coded image. Zeroing out a few AC coefficients of our shift block can improve the visual quality of JPEG coded-images. The proposed scheme not only removes the artifacts easily, but also obtains a good image quality. Compared to other schemes [12-14], the proposed method decreases computational complexity and adapts to real-time applications.

ACKNOWLEDGEMENTS

The author thanks the National Science Council of the Republic of China, Taiwan, for financially supporting this research under Contract No. NSC 98-2221-E-235-010.

REFERENCES

[1] H. Reeve and J. Lim, " Reduction of blocking effect in images," Proc. ICASP'83, 1212-1215 (1983).
[2] B. Ramamurthi and A. Gersho , "Nonlinear space-variant postprocessing 9of block coded images," IEEE Trans. Acoust., Speech, Signal, Processing ASSP-34, 1258-1267 (1986).
[3] J. D. McDonnel, R. N. Shorten, and A. D. Fagan, "An edge classification based approach to post-processing of transform encoded images," Proc. ICASSP. 5, 329-32 (1994).
[4] Y. L. Lee, H.C. Kim, and H. W. Park, "Blocking effect reduction of JPEG images by adaptive filtering,"IEEE Trans. Image Processing, 7, 229-234 (1998).
[5] S.C. Tai, Y. Y. Chen, and S. F. Sheu, "Deblocking filter for low bit rate MPEG-4 video," IEEE Trans. Circuits and System for Video Technology. 15 (6), 733-741 (2005).
[6] J. Luo, C. W. Chen, K. J. Parker, and T. S. Huang, "Artifact reduction in low bit rate DCT-based image compression," IEEE Trans Image Process, 5 (9), 1363-1368 (1996).
[7] Y. F. Hsu and Y. C. Chen, "A new adaptive separable median filter for removing blocking effects," IEEE Trans. Consum. Electron., 2(3), 91-95 (1993).
[8] T. Jarske, P. Haacisto, and I. Dcfcc, "Post-filtcring mcthods for rcducing blocking cffccts from codcd images," IEEE Trans. Consumer Electron. 39, (8), 510-513 (1994).
[9] T.P. O'rourke and R. L. Stevenson, "Improved image decompression for reduced transform coding artifacts," IEEE Trans. Circuits Syst. Video Technol., 5(8), 298-304 (1995).
[10] M. Helsingius and Pauli Kuosmanen, "Image compression using ac prediction," Optical Engineering, 38, 181-184 (1999).
[11] I. R. Ismaeil and R. K. Ward, "Removal of DCT blocking artifacts using DC and AC filtering," Proc. IEEE PACRIM., 229-232 (2003).
[12] B. Zeng, " Reduction of blocking effect in DCT-coded images using zero-masking techniques," Singal Processing, 79 (2), 205-211(1999).
[13] Y. Luo and R. K. Ward, "Removing the blocking artifacts of block-based DCT compressed images," IEEE Trans. On image processing, 12(7), 838-842 (2003).
[14] Y.W. Chang and Y. Y. Chen, "Design of Deblocking Filter for Both Objectvie and Subjective Metrics," IEICE Trans. On Fundamentals, E91-A(8), 2038-2040 (2008).
[15] V. Bhaskaran and K. Konstantinides, [Image and Video Compression Standards Algorithms and Architecture], Kluwer, MA, USA (1997).
[16] Bo Shen, Ishwar K. Sethi, "Direct feature extraction from compressed images," Proc. of the SPIE Conference on Storage and Retrieval for Still Image and Video Databases IV, 404-414(1996).
[17] Jinshan. Tang, Eli Peli and Scott Acton, "Image Enhancement Using a Contrast Measure in the Compressed Domain," IEEE Signal Processing Letters, 10(10), 289-292 (2003).
[18] Z. Wang and H. R. Sheikh and A. C. Bovik, "No-reference perceptual quality assessment of JPEG compressed images," Proc. ICIP'02. I, 477–480 (2002).

Simultaneous detection of randomly arranged multiple barcodes using time division multiplexing technique

Saad Md. Jaglul Haider[a] and Md. Kafiul Islam[b]

Dept. of Electrical and Electronic Engineering[a]; School of Engineering and Computer Science[b]
Bangladesh University of Engineering and Technology[a]; Independent University, Bangladesh[b]
E-mail: haiderpasha2000@yahoo.com[a]; kafiiut@gmail.com[b]

ABSTRACT

A method of detecting multiple barcodes simultaneously using time division multiplexing technique has been proposed in this paper to minimize the effective time needed for handling multiple tags of barcodes and to lessen the overall workload. Available barcode detection systems can handle multiple types of barcode but a single barcode at a time. This is not so efficient and can create large queue and chaos in places like mega shopping malls or large warehouses where we need to scan huge number of barcodes daily. Our proposed system is expected to improve the real time identification of goods or products on production lines and in automated warehouses or in mega shopping malls in a much more convenient and efficient way. For identifying of multiple barcodes simultaneously, a particular arrangement and orientation of LASER scanner and reflector have been used with a special curve shaped basement where the barcodes are placed. An effective and novel algorithm is developed to extract information from multiple barcodes which introduces starting pattern and ending pattern in barcodes with bit stuffing technique for the convenience of multiple detections. CRC technique is also used for trustworthiness of detection. The overall system enhances the existing single barcode detection system by a great amount although it is easy to implement and use.

Keywords: Barcode, Simultaneous Detection, LASER Scanner, Time Division Multiplexing, Bit Stuffing, CRC

1. INTRODUCTION

Barcodes are essential to modern commerce. Consumers, manufacturers, and retailers all get benefit from the use of barcode because it is the least expensive way of providing machine-readable information. All the existing methods of detecting barcode are related with single barcode detection at a time. Time needed in reading barcodes individually in large retail shops or in automated warehouses is not negligible at all. We have used standard existing barcode reader which can give us serial data after receiving reflected LASER beam from the barcodes. We are using linear barcodes only as it is easy to implement. This method would save a lot of time in processing identification of goods and would prevent constructing large queue of customers or goods in large retail shops or warehouses.

Fig 1 shows the system block diagram of the proposed Multiple Detection system. It contains all the modules needed for realization of barcode detection technique such as Clock Pulse Generator, Delay, LASER Beam Sender and Receiver, Data Processing Unit, etc.

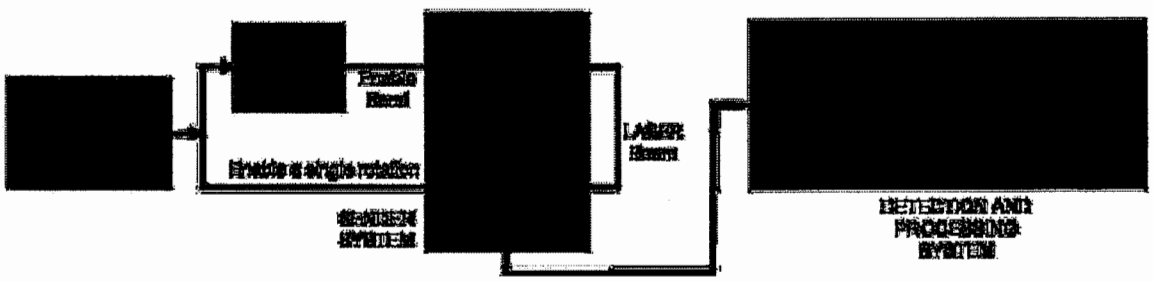


Figure 1. Basic block diagram of the system

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754632 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853224

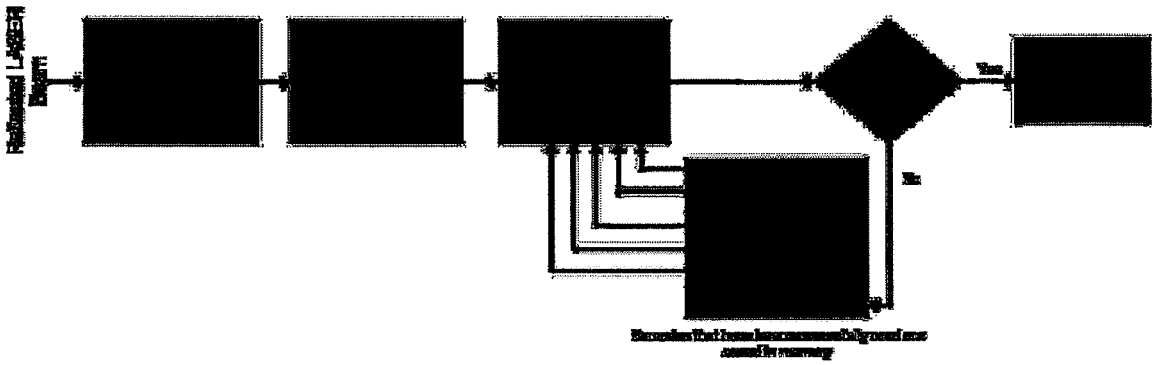


Figure 2. Detection and processing system flow chart

2. OVERVIEW OF THE ALGORITHM

2.1 Sender system

The multiple detection technique is based on the use of time division multiplexing technique. It is done by dividing the scanning process into a number of time slots which is controlled by the "Clock Pulse Generator" in Fig 1. Time Division Multiplexing technique is more clearly illustrated later. The "Sender System" in that figure consists of two parts: the LASER source and the reflector. "Clock Pulse Generator" is used to generate pulses at every time slot to restart its scanning and enabling the rotation of reflector. A delay circuitry is used to make a delay for starting of scanning process with respect to start of rotating the reflector. First of all it would rotate the reflector, then after one delay time it enables the LASER source to send the LASER beam to the reflector. At each time slot it sends LASER beam and scans barcodes and the process repeats at the next time slot with a new position of the reflector.

2.2 Detection and processing system

At the "Detection and Processing System" in Fig 1, we have three blocks: "Receiver", "Data Processing Unit" and the "Output". Details of the "Detection and Processing System" are illustrated by a flow chart in Fig 2.

The receiver receives the LASER beam reflected from the barcodes through the reflector. Then it converts the optical signal into electrical signal which is eventually transferred to the data processing unit where the received data is analyzed using standard barcode reader. After extracting the actual barcode from the digital serial data, cross matching is done with the previous extracted barcodes. If it is a newer data then it is automatically saved in a specific portion of a memory. After analyzing, at last the output is shown on the output section.

3. IMPLEMENTATION OF TIME DIVISION MULTIPLEXING

Time-division multiplexing (TDM) is a type of digital multiplexing in which two or more signals or bit streams are transferred apparently simultaneously as sub-channels in one communication channel, but are physically taking turns on the channel. The time domain is divided into several recurrent time slots of fixed length, one for each sub-channel. A sample byte or data block of sub-channel 1 is transmitted during time slot 1, sub-channel 2 during time slot 2, etc. One TDM frame consists of one time slot per sub-channel. After the last sub-channel the cycle starts all over again with a new frame, starting with the second sample, byte or data block from sub-channel 1, etc.

3.1 Clock pulse generator

The clock pulse generator is introduced to generate clock pulse of specific width and frequency to drive the sender system for implementing the TDM technique. At each clock pulse, a new time slot is implemented. By using this, we able to multiplex the timing of several angles of rotation of the reflector with only one arrangement in order to trigger the same laser source multiple times.

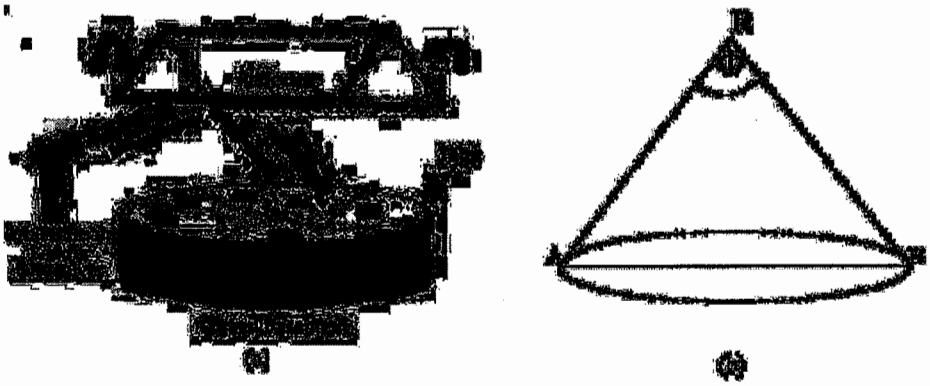


Figure 3. (a) Realization of curve-shaped basement and reflector's mutual position; (b) realization of reflector's rotation angle

3.2 Delay unit

The delay unit is used before the LASER Beam Sender to create one delay time of driving the Beam Sender with respect to driving the Reflector by the clock pulse. The reflector is enabled to start rotating first of all, then after one delay time the LASER source starts sending the beam to the reflector. It would have been difficult for the reflector to work properly if the delay unit would not be used.

4. SENDER SYSTEM DETAILS

To scan multiple barcodes simultaneously, we have used a special arrangement which includes a stationary LASER Beam Sender and Receiver, Reflector and Curve-shaped Basement. The arrangement is shown in Fig 3(a).

4.1 LASER beam sender and receiver

Here we have used active non-contact scanner. The most common form of active non-contact scanner uses a LASER beam that is automatically scanned back and forth across the symbol at a high rate. Active non-contact scanners may be stationary as they are found in grocery stores or they may be hand held. Industrial versions of active non-contact scanners, however, are often kept the beam of light fixed on a single spot and are arranged so that the symbol is moved across the field of view of the scanner. Here we are using the stationary one. The major advantage is that the scanner can read bar code from several feet away. If the symbol is printed large enough, the laser scanner can read the symbol from as far away as 30 feet. In a warehouse this ability can be a definite advantage.

The diameter of the beam will remain small enough to resolve the wide and narrow bars of the bar code even if the reading distance varies. That property allows laser scanners to read bar code on curved surfaces. In our proposed arrangement, ability of reading from 2 feet away is enough. So barcode's bar width should be selected according to the distance of the basement and the reflector.

4.2 Reflector

Reflector is not a fixed device in our arrangement. But it can rotate along a fixed horizontal axis by a predefined angle. The angle is determined by the total coverage angle divided by the number of time slots.

Fig 3(b) shows the basic realization of the reflector's rotation angle. In this figure, R represents the position of the reflector's horizontal rotation axis. Here the diameter of the curve-shaped basement is $AB = D_b$ cm and if we want to scan after each Q cm, the number of time slots, N_t would be

$$N_t = D_b / Q$$

If the total coverage angle is Φ, the angle of rotation per time slot is then determined by,

$$\theta = \Phi / N_t$$

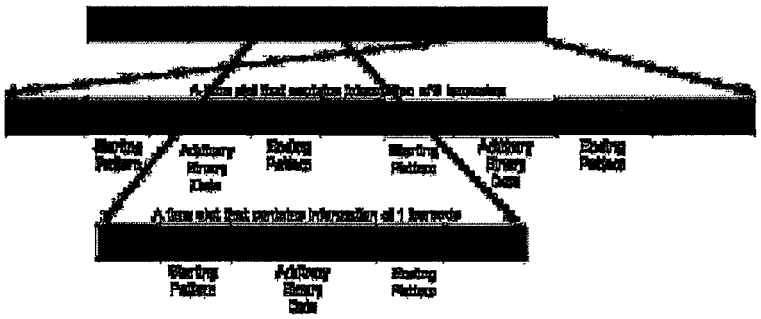


Figure 4. Information in different time slots and idea of starting and ending pattern

4.3 Curve-shaped basement

The main reason behind using curve-shaped (concave) basement is very significant. The LASER beam is incident from a distance about one feet or more. If the basement is flat, then possibility of dispersing the beam to other directions is very high and the possibility of the beam not coming back to the reflector is also very high. The concave basement ensures that the beam is incident at almost 90 degree angle according to the basement and reflected back to the reflector with minimum possibility of dispersing elsewhere.

The curvature of the concave basement is determined by the distance between the center of basement and the reflector. The radius of the curvature should be as much closer to this distance.

5. DETECTION AND PROCESSING SYSTEM DETAILS

Detection and Processing System contains LASER Receiver, Data Processing Unit and Output.

5.1 LASER receiver

LASER Receiver is discussed with the LASER sender system. But it should be mentioned that the receiver receives the reflected LASER beam and convert it to serial data. The serial data is then passed to the Data Processing Unit.

5.2 Data processing unit

This portion of the system is software based. All the inputs are serial data from the receiver. This unit then determines whether the received data is intelligent data or garbage. If it is intelligent, then the intelligent portion is extracted. For the convenience of this extraction and trustworthiness of data, several schemes are described below.

5.2.1 Starting and ending pattern

In our design, a pattern (let it be 01110111) is reserved for starting and ending of barcodes. It is important to do this as multiple barcodes can be detected in a single time slot. If starting and ending pattern is not used, receiver can detect spurious bit between two barcodes as a valid data. Receiver has to distinguish which information is within a barcode and which is not. The receiver program detects the pattern and extracts the information in between the starting and ending patterns.

Fig 4 shows the two possible data sets in two different time slots. One time slot contains data of two barcodes (shown in green), another time slot contains data of only one barcode. If starting pattern is detected but ending pattern could not be detected within a single time slot, then the information after the starting pattern is neglected and deleted.

5.2.2 Bit un-stuffing

In our design of barcodes, starting and ending pattern is used to distinguish two barcodes in a single time slot. But it would be a fault case if the data set of a barcode itself contains a starting or ending pattern. To solve this problem, bit stuffing technique is used to create a barcode. Fig 5(a) shows the technique of bit stuffing. In this technique, if a barcode contains starting and ending pattern in its data portion, then a special bit pattern (let it be 10001000)is inserted into the pattern. Receiver system should be capable of un-stuffing the bit pattern from the data. Again, If a Barcode contains this special pattern, then the same technique of insertion of special bit pattern is applied into it.

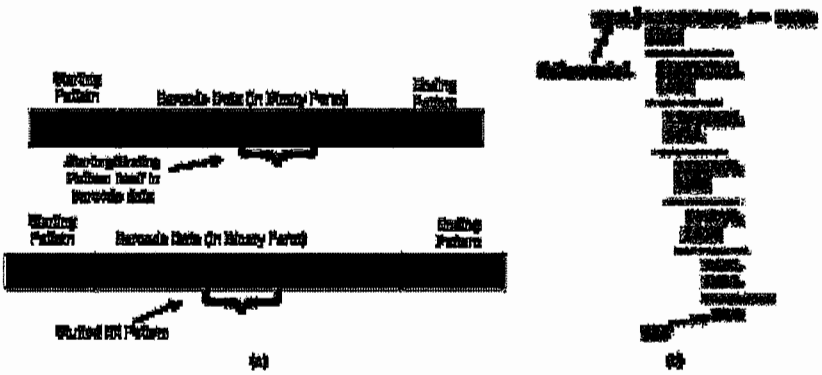


Figure 5. (a) Bit stuffing technique; (b) sample calculation of CRC

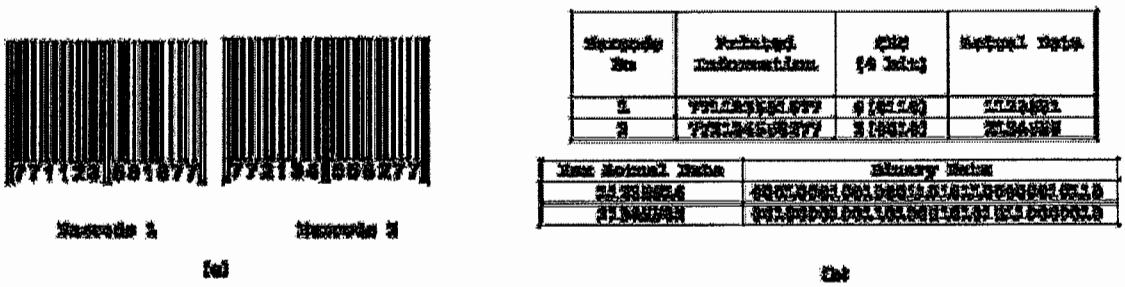


Figure 6. (a) Two sample barcodes which are read for MATLAB program; (b) illustration of sample barcodes

5.2.3 Cyclic redundancy check (CRC)

Cyclic Redundancy Check(CRC) is a well-known technique for the trustworthiness of a received data. As data receiving system is wireless, some errors may occur in receiving the bits. Moreover, poorly printed barcodes can make false reading. A single printing error can give erroneous result. In our design, CRC is appended with the data just before the ending pattern starts. When a message and its CRC are received from the receiver and the CRC matches the message's calculated CRC then it is assumed that, no miss or alteration of data has been occurred.

Fig 5(b) shows the technique of calculating CRC of a sample data. If the input bit above the leftmost divisor bit is '0', do nothing and move the divisor to the right by one bit. If the input bit above the leftmost divisor bit is '1', the divisor is exclusive-ORed into the input. The divisor is then shifted one bit to the right, and the process is repeated until the divisor reaches the right-hand end of the input row.

5.3 Output

Extracted barcodes in a single scan are displayed here. This barcodes are the input of the organization's software for bringing out the related information of the product or material.

6. SIMULATION RESULTS

We are here implementing the software portion specially the 2nd step in the flow chart shown in Fig 2. For simulation purpose, we have used two barcodes shown in Fig 6(a). The information printed on the data are illustrated in the table in Fig 6(b). Both are assumed as hexadecimal numbers for our convenience. Here '77'(Binary 01110111) is the starting and ending pattern.

For CRC, we have used 5 bit Polynomial (10110). So the CRC is 4 bit. We assumed that the two barcodes are read simultaneously in one time slot and hex data are received from the barcode reader serially. A MATLAB program is used to identify the barcodes from the data received. The serial output is converted to binary form and the MATLAB program extracts the barcodes from the binary data, calculate the CRC and match with the received CRC. If match found, then the

Figure 7. Output of the MATLAB program to extract data and check CRC

extracted actual data is forwarded for further proceeding as described in Fig 2. The output of the MATLAB program is shown in Fig 7. Here two barcodes are extracted; CRC is checked and matched with the received CRC.

7. CONCLUSION

The simulation result has shown that it is possible to extract original information from multiple barcode tags simultaneously. This system can be further improved by using dual threshold method [2] which will eventually speed up the scanning and detection process.

REFERENCES

[1] Wakaumi, H. and Nagasawa, C., "Development of a two-dimensional bar-code detection system using multi laser diodes with time-sharing light emission operation", Opt. Rev., vol. 8, pp. 101-106 (2001).

[2] Wakaumi, H. and Nagashaka, C. "A 2D ternary barcode detection system with a dual threshold", IEEE SENSORS EXCO, Daegu, Korea, 1511-1514 (2006).

[3] Hong-ying, S., "The application of barcode technology in logistics and warehouse management", First International Workshop on Education Technology and Computer Science, 732-735 (2009).

[4] Kurose, J. F. and Ross, K. W., [Computer Networking: A Top-Down Approach Featuring the Internet], Dorling Kindersley, India (3rd Edition), Chapter 5.2: Error Detection and Correction (2006).

[5] Tanenbaum, A. S., [Computer Networks], Asoke K. Ghosh, India (4th Edition), Chapter 3.1.2: Framing (Technique of Bit Stuffing)(2007).

[6] Fujiwara, Koyama, S., Kageyama, K. and Hiroshi , " Bar code reader system ", United States Patent 4672184 (1987).

[7] Barcode 1: How Barcodes Are Read, Available at http://www.adams1.com/readers.html

[8] How Do Barcodes Work, Available at http://www.morethanyouthink.com/shopping/barcodes work.html

[9] Webcam Barcode Scanner", Available at http://en.barcodepedia.com

Performance Evaluation of MLP and RBF Feed Forward Neural Network for the Recognition of Off-Line Handwritten Characters.

Rahul Rishi, TITS, Bhiwani, India.
Amit Choudhary, Maharaja Surajmal Institute, New Delhi, India.
Ravinder Singh, Maharaja Surajmal Institute , New Delhi, India.
Vijaypal Singh Dhaka, Institute of Management Studies, Noida, India.
Savita Ahlawat, MSIT, New Delhi, India.
Mukta Rao, Interglobe Technologies Ltd., Gurgaon, India.

ABSTRACT

In this paper we propose a system for classification problem of handwritten text. The system is composed of preprocessing module, supervised learning module and recognition module on a very broad level. The preprocessing module digitizes the documents and extracts features (tangent values) for each character. The radial basis function network is used in the learning and recognition modules. The objective is to analyze and improve the performance of Multi Layer Perceptron (MLP) using RBF transfer functions over Logarithmic Sigmoid Function. The results of 35 experiments indicate that the Feed Forward MLP performs accurately and exhaustively with RBF. With the change in weight update mechanism and feature-drawn preprocessing module, the proposed system is competent with good recognition show.

Keywords- Character Recognition, Image Processing, Feature Extraction, Radial Basis Function.

1 INTRODUCTION

This handwriting recognition domain has got two different problems sets- (1) Recognition of characters as they are written using Tablet PC's, Light Pens, PDAs, etc. And (2) the recognition of already written documents called offline hand recognition.Offline handwritten character recognition is a process where the computer system automatically understands the figure of a handwritten script. Therefore offline recognition is distinguished from on- line where characters' features are available electronically and can be tracked with much ease[1] Although online text capturing process extracts more features in terms of curves, slopes, disconnecting points etc but offline recognition has got significant attraction as it has wide spread application domain for e.g. Bank checks[2,3], Examination forms[4], signature verification[5], Historical documents recognition, remote sensing [6], dynamic modeling and medicine[7]. Additionally, age-old documents and fact sheets stored in the offices have a great necessity to be fed into the computers in text form [12,13,14]. The recognition of numeral strings has gained much attention due to its pivotal role in large number of applications such as reading of text form, postal address interpretation and census forms. The increasing popularity of the neural networks is partly due to their ability to learn and generalization in particular, the feed forward neural network refers to prior assumption about the statistics of input data and can construct complex decision boundaries[8].

Artificial Neural Networks (ANNs) have played a wonderful role in achieving remunerable results. The one essential subset of ANN is Radial Basis Function Network and its generalization. The important aspect of the RBFN is the distinction between the techniques of updating the first and the second layers weights. Various techniques have been proposed in the literature for optimizing the Radial Basis functions such as unsupervised methods like selection of subsets of data points [12], orthogonal least square method [13], clustering algorithm [14], Gaussian mixture models [15] and with the supervised learning method [16].

It has been observed [10] that the use of unsupervised techniques to determine the basis function parameters is not in general an optimal procedure so far as the subsequent supervised training is concerned. The difficulty with the unsupervised techniques arises due to the setting up of the basis functions using density estimation of the input data and takes no consideration for the target levels associated with the data. Thus, it is obvious that to set the parameters of the basis functions for the optimal performance the target data should be included in the training procedure and it reflects the supervised training. Hence, the basis function parameters for regression can be found by the treating the basis function centers and widths along with the second layer weights as adaptive parameters to be determined by minimization of an

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754633 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853479

error function.

The work proposed here is an extension to offline handwriting recognition using Radial Basis Function in Feed Forward Neural Network. This is a form of feed forward multi layer Neural Network[11]. The theory of RBF is still the subject of extensive ongoing research.

In this research work we considered two neural networks architectures (ANN1 & ANN2). The first architecture ANN1 is trained and tested with the conventional back propagation learning algorithm with the incorporation of Doug's Momentum descent term[17]. The ANN2 network architecture has been put into operation with the Radial Basis Function [18] in the hidden layer. The common variable used in these two architectures is the error correction momentum term. This term has also been modified to include second order Steepest Gradient Descent. The results of training and testing sets have been compared. The two types of networks (back propagation and back propagation with RBF) are analyzed to figure out the network that exhibit higher performance results with greater efficiency. Every network is assessed based on the rate of convergence and the speed of determination of the convergence weights for every pattern. So these two things have been the key-focus points of this work. The experiments are conducted with 35 samples of English alphabet. The significant improvement in the recognition of handwritten characters is achieved.

2 THE RADIAL BASIS FUNCTION

As mentioned, RBF networks emulate the behavior of certain biological networks. Basically RBF-MLP is essentially feed forward neural network with three layers namely Input, Hidden and Output. The single hidden layer consists of the locally tuned or locally sensitive units, and the output layer (in most cases) consists of binary responsive units. In the hidden layer units, the unit response is localized and decreases as a function of the distance of input from the unit's receptive field center.

The RBF-MLP uses a static Gaussian function as the nonlinearity for the hidden layer neurons. The Gaussian function responds only to a small region of the input space where the Gaussian is centered. The key to a successful implementation of these networks is to find suitable centers for the Gaussian functions. This can be done with supervised learning. The process starts with the training of input layer. Its function is to obtain the Gaussian centers and the widths from the input samples. Thus achieved centers are then prearranged within the weights of the hidden layer using competitive learning. During this learning process, the widths of the Gaussians are computed based on the centers of their neighboring neurons. The output of this layer is derived from the input samples weighted by a Gaussian combination.

The advantage of the radial basis function network is that it discovers the input to output map using local approximations. Usually the supervised segment is simply a linear combination of the approximations. Since linear combiners have few weights, these networks train extremely fast and require fewer training samples.

In contrast to the classical MLP the activation of a neuron is not given by the weighted sum of its all inputs but by the computation of a RBF. The RBF that we use is Gaussian Function:-

$$\phi(x_i) = e^{-\frac{\|x-\mu_i\|^2}{2\sigma_i^2}}$$

where x is the input to the neuron i, μ_i is the basis of neuron i.

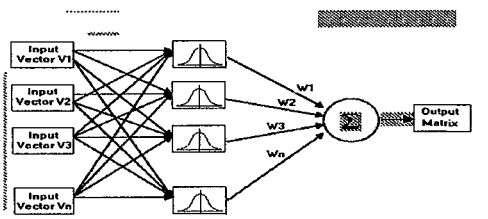


Figure 1: A General MLP-RBF network with one output

The RBF network in Figure 1 is often complemented with a linear part. This corresponds to additional direct connections from the inputs to the output neuron. Each input neuron corresponds to a component of an input vector x.

The hidden layer consists of h neurons. Each node in the hidden layer uses an RBF as its non linear activation function. The input layer has i nodes; the hidden and the output layer have K and J neurons, respectively. The hidden layer performs a non-linear transform of the input. The output layer is a linear combiner, mapping the nonlinearity into a new space. The RBF-MLP can achieve a global optimal solution to the adjustable weights in the minimum MSE by using the linear optimization method. Therefore for an input pattern x, the output of the j^{th} node of output layer can be defined as:

$$y_j(x) = \sum_{k=1}^{K} w_{kj}\phi_k(\|x_i - \mu_k\|) + w_{0j} \quad \text{for all j=1,2.....J.}$$

where $y_j(x)$ is the j^{th} output of the RBF-MLP, w_{kj} as the connection weight from the k^{th} hidden unit to the j^{th} output unit, μ_k is the prototype or centre of the k^{th} hidden unit. The RBF uses the Gaussian function:

$$\phi_k(x) = \exp(-\frac{\|x_i - \mu_k\|^2}{2\sigma_k^2}) \qquad \text{where k = 1, 2... K.}$$

The weight vector between the input layer and the k^{th} hidden layer neuron can be interpreted as the centre μ_k

Where x is the N- dimensional input vector and μ_k is the vector determining the centre of the basis function ϕ_k.

Therefore for an input pattern x, the error of the network can be defined as:

$$E = \frac{1}{2}\sum_{i,k}(o_i^k - y_i^k)^2 \quad \text{where } o_i^k \text{ is the desired output and } y_i^k \text{ is the actual output.} \tag{1}$$

The error function has been considered in "(1)" is the least mean square (LMS). This error will minimize along with the decent gradient of error surface in the weight space between hidden layer and the output layer. The expressions for the derivatives of the error function with respect to the weights basis function parameters for the set of P pattern pairs (xp,yp) can be written as:

$$\Delta w_{ik} = -\eta_1 \frac{\partial E^p}{\partial w_{ik}} \tag{2}$$

$$\Delta \mu_k = -\eta_2 \frac{\partial E^p}{\partial \mu_k} \tag{3}$$

$$\Delta \sigma_k = -\eta_3 \frac{\partial E^p}{\partial \sigma_k} \tag{4}$$

here p=1 to P.

$$y_i^p = \sum_{k=1}^{K} w_{ik}\phi_k(\|x^p - \mu_k^p\|) \tag{5}$$

Since y_i^p is the outcome of Radial Basis Function used and gradient for the network is given by partial differentiation of this error with respect to different parameters. Hence from "(2)" we have,

$$\Delta w_{ik} = \eta_1 \sum_p (o_k^p - y_k^p) * e^{-\frac{\|x^p - \mu_i\|^2}{2\sigma_i^2}} \tag{6}$$

We have from Eq.6, the expressions for change in weight vector to accomplish the learning in supervised way. The setting of the basis function parameters with supervised learning represents a non linear optimization problem which will typically be computationally intensive and may be proven to find local minima of the error function. Thus, for well localized RBF, an input will generate a significant activation in a small region and the opportunity of getting stuck at a local minimum is small.

The RBF training with the decent gradient method is capable to provide the equivalent or better performance compared to that of the multi layer feed forward neural network trained with the back propagation. Now in the following subsection we are presenting the simulation design and implementation details of redial basis function worked as a classifier for the handwritten English character recognition problem and compose the results with the back propagation algorithm for the MLP network.

3 SIMULATION DESIGN AND IMPLEMENTATION DETAILS

The Network used in this work is having RBF imposed on a MLP (Multi Layer Perceptron). The Feed forward MLP used has 36 input neurons, 20 hidden neurons and 26 output neurons. The transfer function used to propagate the weight update among the layers is Radial Basis Function hence MLP-RBF. The standard Feed Forward MLP contains Log-

Sigmoid transfer functions whereas we have used RBF. The 36 input neurons correspond to 36 tangent values of each character's image. The 26 output neurons correspond to 26 letters of English alphabet.

Figure 2: MLP network with RBF in Hidden Layer

3.1 Experimental conditions for both the networks

In total the no of experiments conducted for training the network were 7000 including 3500 experiments for each type of network. The performance testing of the network was done in 200 experiments. The parameters used in the two experiment sets of both type of networks are described in Table 1 and Table 2 below.

Table 1: Parameters used for MLP Network

Parameter	Value
Backpropagation learning Rate (η)	0.1
Momentum Term (α)	0.9
Doug's Momentum Term $(1-(\alpha))^{-1}$	$(1-(\alpha))^{-1}$
Adaptation Rate (K)	3.0
Maximum Error Allowed in Network	0.0001
Initial weights and biased term values	Random Values Between 0 and 1

Table 2: Parameters used for MLP -RBF Network

Parameter	Value
Backpropagation learning Rate	0.1
Momentum Term (α)	0.9
Doug's Momentum Term $(1-(\alpha))^{-1}$	$(1-(\alpha))^{-1}$
Adaptation Rate (K)	3.0
Spread parameter σ	1.0
Maximum Error Exist in the Network	0.00001
Initial weights and biased term values	Random Values Between 0 & 1

3.2 Training and testing patterns preparation

This is the part of Pre-Processing module. In this module the handwritten character samples were gathered mostly from university and school students. Every contributor was given 4 forms to fill in characters into the blocks on the form. The forms were then scanned and individual character's images were extracted. Character's image was binarized and resized to 210×210. Each image was then divided into 36 equal sub-images of the size 35×35. For better understanding the $(9, 9)^{th}$ sub-division is shown in Figure 4.

Figure 4: Dividing the character's image into 36 equal parts

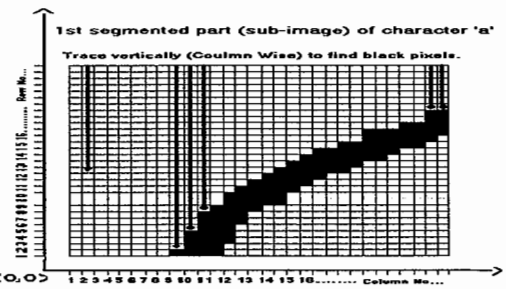


Figure 5: The Column-Based tracing of a binary sub-image.

After this division, the binary sub-image is vertically (Column Based) traced to fit a liner curve. For each of the character images, the following algorithm was applied to get 36 tangent values from each character image:

Algorithm 1 : Pattern Preparation
Input : Character's Image

1. Resize the image ***(I)*** into 210 × 210 pixels, using the K-Nearest Neighborhood algorithm.

2. Divide the image ***(Iresized)*** into 36 equal sub-images using standard IMCROP method of MATLAB. As shown in "Fig. 4".
3. Covert the sub-image ***(ISub1)*** into binary form ***(Ibin1)***
4. Trace ***Ibin1*** vertically (Column - wise) to get the black pixel.
   4.1 Add row number to SumRowCount and Increment *RowCount* by one if a black pixel is found. And store corresponding Column Number in an Array *ArrCoulmnNum.*

   4.2 Calculate the average *RowNum* by dividing *SumRowCount* with *RowCount* when row number is reached to max i.e. 35 and *RowCount* is non-zero.

   4.3 Store non zero values of *RowNum* in an array *ArrAvgRowNum.*

5. Repeat step 4 till the all columns are read.
6. Fit a liner polynomial on *(ArrAvgRowNum, ArrCoulmnNum).*
   for sample test Column Numbers 0:2:34 get the values from polynomial. From these values calculate Tangent value for this polynomial using the PolyVal Function of MATLAB.

7. Repeat steps 3 to 7 for each of 36 sub-images and arrange the obtained tangent values in a matrix of size 36×1 for each character.

Output : 36 tangent values corresponding to the character's image

The tangent values achieved from above mentioned algorithm for each character forms the input sample. Each input sample comprises of 36 tangent values for char 'a', for char 'b', for char 'c'... so on. This is best depicted in following matrix:

aTan1	bTan1	cTan1	dTan1	...	zTan1
aTan2	bTan2	cTan2	dTan2	...	zTan2
aTan3	bTan3	cTan3	dTan3	...	zTan3
aTan4	bTan4	cTan4	dTan4	...	zTan4
aTan5	bTan5	cTan5	dTan5	...	zTan5
aTan6	bTan6	cTan6	dTan6	...	zTan6
⋮	⋮	⋮	⋮		⋮
aTan36	bTan36	cTan36	dTan36	...	zTan36

Table 3: Input sample prepared from the tangent values of individual characters

Parameter Name	Value
Max Epochs	50000
Transfer Function	RBF
Performance Function	MSE
Training Goal	0.001
Show Graph	100 epochs

Table 4. Initial system set-up values for RBF network

3.3 Learning and performance evaluation of the network

The system was trained with 35 samples, with the set of training parameters as shown in table 4. For testing of the network, 5 test samples were created by randomly selecting character images. The results are discussed in the next section.

4. RESULTS AND DISCUSSION

The network was trained with 35 different input samples and the number of epochs required to train the network and the error values for each of the 35 samples are shown in Table 5 and Table 6 respectively.

It is evident that the training process with radial basis transfer function, the number of epochs are very less. This can be best explained by Table 5 below.Table 6 represents the error present in the two network arrangements. The values present in the table clearly depict the lower error values present in case of RBF-MLP as compared to MLP.

Table 5: Comparison of network training iterations (epochs) between the two learning trails in supervised learning.

	Epochs With	Epochs with
Sample 1	46	67
Sample 2	61	88

Table 6. Error level attained by both the networks.

Samples	Error in	Error in
Sample 1	0	0
Sample 2	0	0

Sample 3	84	142
Sample 4	120	204
Sample 5	193	231
Sample 6	247	690
Sample 7	309	763
Sample 8	271	1538
Sample 9	443	916
Sample10	742	1127
Sample11	458	2054
Sample12	957	2348
Sample13	1344	2601
Sample14	1529	3184
Sample15	1942	3812
Sample16	2061	6059
Sample17	2413	6875
Sample18	2570	7426
Sample19	3485	8513
Sample20	5499	9624
Sample21	5316	12198
Sample22	5874	17274
Sample23	6689	20166
Sample24	6701	29312
Sample25	8596	31957
Sample26	8952	36568
Sample27	11603	49465
Sample28	16545	50000
Sample29	18379	50000
Sample30	23004	50000
Sample31	29691	50000
Sample32	40155	50000
Sample33	50000	50000
Sample34	50000	50000
Sample35	50000	50000

Sample 3	0	0
Sample 4	0	0
Sample 5	0	0
Sample 6	0	0
Sample 7	0	0
Sample 8	0	0
Sample 9	0	0
Sample10	0	0
Sample11	0	0
Sample12	0	0
Sample13	0	0
Sample14	0	0
Sample15	0	0
Sample16	0	0
Sample17	0	0
Sample18	0	0
Sample19	0	0
Sample20	0	0
Sample21	0	0
Sample22	0	0
Sample23	0	0
Sample24	0	0
Sample25	0	0
Sample26	0	0.001804
Sample27	0	0.00549
Sample28	0	0.009672
Sample29	0	0.03501
Sample30	0	0.861
Sample31	0	0.91765
Sample32	0	4.64638
Sample33	0.001629	5.5603
Sample34	0.00751	3.491
Sample35	0.020146	7.0042

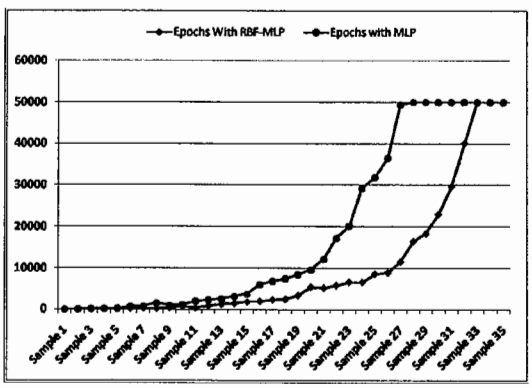


Figure 6. Comparison of network training iterations (epochs) between MLP and MLP-RBF Neural Network.

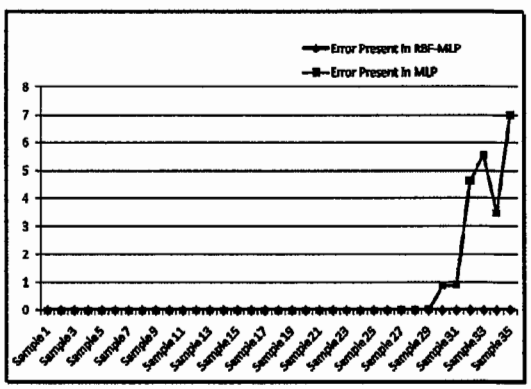


Figure 7. Comparison chart for Error level attained by both the networks.

5. CONCLUSION

The proposed method for the handwritten characters recognition using the Radial Basis Function with Steepest gradient approach showed the remarkable enhancement in the performance as compared to Multi-Layer Perceptron. The results, as shown in table 5, for the different sample of the characters sets, represent the number of iteration as epochs that a sample has considered for the learning of input patterns. The results from last column of table 6 are representing the training with classical method of Multi Layer Perceptron. The second column shows the results of improved version of Radial Basis Function used with Multilayer Perceptron. In this case the smaller error values represent that the error surface is smoother and it will give a good indication of the "true error correction path". The network will converge faster which results in less training time and better recognition accuracy.

6. REFERENCES

[1] Vinciarelli. A., Perrone. M., "Combining Online and Offline Handwriting Recognition," *icdar*, Seventh International Conference on Document Analysis and Recognition (ICDAR'03) - Volume 2, p. 844,(2003).

[2] Rafael. P. and Gupta. A., "A System for Processing Handwritten Bank Checks Automatically,". MIT Sloan Working Paper No. 4346-02, (2002)

[3] Gorski. N., Anisimov. V., Augustin. E., Baret. E., Price. D., Simon. J. C.," Check Reader: A family of bank check recognition systems. Document Analysis and Recognition,". ICDAR apos;99. Proceedings of the Fifth International Conference, 523 – 526,(1999)

[4] Rossant. F., and Bloch. I.,"Robust and adaptive OMR system including Fuzzy modeling, fusion of musical rules, and possible error detection.," *EURASIP J. Appl. Signal Process.* 2007, (1) ,160-160, (2007).

[5] Fierrez. J., Ortega-Garcia. J., Ramos. D., and Gonzalez-Rodriguez. J.," HMM-based on-line signature verification: Feature extraction and signature modeling. *Pattern Recogn. Lett.* 28(16), 2325-2334(2007), DOI= http://dx.doi.org/10.1016/j.patrec.2007.07.012

[6] Biadsy.F., El-Sana. J., Habash. N.,"Online Arabic Handwriting Recognition Using Hidden Markov Models", Tenth International Workshop on Frontiers in Handwriting Recognition [inria-00108306 – version 1], (2006)

[7] Cordella. L. P.,De Stefano. C., Tortorella. F. and Vento. M., "A method for improving classification reliability of multilayer perceptrons," IEEE Transaction on Neural Networks. v6 i5. 1140-1147, (1999).

[8] Szabó. T., and Horváth. G., "An Efficient Hardware Implementation of Feed-Forward Neural Networks," *Applied Intelligence* 21(2), 143-158,(2004).

[9] Ilonen. J., Kamarainen. J. and Lampinen. J., "Differential Evolution Training Algorithm for Feed-Forward Neural Networks." *Neural Process. Lett.* 17(1), 93-105, (2003).

[10] Romero. E. and Alquézar. R., "Heuristics for the selection of weights in sequential feed-forward neural networks: An experimental study." *Neurocomput.* 70, 16-18, 2735-2743, (Oct. 2007).

[11] Dhaka. V. S. and Singh. M. P., Simulating Biological Neural Network Structure in Computers with help of MATLAB for Handwriting Recognition Tasks, Asian J. Exp. Sci., Vol. 21(2), (2007).

[12] Petra. P and Chid. A, "Empirical evaluation of feature subset selection based on a real-world data set", Engineering Applications of Artificial Intelligence, Vol 17, Issue 3, 285-288, (2004)

[13] Bierlaire. M., Thémans. M., "Dealing with singularities in nonlinear unconstrained optimization", European Journal of Operational Research, 196(1), 33-42, (2009).

[14] Jianhua. L. and Laleh. B., "An Effective Clustering Algorithm for Mixed-size Placement", Proc. of ISPD, Austin, (2007).

[15] Shental. N., Bar-hillel. A. and Weinshall. D., "Computing Gaussian Mixture Models with. EM using Side-Information", in Proc. of Int. Conference on Machine Learning, ICML-03, Washington DC, (2003).

[16] Gader. P. D., Mohamed. M. and Chiang. J. H., "Handwritten Word Recognition with Character and Inter-Character Neural Networks", IEEE Transactions on Systems, Man, and Cybernetics-Part B: Cybernetics, Vol. 27, 158-164, (1997).

[17] Mangal. M., Singh. M. P., "Handwritten English Vowels Recognition Using Evolutionary Feed-Forward Neural Network", Malaysian Journal of Computer Science, Vol. 19(2), 169–187,(2006).

[18] Lee. S., "Off-Line Recognition of Totally Unconstrained Handwritten Numerals Using Multilayer Cluster Neural Network". IEEE Trans. Pattern Anal. Mach. Intell. 18(6), 648-652, (1996).

Comparative Wavelet, PLP and LPC Speech Recognition Techniques on the Hindi Speech Digits Database

A. N. Mishra*, M. C. Shrotriya* and S. N. Sharan**

*Department of Electronics & Communication Engineering, BIT, Mesra, Ranchi, India

**Director GNIT, Greater Noida, U.P., India

an_mishra53@rediffmail.com, shrotriya@bitmesra.ac.in, snathsharan@yahoo.com

ABSTRACT

In view of the growing use of automatic speech recognition in the modern society, we study various alternative representations of the speech signal that have the potential to contribute to the improvement of the recognition performance. In this paper wavelet based features using different wavelets are used for Hindi digits recognition. The recognition performance of these features has been compared with Linear Prediction Coefficients (LPC) and Perceptual Linear Prediction (PLP) features. All features have been tested using Hidden Markov Model (HMM) based classifier for speaker independent Hindi digits recognition. The recognition performance of PLP features is11.3% better than LPC features. The recognition performance with db10 features has shown a further improvement of 12.55% over PLP features. The recognition performance with db10 is best among all wavelet based features.

Keywords: PLP, LPC, HMM, Hindi digits recognition

1 INTRODUCTION

Speech is the primary mode of communication. It is a way of sharing facts, thoughts and emotions and also a way of transferring human intelligence from one person to another [1, 2]. Speech recognition is always looked upon as a fascinating field in human computer interaction. Speech recognition is the process of converting an acoustic signal, captured by a microphone or a telephone, to a set of words. The recognized words can be the final results, as for applications such as commands and control, data entry, and document preparation. Speech recognition has been a goal of research for more than four decades but the desired goal is yet to be achieved, namely, a machine that can understand spoken discourse on any subject by all speakers in all environments. A speech recognition system basically comprises of a collection of algorithms from a variety of disciplines, such as statistical pattern recognition, communication theory, signal processing, combinatorial mathematics, linguistic etc. In different recognizers each of these areas is relied on different degrees. Research in speech processing is focused primarily on the man-machine interface. One of the most difficult aspects of performing research in speech recognition by machine is it's inter disciplinary nature and the tendency of most researcher to apply a monolithic approach to individual problem. The Isolated Word Recognition systems [3] were among the first speech recognition systems implemented due to rather straightforward manner in which the basic recognition units – the words can be modelled. Any IDR contains two main phases, training phase and the testing phase.

A speech recognition system has two major components, feature extraction and classification. The recognition performance heavily depends on the performance of the feature extraction block. Thus choice of features and its extraction from the speech signal should be such that it gives high recognition performance with reasonable amount of computation. Various commonly used feature extraction techniques are the Linear Prediction Coefficients (LPC) [3, 4], the Reflection Coefficients (RC) [3], the Linear Prediction Cepstral Coefficients (LPCC) [3] and the Perceptual Linear Predictive (PLP) Coefficients [5]. LPC technique is less effective for representing speech and does not compete with newly proposed technique

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754634 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856318

because it assumes signal stationary [3, 4] within a given time frame and therefore lacks the ability to analyze the localized events accurately. Furthermore, the LPC approach assumes a particular linear (all pole) model of speech production, which strictly speaking is not the case and not able to capture the nasalized and unvoiced sounds properly [3]. Recently, lot of research has been carried out on using wavelet-based features [6, 7] for speech recognition applications. Discrete Wavelet Transform (DWT) has good time-frequency resolution and aims at extracting the localized contributions of the analyzed signal. Due to its energy compaction property [6] better speech recognition accuracy has been obtained as compared to LPC and PLP. In this paper, LPC, PLP and different wavelet based features have been used for Hindi digits recognition. Hidden Markov Model (HMM) [3, 8] based classifier is used for training and testing of Hindi digits.

2 FEATURE EXTRACTION

The raw speech signal is complex and may not be suitable for feeding as input to the automatic speech recognition system; hence the need for a good front-end arises. The task of this front-end is to extract all relevant acoustic information in a compact form compatible with the acoustic models. In other words, the preprocessing should remove all non-relevant information such as background noise and characteristics of the recording device, and encode the remaining (relevant) information in a compact set of features that can be given as input to the classifier. Features can be defined as a minimal unit, which distinguishes maximally close classes. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When the input data to an algorithm is too large to be processed and it is suspected to

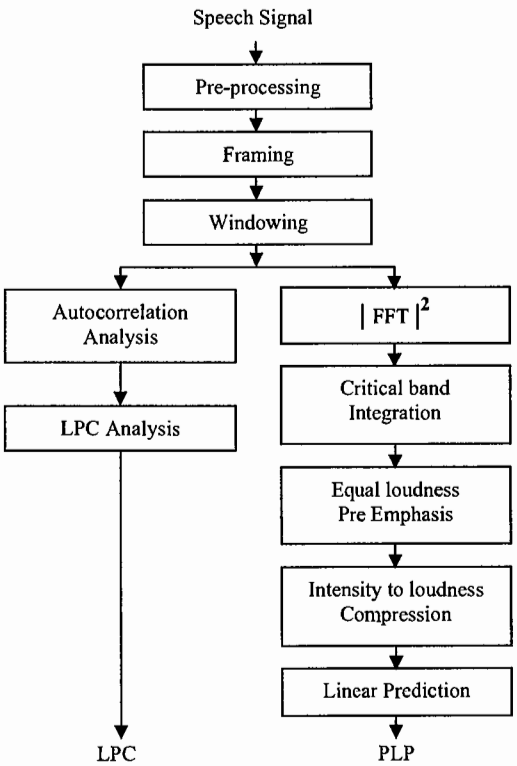


Figure 1: Feature Extraction using LPC and PLP

be redundant (much data, but not much information) then using certain techniques the input set of features is reduced. Here LPC, PLP and Wavelet based feature extraction techniques are used for feature extraction.

2.1 Linear Predictive Coefficients: Both the feature extraction techniques, LPC and PLP are shown in Figure1. The first step of all the three Feature Extraction technique is preprocessing which was applied on all the speech samples. The major steps in pre-processing are pre-emphasis filtering and normalization. The digitized speech signal is passed through first-order FIR filter to spectrally flatten the signal and to make it less susceptible to finite precision effects. A 1st order high-pass FIR filter with filter coefficient 0.98 is used to pre-emphasize the digitized speech signal. The transfer function of the filter is given by:

$$H(z) = 1 - 0.98z^{-1} \tag{1}$$

To compensate for the variation in the amplitudes of different speech samples of same digit, all the speech samples are divided by the sample with highest amplitude. The resulting speech signal is blocked into frames of 25 msec with an overlap of 10 msec. Then Hamming window is applied on each speech frame. These steps are common to all feature extraction techniques used in this paper. The hamming window has the form:

$$w(n) = 0.54 - 0.46 * \cos\left(\frac{2\pi n}{N-1}\right), \quad 0 \le n \le N-1 \tag{2}$$

For calculation of LPC features, each frame of windowed signal is auto-correlated. The highest autocorrelation value is the order of LPC analysis 'p', i.e., which has been taken 13. First thirteen LPC coefficients were chosen for each frame. Finally thirteen LPC coefficients were selected for each sample of Hindi digits by applying vector quantization on features of all frames of each sample of each Hindi digit.

2.2 Perceptual Linear Prediction: PLP inherits its characteristics from LPC as well as MFCC. PLP features can be obtained by following the steps as shown in Figure1. The speech signal is pre-emphasized and normalized similarly as in LPC. The frame duration and overlapping are also taken same as in LPC. The power spectral estimate for the windowed speech signal is computed. The power spectrum is integrated within overlapping critical band filter responses. For obtaining PLP, trapezoidal shaped filters are applied at 1-bark intervals. Restrictions are imposed on higher frequencies of band-pass filter in such a way that pass-band in the domain of critical band modulations is established to a range that appears to be required for speech intelligibility. The elements of critical band spectrum are explicitly weighted. To reduce the effects of amplitude variations for spectral resonances cube root of spectral amplitudes are taken after that IDFT is performed. Durbin's algorithm is used for computing the PLP coefficients. Again 13 features are taken for each speech sample by applying vector quantization on the features of all frames of a sample.

2.3 Wavelet Based Feature Extraction: The speech signal is pre-emphasized and normalized similarly as in LPC. The frame duration and overlapping are also taken same as in LPC. In Wavelet based feature extraction technique the pre-emphasized signal is decomposed into 24 bands using different wavelets. As the database is sampled at 16 kHz, giving an 8 kHz bandwidth signal, first, a full three level Wavelet Packet (WP) decomposition is carried out. This partitions the frequency axis into eight bands each of 1 kHz. The lowest band of 0–1 kHz is further decomposed by applying again full three level WP decomposition. This divides the 0–1 kHz band into eight subbands each of bandwidth 125 Hz. The frequency band of 1–2 kHz is further decomposed by applying two level WP decomposition, thereby, giving four subbands of 250 Hz. The frequency band of 1–1.25 kHz and 1.25–1.5 kHz is further decomposed once, thus increasing the number of bands to six in the 1–2 kHz range. Next, the 2–3 kHz band is selected, complete 2 level decomposition using WP is performed. This gives four bands each of 250 Hz bandwidth. 3–4 kHz band is decomposed once by WP, giving two bands of 3–3.5 kHz and 3.5–4 kHz. The frequency bands of 4–5 kHz, 5–6 kHz, 6–7 kHz, and 7–8 kHz are not further decomposed. This gives a total of 24 frequency bands. The total energy of wavelet coefficients in each sub-band is calculated. Logarithm of the energy of all the

24 frequency bands is computed. Then DCT is applied which arranges the coefficients in ascending order of the weightage of the information contained in the coefficients. First 13 coefficients are selected out of 24 DCT coefficients.

3 EXPERIMENTAL SETUP AND RESULTS

The experimental setup is shown in Figure 2. For speech recognition a suitable database is required. For this purpose, ten repetitions of the Hindi digits from "shunya" to "nau" of 24 speakers of different age groups and dialect were recorded. Then the samples of each digit for all speakers were segmented using cool software. Sampling rate has been taken as 16 KHZ, 16-bits, mono. The database consists of 10 samples of each Hindi digit of each speaker, i.e., total 2400 speech samples (for 24 speakers), 240 samples (for each speaker). The microphone was kept at a distance of 3-4 inches from the speaker's mouth. Microphones from "Intex" and "iball" company were used for recording.

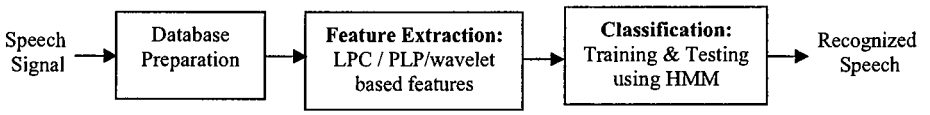


Figure 2: Block diagram of Speech Recognition System

Three different types of features, LPC, PLP and wavelet based features in combination with HMM are used for speaker independent Isolated Hindi digits recognition. For LPC features 13th order LPC analysis was performed on each frame after that first 13 LPC were chosen for each frame. Finally thirteen LPC coefficients were selected for each sample of Hindi digits by applying vector quantization on features of all frames of each sample of each Hindi digit. 2400 feature vectors of all speakers of all samples were stored for the purpose of training and testing. Out of 2400 feature vectors, there are 240 feature vectors of each Hindi digit. For recognition purpose ten HMM models are prepared one for each Hindi digit (0-9) using feature vectors of 200 samples of each Hindi digit as training data. Feature vectors of remaining 40 samples of each digit are used as test data. For recognition of unknown digit the feature vector corresponding to the unknown digit is taken as the observation sequence. The probability of occurrence of that observation sequence is computed for each digit model. The digit whose model gave the highest probability was taken as the recognized digit.

Figure 3: Recognition performance of LPC, PLP and db10 features with HMM

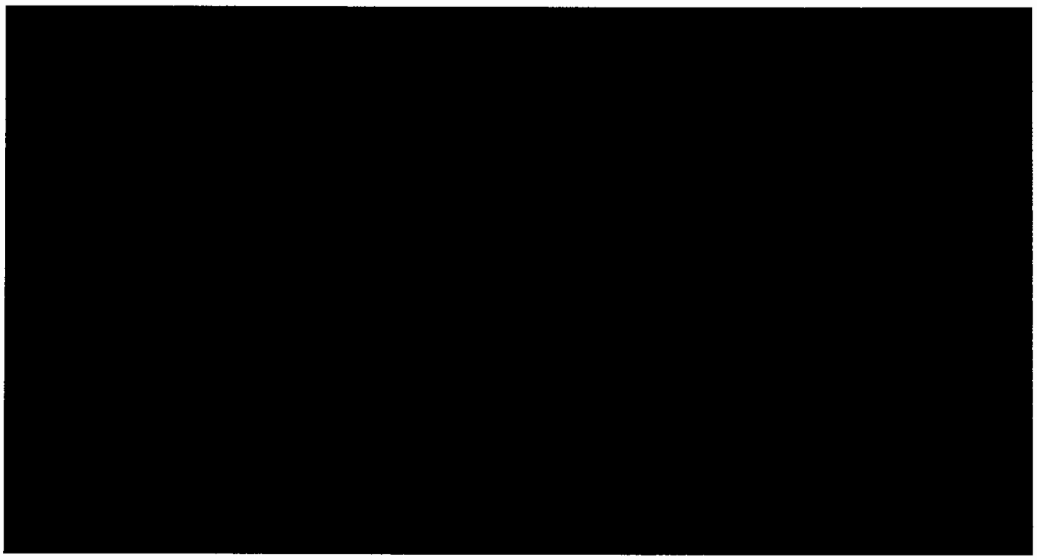


Figure 4: Recognition performance of different wavelet based features with HMM

Similarly 2400 feature vectors were also extracted for PLP and all wavelet based techniques for 24 speakers. Again for recognition purpose ten HMM models are prepared one for each Hindi digit (0-9) using feature vectors of 200 samples of each Hindi digit as training data. Feature vectors of remaining 40 samples of each digit are used as test data. The recognition was done similarly as in the above paragraph.

Comparative speaker independent results for different feature extraction techniques LPC, PLP, and db10 with HMM are obtained as 66%, 77.3% and 89.85% respectively. The results are shown in Figure 3. Comparative speaker independent results for different Wavelet based features coif1, coif2, sym6, sym10, db6, db10 with HMM are found to be 78.30%, 82.75%, 89.60%, 89.70%, 89.67% and 89.85% respectively. These results are shown in Figure 4. The LPC features used for recognition of Hindi digits are usually based on STFT. It is well established now that some plosive sounds are of very short duration ranging only a few milliseconds. Thus window of about 20-50ms used for the Fourier analysis duration [3, 7] will not be able to track these sudden changes, which causes non-stationarity in the signal. The assumption in Fourier analysis of the signal being stationary during 20-50ms duration is specifically violated in case of stops (plosives) phonemes. It has been found that most of the information about the position of articulators is present in first few milliseconds of the onset of burst. To overcome this problem, wavelet transforms has been used with some success for feature extraction in this paper. This is because of its capability to analyse non-stationary along with the stationary signals. Another reason for lower performance of LPC features is that these are all pole based features and not efficient for estimation of unvoiced phoneme and nasals sounds [3].

CONCLUSION

The recognition performance with db10 features was found to be the best among all the three representation. This is due to the fact that for an efficient transmission, the LP coefficients are subjected to quantization and interpolation. Quantization and interpolation of the LP coefficients are problematic because small changes in the coefficients may result in large changes in the power spectrum, and, possibly, in an unstable LP synthesis filter. Perceptual linear prediction, similar to LPC analysis,

is based on the short-term spectrum of speech. In contrast to pure linear predictive analysis of speech, perceptual linear prediction (PLP) modifies the short-term spectrum of the speech by several psychophysically based transformations. Like most other short-term spectrum based techniques PLP is vulnerable when the short-term spectral values are modified by the frequency response of the communication channel. A major concern arising from the classification of spectral data is that the number of variables or dimensionality often exceeds the number of available spectra. This leads to a substantial deterioration in performance of traditionally favored classifiers. It becomes necessary to decrease the number of variables to a manageable size, whilst, at the same time, retaining as much discriminatory information as possible. A new and innovative technique based on wavelets reduces the dimensionality and optimizes the discriminatory information. The discrete wavelet transform has been utilized to produce wavelet coefficients which are used for feature extraction. In conclusion, despite the preliminary stage of our experimental setup in the field of non-linear speech analysis, the results confirmed the hypothesis that using wavelets may bring potential in automatic speech recognition.

REFERENCES

[1] Allen, J. B., "How do humans process and recognize speech" IEEE Transactions on Speech and Audio Process., vol. 2, no. 4, pp. 567–576, 1994.

[2] Lippman, R. P., "Speech recognition by machines and humans," Speech Communication, vol. 22, pp. 1–15, 1997.

[3] Rabiner, L. R. and Juang, B. H., [Fundamental of Speech Recognition], 1st ed., Pearson Education, Delhi, 2003.

[4] Makhoul, J., "Linear prediction: A tutorial review", Proc. IEEE, vol. 63, no. 5, pp. 561–580, 1975.

[5] Hermansky, H., "Perceptual linear prediction (PLP) analysis of speech," Journal of Acoustic Society America 87, pp.1738-1752, 1990.

[6] Siafarikas Mihalis, Mporas Iosif, Ganchev Todor and Fakotakis Nikos, "Speech Recognition using Wavelet Packet Features," Journal of Wavelet Theory and Applications vol.2, pp. 41–59, 2008.

[7] Farooq, O. and Datta, S. 'Mel filter-like admissible wavelet packet structure for speech recognition', IEEE Signal Processing Letters, vol. 8, No. 7, pp. 196-198, 2001.

[8] Toledano Doroteo Torre, Villardebó Jesús Gómez, and Gómez Luis Hernández, "Initialization, Training, and Context-Dependency in HMM-Based Formant Tracking," IEEE transactions on Audio, Speech and Language Processing vol.14, no.2, pp.511-524, 2006.

An improved IHS fusion for high resolution remote sensing images

HU Youjian ZHANG Xiaohua
China University of Geosciences Wuhan, China
Zhangxiaohua861014@163.com

ABSTRACT

Image fusion plays an important role in improving high resolution remote sensing images, as many Earth observation satellites provide both high-resolution panchromatic and multispectral images. To date, many image fusion techniques have been developed. Existing traditional image fusion techniques such as the intensity–hue–saturation (IHS) transform, wavelet transform and principal components analysis(PCA) methods may not be optimal for fusing the new generation commercial high-resolution satellite images such as IKONOS and Quick Bird. However, the available algorithms can hardly meet a satisfactory fusion requirement for high resolution remote sensing images. Among the existing fusion algorithms, the IHS technique is the most widely used one technique. But the color distortion of this technique is often obvious, especially when high resolution multispectral images are fused with its panchromatic images. This study presents a new fusion approach that integrates both IHS and histogram match techniques to reduce the color distortion of high resolution remote sensing fusion results. Different high resolution remote sensing images have been fused with this new approach. The result proves that the concept of the proposed improved IHS is promising, and it does significantly improve the fusion quality compared to conventional IHS transform fusion techniques.

Keywords: Image fusion, IHS transform, An improved IHS transform ,fusion;

1. INSTRUCTION

Remote sensing images such as Multiplatform, Multi-sensor and multi-resolution are arriving at an alarming speed. These images form a multi-source image in the same area. Image fusion is a novel method for combining spectral information of coarse resolution imagery with finer spatial resolution imagery. Image fusion techniques deal with integration of complementary and redundant information from multiple images to create a composite image that contains a better description of the scene[9] .The resulting merged image is a product that synergistically integrates the information provided by various sensors or by the same sensor, which may be found useful for human and the result of image fusion is a new image which is more suitable for human and machine perception or further image-processing tasks such as segmentation, feature extraction and object recognition[1,2,6,7]. Data fusion can reduce the uncertainty associated with the data acquired by different sensors or by same sensor with temporal variation. Further, the fusion techniques may improve interpretation capabilities with respect to subsequent tasks by using complementary information sources.

The fusion of high-resolution panchromatic and multispectral remote sensing images is a very important issue for many remote sensing and mapping applications.[7,8] It is the aim of image fusion to integrate image data recorded at different resolutions, at different time or by different sensors in order to obtain more information than can be derived from only one image alone. The benefit of image fusion has been showed in many practical applications such as land usage classification, map updating and hazard monitoring in national defense

2. IHS FUSION TECHNIQUE

The color system with red, green and blue channels (RGB) is usually used by computer monitors to display a color image. Another color system widely used to describe a color is the system of intensity, hue and saturation (IHS). The intensity represents the total amount of the light in a color (also called brightness), the hue is the property of the color determined by its wavelength, and the saturation is the purity of the color. An intensity image of the IHS system usually appears like a panchromatic image. This characteristic is utilized in the image fusion to fuse a high-resolution panchromatic image into a multispectral image[4].

To conduct an image fusion the three bands of a color image have to be transferred from the RGB space into the IHS space. Before this, the color image should be registered to the high-resolution panchromatic image and resampled to the

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754635 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.852962

same pixel size with the panchromatic image. The intensity image is then replaced by a high-resolution panchromatic image. To have a better fusion quality, the panchromatic image usually needs to be matched to the intensity image before the replacement. After the replacement, the panchromatic image together with the hue and saturation images are reversely transferred from the IHS space into the RGB space, resulting in a fused color image[1]. This process is schematized in the following Fig.1.

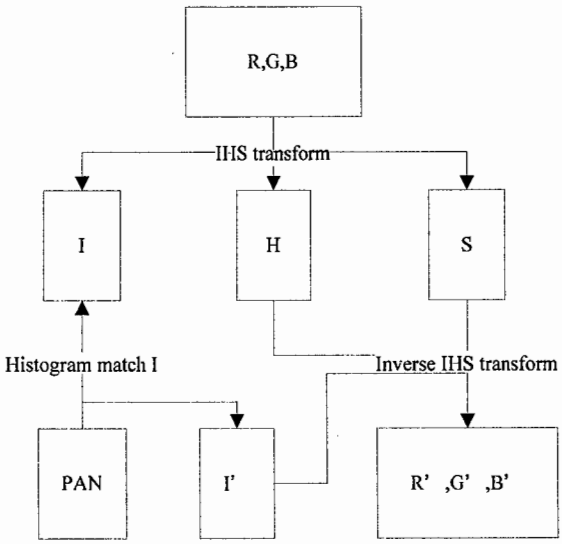


Fig.1 IHS transform fusion

Different transformations have been developed to transfer a color image from the RGB space to the IHS space. One common IHS transformation is based on a cylinder color model which is described by the following equations[2]:

$$\begin{pmatrix} I \\ V_1 \\ V_2 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & \frac{-2}{\sqrt{6}} \\ \frac{1}{\sqrt{2}} & \frac{-1}{\sqrt{2}} & 0 \end{pmatrix} * \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1}$$

$$H = \tan^{-1}\left(\frac{V_1}{V_2}\right) \tag{2}$$

$$S = \sqrt{V_1^2 + V_2^2} \tag{3}$$

Where v1 and v2 are two intermediate values. The corresponding inverse transformation is defined as:

$$\left.\begin{aligned} V1 &= S*\cos(H) \\ V2 &= S*\sin(H) \end{aligned}\right\} \tag{4}$$

$$\begin{pmatrix} R \\ G \\ B \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{6}} & \frac{-1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{-2}{\sqrt{6}} & 0 \end{pmatrix} * \begin{pmatrix} I \\ V_1 \\ V_2 \end{pmatrix} \tag{5}$$

3. THE IMPROVED IHS TRANSFORM FUSION

This fusion is based on the HIS transform, and it is aim to reduce the color distort .The basic idea behind the improved HIS transform method is to modify the input high-resolution panchromatic image so it looks more like the intensity component of the input multispectral image. Instead of using a total replacement of the intensity component, this method uses histogram matching based on IHS transform. First of all, multi-spectral image is decomposed into RGB image, Panchromatic images were then carried out with the RGB histogram matching , Images were obtained after matching and then after matching synthetic RGB images, The composited image instead of Panchromatic image HIS transform, the big difference between HIS reform methods and the improved IHS lies in Panchromatic image of the histogram matching. The following shows the steps in fig.2.

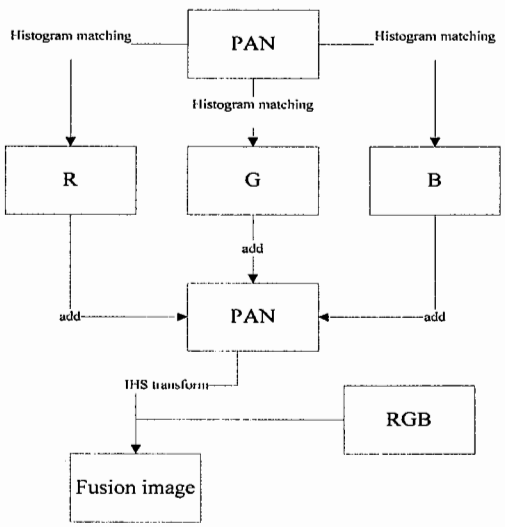


Fig.2 The improved IHS transform fusion

4. EXPERIMENT

To assess the quality of the fused images, fusion results from different techniques should be evaluated visually, spectrally, and spatially Change in spectral characteristics is an indication of the change in radiometric content of the image. Minimizing distortion of the spectral characteristics is important since this ensures that features spectrally separable in the original dataset are still separable in the fused dataset [2,9]. The following pictures can show the results of this experiment. The testing image data consist of an quick-bird data in some area.

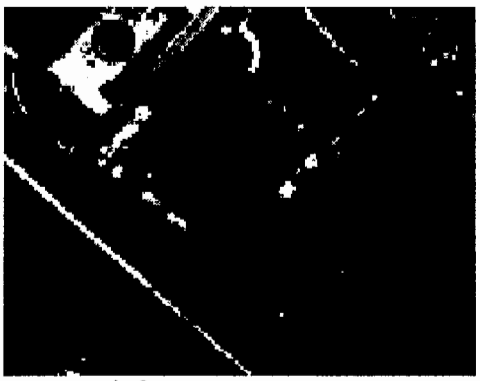


Fig.3 The multispectral image

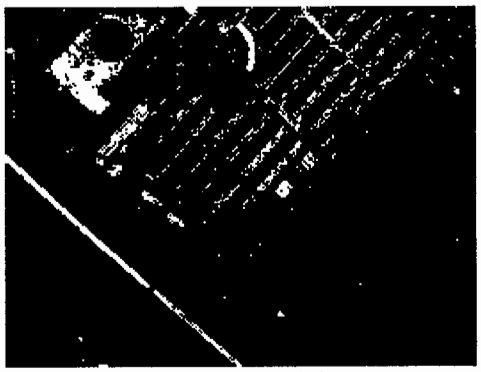


Fig.4 The panchromatic image

Fig.5 The improved HIS transform

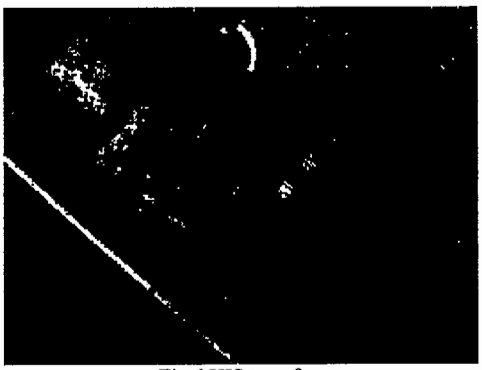


Fig.6 HIS transform

In the IHS fusion results, Compared to Figs. 5 and 6 it can also be seen that the color distortions of the two IHS models. The color of highly reflecting building roofs and other areas are changed from white, in the original quick-bird color image, to purple, in the fusion results while it is distorted from white to light blue, in fusion result, The color of the vegetation areas is also distorted in different directions. In the fusion results, building, corners and other sharp edges can be seen as clear as in the original panchromatic images. This indicates the color distort qualities of the fusion techniques being tested are valuable.

5. STATISTICS

As we all know the Correlation coefficients is one of the most important data which indicates the quality of the fusion images. It can reflect the similarity of the spectral features of the two images. The following is the formula of Correlation coefficients:[12]

$$\rho_{FA} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[F(x_i,y_j)-\bar{f}][A(x_i,y_j)-\bar{a}]}{\sqrt{\sum_{i}^{M}\sum_{j}^{N}[F(x_i,y_j)-\bar{f}]^2[A[x_i,y_j]-\bar{a}]^2}} \tag{6}$$

Where f ,a respectively represent the average value of the fusion image and source image . The correlation coefficients greater, the more fusion images obtained from the source image information and the better result of the fusion.

Table3 The IHS fusion image correlation coefficients evaluation form

correlation coefficients	Band 1	Band 2	Band 3
Band 1	1.0000	0.9757	0.9070
Band 2	0.9757	1.0000	0.9496
Band 3	0.9070	0.9496	1.0000

Table4 The improved HIS fusion image correlation coefficients evaluation form

correlation coefficients	Band 1	Band 2	Band 3
Band 1	1.0000	0.9814	0.9254
Band 2	0.9814	1.0000	0.9570
Band 3	0.9254	0.9570	1.0000

As can be seen from the Table3 and Table4: The correlation coefficient of the improved HIS fusion is bigger than the traditional one. Its spectral degradate smaller, and is more close to the original multi-spectral spectral information. The amount of information of fusion image is more abundant than the original image, and greatly improve the clarity.

6. CONCLUSION

To reduce the color distortion and improve the fusion quality, an improved IHS transform fusion approach is proposed. This approach utilizes the IHS transform to fuse high-resolution spatial information into the low-resolution multispectral images, and uses the histogram matching to reduce the color distortion, in the way of generating a new high resolution *panchromatic* image that highly reduce the color distort. In other words, the results have proved that the concept of the proposed improved HIS transform is promising.

7. REFERENCES

[1] Yun Zhang, Gang Hong.An IHS and wavelet integrated approach to improve pan-sharpening visual quality of natural color IKONOS and Quick-Bird images, Information Fusion 225–234(2005).

[2] Ling Yangrong,Manfred Ehlers, E.Lynn Usery,Marguerite Madden. FFT-enhanced IHS transform method for fusinghigh-resolution satellite images, ISPRS Journal of Photogrammetry & Remote Sensing 381-392(2007).

[3] Lillesand, T.M., Kiefer, R.W., Chipman, J.W. Remote Sensing and Image Interpretation, 5th edition. John Wiley & Sons.

[4] DONG Guangjun , ZHANG Yongsheng , DAI Chengguang. Comparison of fusion algorithms of high resolution imagery. O ptical Technique 32(6) 827-830(2006).

[5] S.K. Pal. T.J. Majumdar, Amit K. Bhattacharya,. ERS-2 SAR and IRS-1C LISS III data fusion: A PCA approach to improve remote sensing based eological interpretation. ISPRS Journal of Photogrammetry & Remote Sensing 61 281-297(2007).

[6] Gonzalo Pajares , Jesus Manuel de la Cruz.A wavelet-based image fusion tutorial. Pattern Recognition 37 1855-1872(2004).

[7] ZHAO Shuhe,ZHANG Xinming ,QU Hongjian. Fusion of multispectral remote sensing image and high resolution spatial panchromatic image .Journal of Geomatics 29 (5) 4-5 (2004).

[8] YANG Xiaohui, JIAO Licheng. Fusion Algorithm for Remote Sensing Images Based on Nonsubsampled Contourlet Transform Acta Automatica Sinica 34(3) 274-281(2.008).

[9] Saraf, A.K.. IRS-1C-LISS-III and PAN data fusion: an approach to improve remote sensing based mapping techniques. International Journal of Remote Sensing 20 (10) 1929–1934(1999).

[10] Simone, G., Farina, A., Morabito, F.C., Serpico, S.B., Bruzzone, L. Image fusion techniques for remote sensing applications. Information Fusion 3 3–15(2002).

[11] Wen, C.Y., Chen, J.K. Multi-resolution image fusion technique and its application to forensic science. Forensic Science International 140, 217–232(2004).

[12] TAN Yongsheng, SHEN Zhangquan, J IA Chunyan,WANG Xinhui, DENG Jinsong, The Study on Image Fusion for Medium and High Spatial Resolution .Remote Sensing Images .Remote Sensing Technology Application 22(4) 536-541(2007).

[13] WANG Wenjun, QIN Qiming, CHEN Sijin, FU Wei. A New RS Image Fusion Method Based on Feature. Remote Sensing Technology Application 19(4) 258-261(2004).

Image enhancement by Curvelet, Ridgelet & Wavelet transform.

Vinay Mishra (Nokia Siemens Networks, India)

Pallavi Parlewar (R. K. N. E. C., Nagpur, India)

ABSTRACT

Image Processing always aims at extracting maximum information from an image. To achieve this we have to analyze the image completely along its periphery. But the parts of an image are hardly straight, they contain continuously varying slopes. Wavelet based image processing gives low resolution when the image has largely varying slopes and they give redundant coefficients. If we tile the whole image, we get curve-lets meaning 'small curves'. If this tilling is optimum, we get parts of the curve which resemble to the straight lines. These straight lines are then analyzed and reconstructed using 'Curvelet Transform'. Curvelet Transform represents edges of a curve better than Wavelet Transform. This transform uses 'Ridgelet Transform' as a main processing. Ridgelet Transform is a two step process using Radon Transform and DWT. Radon transform analysis involves the mapping of rectangular coordinates into the polar or angular coordinates. With the increasing need for higher speed and lower memory requirement, we, in this paper propose to compute the Ridgelet coefficients without involving the conversion to angular coordinates. We have used Radon transforms our basic building block. As it will be seen taking 1-D DWT on Radon Transform results in Ridgelet Transform. At the end of the paper the images having many 'ridges', our transform gives better PSNR than Wavelet transform and many others. It also saves computational time by using fast FFT algorithm and avoiding operating on Tiles having less variation of pixels. The PSNR also depends on the algorithm used to perform DWT.

Curvelet, Ridgelet, Wavelet, modulus, filtering, downsampling.

1. INTRODUCTION

The need for good picture quality is increasing day by day. It may be a satellite image or even an internet picture; it should be a denoised one. Time required to process an image is also a factor of tradeoff. The earlier used transforms are therefore becoming redundant now. The recently used Curvelet transform appears to be promising and is an area of research now. We in this paper are presenting the optimum way to apply Ridgelet transform & the comparing analysis of Ridgelet and Wavelet transform. The results of Ridgelet transform have shown a high denoising and enhancing capability. The technique for minimizing the calculation lies in identifying the tiles with maximum information. (Refer the program results at last)

2. CURVELET TRANSFORM

Curvelet Transform is an extension of the most commonly used transform, Wavelet transform. Since Wavelet operates on point singularities, the enhancement of curves is not possible with Wavelet transform. Thus the Curvelet transform comes into picture. Curvelets are smooth, highly anisotropic elements ideally suited for detecting and synthesizing curved edges. This transform uses Wavelet and radon transform as we will see. The result of Curvelet transform is still under improving stage. Basically, Curvelet transform consists of four parts:

a) Subband Decomposition
b) Smooth Partitioning
c) Renormalization
d) Ridgelet Synthesis

In *Subband Decomposition*, various sets of Bandpass filters are used on the image. Each filter will result into decomposition of the curve into fine-scale subband. *Smooth Partitioning* of each of the subband into `squares' will result

v.mishra21@rediffmail.com
pallavi_parlewar@rediffmail.com

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754636 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853817

in either an `empty square' or a ridge fragment. Such nearly straight ridge fragments are precisely the desired input for the Ridgelet transform after renormalizing the image pixels intensities into the desired range.

Ridgelet Synthesis:

There are several transform that are available for processing on the one dimension (pixels) and two dimension (lines or curves). In our program we have used various transforms that process the images. They are:

1) Radon Transform
2) Wavelet Transform

2.1 RADON TRANSFORM

The Finite Radon transform (FRAT) is defined as summations of image pixels over a certain set of "lines". Those lines are defined in a finite geometry in a same way as the lines for the continuous Radon transform in the Euclidean geometry denote Zp = {0, 1... p-1}, where p is a prime number. The FRAT of a real function f on the finite grid Z^2p is defined as

$$rk[l]=FRATf(k,l)\frac{1}{\sqrt{p}}\sum f[i.j] \qquad \ldots\ldots 1$$

Here $L_{k,l}$ denotes the set of points that make up a line on the lattice $Z^2{}_p$, or more precisely

$$Lk,1=\{(i,j): j=ki+l(\text{mod } p), i\in Zp\}, 0\le k<p \qquad \ldots\ldots 2$$

$$L_{p,l}=\{i,j): j\in Z_p\}$$

For this we used finite field in which the Radon Transform is computed on the Finite Field (which means finite length of signals). A finite field is a set of finite number of continuous numerals in which the length of set is equal to a prime number. The real numbers are not used because they are of infinite length. Any operation done on real number is equivalent to the same operation done on finite field set of numbers. Moreover, finite field has all the properties identical to the real numbers.

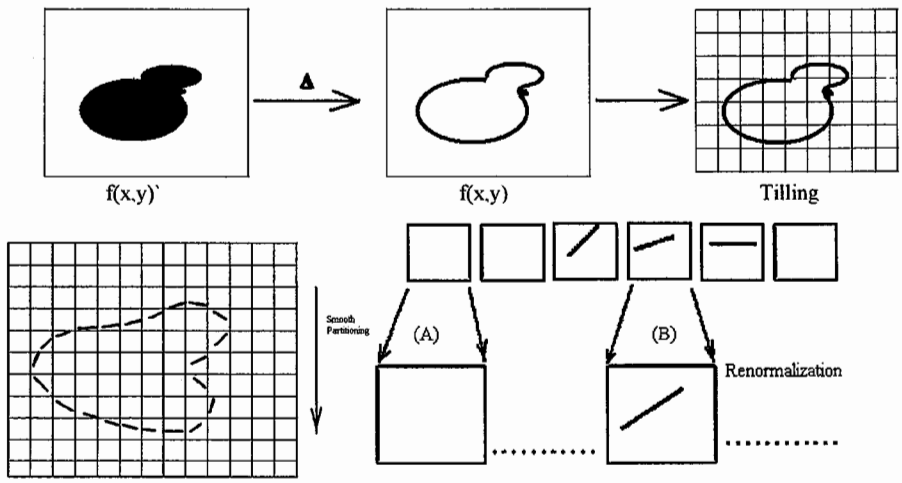


Fig. 1. Basic steps for image processing

2.1.1 Forward radon transform (FRAT).

We observe that in the FRAT domain, the energy is best compacted if the mean is subtracted from the image f [i, j] prior to taking the transform given in (1), which is assumed in the sequel. We also introduce the factorZ_p in order to normalize the image between the input and output of the FRAT.

Just as in the Euclidean geometry, a line on the affine plane is uniquely represented by its slope or direction (k = p corresponds to infinite slope or vertical lines) and its intercept. One can verify that there are lines defined in this way and every line contains p points. Moreover, any two distinct points on belong to just one line. Also, two lines of different slopes intersect at exactly one point. For any given slope, there are p parallel lines that provide a complete cover of the plane. This means that for an input image f[i,j] with zero-mean, we have

$$\sum_{l=0}^{p-1} rk[l] = \frac{1}{\sqrt{p}} \sum f[i,j] = 0 \qquad k \in Z_p \qquad \text{.....3}$$

This explicitly reveals the redundancy of the FRAT: in each direction, there are only p - 1 independent FRAT coefficients. Those coefficients at p+1 direction together with the mean value make up totally of

$$(p+1)(p-1)+1 = p^2$$

independent coefficients (or degrees of freedom) in the Finite Radon domain, as expected. By analogy with the continuous case, the Finite back projection (FBP) operator is defined as the sum of Radon coefficients of all the lines that go through a given point, that is

$$FBP_r(i,j) = \frac{1}{\sqrt{p}} \sum rk[l] \qquad (i,j) \in Z^2{}_p \qquad \text{.....4}$$

Where $P_{i,j}$denotes the set of indices of all the lines that go through a point $(i,j) \in Z^2{}_p$ More specifically, using (2) we can write

$$P_{i,j} = \{k,l) : l = j - ki \,(\mathrm{mod}\, p), \qquad k \in Z_p \qquad \text{.....5}$$

The below figure (Fig 1) shows a 7X7 tile as an example, & calculation of Radon coefficient for slope zero. For slopes other than zero, wrap-around effect takes place due to modulus operation as shown in (Fig.3)

2.2 To find radon coefficients:

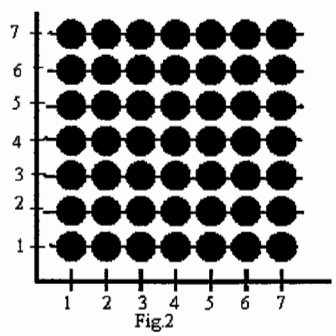


Fig.2

Fig.2 j=ki+l, k is the slope =0 & l is intercept varying from 0 to 6

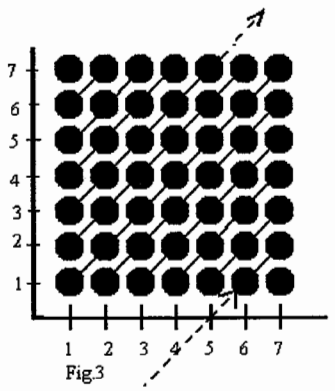


Fig.3

Fig.3 Due to modulus operation periodicity of line takes place & I(5,7) becomes I(5,0), also I(6,8) becomes I(6,1)

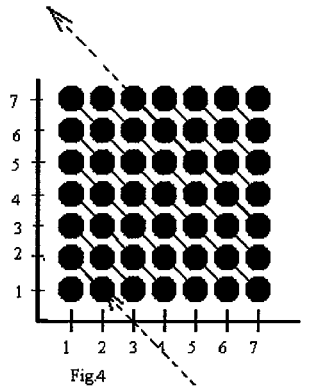


Fig.4

Fig.4 Due to modulus operation periodicity of line takes place & I(1,7) become I(1,0), also I (0,8) becomes I(0,1)

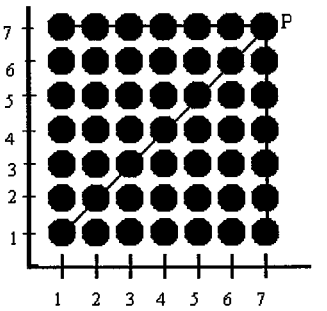


Fig.5 Calculation of FBP at point P

After plotting all 49 lines FFBP is calculated as summation of all radon coefficient passing through a given point as seen above in Fig.4. From the property of the finite geometry Z_p^2 that every two points lie on exactly one line, it follows that every point in Z_p^2 lies on exactly one line from the set $P_{i,j}$, except for the point (i, j) which lies on all p + 1 lines. Thus substituting equation 1 in equation 4 we get

$$FB\,\Pr(i,j)=\frac{1}{p}\sum\sum f[i',j'] \qquad \text{.....6}$$

The above summation is over the line j = k i + l (mod p).

$$=\frac{1}{p}(\Sigma f[i',j']+p.f[i,j]) \qquad \text{.....7}$$

= f[i,j], where f [i,j] is an input image.

So the back-projection operator defined in indeed computes the inverse FRAT for zero-mean images. Therefore we have an efficient and exact reconstruction algorithm for the FRAT. Furthermore, since the FBP operator is the ad joint of the FRAT operator, the algorithm for the inverse of FRAT has the same structure and is symmetric with the algorithm for the forward transform.

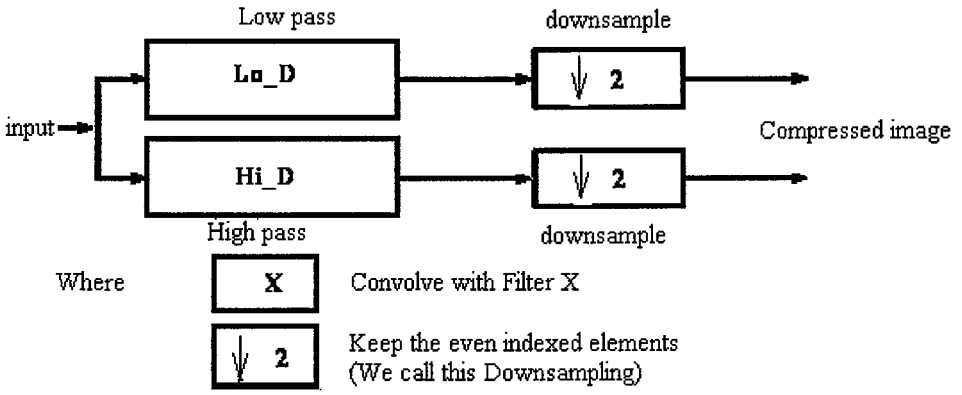
3. MULTI RESOLUTION ANALYSIS:

3.1 DISCRETE WAVELET TRANSFORM:

Although the discretized continuous Wavelet transform enables the computation of the continuous Wavelet transform by computers, it is not a true discrete transform. As a matter of fact, the Wavelet series is simply a sampled version of the CWT, and the information it provides is highly redundant as far as the reconstruction of the signal is concerned. This redundancy requires a significant amount of computation time and resources. The discrete Wavelet transform (DWT), on the other hand, provides sufficient information both for analysis and synthesis of the original signal, with a significant reduction in the computation time. The Wavelet series expansion maps a function of continuous variable into a sequence of variable.

3.1.1 STEPS IN DISCRETE WAVELET TRANSFORM:

Suppose the size of the image is 256 X 256 (DWT) Referring Fig.5, we can see that the basic steps in DWT are Filtering & down sampling along rows & column. Follow all the steps in reverse order to calculate Inverse DISCRETE WAVELET TRANSFORM (IDWT).



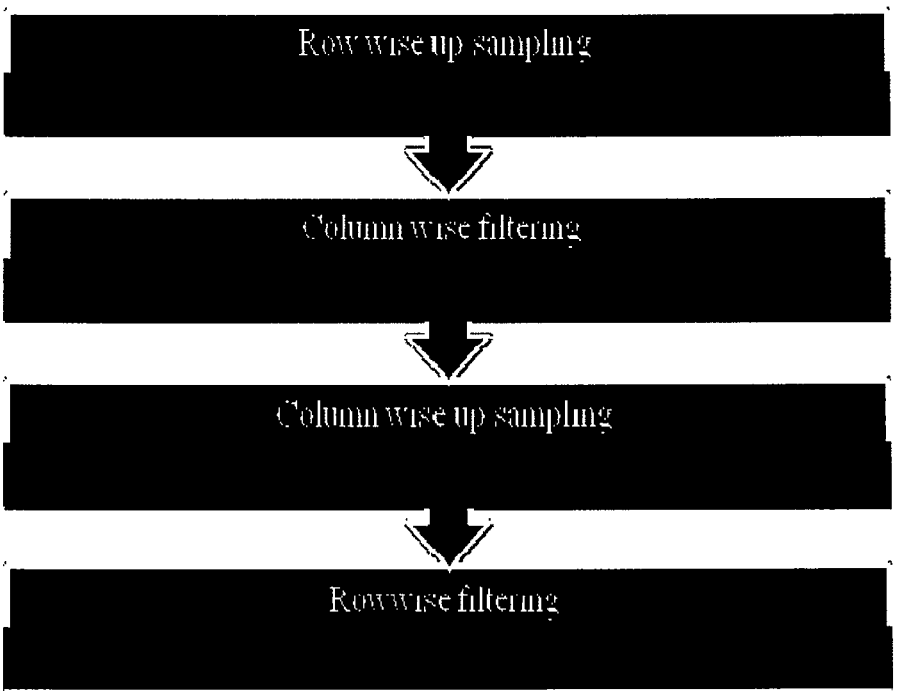


Fig.6 Algorithm (As seen, the size of the image is compressed to 128X128)

4. ALGORITHM FOR RIDGELET TRANSFORM

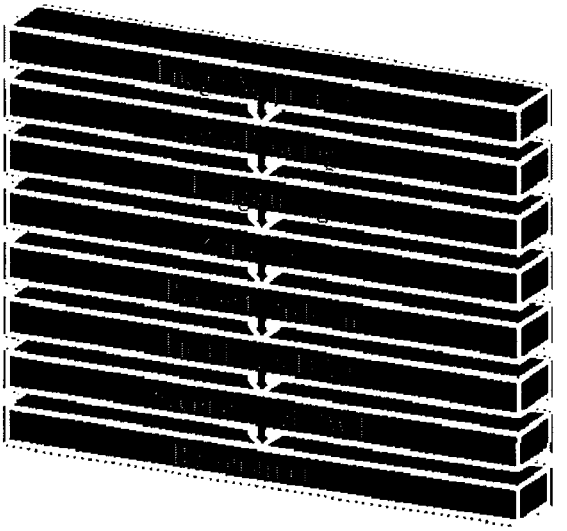
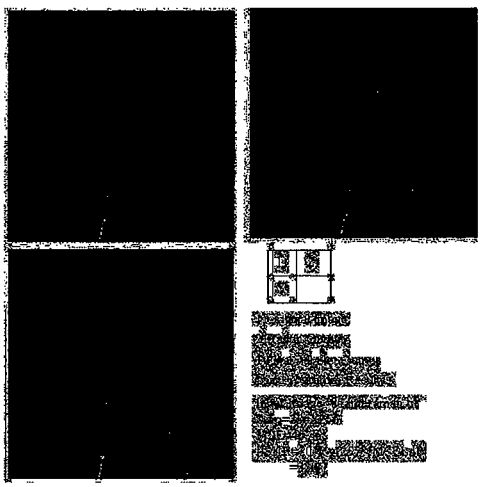


Fig.7

5. RESULT:



References

[1] Rafael Gonzalez,[Book on Digital Image Processing using MATLAB]

[2] Candes J. and Donho L.,"The Curvelet Transform for Image Denoising" ,2002

[3] Greval G.S [A Book on Mathematics]

[4] Proakis ,[Book on Digital Signal Processing]

A Novel Relay Selection Algorithm in Wireless Cooperative Networks Based on PSO

Allam Maalla, Chen Wei
School of Information Engineering, Wuhan University of Technology, Wuhan 430070.Hubei, China
Email:allammaalla@yahoo.com

ABSTRACT

Cooperative networks have been appreciated for their spatial diversity benefits in wireless communications. In this paper a novel scheme of cooperative networks depending on the number and locations of relays in the network. The effect of relay number and locations are investigated by considering energy optimization. First selects the optimal relay from a set of available relays and then uses this "optimal" relay for cooperation between the source and the destination. The simulation-based performance analysis confirms that the cooperative relaying scheme has an advantage of diversity gain thus improving the bit error ratio performance. The simulation results demonstrate that the proposed cooperative relay node selection algorithm can improve performances by achieving the cooperative gain.

Keywords- Cooperative Relay Node Selection, distributed relay selection; wireless multi-hop networks

1. INTRODUCTION

Cooperative relaying is a promising technique for multihop wireless networks to exploit spatial diversity. In addition, using relays in cooperative wireless networks can potentially lead to significant capacity increases, which is a critical task of network planning and deployment to achieve an efficient and scalable network design [1].
Many researches focus on fundamental relaying problems of determining who should be relays for whom and how to perform resource allocation for these relaying schemes jointly with routing such that the system performance. While most previous relay placement related studies only consider data forwarding through traditional multi-hop relaying [2,3]. Network coding extends store-and-forward-based routing solutions and allows intermediate relay nodes to code over the incoming packets [5]. Since the received signals at the relay nodes are a superposition of transmitted signals, the simple amplify-and-forward (AF) scheme is a form of network coding at the signal level [5,6].

Although relay problem remains unsolved in terms of the optimal communication scheme, the studies above have taken significant steps in quantifying the performance gains obtained from cooperation. However, research in this subject has been mainly concentrated on information theoretic considerations, and only a few studies have focused on the system level performance of cooperation such as [7], in which Mergen *et al* study the problem of cooperative broadcast under a single-source single-destination setup with multiple levels of cooperation, where the broadcast performance is quantified by finding a signal decoding threshold above which a message is propagated in the network. Nabar *et al* [8] further evaluate the performance of cooperative schemes in the case of single source, single relay, and single destination and prove that full diversity can be obtained. In [9] Sankaranarayanan, Kramer and Mandayam further consider the case where multiple sources send their message to a relay, and the relay either simply forwards the data or first decodes and then forwards.

This paper fully exploit the performance benefits of cooperative relaying in wireless network and it shown that optimized system performance in terms of Relay selection through a heuristic approach based on Particle swarm optimization (PSO). An arbitrary number of relays and transmit energy on the error performance are considered with relay locations. Based on the proposed theoretical models and considering the computation complexity, the sub-optimal cooperative relay node selection algorithm is determined to improve wireless efficiency from the global viewpoint of heterogeneous wireless communication systems under the conditions of guaranteeing the equivalent transmission rates between the first and second hops.

The rest of this paper is organized as follows. Section II introduces the model of cooperative relay node based heterogeneous wireless communication system, and analyzes its challenging behaviors. The optimization framework for the utility-power tradeoff is presented in Section IV. Section IV presents the simulation results and demonstrates the performance gains of the proposed algorithm. The conclusions are stated in Section VI.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754637 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856319

2.Related Work

The relay selection algorithms for larger wireless networks are complex and difficult to be solved, in which the calculation of multi-hop route jointly with cooperative relay node selection remains an interesting and challenging open problem. The key capacity results for the case of a single relay were introduced by Cover and El Gamal in [10]. The capacity region for the relay channel with M relays is not known to date. However, Gastpar and Vetterli [11] have obtained upper and lower bounds on the capacity scaling under Gaussian noise and show that these bounds meet in the limit of large number of nodes. The existing work on the resource allocation has been done in terms of optimal energy, location, and joint energy-location for the system resource [12, 13, 14]. Analysis and optimization of wireless systems using different scheduling policies were conducted in [15]. Cross-layer design and optimization for emerging multihop networks such as wireless mesh networks [16]. Among these optimizations, it has been shown that the location of relay is critical to determine the system performance. Furthermore, the location information may be useful for higher layer applications such as location based routing. Hou *et al.* [17] formulated the joint energy provisioning and relay node placement problem into an MINLP, and a heuristic was developed to solve it. Numerous research works from the information theory respective have focused on the design of cooperation protocols and analysis. The capacity bound in wireless relay channel was investigated in [18]. The achievable rate formula and a coding scheme for the multiple-level relay channel was presented in [19].

3.Wireless Network Model and Cooperative Scheme

In this section, we explain the network model under consideration and present a generic relaying architecture.

A. SYSTEM MODEL

As shown in Fig.1, in our wireless network, communication between a source and a destination terminal is facilitated by direct graph $G = (V,L)$, where V is the set of nodes and L is the set of directed links. There is a link between a pair of nodes if the communication link can be established between them. With utilizing the cooperative technique in the heterogeneous relay nodes, the high transmission data rate and the extended coverage can be sufficed, which will achieve the performances of seamless convergence between the heterogeneous wireless access networks. The number n (≥ 0) of the relays which can successfully decode the received symbol is a random variable, and n has a special distribution such as Poisson distribution, normal distribution, and so on. The proposed system model is depicted in Fig.1.

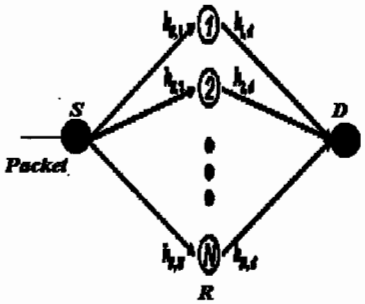


Fig.1 Cooperative System with Random Relay Model

B. COOPERATIVE SCHEME

Under the proposed scheme, it is shown in Fig.1 that the multihop cooperative relaying networks. Signals transmitted by the different relay station are delayed copies of the same signal. The received signal at the destination is demodulated using estimation method. The source station transmits the data streams $s(k)$.

The received signal y_D at the destination station is given by

$$yD = \sum_{i=1,2} h_{RiD}\sqrt{E_{Ri}^t}\acute{s}(k-i+1) \qquad \text{.....(1)}$$

where h_{R_1D}, h_{R_2D} represent complex channel coefficients between the destination station and the relay stations R_1, R_2, respectively. E_{R1}^t, E_{R2}^t denote the transmission power of R_1, R_2, respectively.

It means that each relay station retransmits the received packet only if it is correctly decoded. The transmission power and noise performance of each station are assumed to be equal. The relay route and timing synchronization are assumed to be established in advance.

Let the transmission power for direct transmission on link (i, j) be $P_d(i, j)$, and the total interference at the receiving side of this link be $N_0(i, j)$. Then, the achieved rate (b/s/Hz) for direct transmission is

$$r_d(i,j) = log_2\big(1 + g(i,j)P_d(i,j)\big) \quad \text{.....(2)}$$

The achieved data rate at relay node k in the first time slot and at node j in the second time slot after using maximum ratio combining are given, respectively, by

$$r_r(i,j) = log_2\big(1 + g(i,k)P_d(i,j)\big) \quad \text{.....(3)}$$

$$r_c(i,j) = log_2\left(1 + g(i,j)P_d(i,j) + g(k,j)P_{r,(i,j)}(k,j)\right)$$

Note that for relaying to be useful, the achieved rate at relay node k must be higher than that due to direct transmission. Let the maximum achievable data rate on link (i,j) be $r(i,j)$. This achievable rate on each link depends on the transmission power, link gain, and the transmission strategy, the achievable transmission rate on link (i,j) for different transmission strategies can be written as

$$r(i,j) = \begin{cases} r_d(i,j) & for\ dirct\ transmission \\ r_c(i,j) = r_c(i,j) & for\ cooperative\ transmission \end{cases} \quad (4)$$

III. Proposed Algorithm

For this problem, the source rates S_i are assumed to be fixed and the optimization problem, in essence, is the joint routing, relay selection, and power allocation in a multihop wireless network. The goals are to find optimal transmission strategy for each link (i.e., either direct or cooperative transmission), the optimal power allocation for the chosen strategy, and the link flow to route data generated by source nodes to the corresponding destination node. This is a cross-layer design problem for both physical layer (i.e., relay selection, power allocation) and network layer (i.e., routing of traffic flows). The problem can be stated as follows:

$$\begin{aligned} &minimize && \textstyle\sum_{(i,j)\in L} P_t(i,j) \\ &subject\ to && \textstyle\sum_{j\in O(i)} x(i,j) - \sum_{j\in I(i)} x(i,j) = S_i, i \in V \quad \text{.....(5)} \\ & && x(i,j) \le r(i,j), (i,j) \in L \end{aligned}$$

$$x \ge o, P_{min} \le P \le P_{max}$$

Where $\le, \ge$ denote component-wise inequalities, $\boldsymbol{x}$ is the vector of link flows, P_{min} (P_{max}) denotes the lower (upper) limit for the power vector with element $P_{min}(i,j)$ ($P_{max}(i,j)$) being the minimum (maximum) allowed power to transmit or relay packets on link (i,j). We assume that this problem is feasible. The feasibility is assumed to be maintained by an appropriate admission control algorithm.

PSO has been successfully used to perform a wide range of optimization problems [10]. Each particle in the swarm represents a multicast tree. The population of permutation-based chromosomes $\{c_p = [c_p^1, \ldots\ldots, c_p^N], p = 1, \ldots., P\}$ is initialized by employing a hybrid of random and deterministic approaches, where P is known as the population size. In order to produce the deterministic solution, the MPM algorithm is run first and the corresponding node subsequently added to the broadcast tree is recorded into a node sequence, which serves as the deterministic population for the proposed PSO approach. Based on the node sequence, the construction of the broadcast tree can be carried out using PSO algorithm as follows:

Input: source node s and chromosome $c_p = [c_p^1, \ldots\ldots, c_p^N]$;

Output: broadcast tree $\boldsymbol{T}$ and the corresponding consumption power P_T. For clarity, the flow of these operations is illustrated in Fig.2.

1) Initialize P_T=0, mark the source node s in the permutation c_p to avoid the subsequent selections;

2) Obtain the first unmarked node z in permutation c_p, construct a link between source nodes s and z, and assign the required power p_{sz} to the source node;

3) Check all the remaining unmarked nodes in the permutation to see if any other node is already covered by the above transmission between nodes s and z, mark those nodes to avoid the subsequent selections;

4) Go through the unmarked node z' in permutation sequentially. Employ the rules of the minimum power multicase algorithm to add node z' to the existing tree and assign a power $p_{s'z'}$ to the corresponding source node (assuming it is s');

5) Check all the remaining unmarked nodes in the permutation to see if any other node is covered by the above transmission between nodes s' and z', mark those nodes to avoid subsequent selections;

6) If not end of permutation, go to 5;

7) Calculate P_T by eq.(5), and form a Multicast tree $\boldsymbol{T}$ in which all constructed links correspond to the edges in $\boldsymbol{T}$.

4.Numerical Analysis And Simulation Results

In this section, present numerical results based on the proposed analytical framework and compare them with the simulation results by consider the effect of relay locations and numbers with in three scenarios:

First scenario: the convergence of the simulation and proposed algorithm are evaluated by assume that all relay nodes located as shown in Fig.3;
Second scenario: the performance gain of with difference location and fix no. of relays are verified.
Thrid scenario: the performance gain of with Fix location and different no. of relays are verified.
First scenario: To show the convergence of the proposed algorithms, a wireless network with 25 relays distributed in an area of 200m x 200m. Two topologies was proposed namely, grid and random topology (Fig.3). For the both topology, Fig.3a shows total 25 nodes fix at four corners and one node at the center of the area; and Fig.3b shows total 25 nodes fix as follows 20 nodes are positioned randomly with 5 nodes in each area of 100m × 100m and power limits on each link (i, j) are $P_{\min}(i, j) = 0$ mW, $P_{\max}(i, j) = 50$ mW for transmissions from both source nodes and relay nodes.

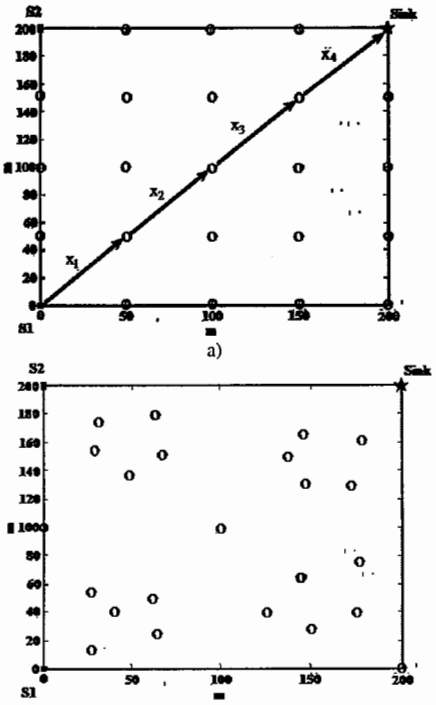


Fig.3. Distribution Topology 25 relays and routing path from a source relay to the destination node a) Grid topology b) Random topology.

The total power consumed versus the source rate for both the network topologies with and without cooperative diversity implementation is shown in Fig.4. For the random topology actually requires a bit higher power consumption than the grid topology for these particular source-sink pairs. This performance gain is achieved without sacrificing the distributed nature of the proposed algorithm because only local information is needed to search for the best relay together with the optimal power allocation.

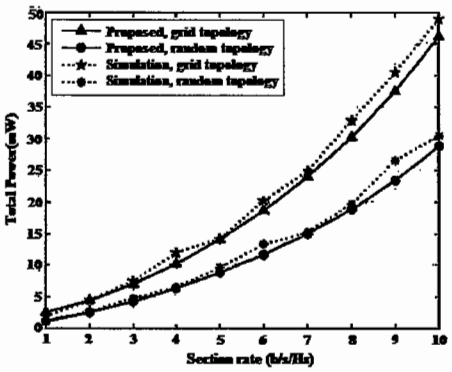


Fig.4 Total power consumption same source rate with grid and random.

Second scenario: the wireless network performance of ten cases with 20 relays and different allocation are verified. The Figs.5 indicate that with same transmit energy deciding packet movement depends on only the node allocation where this packet moves towards the source node. Therefore, a cost-effective of relay allocation can be obtained using our proposed formulation based on optimum location.

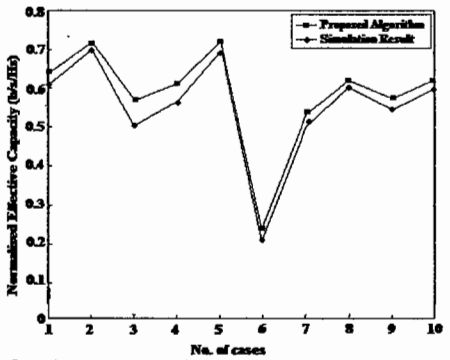


Fig.5.Performance comparison of ten different allocation cases.

Third scenario: the wireless network performance of five cases with different (12, 16, 20, 24, 28 relays) and gird node distribution are verified. Fig. 6 shows the maximum network sum-rate is depicted for different number of relays within the network. It can be observed that increasing the number of relays leads to sum rate increase as expected. In addition, having more nodes in the network results in more sources and, therefore, more cooperative regions to be scheduled during each transmission.

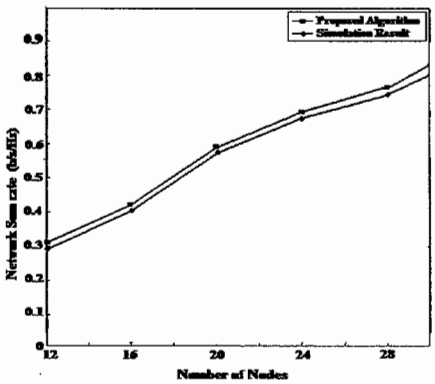


Fig. 6 Network sum rate vs. different number of relays

5.CONCLUSIONS

By applying heterogeneous cooperative relay network based cooperative relay node selection and PSO algorithm can successfully decode the received symbol is a random variable. The theoretical optimization model and proposed algorithm are not only available for the heterogeneous relay based wireless networks, but efficient for the homogeneous relay based wireless networks. Numerical results based on the proposed analysis provide design guidelines for optimal relaying that they can improve the system performance dramatically. Further work should consider the cooperative scheduling algorithms and the mechanisms of QoS guaranteeing.

REFERENCES

[1] J. N. Laneman and G. W. Wornell, "Energy-efficient antenna sharing and relaying for wireless networks," in *Proc. of WCNC*, vol. 1, 2000, pp. 7–12

[2] A. So and B. Liang, "An efficient algorithm for the optimal placement of wireless extension points in rectilineal wireless local area networks," in *Proc. QShine'05*, pp. 25-33, Orlando, Florida, Aug. 2005.

[3] Y. T. Hou, Y. Shi, H. D. Sherali and S. F. Midkiff, "On energy provisioning and relay node placement for wireless sensor networks," *IEEE Trans. on Wireless Commu.*, vol. 4, no. 5, pp. 2579-2590, Sep. 2005.

[4] R. Ahlswede, N. Cai, S.-Y. R. Li, and R. W. Yeung, "Network Information Flow," IEEE Trans. Inform. Theory, vol. 46, no. 4, pp. 1204–1216,Apr. 2000.

[5] S. Katti, S. Gollakota, and D. Katabi, "Embracing Wireless Interference: Analog Network Coding," in ACM SIGCOMM, Kyoto, Aug. 2007.

[6] S. Zhang, S. C. Liew, and P. P. Lam, "Hot Topic: Physical-Layer Network Coding," in Proc. ACM MobiCom, Los Angeles, 2006.

[7] B. S. Mergen, A. Scaglione, and G. Mergen, "Asymptotic analysis of multi-stage cooperative broadcast in wireless networks," IEEE Trans. Inform. Theory, vol. 52, pp. 2531–2550, June 2006.

[8] R. U. Nabar, H. B¨olcskei, and F.W. Kneub¨uhler, "Fading relay channels: Performance limits and spacetime signal design," IEEE J. Select. Areas Commun., vol. 22, pp. 1099–1109, Aug. 2004.

[9] L. Sankaranarayanan, G. Kramer, and N. B. Mandayam, "Hierarchical sensor networks: capacity bounds and cooperative strategies using the multiple-access relay channel model," in Proc. IEEE SECON, pp. 191–199, Oct. 2004.

[10] T. M. Cover and A. A. Elgamal, "Capacity theorems for the relay channel," IEEE Trans. Inform. Theory, vol. 25, pp. 572–574, Sept. 1979.

[11] M. Gastpar and M. Vetterli, "The capacity of large Gaussian relay networks," IEEE Trans. Inform. Theory, vol. 51, pp. 765–779, Mar. 2005.

[12] R. Cao and L. Yang, "Practical issues in resource optimization of relay networks," in Proc. of CISS, 2008.

[13] W. Cho, R. Cao, and L. Yang, "Optimum resource allocation for amplify and- forward relay networks with differential modulation ," IEEE Trans. on Signal Processing, vol. 56, no. 11, pp. 5680–5691, Nov. 2008.

[14] W. Cho and L. Yang, "Optimum resource allocation for relay networks with differential modulation," IEEE Trans. on Communications, vol. 56, no. 4, pp. 531–534, Apr. 2008.

[15] L. B. Le, E. Hossain, and A. S. Alfa, "Delay statistics and throughput performance for multi-rate wireless networks under multiuser diversity,"*IEEE Trans. Wireless Commun.,* vol. 5, no. 11, pp. 3234–3243, Nov.2006.

[16] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks" IEEE Commun. Mag., vol. 43, no. 9, pp. 23–30, Sept. 2005.

[17] A. Host-Madsen and J. Zhang, "Capacity bounds and power allocation for wireless relay channel," IEEE Trans. on Inf. Theory, vol. 51, no. 6, pp. 2020-2040, June 2005.

[18] L.-L. Xie and P. R. Kumar, "An achievable rate for the multiple-level relay channel," IEEE Trans. on Inf. Theory, vol. 51, no. 4, pp. 1348-1358, Apr. 2005.

[19] R. Bose, "A smart technique for base-station locations in an urban environment," IEEE Trans. on Vehi. Tech., vol. 50, no. 1, pp. 43-47, Jan. 2001.

A Study of Image Encryption Aritlunetic Based on Chaotic Sequences

Huang Xiaolong
(Department of Mathematics & Computer Information Engineering, Baise University, Baise, Guangxi 533000, China)

ABSTRACT

The multimedia information, especially video and audio information, regarded as a common data stream, with the traditional encryption technology encrypted, ignoring the characteristics of multimedia data, has some limitations. On the other hand, chaotic sequences have several good properties including the ease of their generation, their sensitive dependence on their initial condition and so on. Therefore, this paper discussed image encryption arithmetic on the basis of chaotic sequences through dispersing the real number value chaotic sequences into symbol matrix and transformation matrix, and then encrypted the image. Preliminary results proved that the image encryption arithmetic based on chaotic sequences possesses the traits, namely fast speed for encryption speed, perfect results for encryption.

Key words: chaos, chaotic sequences, array transformation, encryption

1. INTRODUCTION

With the Internet technology and the rapid development of multimedia technology, multimedia communications have become an important means of information exchange among people. Through a variety of networks, people exchange information, conduct online trade, voice dialogue and online video. Digital image has the advantages of visual, intuitive and vivid as one of the most important forms of message expression in the multimedia message. In particular, after overcoming the past problem due to large amount of image data storage and transmission, it is becoming or has become the mainstream of message expression forms at the present time or even in the future. Digital image can facilitate people's lives. Meanwhile, it also has many security risks. Image encryption becomes the core technology of image security protecting. It is a direct and effective means to protect the image security. Therefore, as multimedia technology has been increasingly used today, it is important theoretically and practically to study the image encryption.

2. Related research and analysis

With the common action of the encryption key and encryption functions, image encryption changes an image into a chaotic encrypted image which is similar to noise so as to make the true image information not be intuitively felt. At present, the research on the image encryption of domestic and foreign experts or scholars mainly focuses on the following aspects:

i) Spatial domain image encryption [8,9]. There are mainly two ways of encryption: one is by changing the position relationship between the image pixels, which is called image encryption based on scrambling image. The other is by using a certain encryption rule, changing the pixel value, and smoothing the encrypted image histogram, making the information entropy close to the maximum. Such encryption is called image encryption based on information entropy. In this aspect, Ma Zaiguang [8], Chen Guanrong[10] and other experts obtained two-dimensional and three-dimensional transformation which apply for digital image encryption by dispersing the Arnold transformation. After repeated transformation image has become disorganized and got a more satisfactory result of scrambling, this kind of encryption has been widely used in image encryption, information hiding and digital watermarking.

ii) Transform domain image encryption. Image relative to the text has the features of a large amount of data, redundancy and strong correlation. For the need of storage and transmission, encoding processing is often made to an image after some kind of orthogonal transform. If the image data is encrypted with simultaneous encoding, such method of image encryption is called transform domain image encryption. At present mainly is including image encryption based on tree structure, such as: binary tree, quad-tree and wavelet zero-tree and other tree structure which have been widely used [11].

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754638 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.854080

There is also an image encryption based on neural network, an image encryption based on cellular automata, an image encryption for quantum cryptography and so on.

This paper, based on the principle of chaotic systems, combined with image scrambling (rotation) transformation of innovative ideas, put forward the image on the basis of chaotic sequence scrambling algorithm, which takes full advantage of the features of the chaotic system of initial value sensitivity, parameter sensitivity and randomicity. Through Image Scrambling (rotation) transform ideas combined this algorithm, image encryption can effectively disorder images, but also has good security.

The multimedia information, especially video and audio information, regarded as a common data stream, with the traditional encryption technology encrypted, ignoring the characteristics of multimedia data, has some limitations. Image scrambling (array) transformation is a common image encryption method [1]. Some literature [2] has adopted parametric two-dimensional chaotic maps to arrange each pixel of images in the spatial domain. Arrangement is iteration, thus the number of iterations, as part of the key, has a perfect encryption effect. Literature [3] presented an image scrambling transform algorithm based on geometric computing. Literature [4] introduced the magic square based on the image scrambling (rotation) transformation. Set image $I = N \times N$, for a given magic square $A = N \times N$, then put the elements of I and A in columns and rows in correspondence. Move the element with the value of one in the A to the position of the element with the value of two, then move the element with the value of two in the A to the position of the element with the value of three ..., move n^2 to the position of the element with the value of one, and thus get a rotation of p, A is changed into A_1, $A_1 = p$ (A). Because I was with the A in correspondence, as A is converted to A_1, I converted to I_1. When the repeated use of permutation p, a series of image scrambling can be obtained. With the above-mentioned ideas and the characteristics of chaotic dynamics, this paper proposes an image encryption algorithm based on chaotic sequences.

3. Principles of mathematical description of chaotic systems

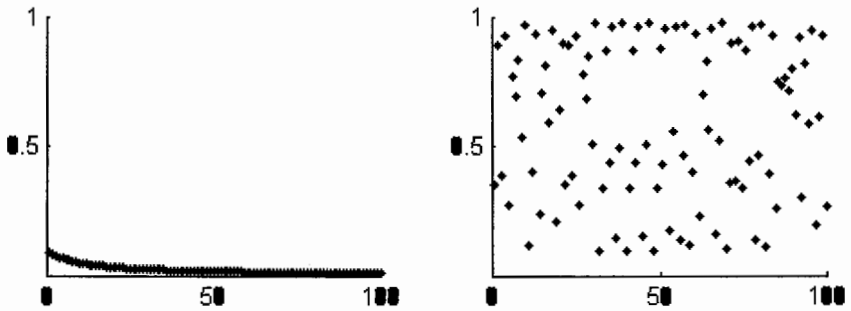
Chaos appears in a similar random process deterministically, in a nonlinear dynamic system. This process is neither a cycle nor a convergence, and depends on the initial value very sensitively.

3.1 One-dimensional Logistic Mapping

A very simple and extensively-studied dynamic system is logistic mapping, which is defined as follows.

$$x_{k+1} = \mu x_k (1 - x_k) \qquad (1)$$

Here, $0 \le \mu \le 4$ is called the branching parameter, $x_k \in (0,1)$, the definition is Ibid. Chaotic dynamical systems research pointed out that when $3.5699456 \ldots < \mu \le 4$, logistic mapping is in the chaotic state. In other words, from the initial conditions x_0 in the logistic mapping, the sequences generated under the action of (x_k; k = 0,1,2,3 ... ..) are non-cyclical, non-convergence, and are very sensitive to the initial value. We can design experiments to test these: X0 is the initial value of 0.1, and the parameter μ, is 1 and 3.8 respectively, the following comparison chart is available:



The logistic map can also be defined in the interval (-1, 1). Its form is as follows:

$$x_{k+1} = 1 - \lambda x_k^{\,2} \qquad (2)$$

Here,$\lambda \in [0, 2]$. Under the conditions that $\lambda = 2$ and full mapping, the generated PDF (probability density function) is as follows.

$$\rho(x) = \begin{cases} \dfrac{1}{\pi\sqrt{1-x^2}} & -1 < x < 1 \\ 0 & else \end{cases} \qquad (3)$$

For logistic mapping in the formula (1), if $\mu=4$, PDF can be rewritten as

$$\rho(x) = \begin{cases} \dfrac{1}{\pi\sqrt{x(1-x)}} & 0 < x < 1 \\ 0 & else \end{cases} \qquad (4)$$

By $\rho(x)$, we can easily obtain some significant statistical characteristics of chaotic sequences generated by logistic mapping [5]. For example, the average time of x is the average of the chaotic sequence trajectory point:

$$\bar{x} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} x_i = \int_0^1 x\rho(x)dx = 0 \qquad (5)$$

With regard to correlation function, select two initial values x_0 and y_0 independently, then the sequence of cross-correlation function is:

$$\begin{aligned} c(l) &= \lim_{N\to\infty} \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{x})(y_{(i+l)} - \bar{y}) \\ &= \int_0^1 \int_0^1 \rho(x,y)(x-\bar{x})(\tau^l(y) - \bar{y})dxdy = 0 \end{aligned} \qquad (6)$$

Pay attention to combining pdf $p(x,y) = p(x) * p(y)$.
The above characteristics of logistic sequence show that despite the chaotic dynamical system is deterministic which is equivalent to traverse the statistical properties of white noise, the logistic sequence has many traits: a simple form, the initial conditions and sensitivity of the statistical properties of white noise.

We can study an avalanche effect of Logistic further, design experiments: parameters μ are 3.999, the initial value is 0.663489 and 0.6634891 respectively, 22 iterations give the following data:

initial conditions: 0.663489		initial conditions: 0.6634891	
1-0.892862116169121	2-0.382541771358801	1-0.892861985410579	2-0.382542182216613
3-0.944578054533363	4-0.209349063355894	3-0.944578440505217	4-0.209347690941557
5-0.661922610078587	6-0.924292931507899	5-0.661919419720743	6-0.924296306472932
7-0.279202278007008	8-0.802980980191195	7-0.279190850810577	8-0.80296084534063
9-0.631228077313107	10-0.928788974981573	9-0.63127675750006	10-0.928737987631679
11-0.263898659587664	12-0.775082066651439	11-0.264073114466537	12-0.775410633683096
13-0.695576127861026	14-0.844882413061441	13-0.69485444158856	14-0.84600666890691
15-0.522913923432061	16-0.995405058930669	15-0.519814746437056	16-0.995933429536304
17-0.0182495720677421	18-0.0714869354964746	17-0.0161596335387537	18-0.063435014132761
19-0.264842448663692	20-0.776856694937551	19-0.237049942327802	20-0.72162049600905
21-0.69166797816834	22-0.850920910717995	21-0.801528579437324	22-0.634731257972212

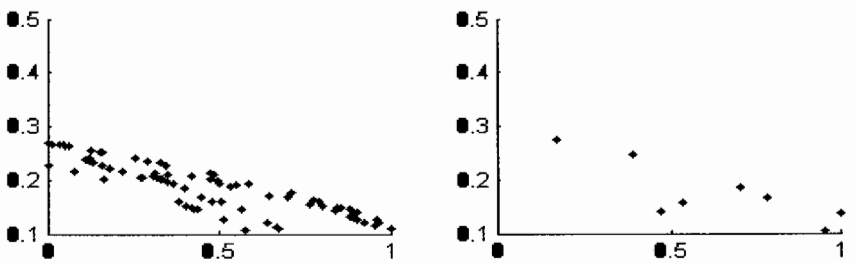
From the above experimental data, we can see, when two initial values are close, their previous N iterations of the results are very similar, until the first N +1 times, the result can only have a very different result, that is, until an avalanche occurs effect, this result is that we can not accept. For a further study, we repeated the above experiment by modifying the initial value and got large number of experimental data, from the experimental data to conclude his avalanche effect of a law that a maximum of 20 iterations, after the avalanche effect is better. Therefore, in order to get a better effect of random sequence, we can re-iterate, after 20 whichever is the result.

3.2 Two-dimensional Henon Mapping

The two-dimensional Henon mapping is an inspiration the astronomer Henon got from the studying globular clusters group, as well as from the Lorenz attractor. In 1976, Henon proposed non-linear two-dimensional chaotic systems, which has the similar nature of the one-dimensional maps, but also the existence of the similarity of the phase space. The equation is as follows:

$$\begin{cases} x(n+1) = 1 - ax(n)^2 + y(n) \\ y(n+1) = bx(n) \end{cases} \quad (7)$$

In (7), a and b are arbitrary constants. When b = 1 area preserving mapping, b <1 dissipative mapping, the parameters of 1.05 <a <1.8, b = 0.3, the working of Henon Mapping is in the chaotic state [7], while the combination of the classical parameters is a <1.4, b = 0.3[6]. For this, we can design experiments to test. The first group: when a = 1.4, b = 0.3, then x0 = 0.3, y0 = 0.1. Group II: when a = 1.6, b = 0.35, then x0 = 0.3, y0 = 0.1. 300 times, respectively, after two iterations available following charts:



The above charts show that two-dimensional Henon mapping chaotic pseudo-random sequences fully meet the uniform distribution, δ-link of the Autocorrelation, O cross-correlation properties. The application of Henon

mapping is mainly based on three reasons: a) A theory of more in-depth study of chaotic behavior; b) it has a good cryptographic properties; c) Henon mapping is a two-dimensional, providing a relatively large dense key space.

4. The Image Encryption Aritlunetic Based on Chaotic Sequences

4.1 Algorithm theory

In the traditional iterative password system, the main task of arranged operator is to rearrange the elements in the plaintext data block (is also called "scrambling"), which makes the cipher text blocks random. However, these operators are usually arranged in advance, which has nothing to do with the key. This obvious defect makes the product of some iteration password system particularly vulnerable to differential cryptanalysis attacks, while the key-based arrangement will improve security significantly. In the key-based-arrangement algorithm, the key is the parameters, which can be arranged to determine the nature only.

The arrangement based on keys can be carried out in the frequency domain or space domain. Array transformation can be partial, or global. The completion of an encryption algorithm in spatial arrangement is relatively simple, because it does not require the use of the transformation from spatial domain to the frequency domain which the general frequency-domain algorithm needs necessarily, and the calculation is relatively small. However, the partial random scrambling effect of the spatial domain is not very good. The advantage of the frequency domain algorithm is that each point changes in the frequency domain will have some impact in the entire data set. If such image data has a change in DCT coefficients through the DCT transform, the IDCT inverse operation will be reflected in all of the pixel points. Compared to the air space algorithm, frequency-domain encryption algorithm is more efficient.

Algorithm proposed in this paper is based on the above theory. In the algorithm, the user uses super-key to generate chaotic sequences, and then uses chaotic sequences to produce a transformation matrix.
Set image as I, the size of I is $n = M \times N$, conducted in the DCT domain: ①change images through DCT transform, I_D = DCT (I); ②change the value of I_D by the use of real-valued chaotic sequence and get I_{DS}; ③ arrange the transformation of I_{DS} and get I_{DST}; ④ put I_{DST} with the dot matrix symbol and get I_{DSTS}; ⑤ put I_{DSTS} into anti-transformation, $I_E = DCT^{-1}$ (I_{DSTS}) the completion of image encryption.

4.2 Generate chaotic sequence

Chaotic sequence generated from the logistic mapping approach can be summarized as follows.
4.2.1 Real-valued sequences, that is, $\{x_k; k = 0,1,2,3 ...\}$, are formed through a chaotic mapping of the trajectory points.
4.2.2 Binary sequence, we can define a threshold function Γ by the above-mentioned real-valued chaotic sequence,

$$\Gamma(x) = \begin{cases} 0 & -1 \le x < 0 \\ 1 & 0 \le x \le 1 \end{cases} \qquad (9)$$

Binary chaotic sequence will be $\{\Gamma(x_k); k = 0,1,2,3 ...\}$
4.2.3 Bit sequence: obtain the same sequence from the real chaos. The difference is that bit sequence is by$\{x_k; k = 0,1,2,3...\}$ rewriting the x_k for the floating-point numbers of the L-bit to receive the form.

$$|x_k| = 0.b_1(x_k)b_2(x_k)\cdots b_i(x_k)\cdots b_L(x_k) \qquad (10)$$

Here $b_i(x_k)$ is $|x_k|$'s the i bits. The sequence is $\{b_i(x_k); i = 0, 1, 2 ... L; k = 0,1,2,3 ...\}$.
3.2.4 Four value chaotic sequences: the generation of this chaotic sequence and the one of binary chaotic sequences are very similar.

With two-dimensional Henon map chaotic sequence generating method, we can first select the initial value of Henon map x0, y0, after 20 iterations, obtaining xn and yn, set M = xn×yn. Then the obtained sequence is S = | M | - [| M |]. If S = 0, then return and re-enter the initial values or parameters iteration. [7]

4.3 Design algorithm

This algorithm uses pairs of chaotic map sequences, the S will be the number of logistic maps, as the initial value x0. [7]Set up the original image as I, I = M×N. Using the key value of x_0, we use formula (4) to generate real-valued chaotic sequence x_k, the use of formula (2) to generate real-valued chaotic sequence y_k. In this algorithm, we do not use the initial segment of this sequence.

Generate symbol matrix

The use of real-valued chaotic sequence x_k, k = 1,2, ..., by defining a threshold function Sign we obtain a symbol sequence, then the sequence according to the order of rows or columns constituting a symbol matrix S, S = M×N.

$$Sign(x_k) = \begin{cases} -1 & -1 \le x_k < 0 \\ 1 & 0 \le x_k \le 1 \end{cases} \quad (11)$$

Therefore, the symbol sequence also has chaotic characteristics.

4.3.1 Generate permutation matrix

The use of real-valued chaotic sequence to generate the corresponding permutation matrix P, P = M×N. For P, either element of P Pij∈[1,2, ..., M×N], and $P_{ij} = P_{kl}$, if and only if i = k, j = l.

4.3.2 Replacement rule *p*

We use the non-linear displacement (rotation) method of magic square. The image I = M×N and the permutation matrix P = M×N to correspondence of elements in columns and rows. Move the value of Element 1 to the position of the value of 2, then move the value of Element 2 to the position of the value of 3, ..., move M×N to the position of the value of 1, thus get a rotation of *p*, P is changed into P_1, $P_1 = p$ (P). Because I and P are in correspondence, with the P converted to P_1, I will also be converted to I_1.

4.3.3 Encryption Algorithm

(1) Put parameters: *. The file name of original image is InImage; *. The file name of results image is OutImage; *. key (x0, n);

(2) By the key values of x0, generate real-valued chaotic sequence x_k and y_k, k = 1, 2...

(3) Generate symbol matrix S and the permutation matrix P;

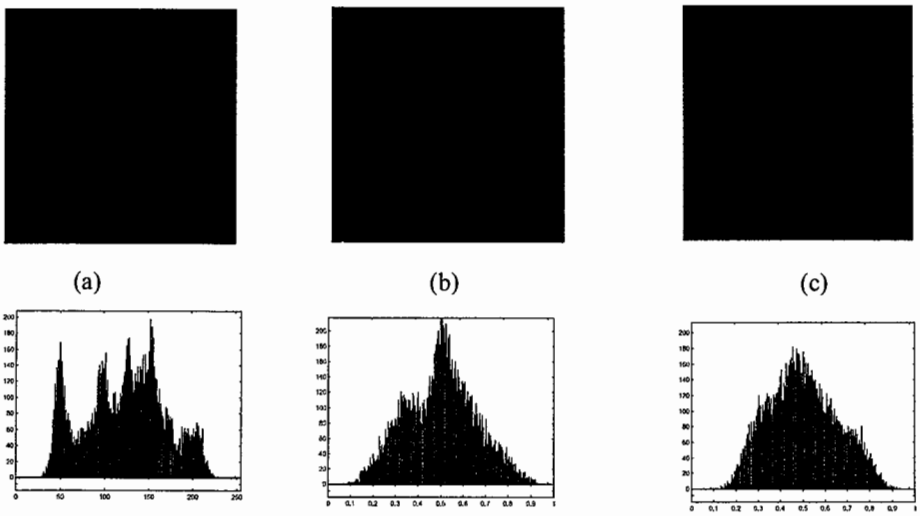
(4) Exchange the image I and DCT, and then change the part of DTC coefficient by y_k, get I_D;

(5) Change the I_D according to the rules *p*, and make I_{DT} with the symbol matrix S for scalar multiplication.

4.3.4 Decryption Algorithm

Users must enter the correct key (the initial value x0 chaotic sequence and iteration n), put the encryption algorithm into reverse operation, decrypted image can be obtained.

5. Experiment Results

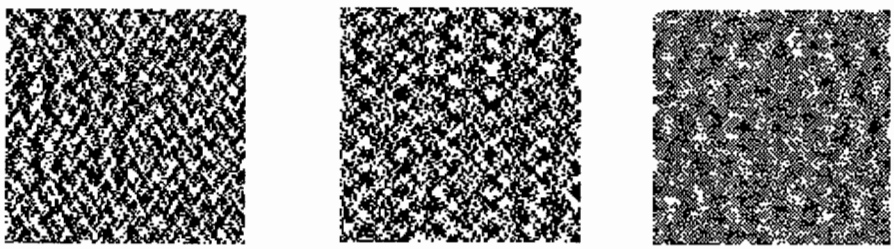


(a) (b) (c)

(e) (f) (g)

Figure 1 (a) is lean's origin(128×128), (b)and (c) are encrypted image,

(e), (f), (g) are the corresponding histogram.



(h) (i) (j)

Figure 2 Decryption result (h) (x0=0.60772,n=1) is (b)'s decryption in Figure 1
(i) (x0=0.60772,n=1) (j) (x0=0.60772,n=4) is (c)'s decryption in Figure 1

Adopt the method mentioned in this paper to test multiple images. Figure 1 is encryption results of lena image in DCT domain by using the algorithm (The key for Figure b is x0 = 0.60773, n = 1, the key for Figure c is x0 = 0.60773, n = 4). The algorithm for image encryption is carried out in the DCT domain, and its features are: first, in the DCT domain, employ real-valued chaotic sequence to scale DCT's coefficients, which brings great damage to the image quality, and the first encryption is completed; second, arrange its non-linear transformation, for a second encryption; the final adoption of the symbol matrix of random changes the sign of its DCT coefficients to achieve the effect of image encryption. The above three-layer encryption can be partially encrypted, the overall can also be encrypted, and the number of iterations can be used as the key part. If we do not know the key, we can not generate the corresponding proportion of matrix, transformation matrix and the symbol matrix. As chaotic sequence is very sensitive to the initial value, even though there are minor changes in the key values, we will also get a completely different decryption results (Figure 2) and thus will not be able to decrypt the image.

6. Conclusions

A large number of experiments show that this algorithm possesses the traits, namely fast speed for encryption speed, perfect results for encryption. For the generated chaotic sequences, it is best policy not to choose the initial partial sequence of paragraphs, which can strengthen the encryption.

References

[1]Qi Dongxu, Ding Wei, Li Huashan. A new image transformation scheme and digital approximation between different Pictures. In : ChenZhongying , Li Yuesheng , *Charles Micchelli* A , et al . , ed. Advances in Computational Mathematics. Lecture Notes in Pure and Applied Math, 202. New York : Marcel Dekker , Inc , 1999. 465～469

[2]Huang Feng, Feng Rong. Image encryption approach based on a new invertible two-dimensional map. Optical Technique, 2007,Vol.33, No. 6:823-826.

[3] Zou Jiancheng, Tie Xiaoyun, Digital image of the Arnold transform and its cyclical, the collected works of China's Conference on Computer Graphics, Hangzhou, 2000.

[4] Kwok H S, Tang Wallace K S. A Fast Image Encryption System Based on Chaotic Maps with Finite Precision Representation. Chaos, Solitons and Fractals, 2007, 32(4):1518-1529.

[5] Gao H J, Zhang Y S, Liang S Y, et al. A New Chaotic Algorithm for Image Encryption. Chaos, Solitons and Fractals, 2006, 29(2):393-399.

[6]Zheng Weimao. ***Practical Symbolic Dynamics***. Shanghai: Shanghai Science and Technology Education Press.1994.159-167.

[7]Li En,Wu Min,Xiong Yonghua. ***Chaotic map based on double encryption algorithm design and application***. Sichuan: **Computer Application Research**.2009:26(4)1512-1514.

[8]GAO HJ, ZHANG YS, LIANG SY,etal.A new chaotic algorithm for image encryption. Chaos, Solutions and Fractals, 2006, 29(2): 393-399.

[9]GUAN ZH, HUANG FJ, GUANWJ. Chaos-based image encryption algorithm. PhysicsLettersA, 2005, 346: 153-157.

[10]Guanrong Chen,Yaobin Mao and Charles K.Chui.A symmetric image encryption scheme based on 3D chaotic cat maps[J],Chaos,Solitions and Fractals,2004,21(3):749-761.

[11]Lian Shiguo, Wang Zhiquan, Li Zhongxin. Multimedia Coding Scheme Security Based Quadtree. Chinese Journal of Image and Graphics,2004，9(3):353-359.

About the author

A NEW THRESHOLD BASED MEDIAN FILTERING TECHNIQUE FOR SALT & PEPPER NOISE REMOVAL

Geeta Hanji [#1], M.V.Latte[*2]

E & CE, E &CE

P.D.A. Engineering College Gulbarga,-585102 Karnataka, INDIA

Bhagavan Mahavir Jain Engineering College Bangalore,Karnataka, INDIA

hanjigeeta@yahoo.com,mvlatte@rediffmail.com

ABSTRACT:

Removing Noise from the image is a challenging problem for the researchers. This paper proposes a two phase threshold based median filtering technique for salt & pepper impulse noise removal. It is implemented as a two pass algorithm: In the first pass corrupted pixels are perfectly detected using min-max strategy & an adaptive working window based on estimated noise density. Second phase is a threshold based filtering technique to correct the corrupted pixels by a valid median. Experimental results have shown that the proposed technique performs far more superior than many of the efficient median based filtering techniques reported in the literature in terms of Peak Signal (PSNR) & visual perception of the images corrupted by impulse noise even to the tune of seventy percent.

Key words:

Threshold value, Salt & Pepper noise, Peak Signal to Noise Ratio (PSNR), Mean Square Error (MSE), Mean Absolute Error (MAE).

I. INTRODUCTION

Noise is any undesirable information that contaminates an image. Noise appears in an image from a variety of sources. Removing Noise from the original signal is still a challenging problem for researchers. The salt & pepper type noise is typically caused by malfunctioning [1] of the pixel elements in the camera sensors, faulty memory locations etc. For the images corrupted by salt & pepper noise, the noisy pixels can take only the maximum & minimum values in the dynamic range.

To recover the images from salt & pepper noise, many nonlinear filtering techniques are reported in the literature. One of the robust nonlinear filters is the standard median (SM) filter [2] which exploits the rank order information of the pixel intensities within a working window & replaces the center pixel with the median value. Due to its effectiveness in noise suppression & simplicity in implementation, various modifications of the SM filter have been proposed, such as weighted median (WM) [3] filter & the center weighted (CWM) [3] filter. These conventional median filtering techniques apply the median operation to each pixel unconditionally, that is without considering whether it is corrupted or uncorrupted. As a result even the uncorrupted pixels are filtered & this causes the image degradation. An intuitive solution to overcome this problem is to implement an impulse noise detection mechanism prior to filtering. Hence only those pixels identified as "corrupted" would undergo the filtering process. By incorporating such noise detection mechanism into the median filtering frame work, the so called switching median filters [4]-[6] had shown significant performance improvement.

The most popular approaches for dealing with such noise have been based on threshold based median filtering techniques [7],[.8],[9].These filters have demonstrated excellent performance for noise density in the range 5-45% ,but fail to provide good results at higher noise densities, i.e. in the range 50- 70%. To solve this problem, a two phase threshold based median filtering method for impulse noise detection & filtering is proposed in this paper. Our new scheme exhibits improved performance in removing impulse noise while preserving fine details of the 2-D image structure.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 754639 · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856333

The rest of the paper is organized as follows: Noise models considered for our work are described in section II. Our proposed approach is described in section III.In section results are presented and discussed in. Finally the paper is concluded in section V.

II. NOISE MODELS

Two impulse noise models are implemented, for extensively examining the performance of proposed filter with the consideration of practical situations. Each model is described in detail as follows.

I) **Noise Model 1**: Noise is modeled as a salt-and-pepper impulse noise. Pixels are randomly corrupted by two fixed extreme values, 0 and 255 (for 8-bit monochrome image), generated with the same probability. That is, for each image pixel at location (i, j) with intensity value $S_{i,j}$, the corresponding pixel of the noisy image will be $X_{i,j}$, in which the probability density function of $X_{i,j}$ is

$$f(x)=\begin{cases}\frac{P}{2} & for\ x=0\\ 1-P, & for\ x=s_{i,j}\\ \frac{P}{2} & for\ x=255\end{cases} \qquad [Model\ 1] \tag{2.1}$$

where P is the noise density.

II) **Noise Model 2**: For the Model 2, it is similar to Model 1, except that each pixel might be corrupted by either "pepper" noise (i.e., 0) or "salt" noise with unequal probabilities. That is

$$f(x)=\begin{cases}\frac{P_1}{2} & for\ x=0\\ 1-P, & for\ x=s_{i,j}\\ \frac{P_2}{2} & for\ x=255\end{cases} \qquad [Model\ 2] \tag{2.2}$$

Where $P=P_1+P_2$ is the noise density and $P_1 \neq P_2$.

III. PROPOSED APPROACH

Our approach is a two phase threshold based median filtering technique for impulse noise detection and filtering. It uses an adaptive window and min-max technique for impulse noise detection & a properly chosen threshold value for filtering. The proposed algorithm is described below:

3.1 Impulse Detection & adaptive Threshold Based Filtering

The proposed algorithm first detects the impulse noise in the image. The corrupted & uncorrupted pixels in the image are detected by checking the pixel element value against the minimum & maximum values in the selected adaptive window maximum & minimum values that the impulse noise takes will be in the dynamic range (0,255) [10].If the pixel being currently processed has a value within the minimum & maximum values in the window of processing, then it is an uncorrupted pixel & no modification is made to that pixel. If the value doesn't lie within the range then it is a corrupted pixel. Our algorithm computes two median values: Median M1 of the pixels of the sub window ignoring the pixels with the maximum (255) and minimum value (0) and the median M2 of the whole window.

If the difference between the calculated median (M1 or M2) & the intensity of the considered pixel is greater than the threshold, we replace the intensity of the considered pixel by the valid median. Otherwise the intensity of the considered pixel is unchanged.

3.2 Algorithm Steps

1. For each pixel p in the image do

   i. Select a working window of size m×n based on estimated noise density around that pixel.
   ii. Find out the Median M1 of the pixels of the sub window ignoring the pixels with the maximum (255) and minimum value (0).
   iii. Calculate median M2 of the whole window.
   iv. Find the Count 'C'of the number of pixels obtained after ignoring pixels of minimum and maximum value within the mxn sub window. If this count is greater than or equal to
   v. 1/3 rd of m×n then consider the Median M1 value otherwise consider median M2 value for the replacement strategy.
   vi. Compute the difference, Diff = Difference between M1 (or M2) and the intensity of pixel p.
      a) If Diff ≥ Threshold, then replace intensity of p by M1 or M2.
      b) Otherwise leave the pixel value unchanged.

   Working window is decided based on the estimated noise density adaptively as given in table below:

TABLE 3.1 Suggested Working Window size for the estimated noise density

Noise density P	Working window (W×W)
P < 30%	3 × 3
30% ≤ P ≤ 50%	5 × 5
P > 50%	7 × 7

3.3 The significance of the Threshold value & Adaptive Threshold Selection

The sanctity of a pixel is decided solely by the threshold. If a predetermined parameter of a test pixel exceeds the threshold value, it is termed as contaminated.

Solution to image restoration problem depends very much on the type of image, characteristics and density of noise. It is observed that a single threshold value does not serve the purpose well in different noise conditions. For Salt and Pepper noise the value of the distorted pixel is 0 or 255. So we find a significant difference between the median and the value of the distorted pixel. Replacing only the distorted pixel will provide us with better results than replacing all the pixels. Hence a properly chosen threshold value ensures that only the distorted pixels are replaced.

It is in general observed that, there exists an optimum threshold for every image and for a particular noise density. Even these values may differ from image to image for the same noise density. There exists a nonlinear relationship between optimum threshold and noise densities as well as MSE.

In our experimentation it is observed that for a LENA image corrupted with salt & pepper noise of noise density in the range 70-80%, if we take small threshold value such as 5 or 10, then not only the distorted pixels but also the other pixels are replaced. If we increase the threshold value, then PSNR increases. If we take a very large threshold value, some distorted pixels are not replaced which decreases the PSNR.A threshold value of 60 is found to provide satisfactory results for the image corrupted with salt & pepper noise of noise density in the range 60-70%.For the LENA image corrupted with different values of noise densities, the threshold values suggested are tabulated in Table 3.2 below:

TABLE 3.2 Suggested Threshold Values for the estimated noise density

Noise density P	Threshold value
P < 30%	35
30% ≤ P ≤ 50%	45
P>50	60

IV. SIMULATION RESULTS AND DISCUSSION

Intensive simulations were carried out using Images such as Lena of size 512 × 512 with noise densities such as low (20%), medium value (40%) & high noise (70%). Performance of the proposed algorithm is measured using quantitative performance measures such as Peak Signal to Noise Ratio (PSNR), Mean absolute Error (MAE), Mean Square Error (MSE) as well as the visual quality of the restored images. These parameters are defined as follows.

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

$$MSE = \frac{1}{MN}\sum_i \sum_j \left(R_{i,j} - X_{i,j}\right)$$

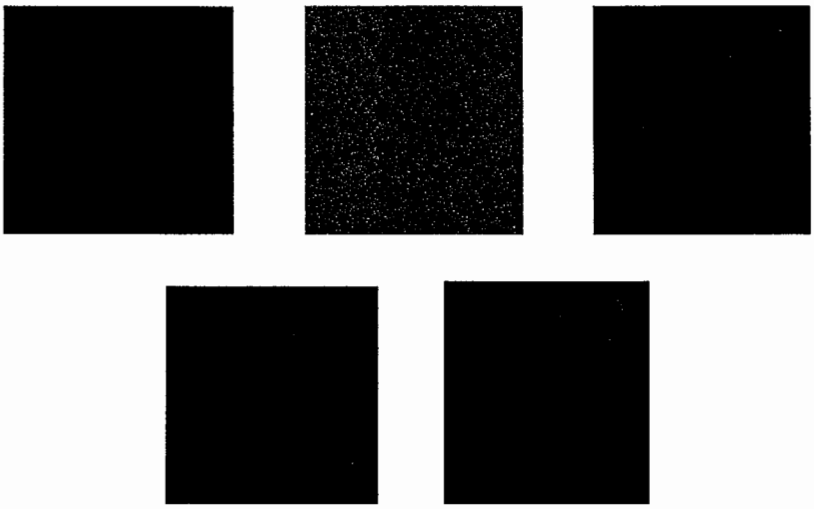
$$MAE = \frac{1}{MN}\sum_i \sum_j |R_{i,j} - X_{i,j}|$$

Where $R_{i,j}$ and $X_{i,j}$ denote the pixel values of restored image and the original image respectively and M x N is the size of the image. Filtering window used for median filters and decision based filters is 3x3. The Thresholds values used for multiple threshold filter are tuned to give best performance in terms of PSNR, MAE and MSE. The Tolerance value used for TSAMFT is 35 for 512x512 Lena image. All the filters are implemented in MATLAB 7.5.

Figures (4.4), (4.5), (4.6) and tables (4.4), (4.5), (4.6) show the performance results of the various filters, on the image LENA at low (20%), medium (40%), and high (70%) noise densities, respectively. The results of PA indicate better PSNR values compared to Adaptive Median Filter(AMF),Standard Median Filter (SMF), Weighted Median Filter (WMF), Center Weighted Median Filter (CWMF), Tolerance Based Selective Arithmetic Mean filtering (TSAMFT), Multiple Threshold based Filter (MULTHF). For lower noise density up to 30% almost all algorithms performs equally well in removing the salt and pepper noise completely with edge preservation. For the case of noise density 40% and above the standard algorithms such as SMF, WMF, CWMF, TSAMFT fails to remove the salt and pepper noise completely. In the case of high noise density such as 70% of salt and pepper noise the standard methods are very poor in noise cleaning and details preservation, where as the proposed technique performs well in terms of noise removal and edge preservation.

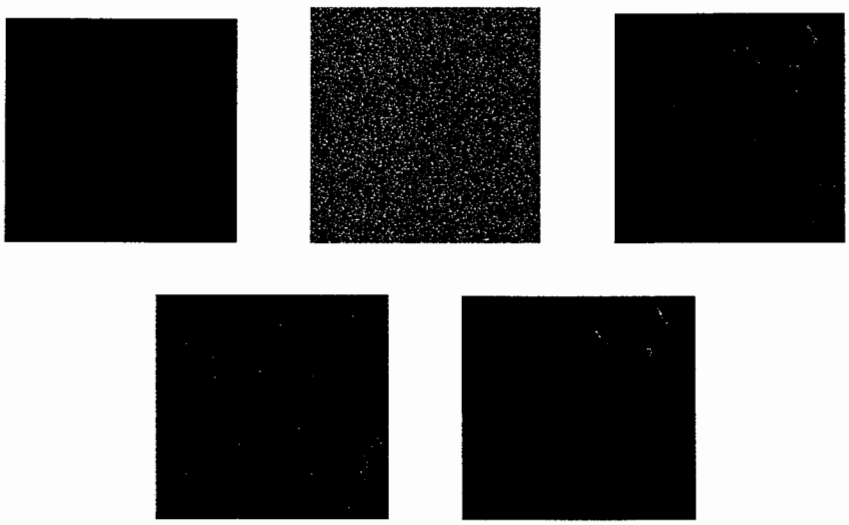
Figures (4.4), (4.5), (4.6) and tables (4.4), (4.5), (4.6) show the performance results of the various filters, on the image LENA with unequal noise densities. The maximum window size of 17x17 is selected for AMF to give better results at high noise density level. The visual quality, PSNR, MAE and MSE results clearly show that the proposed technique out performs among all.

Table 4.7 shows the computation time in seconds required for denoising the LENA grayscale image that is corrupted with salt and pepper noise of density 70%. The computation time of the proposed technique is slightly higher than the other techniques since it uses an adaptive filtering window of size 7x7.



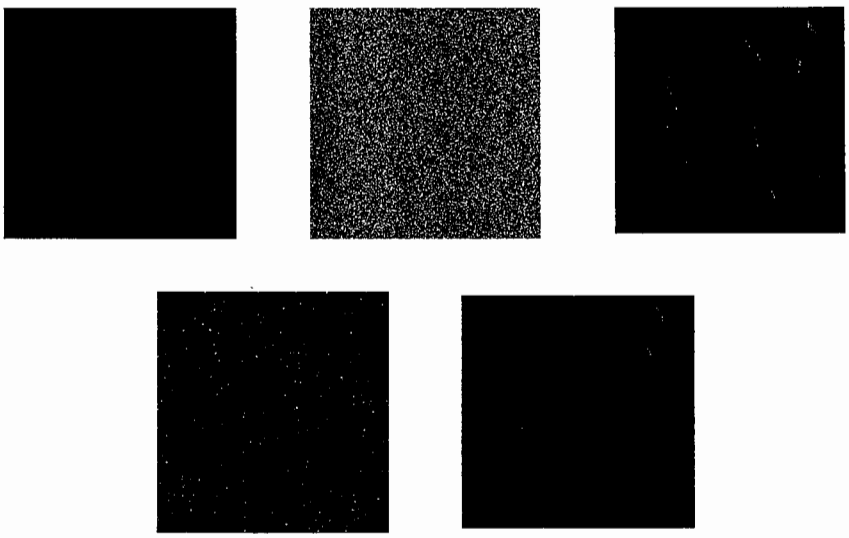
a	b	c
d	e	

Fig.4.1 a) Original Lena image b) Noisy image with 20% Noise density & PSNR of 12.43 c) TSAMFT with PSNR of 27.33 d) MULTHI with PSNR of 36.47 e) PA with PSNR of 38.31



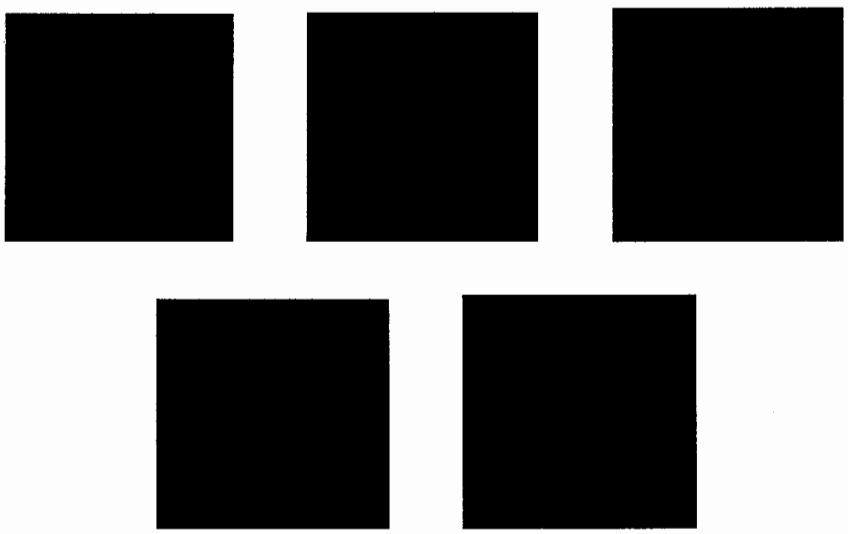
a	b	c
d	c	

Fig.4.2 a) Original Lena image b) Noisy image with 40%Noise Density & PSNR of 9.43 c) TSAMFT with PSNR of 25.24 d) MULTHI with PSNR of 31.73 e) PA with PSNR of 33.47



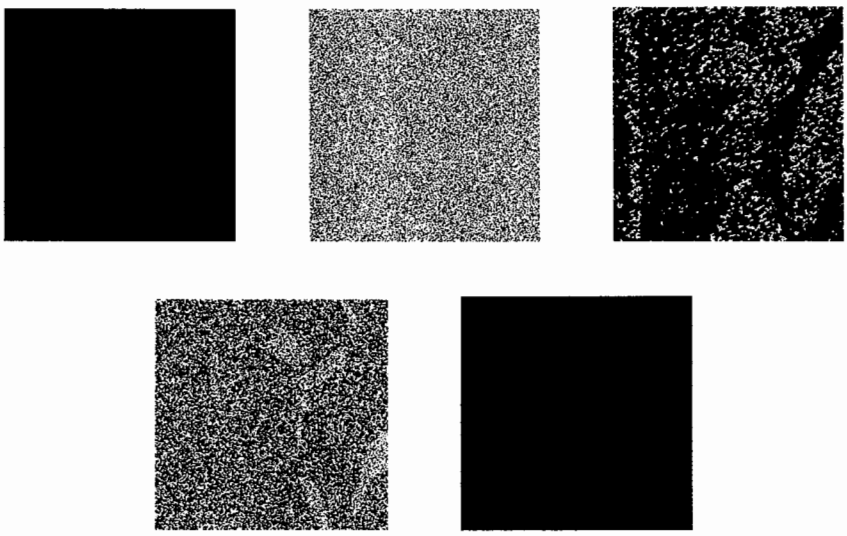
a	b	c
d	e	

Fig.4.3 a) Original Lena image b) Noisy image with 70% Noise Density & PSNR of 6.99 c) TSAMFT with PSNR of 20.50 d) MULTHI with PSNR of 23.29 e) PA with PSNR of 29.41



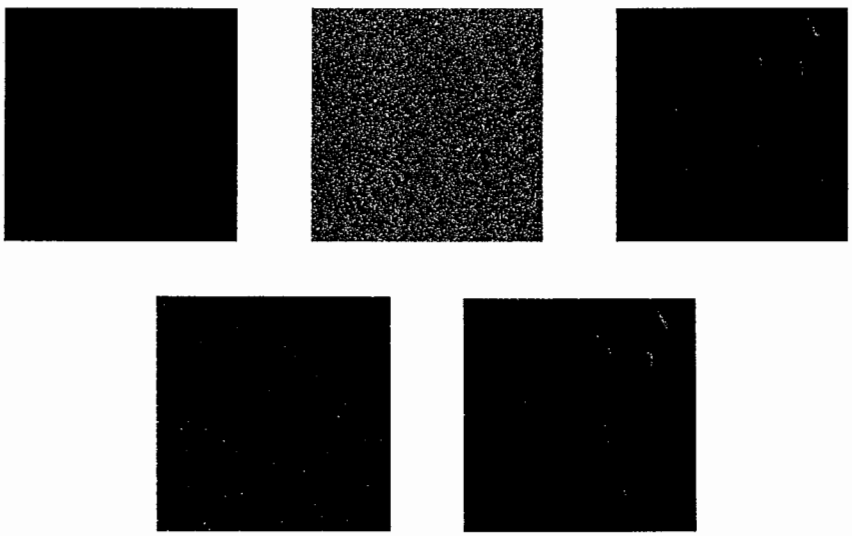
a	b	c
d	e	

Fig.4.4 a) Original Lena image b) Noisy image with 0% Salt ,70% Pepper Noise & PSNR of 7.20 c) TSAMFT with PSNR of 14.99 d) MULTHI with PSNR of 8.65 e) PA with PSNR 28.92



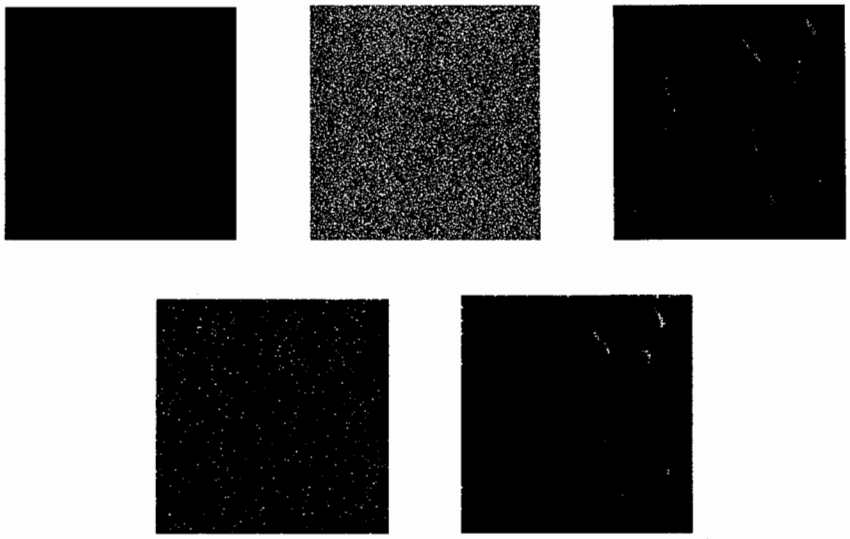
a	b	c
d	e	

Fig.4.5 a) Original Lena image b) Noisy image with 70% Salt, 0% Pepper Noise & PSNR of 8.79 c) TSAMFT with PSNR of 14.58 d) MULTHI with PSNR of 8.25 e) PA with PSNR of 28.81



a	b	c
d	e	

Fig.4.6 a) Original Lena image b) Noisy image with 30% Salt, 40% Pepper Noise & PSNR of 7.02 c) TSAMFT with PSNR of 20.47 d) MULTHI with PSNR of 22.01 e) PA with PSNR of 29.48



a	b	c

Fig.4.7 a) Original Lena image b) Noisy image with 40% Salt, 30% Pepper Noise & PSNR of 6.96 c) TSAMFT with PSNR of 20.26 d) MULTHI with PSNR of 21.81 e) PA with PSNR of 29.29

TABLE 4.1
Comparative results for various filters in terms of Peak Signal to noise ratio for "Lena" (512x512 grayscale) image

ND	Noisy image	SMF	WMF	CWMF	AMF	TSAMT	MULTH	PA
10%	15.44	33.12	33.89	34.52	38.11	27.52	39.64	40.97
20%	12.43	28.76	29.94	29.46	35.93	27.33	36.47	38.31
30%	10.69	23.64	25.10	24.41	33.81	26.85	34.09	35.12
40%	9.43	19.08	20.84	19.94	32.06	25.24	31.73	33.47
50%	8.44	15.29	17.22	16.47	30.18	22.21	28.38	32.01
60%	7.66	12.35	14.03	13.24	28.58	21.70	25.73	30.78
70%	6.99	10.0	11.32	10.72	26.80	20.50	23.29	29.41

TABLE 4.2
Comparative results for various filters in terms of Mean Absolute Error for "Lena" (512x512 grayscale) image

ND	Noisy image	SMF	WMF	CWMF	AMF	TSAMT	MULTH	PA
10%	12.72	2.79	2.09	1.88	0.96	2.36	0.48	0.40
20%	25.48	3.57	2.86	2.58	1.27	2.86	0.97	0.80
30%	38.06	5.32	4.33	4.16	1.73	3.50	1.53	1.38
40%	50.89	9.21	6.89	7.54	2.33	4.74	2.16	1.93
50%	63.83	16.81	12.24	13.64	3.04	8.05	3.08	2.50
60%	76.51	29.77	21.24	24.06	3.94	9.21	4.35	3.13
70%	89.28	47.27	36.68	42.0	5.15	11.88	6.05	3.90

TABLE 4.3
Comparative results for various filters in terms of Mean Square Error for "Lena" (512x512 grayscale) image

ND	Noisy image	SMF	WMF	CWMF	AMF	TSAMT	MULTH	PA
10%	1850	30.44	26.55	22.70	10.09	115.53	7.06	5.18
20%	3707	86.64	70.45	63.84	16.28	119.40	14.63	9.59
30%	5530	292.2	216.84	227.06	26.59	132.79	25.31	19.97
40%	7400	821.4	519.56	659.97	41.74	195.85	43.58	29.18
50%	9310	1890	1250	1510	60.35	390.01	94.29	40.91
60%	11120	3830	2610	2980	87.45	438.56	173.49	54.27
70%	12990	6430	4780	5670	135.6	585.76	304.59	74.35

TABLE 4.4 Comparative results for various filters in terms of Peak Signal to noise ratio for "Lena" (512x512 grayscale) image with Unequal Noise Densities.

% of salt noise	% of pepper noise	Noisy image	SMF	WMF	CWF	AMF	TSAMT	MULTH	PA
0%	70%	7.20	6.09	5.92	6.01	6.13	14.99	8.65	28.2
70%	0%	6.79	5.70	5.54	5.60	5.73	14.58	8.25	28.1
30%	40%	7.02	9.85	10.92	10.33	26.9	20.47	22.01	28.2
40%	30%	6.96	9.64	10.72	10.05	26.2	20.26	21.81	28.4
20%	50%	7.07	8.43	8.54	8.49	14.7	19.59	16.78	28.6
50%	20%	6.90	8.08	8.14	8.14	13.2	19.65	16.61	28.2

TABLE 4.5 Comparative results for various filters in terms of Mean Absolute Error for "Lena" (512x512 grayscale) image with Unequal Noise Densities.

% of salt noise	% of pepper noise	Noisy image	SMF	WMF	CWMF	AMF	TSAMT	MULTH	PA
0%	70%	86.81	112.1	116.3	114.63	111.9	23.26	73.17	4.24
70%	0%	91.78	118.4	122.9	121.08	117.8	25.02	76.51	4.24
30%	40%	88.84	49.49	40.04	44.80	5.51	11.82	7.05	4.22
40%	30%	89.62	50.91	41.00	46.40	5.50	12.13	7.07	4.26
20%	50%	88.26	67.52	66.34	66.65	22.91	12.99	15.62	4.22
50%	20%	90.27	70.42	70.01	69.56	25.08	12.85	15.62	4.26

TABLE 4.6
Comparative results for various filters in terms of Mean Square Error for "Lena" (512x512 grayscale image ith Unequal Noise Densities.

% of salt noise	% of pepper noise	Noisy image	SMF	WMF	CWMF	AMF	TSAMT	MULTH	PA
0%	70%	1230	1590	1660	16270	1580	2058	8870	100
70%	0%	1360	1740	1810	17900	1730	2260	9710	110
30%	40%	1290	6710	5250	60200	159.2	583.51	408.40	73.3
40%	30%	1300	7060	5500	6410	158.4	612.26	427.89	76.6
20%	50%	1270	9320	9080	9180	2540	714.53	1363	78.0
50%	20%	1320	1000	9970	9970	3020	703.98	1418	73.5

TABLE 4.7
Computational Time in Seconds for denoising 70% salt & pepper noise corrupted LENA (512x512 gray) image for different denoising techniques.

FILTER	TIME
Standard Median Filter	0.08
Weighted Median Filter	6.5
Centre Weighted Median Filter	6.46
Adaptive Median Filter	2.91
Tolerance Based Selective Arithmetic Mean filter	8.82
Decision Based Filter	2.83
Improve Decision Based Filter	3.08
Multiple Threshold Filter	8.02
Proposed Algorithm	14.87

V CONCLUSIONS AND FUTURE PLAN

In this paper we have proposed a two phase threshold based impulse noise removal filter that can show better performance than the existing mean & median based filters. Our scheme uses min-max strategy to detect the noisy pixels & an adaptive working window based on estimated noise density in the filtering process. Our filtering scheme is a threshold based technique to correct the corrupted pixels by a valid median. Experimental results have shown that the proposed technique performs far more superior than many of the efficient median based filtering techniques reported in the literature in terms of Peak Signal (PSNR) & visual perception of the images corrupted by impulse noise even to the tune of seventy percent. This filter can be further improved to apply for the images corrupted with high density impulse noise up to 90% and random valued impulse noise.

References

[1] Z.Wang and D. Zhang, "Progressive Switching Median filter for the Removal of Impulse Noise from highly Corrupted Images", IEEE Transaction on Circuits and Systems II, Volume 46, pp 78-80, 1999.

[2] Abreu E, Lightstone M. and Mitra S.K., "A New Efficient Approach for the Removal of Impulse Noise from Highly Corrupted Images", IEEE Transaction on Image Processing, Volume 5 No 6, pp 1012-1025, 1996.

[3] Ko S.J and Lee Y.H, "Centre Weighted median filters and Their Applications to Image Enhancement", IEEE Trans Circuit Systems Volume 38 No 9 pp 984-993.

[4] Brownrigg D.R.K, "The Weighted Median Filter," Communication, ACM Volume 27, No 3 pp 807-818 1984. Arce G. and Prades J., "Recursive Weighted Median Filters Admitting

[5] Negative Weights and Their Optimization", IEEE Transaction on Signal Processing, Volume 48, No 3, pp 768-779, 2000.

[6] Hwang H and Haddad R.A, "Adaptive Median Filters new and results", IEEE Transaction on Image Processing Volume 4 No 4 pp 499-502, 1995.

[7] Rajoo Pandey," An improved Switching Median Filter for Uniformly Distributed Impulse Noise removal",PWASET VOLUME 28,2008.

[8] Shahriar Kaisar, Md Sakib Rijwan, Jubayer Al Mahmud and Muhammed Mizanur Rahman, " Salt and Pepper Noise Detection and removal by Tolerance based Selection Arithmetic Mean Filtering Technique for Image Restoration", IJCSNS International Journal of Computer Science and Network Security Volume 8, No 6 ,2008.

[9] Tzu-Chao Lin and Pao-Ta Yu, "Salt-Pepper Impulse Noise Detection and Removal using Multiple Thresholds for Image Restoration", International Journal of information Science and Engineering Volume 22, 2006 pp No 189-198.

[10] Raymond H. Chan, Chung-Wa Ho, and Mila Nikolova,4 "Salt –and –Pepper Noise Removal by Median-Type Noise Detectors and Details-preserving Regularization", Image Processing, IEEE Trans Volume 14 pp 1479-1485, Issue 10, Oct.2005.

[11] Pitas I. and Venetsanopoulos A.N., "Nonlinear Digital Filters: Principal and Application", Boston, MA: Kluwer Academic, 1990.

[12] Gonzalez R.C and Woods R.E, "Digital Image Processing", Addison-Wesley Publication Company, 2002.

[13] Bovik A.C, "Handbook of Image and Video Processing", Academic Press, 2000.

[14] Pratt W.K, "digital Image Processing", New York Wiley, 1991.

Design of Knight LED System

Zheng wen, Lou yuna, Xiao zhihong

(Department of Electronics Information, Zhe jiang *Wan Li University, ningbo,315100*)

Zhengwen0722@163.com louyuna@163.com jhxiaozh@hotmail.com

ABSTRACT

This design introduces a used car on the design of LED decorative light strip. This LED named Knight LED. In This system we use ATMEGA8 as the Master MCU Chip. Through the microcontroller to implement the wireless remote control receiver and the LED lights of different modes of switching, different brightness control. Also we use ULN2803 as the LED driver.

Knight LED; Wireless; LED driver

1 DESCRIPTION

With the LED's development, from lighting to decorative, from the family to the square can be seen everywhere LED-related products. LED energy saving long life to the dominant, in-depth types of decorative lighting and other occasions. Generally high-rise buildings can be achieved with LED guardrail lights. In the family can use LED lights to achieve the lighting. In the car, many new cars use high-power high-brightness LED as a lighting. Also it used High-Brightness LED lights as brake lights .It is Installed in the front or rear of the car. There are different modes of LED display. There are different modes of LED display, and through wireless remote control to select the switch model, that can do decorative, but also to increase safety, the visual effects to enhance the awareness of other vehicles.

2 DESIGN

This system is mainly composed of three parts, the wireless transmitting and receiving part, the MCU control part and the part of LED driver display circuit. The wireless part of the circuit mainly by the launch of the remote control and receiver module. In this design we use 315M common frequency bands for wireless data transfer. We use the DF radio transmitter module and the super-regenerative radio receiver modules to achieve wireless data transmission. It use MCU

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463A · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853567

to handle the received wireless data, through software decoding and save the wireless remote control key to the MCU's internal EEPROM. In the drive circuit design, the transistor can be used to drive the LED, can also be used to achieve special driver chip. This design uses a ULN2803 driver chip to achieve the LED driver is divided into 8-way, each way each driver 6 LED.

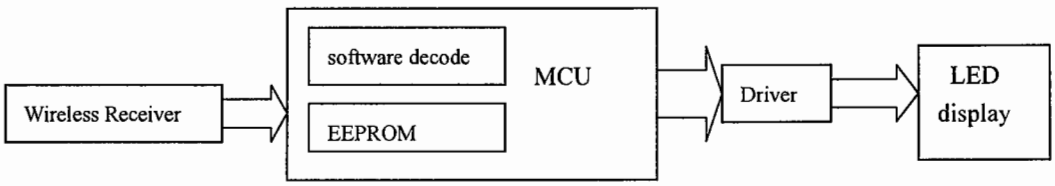


Figure 2-1 :Design Diagram

3 HARDWARE DESIGN

In this Design we use ATMEGA8 as the master control chip. ATMEGA8 is a High-performance, Low-power AVR® 8-bit Microcontroller. It has 8K Bytes of In-System Self-Programmable Flash,512 Bytes EEPROM and 1K Byte Internal SRAM[1]. All of this can meet the needs of this design.

The wireless data receiving part of this design uses software decoder to achieve. Generally are based on the codec chip PT2262 PT2272 such as data encoding and decoding[2]. But there is one drawback is that the same address code can be matched. We welding address code increased the amount of labor, also has more Repetition coding. In this design we use coding chip named ev1527.

eV1527 is one that can be designed and manufactured by the CMOS pre-burn within yards of the study coding IC, by the software decoder: A total of 20 bits within the code may be pre-burn 1048576 Group (220) yards inside the portfolio, reducing the chances of repeating the use of codes. The advantage of using software decoding is: to save the hardware cost, because a microcontroller can be achieved by programming a variety of functions, single-chip decoding processing can be done, while other operations can be achieved.

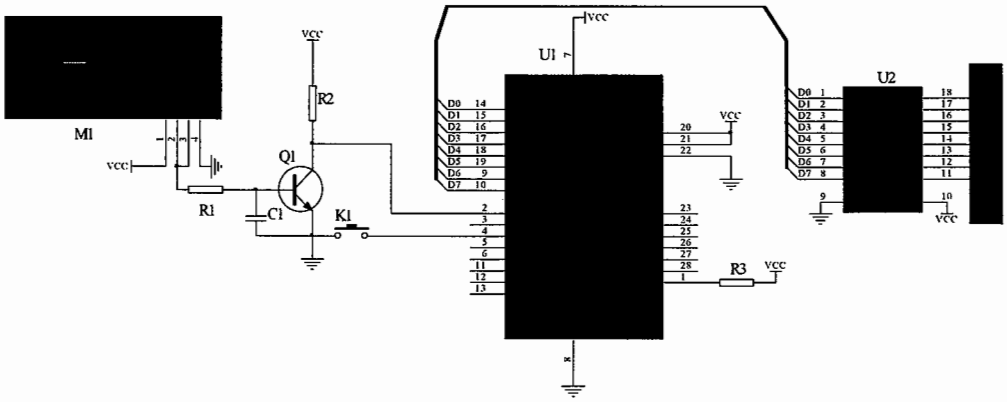


Figure 2-2: Schematic

Circuit diagram and analysis:

Where M1 is a wireless receiver module, using 315M band, receive post-reverse through the transistor and the shaping and ultimately into the microcontroller for data processing. First, the wireless signals must be to learn from well encoded, then stored in the EEPROM, the data can be powered down to save. Remote key function is to switch machines to achieve, mode switch, display brightness and speed of adjustment, and can power-down mode to save the current state.ULN2803 is mainly used as LED drivers. LED lights strip can be designed into different styles.

4 SOFTWARE DESIGN

Software design is divided into two parts: One is wireless data software decoding, and data encoding format to learn and to be stored; second multi-mode Knight LED control design. LED brightness can be used to adjust the brightness of PWM wave, this design is divided into 100 brightness levels to adjust the brightness level.

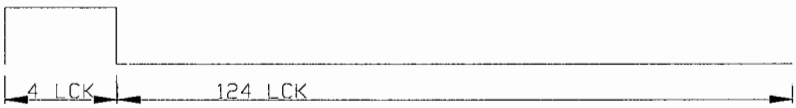
The key to this design is a wireless data software decoding, we must first identify the synchronization code, when receiving the correct code after the receive data synchronization code and key code[3].

The software decoding, we must understand the encoding format to receive.

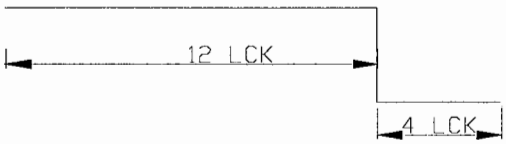
Here is the encoding format ev1527:

Synchronous Code	InternalCode (C0—C19)	Key Code (D0-D3)

Synchronous Code



DATA (H)



DATA (L)

Figure 2-3: encoding format

Definition:1 CLK=8 OSC CLOCK

How to recognition the Synchronous code? This article is an adaptive approach to software decoding .In this article, We propose a decoding idea. Decoding to note that, with a great radio reception interference, it is not applied in dealing with decoding with the MCU's interrupt, because this could lead to frequent interruption of access, resulting in crashes from happening microcontroller. so this design approach used in queries decoding process. Synchronous code identification consists of two parts, synchronization code yards high and synchronous low recognition. We known ev1527 has 20 Internal Code and four key code ,so after the MCU recode the Synchronous code ,it also should recode data code and key code.

This design has 18 kinds of LED display modes, including six kinds of PWM modes and 12 normal modes. we can use Wireless Remote Control Device to choose the mode and change the led display speed of the mode. In programming , we Save the led mode data into the code ROM of ATMEGA8 ,but save the speed data and mode num data into the EEPROM, so the data can be saved after power-down.

Figure 2-4 is the program flow chart. The left image shows Synchronous decoding and the right image shows Internal

code and key code decoding.

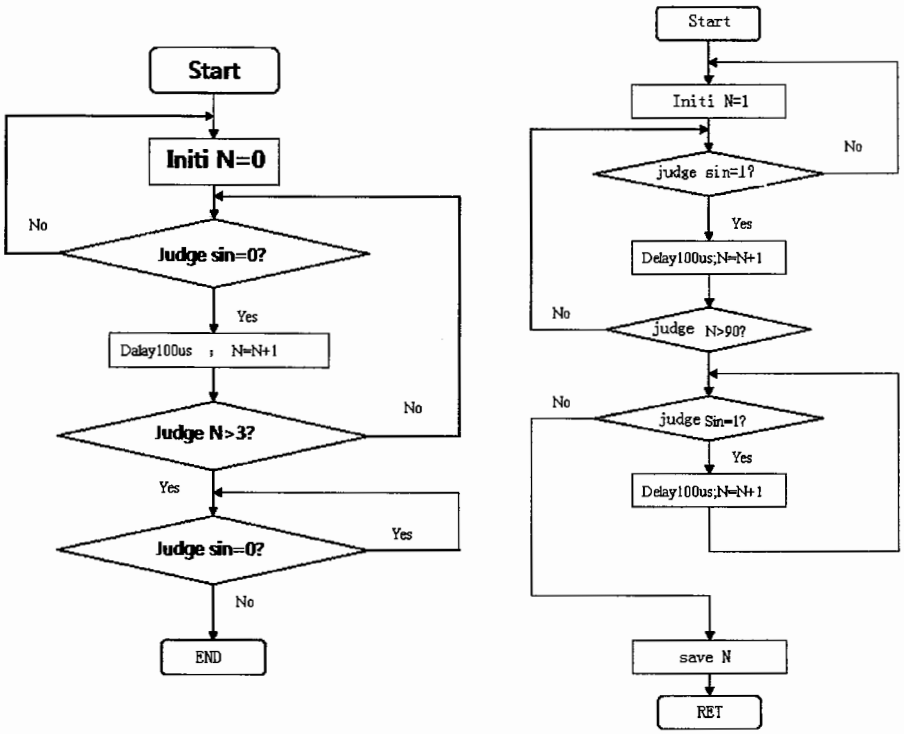


Figure 2-4: Program flow chart

5 SUMMARY

This design introduced Knight LED decoration can be applied directly to the car up, easy to install works well. Using a wireless remote control to achieve the mode switch machines and other functions, greatly facilitates the use of. Useful data is stored in the MCU's internal EEPROM, carried out through wireless data transfer mode, you can very easily update the LED display mode.

References

[1] Atmega Corporation.Atmega8 datasheet.http://www.alldatasheet.com/datasheet-pdf

[2] Weiqi Wu.PT2262 Encoding chip Software decoding [J]. Micro-Computer Interest,2004.20(7):110-112

[3] JingQian Zhou, RenNong Fang. ev1527 coding and decoding chip applications [J]. Research and Development , 2007. 26 (6): 35-38.

Infrared Image Denoising Based on Stationary Wavelet Transform

Xiao Zhihong Shi Jiale Guan Zongqi

(Department of Electronics Information, Zhejiang Wanli University, Ningbo, 315100, China)
jhxiaozh@hotmail.com 519641927@qq.com guanzongqi@163.com

ABSTRACT

Firstly, infrared image is decomposited using stationary wavelet transform, it is proposed based on stationary wavelet transform with Interscale and Intrascale Dependencies for infrared image denoising. Then the minimum mean square-error estimation is applyed to estimated coefficient. The wavelet coefficients are revised using the correlations between coefficients at the same scale. The denoised image is obtained through inverse wavalet transform. The experimental results show the infrared image can be denoised better than the method neglecting the correlations between Intrascales and have a well SNR as well as the visual quality.

Infrared images denoise ; stationary wavelet transform ; minimum mean square-error estimation ; Interscale and Intrascale Dependencies

1 INTRODUCE

The infrared imaging technology has been widely used in missile guidance and remote sensing for military and civilian spheres. However, as the external environment and the inherent characteristics of the instrument,the noise will be inevitable in the acquisition, transmission and storage of infrared image process. So, Study of infrared image denoising algorithm has a great significance. Wavelet transforms has the characteristics of low entropy, decorrelate and flexibility in the choice of wavelet base, And the noise are whitening after changing. So, the wavelet domain denoising more flexible than the spatial domain [1],The wavelet denoising has become the one of the principal means for image denoising [2-4]。Wavelet denoising method includes two types: Denoising Method of the module maxima by using wavelet transform, filtering method correlation coefficient based on wavelet transform domain and Wavelet Threshold Denoising. Compared with the orthogonal wavelet transform, Stationary wavelet transform has obvious advantages in noise reduction model and threshold method, It inhibits the Gibbs phenomenon caused by denoising of image signal by orthogonal wavelet thresholding. [1-2]。A methods using stationary wavelet transform with Inter-scale and Intra-scale Dependencies for infrared image noise-free is proposed in this paper. , Firstly, infrared image is dc-composited using stationary wavelet transform. For the high frequency component of image decomposition, the wavelet coefficient vectors are formed with the different scales. Then the minimum mean square-error estimation is applied to estimated coefficient. The wavelet coefficients are revised using the correlations between coefficients at the same scale. The noise-free image is obtained through inverse wavelet transform.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463B · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853569

2 STATIONARY WAVELET TRANSFORM (SWT)

Stationary wavelet transform is modified through the wavelet transform , It is no longer the next sampling to low-pass and high-pass filter output, Compared with the Discrete orthogonal wavelet transform(DWT), SWT has a "translation invariance", the number of wavelet coefficients is equal to the number of pixels in original image after the layers of sub-band decompositio. In addition, it is need for the upsampling filters. The decomposition formula is :

$$\begin{cases} s_{j+1,k_1,k_2} = \sum_m \sum_n H_j(m-2k_1)H_j(n-2k_2)s_{j,m,n} \\ w^H_{j+1,k_1,k_2} = \sum_m \sum_n H_j(m-2k_1)G_j(n-2k_2)s_{j,m,n} \\ w^V_{j+1,k_1,k_2} = \sum_m \sum_n G_j(m-2k_1)H_j(H-2k_2)s_{j,m,n} \\ w^D_{j+1,k_1,k_2} = \sum_m \sum_n G_j(m-2k_1)G_j(n-2k_2)s_{j,m,n} \end{cases} \quad (1)$$

Among them , s_{j+1,k_1,k_2} , w^H_{j+1,k_1,k_2} , w^V_{j+1,k_1,k_2} , w^D_{j+1,k_1,k_2} are the scale $j+1$ of the low frequency components, high frequency horizontal component, high frequency vertical component and diagonal component after SWT, filters H_j , G_j were inserted $(2^{j-1}-1)$ zero between the original filter coefficients H_0 , G_0 . The corresponding reconstruction algorithm is:

$$\begin{aligned} s_{j,m,n} = \frac{1}{4}\sum_{i=0}^{3}\{ & \sum_m \sum_n \tilde{H}_j(m-2k_1-i)\tilde{H}_j(n-2k_2-i)s_{j+1,m,n} \\ & + \sum_m \sum_n \tilde{H}_j(m-2k_1-i)\tilde{G}_j(n-2k_2-i)w^H_{j+1,k_1,k_2} \\ & + \sum_m \sum_n \tilde{G}_j(m-2k_1-i)\tilde{H}_j(n-2k_2-i)w^V_{j+1,k_1,k_2} \\ & + \sum_m \sum_n \tilde{G}_j(m-2k_1-i)\tilde{G}_j(n-2k_2-i)w^D_{j+1,k_1,k_2}\} \end{aligned} \quad (2)$$

$\tilde{H}_j, \tilde{G}_j$ are dual filters of H_j and G_j .

3 INFRARED IMAGE NOISE-FREE BASED ON THE STATIONARY WAVELET TRANSFORM USING INTER-SCALE AND INTRA-SCALE DEPENDENCIES

3.1 Based on wiener estimation of wavelet coefficients

IR imagc modcl as follows: :

$$s = x + n \quad (3)$$

Among them, $x \hat{I}\ R^{M' N}$ is the real images, $s \hat{I}\ R^{M' N}$ **is the** observe images, $n \hat{I}\ N(0, s^2)$, assumptions may not be relevant x . using stationary wavelet transform, j Scale in the wavelet domain described as:

$$w_j = u_j + v_j \quad (4)$$

Where w_j is s in the wavelet coefficients at j scale, Where u_j is x in the wavelet coefficients at j scale, Where v_j is n in the wavelet coefficients at j scale.

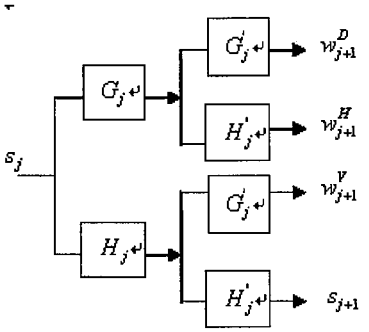


Fig1. two-dimensional image stationary wavelet transform

As shown in Figure 1, using filter convolution is expressed stationary wavelet transform, $w_{j+1}^D = s_0 * L_j^D$, where * denotes the convolution , $L_j^D = H_0 * H_0^{'} * \ldots * H_{j-1} * H_{j-1}^{'} * G_j * G_j^{'}$; Similar can be found: $w_{j+1}^H = s_0 * L_j^H$, $w_{j+1}^V = s_0 * L_j^V$, Among them, $L_j^H = H_0 * H_0^{'} * \ldots * H_{j-1} * H_{j-1}^{'} * G_j * H_j^{'}$

$L_j^{V} = H_0 * H_0^{'} * \ldots * H_{j-1} * H_{j-1}^{'} * H_j * G_j^{'}$

Formula (4) using Wiener filter to estimate the wavelet coefficients[3,6] ,

$$\hat{u}_j = c \cdot w_j \qquad (5)$$

Among them, $c = \dfrac{\sigma_{u_j}^2}{\sigma_{u_j}^2 + \sigma_j^2}$, $\sigma_{u_j}^2$ is the variance of u_j , σ_j^2 is the variance of v_j . the noise variance at scale j is: $\sigma_j = \left\| L_{j-1} \right\| \sigma$, L_{j-1} correspond to vertical, diagonal or horizontal filter, $\|\bullet\|$ is norm operation. The standard deviation of the noisy-free image is to estimate through $\hat{\sigma}_{u_j}^2 = \sigma_{w_j}^2 - \sigma_j^2$, Among them, $\sigma_{w_j}^2 = \dfrac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} w_j^2(m,n)$ 。

3.2 Noise-free algorithm based on the intra- scale coefficients of correlation

As the stationary wavelet transform is a non-orthogonal wavelet transform, it has redundancy, the image signal has a strong correlation between the wavelet coefficients in the adjacent scales and the adjacent scales, and Noise wavelet coefficients are weakly related or unrelated. Using the inter-scale of correlation to estimated wavelet coefficients:

$$\begin{cases} \vec{w}_j(m,n) = \left[w_j(m,n) \quad w_{j+1}(m,n) \right]^T \\ \vec{u}_j(m,n) = \left[u_j(m,n) \quad u_{j+1}(m,n) \right]^T \\ \vec{v}_j(m,n) = \left[v_j(m,n) \quad v_{j+1}(m,n) \right]^T \end{cases} \qquad (6)$$

By the formula (6), there are $\vec{w}_j = \vec{u}_j + \vec{v}_j$, Then the minimum mean squared error (MMSE) estimator:

$$\hat{\vec{u}}_j = P_j (P_j + R_j)^{-1} \vec{w}_j \qquad (7)$$

Among them,

$$P_j = E\left[\vec{u}_j \vec{u}_j^{\,T}\right] = E\begin{bmatrix} u_j^2 & u_j u_{j+1} \\ u_j u_{j+1} & u_{j+1}^2 \end{bmatrix} \quad R_j = E\left[\vec{v}_j \vec{v}_j^{\,T}\right] = E\begin{bmatrix} v_j^2 & v_j v_{j+1} \\ v_j v_{j+1} & v_{j+1}^2 \end{bmatrix}$$

To do this requires the following estimates:

$$E[w_l w_k] = \frac{1}{M \times N} \sum_{m=1}^{M} \sum_{n=1}^{N} w_l(m,n) \cdot w_k(m,n)$$

$$E[u_l u_k] \approx E[w_l w_k] - E[v_l v_k]$$

$$E[v_j v_{j+1}] = \rho_{j,j+1} \sigma_j \sigma_{j+1}$$

Among them, $l,k = j, j+1$, then we get $\hat{\vec{u}}_j$. Estimated from $\vec{u}_j$, $\vec{u}_{j+1}$ can estimates from $\hat{\vec{u}}_{j+1}$. As the noise v_j and v_{j+1} is the noise at different scale wavelet space mapping, the correlation coefficient $\rho_{j,j+1}$ between them is estimated by the following formula [3],

$$\rho_{j,j+1} = \frac{\sqrt{\sum_l \sum_k L_{j-1}(l.k) L_j(l.k)}}{\|L_{j-1}\| \cdot \|L_j\|} \tag{8}$$

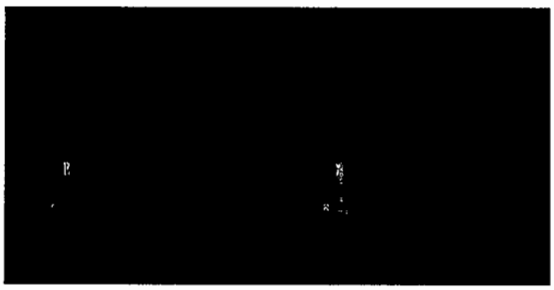
3.3 **Denoising algorithm based on the inter- scale and intra-scale dependencies**

It can be seen in formula (7), estimate of the wavelet coefficients do not take into account same scales correlation of the wavelet coefficients. If a coefficient is greater difference between the mean than its neighborhood coefficients, we need to adopt the following formula to amendment the neighborhood of wavelet coefficients.

$$U_{j,m,n} = \begin{cases} \hat{u}_{j,m,n}, & |\Delta| < 3\delta \\ \bar{u}_{\kappa s(m,n)} + \hat{u}_{j,m,n} \cdot sign(\Delta) \cdot \exp\left\{1 - \dfrac{|\Delta|}{3\delta}\right\}, & |\Delta| \geq 3\delta \end{cases} \tag{9}$$

$sign(x) = \begin{cases} 1, & x \geq 1 \\ -1, & x < 0 \end{cases}$, $\hat{u}_{j,m,n}$ indicated the Estimated Wavelet Coefficients in formula (7), $\kappa s(m,n)$ indicated the neighborhood of (m,n) ,the neighborhood identification is critical. As the details of different directions are different, The location (m,n) of the estimated wavelet coefficients $\hat{u}_{j,m,n}$ which are adjacent in the horizontal it is the horizontal component of the neighborhood; Similarly, the diagonal components and the vertical component is two adjacent point of vertical and diagonal directions; $\bar{u}_{\kappa s(m,n)}$ indicated that in (m,n) neighborhood $\kappa s(m,n)$ three wavelet estimate coefficient mean value ; δ indicated tha in neighborhood $\kappa s(m,n)$ three wavelet estimate coefficient standard mean-square deviation. $\Delta = \hat{u}_{j,m,n} - \bar{u}_{\kappa s(m,n)}$ in formula (9), it will effectively remove the estimated wavelet coefficients inaccuracy,uses the neighborhood information to amend $\hat{u}_{j,m,n}$.

4 SIMULATION RESULTS



(a) Infrared Image Denoising (b) Noisy Image (SNR:12.6236dB)

(c) Denoising Results (SNR:23.8337dB) (d) Denoising Results in this paper (SNR:24.7336dB)

Fig2. Simulation Results when Noise Variance=20

(a) Noisy Image (SNR:7.7628) (b)Denoising Results (SNR:21.6278)

(c) Denoising Results in this paper (d) Wiener Denoising Results

(SNR:22.9220) (SNR:14.9591)

Fig3. Simulation Results when Noise Variance=35

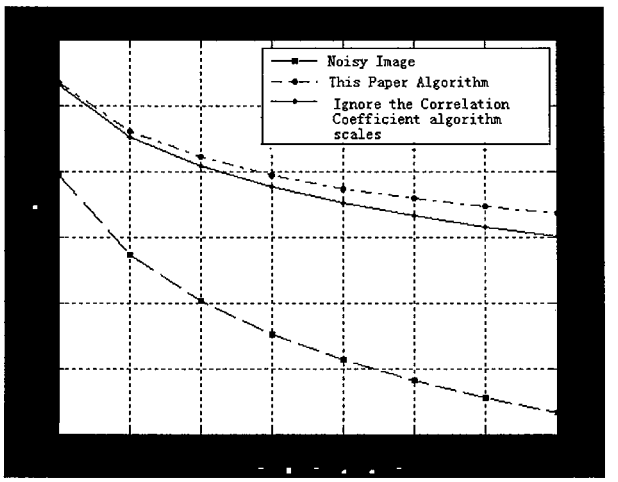


Fig4. SNR curve with the Noise Variance

5 CONCLUSION

A methods using stationary wavelet transform with Inter-scale and Intra-scale Dependencies for infrared image denoising is proposed in this paper. In the calculation of the estimated wavelet coefficients, considered its neighborhood relations between wavelet coefficients and thereby enhanced the estimation accuracy of signal wavelet coefficients. The experimental results show the infrared image can be denoised better than the method neglecting the correlations between Intrascales and have a well SNR as well as the visual quality.

REFERENCES

[1].Zhang Changjiang,Fu Jinyin,Jin Mei. Based on discrete orthogonal wavelet transform infrared image denoising method [J], Infrared and Laser Engineering ,2003, 32(4):401-406.

[2].ZHANG L, PAUL B.WU X. Multiscale LMMSE-based image denoising with optimal wavelet selection[J]. IEEE Trans Circuits and Systems for Video Technology, 2005, 15(4):469-481.

[3].PORTILLA J, STRELA V, WAINWRIGHT M J, et al. Image denoising using scale mixtures of gaussians in t he wavelet domain[J]. IEEE Trans. Image Processing, 2003, 12(11): 1338-1351.

[4].Hou Jianhua,Tian Jinwen,Liu Jian. Local wavelet domain Wiener filter estimation error analysis and image denoising [J], Photon Journal ,2007,36(1): 188-191.

[5].C. Chaux, A. Benazza, P. Jean. A block-thresholding method for multispectral image denoising[C], Proceedings of the SPIE, 2005, Volume 5914: 495-507.
[6].Blu. T, Luisier.F, Lausanne, The SURE-LET Approach to Image Denoising[J], IEEE Transactions on Image Processing, 2007, 16 (11) : 2778-2786.

Research of Digital Controlled DC/DC Converter Based on STC12C5410AD

Chen Dan-jiang Jin Xin Xiao Zhi-hong

Telecommunication Academe of Zhejiang Wanli University, Ningbo Zhejiang 315100, China
Email:Cdj02@163.com, jin06011609@163.com, jhxiaozh@hotmail.com

ABSTRACT

In order to study application of digital control technology on DC/DC converter, principle of increment mode PID control algorithm was analyzed in the paper. Then, a SCM named STC12C5410AD was introduced with its internal resources and characteristics. The PID control algorithm can be implemented easily based on it. The output of PID control was used to change the value of a variable that is 255 times than duty cycle, and this reduced the error of calculation. The valid of the presented algorithm was verified by an experiment for a BUCK DC/DC converter. The experimental results indicated that output voltage of the BUCK converter is stable with low ripple.

Digital control; DC/DC converter; PID algorithm

1 INTRODUCTIONS

Switching mode power supply (SMPS) is a new type of power supply which stabilizes its output voltage by changing on and off time of a power switch based on modern power electronic technology. Since the age of 1990th, SMPS is used widely in electronic device, communication power supply, and so on. SMPS controlled by microcomputer is more intelligent and has more functions, such as real-time monitoring, automatically displaying the power status, programmed by keys, and so on. At the same time, it is easy to limit the output current and achieve soft-start by microcomputer. [1][2]

2 CONSTRUCTION OF DIGITAL CONTROLLED DC/DC CONVERTER

As shown in figure 1, main circuit is a BUCK converter. Output voltage of the converter, V_{out}, is divided by two resistors. The sampling voltage across R_2 is send to the ADC unit of microcomputer and the corresponding output signal is a digital voltage V_c.Reference voltage V_{ref} is connected to I/O port of the microcomputer and can be modified by keyboard, so it is not necessary to add another ADC unit for this voltage. Both of the voltages are used as input signals of PID algorithm, and output of the algorithm is used to vary the duty cycle of PWM wave, which drive MOSFET of the

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463C · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853570

BUCK converter and keep the output voltage V_{out} stable.

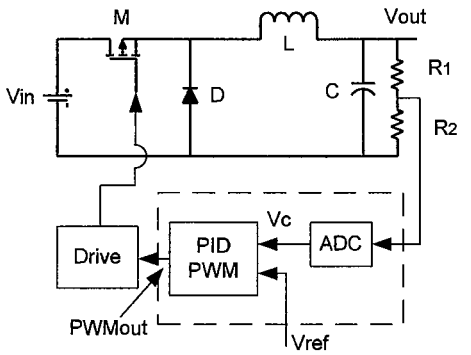


FIG 1 Diagram of digital controlled DC/DC converter

3 INCREMENT DIGITAL PID CONTROL ALGORITHM

PID control is the most popular algorithm in the analog control system. The corresponding one in computer control system is digital PID Increment digital PID control algorithm is used in this system and a typical rule is as follows:[3][4]

$$e(k)=r(k)-c(k) \qquad (1)$$

Where $c(k)$ denotes the kth sampling voltage, like V_c in figure 1, and $r(k)$ the reference voltage at the same time. The diagram of the increment digital PID control algorithm is shown in figure 2.

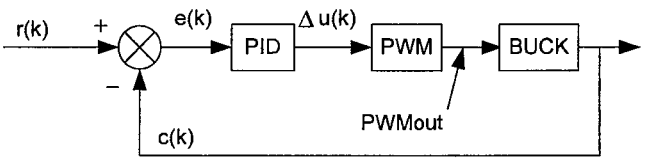


Fig 2 Diagram of the increment digital PID control algorithm

The detailed algorithm for PID is given by:

$$\Delta u(k)=K_P[e(k)-e(k-1)]+K_I e(k)+$$

$$K_D[e(k)-2e(k-1)+e(k-2)] \qquad (2)$$

K_P, K_I and K_D are the proportion, integral and differential coefficient respectively. $e(k-1)$ and $e(k-2)$ are the (k-1)th and (k-2)th errors between the reference voltage and the sampling output voltage. $\Delta u(k)$ is the result of the digital PID algorithm, which will be discussed later. The detailed algorithm given by equation (2) is shown in figure 3.

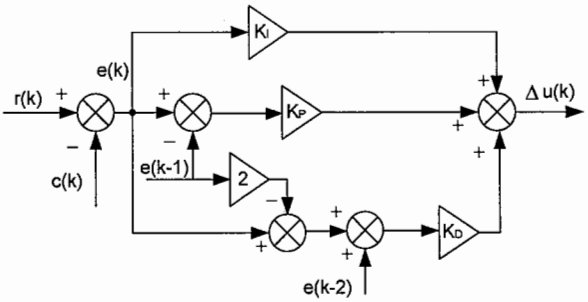


Fig 3 Detailed algorithm of increment digital PID

4 IMPLEMENT OF THE INCREMENT DIGITAL PID CONTROL ALGORITHM

The MCU adopted in this system is STC12C5410AD, and it has single clock/machine period, enhanced 51-core. Average speed of it is 8-12 times faster than that of 8051. There are eight channels 10-bits ADC unit and four channels PWM units in the MCU. Therefore, it is very suitable that works as a control center for the digital controlled converter. [5]

There are four groups 8-bits registers named CCAPxL and CCAPxH (contr0l registers of PWM unit, value of x is from 0 to 3) in the MCU. Value of these registers decides duty cycle of the PWM waveform. In software, when CCAP2L=CCAP2H=PWMDATA , duty cycle of the PWM wave can be given by:

$$D = \frac{255 - PWMDATA}{255} \qquad (3)$$

By varying PWMDATA from zero to 255, the value of duty cycle can be modified from one to zero. $\Delta u(k)$ is output of PID control shown in figure 3, and it is used as an increment to change value of the variable PWMDATA. The benefit of this algorithm is obviously. When a constant sampling period is adopted, if variables K_P, K_I and K_D are determined, value of $\Delta u(k)$ can be got from equation (1) and equation (2). The value of $\Delta u(k)$ is an increment of PWMDATA and is a very small number. If value of $\Delta u(k)$ is not reasonable, it can be deleted by software, so misoperation can be avoided easily. In addition, $\Delta u(k)$ is used as increment of PWMDATA instead of duty cycle. The reason is that value of PWMDATA is 255 times than duty cycle, and this can reduce the error in calculation.

5 EXPERIMENT RESULTS

An experiment prototype is set up based on the theory above to verify it. Switching frequency is 45KHz. PID parameter are as follows: $K_P = 0.175$, $K_I = 0.00625$ and $K_D = 0.0875$. Drive waveform of MOSFET in figure 1 is shown in figure 4.

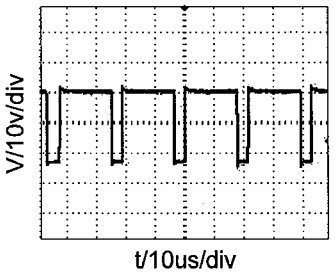


Fig 4 Drive waveform of MOSFET

Figure 5 and Figure 6 is the input and output voltage waveforms of the converter. The upper waveform is input DC source of the converter, as shown in figure 1. The lower waveform is output voltage of the converter and its average value is about 12V. As can be seen from these two Figures, the output voltage ripple is about 400mV. When increasing the input voltage ripple, we can see that the output voltage is almost keeping stable.

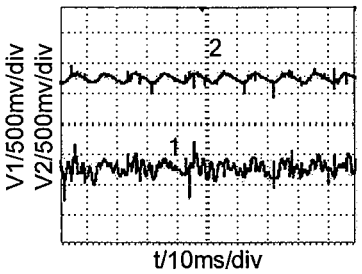


Fig 5 Input and output voltage of the converter (input ripple is small)

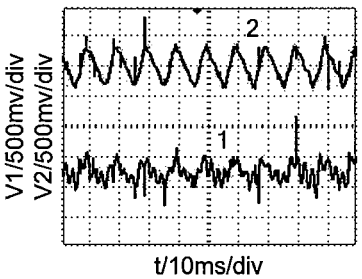


Fig 6 Input and output voltage of the converter (input ripple is large)

6 CONCLUSIONS

Principle of increment PID control algorithm and its implementation method were analyzed in the paper. Results of an experiment prototype based on MCU STC12C5410AD and BUCK converter verify the valid of the theory proposed. The experiment results show that output voltage of the converter controlled by digital PID algorithm has small ripple. The waveforms of input and output voltage of the experiment show that output voltage is stable even the input voltage's ripple is changed.

Reference

[1]Wang xiaoshuang. Switching mode power supply controlled by single chip microcomputer 89C51[J]. Telecom Power Technologies. 2004, 21(4):31-36.

[2]Hang lijun, Yao wen-xi, Lu zhengyu. Implementation of digital increment mode pi algorithm in bi-directional dc/dc converter[J]. Power Electronics. 2004, 38(3):42-44.

[3] Rong xiaofei, Wen yingying. The design of a digital pi controller[J]. Industry Control and Application . 2008, 27(10), 42-44.

[4]Tao yonghua. New PID control and its application[M]. China Machine Press. 2003, 1-7.

[5] Datasheet of STC12C5410AD. www.mcu-memory.com.

[6]Chen chuanshuo, Tian lihua. Tuning method of PID parameter-contorl[J]. Journal of Changchun Post and Telecommunication Institute. 1994, 12(1),9-16.

[7] Jin qi, Deng zhijie. PID control principle and technique of parameter tuning[J]. Journal of Chongqing Institute of Technology (natural science). 2008, 22(5), 91-94.

The Realization of Data Communication in the Intelligent Vehicle Dispatching System

Guan Zongqi Jiang Liulu Xiao Zhihong

(Department of Electronics Information, Zhejiang Wanli University, Ningbo, 315100, China)
guanzongqi@163.com 344903483@qq.com jhxiaozh@hotmail.com

ABSTRACT

The vehicle dispatching system (VDS) is a kind of real-time management system for moving objects. It is developed to meet with the requirement of vehicle orientation and dispatching. The GPS, mobile communication, data communication, and computer and GIS techniques are integrated into this system. With these techniques, the VDS is described, and its functions are analyzed. Then the communication between mobile terminals and the server is designed based on Winsock and Java multithreads technique.

Vehicle Dispatching System ; Winsock ; multithread ; Java

1 INTRODUCTION

With the accelerating pace of people's lives, city transport problems have become increasingly of concern to people. Improving the urban problems of traffic congestion, improving transit operational efficiency, holding operation of vehicles in time, all above requests promoted ITS (Intelligent Transportation System) development. The paper carried out research and exposition aimed at the core issue in ITS —— Vehicle Scheduling. And take advantage of Java programming and Winsock data communications to carry out the design and implementation.

2 SYSTEM OVERVIEW

The system is divided into two parts from the macro level: vehicle terminals and services dispatch center. The system block diagram is shown in Figure 1:

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463D · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853571

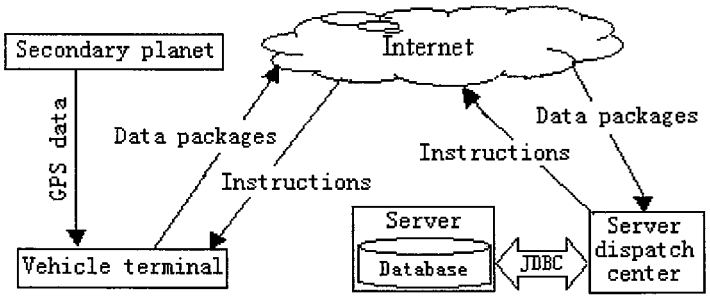


Fig1. The sketch map of system composing

2.1 **Vehicle Terminals**

The vehicle terminals is composed of MCU、GPS receiving and processing module、GSM/GPRS communication module、liquid crystal display module、voice module and electronic map. With these modules, the terminals can complete the functions of location information collection、data transmission、voice calls and location display.

First, use MCU to control GPRS to make vehicles "real-time" online. Because GPRS service charged according to data traffic, so when no data traffic happened, there will be no fees even user hanged in the line. Use the serial of MCU to control GPS module to get GPS data and store it in a fixed area. Then the current location information will be displayed on the electronic map. The service dispatch center will establish communication links with the terminal by sending IP address and port number. The terminal packed GPS data into UDP/IP pack, sent it to the dispatching centre. So the centre can get the terminal's location information. At the same time, the terminal can receive server command and start voice module to promote. It also can carry out GSM voice communication according to need by controlling the communication module.

2.2 **Service Dispatch Center**

Vehicles sent the address information to the server. The server stored the information to the database and continuously updated the address information of vehicles. In this way, the center can consider both arrangements for the bus lines and information sent by the vehicles (such as traffic congestion, locomotive fault, etc), then send dispatch command to the drivers. The drivers can timely adjust vehicle operation. The system has inquiry functions of relevant data about vehicles、routes and roads, which will help to achieve effective management.

In the system, the program of server is written in Java language. Java has increasingly won the favor of people since Sun corporation released Java specification in May 1995. It's not just a language, but has also become a technology-Java technology.

Java is a kind of object-oriented cross-platform programming language, it is applied to distributed computing environment. Java supports multi-threading. It has good features such as portability、security、stability and distributing. Java also provides a Swing class libraries and web libraries, it can run on a platform developed Java Virtual Machine (JVM). In this way, it can satisfy people the requirements of visiting the Intranet on different hardware and software

platforms.

3 THE PROGRAM DESIGN OF SERVICE DISPATCH CENTER

The server-side used ordinary computer, dialed-up internet, had an independent IP address. This section mainly completed data sending and receiving. The terminal connected with the server through JDBC (Java Database Connectivity). The following will describe the feature implementation of various parts.

3.1 Data transmission protocol

UDP protocol didn't provide submission and copy protection. It is an unreliable protocol compared to the TCP protocol. But TCP protocol needs to establish a specific transmission channel to control data transmission when it sends data. The data of UDP protocol comes with the transmission control information. Thus, the system spending can be saved by using UDP to transmit data. And UDP is efficient, flexible and easy to implement. Therefore, on the network, UDP is frequently used when the data needs to be retransmitted after a certain time. This condition is just consistent with the method of sending GPS data from vehicle terminal to the server. So we are using the UDP protocol when data transmission.

3.2 Data transmission

Data transmission is completed by transport layer. For the software of transport layer, the message to transmit is destination address and former address priority. As for the message content, the transport layer software doesn't care, just waiting for the database to process.

When the sink received the data, it will send the receipts of successfully handled database to the sender. Then the data transfer flag of sender will be set successful transmission which can be reflected at the interface.

As using GSM and Internet networks, relying on GPRS services, all these made no involving of the underlying data communication problems during exploitation. The exploitation of entire communication protocol is equivalent to the application layer. Each network layer under application layer is "transparent" to communication.

3.3 Unpacking and processing of the data

After the sink received UDP data package, it take out the data information which signed IP address and port o f the sender. The information is taken out by using method of getAddress( ) and getPort( ) and stored in the String variable. Finally the data will be stored in the database through data interface JDBC.

3.4 Code Demo

Due to the space limitations, the following only gives the core part of the code.

3.4.1 The following code mainly achieved to send data packets

```
...
/*Extract system time and accurate to the level of millisecond*/
Calendar calendar = new GregorianCalendar();
Date trialTime = new Date();
calendar.setTime(trialTime);
timel=calcndar.gct(Calendar.HOUR);
...
```

```
/*Put the instructions for sending into the DatagramPacket, send it by the DatagramSocket*/
Datagram=jTextField.getText()+"    "+time;
byte msg[]=Datagram.getBytes();
pac=new DatagramPacket(msg,Datagram.length(),RemoteHost,RemotePort);
UDPServerSocket.send(pac);
...
```

3.4.2 **The following codes fragment handle the events of user clicking button**

```
...
public void actionPerformed(ActionEvent e){
if(e.getSource()==jButton){
/*call send thread*/
      Thread send=new Thread(new Send(),"thread");
      send.start();
      }
}
...
```

3.4.3 **The following codes fragment complete receiving and establish a connection with the database**

```
ServerPacket=new DatagramPacket(dataBuf,512);
UDPServerSocket.receive(ServerPacket);
RemoteHost=ServerPacket.getAddress();
RemotePort=ServerPacket.getPort();
Datagram=new String(dataBuf,0,0,ServerPacket.getLength());
...
/*establish JDBC-ODBC bridge*/
try {
   Class.forName("sun.jdbc.odbc.JdbcOdbcDriver");
}catch(ClassNotFoundException e) {}
/*connecting to the database*/
try{
Connection   con = DriverManager. getConnection("jdbc:odbc:mydata", "myLogin", "myPassword ");
}catch(SQLException e) {}
...
```

3.5 **Database Design**

Java can access a variety of database formats, we choose SQL Server 2000. The database has diary、driver and boot database. In the diary database, it records events happened in scheduling with fields of date、time and event types included. In the boot database, it records boot and shutdown time of dispatcher with fields of date、time and event included. In the driver database, it records the information of drivers with fields of identity code, vehicle number, driver's

name、telephone number、address and the other information.

4 CONCLUSION

The paper have initially completed the vehicle scheduling system to achieve the above functions through design and experiments. The data transmission is very stable as a result of Winsock transmission of data. The system has good support to the internal networks of bank, government and corporate. It can also be applied to taxis, postal vehicles, the management of navigation lights and the monitoring of ship.

REFERENCES

[1] Yin Zhaolin. Java network programming[M]. National Defense Industry publishing. Beijing. 105-179 (2001).

[2][America] Dan D.Gtierrez. Web database exploitation technology and example explain. Electronics Industry Publishing. Beijing. 147-204 (2000).

[3] JavaTMRemoteMethodInvocationSpecification,JDK1.2. http://Java.sun.com. (1997).

[4] (America) Wayne Tomasi, Electron communications system(Version 4), Electron Industry Publishing, Beijing, (2002).

Effect of selected attribute filters on watermarks

Florence Tushabe[1] and M. H. F. Wilkinson[2]

[1]Faculty of Computing and IT, Makerere University, P. O. Box. 7062, Kampala, Uganda.
[2]Institute of Mathematics and Computing Science, University of Groningen,
P. O. Box 407, 9700 AK Groningen, The Netherlands

ABSTRACT

This paper shows the effect that selected attribute filters have on existing watermarks of an image. Seven transform domain watermarking algorithms and five attributes have been investigated. The attributes are volume, gray-level, power, area and vision. Apart from only one, all of the filters have been found not to affect the underlying watermarks.

Keywords: Attribute Filtering, mathematical morphology, robust watermarking, attacks to watermarks, power gray-level attribute, volume attribute, vision attribute, area attribute.

1. INTRODUCTION

Attribute filtering is emerging as a powerful tool for enhancing various applications within the image processing domain. Tushabe and Wilkinson[1] report how attribute filters improve compression results in terms of both size and quality. They also greatly improve retrieval results for large scale content-based image retrieval[2], for diatom analysis and classification[3] and early skin cancer detection[4]. An attribute filter removes elements of an image that do not satisfy a given criteria, T, of a given attribute like area, perimeter, volume, circularity, etc [5]. The remaining components are then left untouched and this ensures edge preservation. Attribute filtering is an attractive option for users because it offers more flexibility in terms of what parts of an image to remove, the choice of attributes and filtering rules to use, as well as the guarantee of shape preservation for the remaining parts.

With the increasing adoption of attribute filters as a pre-processing step for various applications, concerns emerge about the impact this has on existing watermarks. A watermark is data that is embedded into a media or multimedia object to enhance or protect its value[6]. When extracted from an image, a watermark can prove ownership, verify integrity and convey object specific information like the buyer or seller, date taken, number of copies distributed and so forth.

This work investigates the effect that attribute filtering has on watermarked images. The study sought to find out how the attributes of gray-level, area, power, volume and vision affect watermark signatures if they are applied to an image that remains visually lossless. These attributes were selected because they have successfully been used in previous applications. For example, area was used for shape analysis[2], the improvement of image retrieval[3] and the identification of suspicious moles[4]; volume, vision and power for the improvement of compression[1] and gray-level within watermark insertion.

In this paper, we show that attribute filtering and in particular, filtering by the selected attributes does not compromise existence of watermarks in images. In Section 2 and 3, we briefly discuss the theory behind watermarking and attribute filtering respectively. Section 4 presents the results of experiments conducted and some concluding remarks are given in Section 5.

2. WATERMARKING

Image watermarking is performed either in the spatial or transform domain[7]. Spatial domain techniques directly manipulate pixel contents through either masking approaches or least significant bit modification. While transform domain techniques convert the image into frequencies of various coefficients. Discrete cosine transforms (DCT) and Discrete wavelet transforms (DWT) are the most common representations of image watermarking schemes in the

[1] www.cit.mak.ac.ug/staff/tushabe; tushabe@cit.mak.ac.ug
[2] www.cs.rug.nl/~michael/; m.h.f. wilkinson@rug.nl

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463E · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855420

transform domain. Generally, for an NxN image, an NxN transformation is implemented to produce NxN coefficients. The watermark w is converted into an n length vector following a preferred distribution like a Gaussian or normal distribution and then added to chosen n coefficients $n \leq N$ to produce the watermarked image.

Watermarks can either be robust or fragile. A robust watermark is one that will not alter or be removed after a slight modification on the image is performed [6]. This could be common image processing techniques like re-scanning, addition or multiplication of uncorrelated noise, multiple watermarking, inverse deduction and conspiracy attacks. Others can be estimation of the original image and blind modification through operations including quantization, re-sampling, analog-to digital conversion, rotation, scaling and cropping e.t.c [8].

On the other hand, fragile watermarks are those designed to alter or be destroyed after slight modifications to the image[8]. Fragile watermarks are preferred in circumstances of image authentication or integrity checking like for fingerprint database forensics. While robust watermarks are preferred for copyright protection especially proving ownership.

This work tested one DCT[9] and six DWT[10, 11, 12, 13, 14, 15] embedding techniques because of their robustness to attacks. Cox et al [9] embeds the signature onto the largest n DCT coefficients. Corvi and Nicchiotti[10] embed the watermark into a multi-resolution approximation image using using an additive embedding formula while Wang et al[13] adopt design principles of successive subband quantization (SSQ) and bit-plane coding. Others add the watermark to a few selected significant coefficients after a three-level decomposition. They include Dugad et al[11] that uses Daubechies-8 filters and Kim and Moon[12] that apply bi-orthogonal filters and a level-adaptive thresholding scheme. Two-level decompositions are represented by Xia et al[14] in which the Haar wavelet filter produces the watermark that is added to the large coefficients of the high and middle frequency bands of the DWT. While Zhu et al[15] performs a four-level wavelet decomposition in which all high-pass subband coefficients are selected and modified using the additive embedding formula.

3. ATTRIBUTE FILTERING

The attribute-based approach to morphological filtering is a method used to filter out desired parts of an image without changing the shapes of the remaining ones[5]. This is of particular interest to the process of image authentication / validation because shape preservation and hence visual quality is maintained or even improved. The image is decomposed into sets of connected components, each of which adopts a single attribute value, r, and is considered for further processing only when r satisfies a given criterion. The attributes are chosen as deemed fit for the application and examples include area, perimeter, elongation, volume, circularity, etc.

Attribute filtering consists of binary attribute openings, binary attribute thinnings, gray-scale attribute openings and gray-scale attribute thinnings. A binary attribute opening is a trivial opening of the connected opening of a set X. Breen and Jones[5] define a binary attribute opening, T^T of a set X with increasing criterion T as:

$$T^T(X) = \bigcup_{x \in X} T_T(T_x(X)) \tag{1}$$

Where $T_x(X)$ is a connected opening of a set X at point x. It is the connected component of X that contains x if $x \in X$ and ϕ otherwise. T_T represents the trivial opening of a set C if $C \subseteq E$ and T is an increasing criterion. It is C if C satisfies criterion T and ϕ otherwise.

Attribute openings are characterized by being increasing $(C \subseteq D \Rightarrow \psi(C) \subseteq \psi(D))$, idempotent $(\psi\psi(C) = \psi(C))$ and anti-extensive $(\psi(C) \subseteq C)$. Example attributes include area, perimeter and moment-of-inertia.

On the other hand, a binary attribute thinning is a trivial thinning of the connected opening of a set X. It is defined as [5]:

$$\varphi^T(X) = \bigcup_{x \in X} \varphi_T(T_x(X)) \tag{2}$$

Where φ_T represents a trivial thinning of a set $X \subseteq E$ with T being any criterion. It is C if C satisfies T and ϕ otherwise. Attribute thinnings are characterised by being idempotent, anti-extensive and non-increasing $(C \subseteq D \Rightarrow \psi(C) \subseteq \psi(D))$. Example attributes are length, compactness, non-compactness, circularity and entropy.

3.1 The Max-tree

The max-tree approach[16, 17] which was used to implement the attribute thinnings can be briefly described as follows. Let the peak components, P_h^k of an image represent the connected components of the threshold set at gray level h with k from some arbitrary index set. These peak components are arranged into a tree structure and filtered by removing nodes whose attribute values are less than a pre-defined threshold, T. Thus, the max tree is a rooted tree in which each of its nodes, C_h^k, at gray-level h corresponds to a peak component, P_h^k [17]. An example is shown in Figure 1 which illustrates the peak components, P_h^k, of a 1-D signal, the corresponding C_h^k at levels $h = 0,1,2,3$ and the resultant max-tree. Further details can be found from [16, 17]

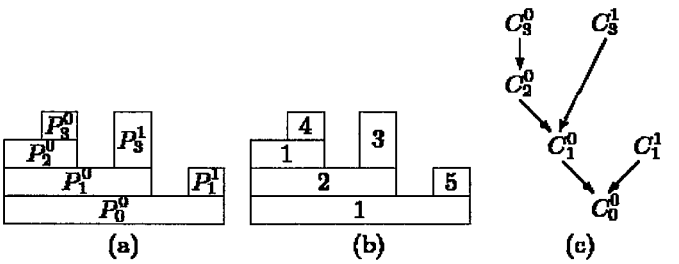


Figure 1: Peak components (P_h^k)(a), possible attribute values (b) and corresponding max-tree - (C_h^k) (c).

3.2 Tested Attributes

In order to check robustness to watermarks, five filters have been analysed:

3.2.1 The Grayscale Attribute: We define the grayscale attribute as the average value of the pixels in a given component of a grayscale image. It represents how dark or light the color of that component or region is. There are 255 possible grayscale values. The grayscale attribute is given by:

$$G(X) = \frac{\sum_{x \in X} V(x)}{\sum(x)} \tag{3}$$

Where X is the set of pixels in the region and $V(x)$ the gray-level value of pixel x.

3.2.2 The Area Attribute: The area attribute is the size of a connected component[18]. It is given by:

$$A(X) = \sum_{x \in X} x \tag{4}$$

Where X is the set of pixels in the region

3.2.3 The Power Attribute: The power attribute[24] measures the power in the components to be removed. The components are selected by considering both the area where the changes will occur and the change in intensity values that will result. It is based on the theory that a small change over a large area will cause the image to seem visually lossless. The formula for calculating the power removed is:

$$P(X, Y, \alpha) = \sum_{x \in X} (Y(x) - \alpha)^2 \tag{5}$$

Where x is the set of pixels in the region, y is the original image and α is the new intensity value of the region.

3.2.4 The Volume Attribute: The volume attribute[19] behaves in a very similar manner to how the power attribute works. Previous experiments have shown that the volume attribute removes more pychovisually redundant information at a

lesser computational time than the power attribute[1]. It is calculated based upon change in intensity over area for P_h^k components. The volume attribute is given by:

$$V(X,Y,\alpha) = \sum_{x \in X} (Y(x) - \alpha) \tag{6}$$

Where x is the set of pixels in the region, y is the original image and α is the new intensity value of the region.

3.2.5 The Vision Attribute: The vision attribute[1] works in a similar manner to the volume attribute, but calculates the change in intensity over area for C_h^k components instead of P_h^k and hence removes all C_h^k nodes that do not comply with the conditions.

4. EXPERIMENTAL RESULTS

We tested the robustness of watermarks against the effects of binary attribute openings and thinnings. By using the direct criteria of the max tree approach (removes nodes if and only if $(r < T)$, the attributes gray-level, area, power, volume and vision were assessed[16, 22]. Twenty two (22) grayscale test images obtained from [20] containing 12 natural small grayscale images and 10 natural medium grayscale images were embedded with watermarks of lengths 100 each. Seven publicly available watermarking algorithms[9, 10, 11, 12, 13, 14, 15] were then investigated using the method below:

-For each watermarking algorithm, create one signature of length 100 and embed within the 22 images;

-Filter each image using any of the five attributes of gray-level, area, power, vision and volume. Visual lossless-ness was ensured by having $PSNR > 26$ and crosschecked with physical examination. For each image, both attribute openings and thinnings were applied (at the same thresholds) in order to improve pyschovisual lossless-ness;

-Extract the watermark from all the images and compare it with the original signature. A correlation coefficient of 0.2 between the extracted and embedded signature means that the watermark has been obtained[23].

4.1 Filtering by Gray-level

Filtering by gray-level removes all components whose gray-level is less than or equal to a given threshold. By using a low parameter of $T = 50$, visual lossless-ness is ensured for all images. Five of the algorithms reported high detection rates of more than 80% while Dugad[11] showed non-robustness with only 36% accuracy. When a high parameter value $T = 150$ was applied, the original image was significantly changed and only one (Xia[14]) of the seven algorithms detected the watermarks. Table 1 summarizes the percentage of images (out of 22) that returned positive detection after filtering by gray-level.

Table 1: Detection of watermarks after gray-level filtering

T/Algorithm	Xia	Cox	Dugad	Zhu	Corvi	Wang	Kim
T = 50	100	100	36	86	73	95	82
T = 100	100	86	14	64	36	73	14
T = 150	95	41	14	59	18	41	0

4.2 Filtering by Area

Filtering by area removes components of the image whose size is less than a given threshold. By using personalised thresholds for all the images, their visual quality was preserved. Four of the algorithms showed perfect accuracy and Kim[12] performed just averagely (55%). Table 2 summarizes the overall performance.

Table 2: Detection of the watermarks after filtering

T/Algorithm	Xia	Cox	Dugad	Zhu	Corvi	Wang	Kim
Area	100	100	73	86	100	100	55
Power	100	100	68	86	100	100	86
Volume	100	91	45	86	73	82	9
Vision	100	100	68	95	100	100	68
Gray-level	100	100	36	86	73	95	82
Average	100	98.2	58	87.8	89.2	95.4	60

4.3 Filtering by Power

Filtering using power was performed by selecting low parameter of $T = 1000$ that maintains visual lossless-ness for all the images. Four of the algorithms reported perfect detection rates (100%) and the others also performed quite well, as shown in Table 2.

4.4 Filtering by Volume

Filtering by volume removes components of the image whose volume is less than a given threshold. By using a general threshold of $T = 5000$ for all the images, their visual quality was heavily degraded and clearly different from the originals. In spite of the visual lossyness, Xia[14] still performed perfectly while Kim[12] detected the watermarks in only two of the images. Table 2 shows further details of the overall performance. Please note that at visual lossless-ness, volume performs similarly to area.

4.5 Filtering by Vision

Personalised thresholds were obtained for all the images in order to ensure preservation of the visual quality. This is largely because, for vision, a small change in threshold causes a big visual difference. Four of the algorithms showed perfect accuracy and the rest also performed quite well too as illustrated in Table 2.

5. DISCUSSION

The results of the experiments conducted above show that the watermarking algorithms tested are quite robust against the attribute filters of gray-level, area, power, volume and vision. It is especially true for Xia[14], Cox[9], Wang[13], Corvi[10] and Zhu[15]. This is evidenced when the volume filter (and most likely area and vision) detected the signatures even after severe degradation. The average performance for Kim[12] (60%) and Dugad[11] (58%) should be received with a little less excitement though. The robustness of attribute filtering to these schemes can be attributed to the preservation of parts with strong signals like the edges. It is worth noting that Dugad[11] is not robust against gray-level filtering since only 8 of the 22 images returned a positive detection at visual lossless-ness.

6. CONCLUSION

This work has summarised the effect of selected attribute filters on watermarks. A marked image was modified by applying a dual attribute opening and thinning filter in order to improve and maintain visual lossless-ness. It can be concluded that of the seven watermarking algorithms tested, only one (Dugad[11]) is not robust against only one filter, the gray-level one. In other words, area, volume, vision, power and gray-level filters are robust to transform domain watermarking schemes. As expected, signature detection becomes harder when large threshold limits are used because more image detail has being discarded. It is interesting to note that even at high threshold levels, most of the algorithms detected the signatures, albeit with lower correlation coefficients. This means that these filters can safely be incorporated as a pre-processing step within applications like compression[1], content-based image retrieval[2] etc without fear of compromising existing signatures.
In future, the robustness of other attributes can be investigated for conclusive remarks about their effect on watermarked images. Spatial domain watermarking algorithms can also be examined.

Acknowledgements

We would like to thank Dr. Peter Meerwald for writing the watermarking codes that were partly used to conduct this research and Nuffic under the NPT project on *Building sustainable ICT Training Capacity in Four Public Universities in Uganda* for sponsoring this research.

References

[1] Tushabe, F., Wilkinson, M. H. F., "Image Preprocessing for Compression: Attribute Filtering", Proc. International Conference on Signal Processing and Imaging Engineering (ICSPIE'07), (2007).

[2] Tushabe, F., Wilkinson, M. H. F., "Content-based Image Retrieval Using Combined 2D Attribute Pattern Spectra", In Advances in Multilingual and Multimodal Information Retrieval, Revised Selected Papers of the 8th Workshop of the Cross-Language Evaluation Forum, CLEF 2007, Budapest, Hungary, September 19-21, 2007, Peters, C.; Jijkoun, V.; Mandl, Th.; Müller, H.; Oard, D.W.; Peñas, A.; Petras, V.; Santos, D. (eds.), Vol. 5152/2008, Springer (2008).

[3] Urbach, E. R., Roerdink, J. B. T. M., Wilkinson, M. H. F., "Connected Shape-Size Pattern Spectra for Rotation and Scale-Invariant Classification of Gray-Scale Images", Trans. Pattern Analysis and Machine Intelligence, 29(2), 272-285 (2007).

[4] Naegel, B., Passat, N., Boch, N., Kocher, M., "Segmentation using vector-attribute flters: methodology and application to dermatological imaging", Proc. 8th ISSM, vol. 1, 239 - 250, (2007).

[5] Breen, E.J., Jones, R., "Attribute Openings, Thinnings and Granulometries", Computer Vision and Image Understanding, 64(3), 377-389, (1996).

[6] Mintzer, F., Braudaway, G. W., Bell, A. E., "Opportunities for Watermarking Standards", Communications of the ACM, 41(7), 57 - 64, (1998).

[7] R. Wolfgang., C. Podilchuk., Delp, E., "Perceptual Watermarks for Digital Images and Video", Proc. IEEE Special Issue Identification and Protection of Multimedia Information, 87(7), 1108 - 1126, (1999).

[8] Lin, E.T., Delp, E.J., "Review of Fragile Image Watermarks", Proc. Multimedia and Security Workshop (ACM Multimedia '99), 25-29, (1999).

[9] Cox, J., Kilian, J., Leighton, T., Shamoon, T. G., "Secure Spread Spectrum Watermarking for Multimedia", Proc. IEEE International Conference on Image Processing (1997), 6, 1673-1687, (1997).

[10] Corvi, M., Nicchiotti, G., "Wavelet-based Image Watermarking and Copyright Protection", Proc. Scandinavian Conference on Image Analysis (SCIA '97), (1997).

[11] Dugad, R., Ratakonda, K., Ahuja, N., "A New Wavelet-based Scheme for Watermarking Images", Proc. 5th IEEE International Conference on Image Processing, (ICIP 98), (1998).

[12] Kim, J.R., Moon, Y. S., "A Robust Wavelet-based Digital Watermark Using level- Adaptive Thresholding", Proc. 6th IEEE International Conference on Image Processing (ICIP 99), (1999).

[13] Wang, H-Y., Su, P-C., Jay Kuo, C-C., "Wavelet based digital image watermarking", Optics Express, 3, 491 - 497, (1998).

[14] Xia, X-G., Boncelet, C. G., Arce, G.R., "Wavelet Transform based watermark for digital images", Optics Express 3, 477 - 523, (1998).

[15] Zhu, W., Xiong, Z., Zhang, Y-Q., "Multiresolution watermarking for images and video: A unified approach", Proc. 5th IEEE International Conference on image processing, (ICIP 98), (1998).

[16] Salembier, P., Oliveras, A., Garrido, L., "Antiextensive connected operators for image and sequence processing", IEEE Transactions In Image Processing, 7(4), 555-570, (1998).

[17] Meijster, A., Wilkinson, M. H. F., "A comparison of Algorithms for Connected Set openings and Closings", IEEE Transactions. Pattern Analysis and Machine Intelligence, 34(4), 484-494, (2002).

[18] L.Vincent., "Greyscale area openings and closings, their efficient implementation and applications", Proc. 1st Workshop on Mathematical Morphology and its Application to Signal Processing, 22 – 27, (1993).

[19] Vachier, C., "Utilisation d'un critère volumique pour le filtrage des images", Proc. reconnaissance des formes et intelligence artificielle, 307 – 315, (1998).

[20] Waterloo Factual Coding Project, University of waterloo image set.
http://links.uwaterloo.ca/oldwebsite/bragzone.base.html

[21] Urbach, E. R., Wilkinson, M. H. F., "Shape-Only Granulometries and Grey-Scale Shape Filters", Proc. ISMM 2002, 305-314, (2002).

[22] Meerwald, P., "Digital Image Watermarking in the Wavelet Transform Domain", Masters Thesis, Department of Scientific Computing, University of Salzburg, Austria, (2001).
http://www.cosy.sbg.ac.at/~pmeerw/Watermarking.

[23] Young, N., Evans, A. N., "Psychovisually Tuned Attribute Operators for Pre- Processing Digital Video", Journal of Vision, Image & Signal Processing (VISP), 150(5), 277 -286, (2003).

Biometric Image Enhancement using Decision Rule Based Image Fusion Techniques

G Mary Amirtha Sagayee,
PhD Scholar, Anna University, Chennai, Tamilnadu, India, gmasagayee@gmail.com
S Arumugam,
Principal, Nandha Engineering College, Erode, Tamilnadu, India, arumugamdote@yahoo.co.in

ABSTRACT

Introducing biometrics into information systems may result in considerable benefits. Most of the researchers confirmed that the finger print is widely used than the iris or face and more over it is the primary choice for most privacy concerned applications. For finger prints applications, choosing proper sensor is at risk. The proposed work deals about, how the image quality can be improved by introducing image fusion technique at sensor levels. The results of the images after introducing the decision rule based image fusion technique are evaluated and analyzed with its entropy levels and root mean square error.

Keywords: Finger Print Image, Wavelet Neural Network, Image Fusion, Entropy, RMSE.

1. INTRODUTION

Biometric is the measurement of personal physical features, actions or behavioral characteristics that distinguish between individuals. The biometric systems are socio-technical systems, comprising both the technology and the environment in which technology operates. Biometrics can be applied to two types of applications: access control and identification. Especially in identification applications the error rates are high level due to the environment where it is applied or used. In such cases, the uncertainty will arise in measuring process itself. Also, this systems have social, ethical, and acceptance issues. As such there are many factors that need to be considered. Because of its uniqueness, it is widely accepted.

2. BIOMETRIC SYSTEM

It consists of many components like data collection, data storage, signal processing, a decision rules and transmission techniques. Each components of this system has several functions like the sensor function which acquires the biometric data or signal, the feature extraction system which extracts features and represent it as signal for storage, and the matching function then processes the matching scores which determines or verifies the identity of an individual. Biometric system has two distinct applications such as identification and authentication. For example, an Automatic Fingerprint Identification and Verification system have 2 modes of operation such as criminal/searching in database for a match and Civilian/One-to-one matching from captured big metrics/templates creation for a database. The performance measuring metrics for this system are accuracy and speed respectively.

2.1 Finger Print:

Finger print recognition is the leading biometric technology and it is the most popular in the market , comprising 32% of the total market[3]. Because it is a proven technology and it is capable of showing very high levels of accuracy. Also, the cost of implementations is very low when compared to other biometric based applications. The major drawback of finger print recognition systems is contact nature of sensors such as inability of the sensing process to accommodate dirt and other environment. The ability of the system to perform well (within the limits of its design) is based almost solely upon the quality of the biometric captured. A well captured biometric is rich in distinguishing

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463F · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855858

information, which in turn gives the feature extraction algorithms the best chance of finding a match with existing records.

2.2 Error rates

In general, all the biometric systems will produce two types of errors due to false positive and false negative occurrence. The accuracy of a biometric system can be measured from two statistical measurements:

(i) The False Match Rate (FMR), also known as a Type I error or False Acceptance Rate (FAR) is a measure of the readings that the system incorrectly matches to its database. The lower the FAR, the better a system's security.
(ii) The False non-Match rate (FnMR), also known as a Type II error or False Rejection Rate (FRR) is a measure of the readings that the system incorrectly fails to match to its database. The lower the FRR, the easier a system will be to use.

The research survey states that most biometric vendors claim a FMR and FNMR for Fingerprint is typically 0.2%. [3] There is a trade off between these two errors, based on decision threshold which has been chosen. It should be chosen in such a way that this decision threshold should reduce the FMR and increase FNMR or vice versa. Also, it varies between technologies based on the discrimination of the biometric data measured.

3. METHODOLOGY

Multi-modal biometrics, or biometric fusion, is the process of combining information from multiple biometric readings, either before, during or after a decision has been made regarding identification or authentication from a single biometric.

Fused Biometric image has the possibility to make identification more secure and more accurate than single biometric systems. How would one replace or overcome the problem of a compromised fingerprint? Biometric Image Fusion could help – by not relying on a single biometric, as could weighting – by more heavily relying on those biometrics not compromised. Proposed design will act intelligently by adopting Artificial Neural Network with layer concepts.

3.1 Image Fusion

The fusion process can take place at different levels of information representation; a generic categorization is to consider the different levels as, sorted in ascending order of abstraction: signal, pixel, feature and symbolic level. In this proposal, we focus on the so-called pixel level fusion process, where a composite image has to be built of several input images. Currently, the result of pixel level image fusion is considered primarily to be presented to the human observer, especially in image sequence fusion (where the input data consists of image sequences).

A possible application is the fusion of forward looking infrared (FLIR) and low light visible images (LLTV). In pixel-level image fusion, some generic requirements can be imposed on the fusion result. The fusion process should preserve all relevant information of the input imagery in the composite image (pattern conservation) The fusion scheme should not introduce any artifacts or inconsistencies which would distract the human observer or following processing stages .The fusion process should be shift and rotational invariant, i.e. the fusion result should not depend on the location or orientation of an object the input imagery .

In case of image sequence fusion arises the additional problem of temporal stability and consistency of the fused image sequence. The human visual system is primarily sensitive to moving light stimuli, so moving artifacts or time depended contrast changes introduced by the fusion process are highly distracting to the human observer. So, in case of image sequence fusion the two additional requirements apply. (i)Temporal stability: The fused image sequence should be temporal stable, i.e. gray level changes in the fused sequence must only be caused by gray level changes in the input sequences, they must not be introduced by the fusion scheme itself; (ii)Temporal consistency: Gray level changes occurring in the input sequences must be present in the fused sequence without any delay or contrast change.

The following sections summarize several approaches to the pixel level fusion of spatially registered input images. Most of these methods have been developed for the fusion of stationary input images. Due to the static nature of the input data, temporal aspects arising in the fusion process of image sequences, e.g. stability and consistency are not addressed. A generic categorization of image fusion methods are linear superposition, nonlinear methods, optimization approaches, artificial neural networks, image pyramids, wavelet transform, generic multi resolution fusion scheme. [2]

3.2 Wavelet Transform

Wavelets are mathematical functions defined over a finite interval and having an average value of zero that transform data into different frequency components, representing each component with a resolution matched to its scale. The basic idea of the wavelet transform is to represent any arbitrary function as a superposition of a set of such wavelets or basis functions. These basis functions or baby wavelets are obtained from a single prototype wavelet called the mother wavelet, by dilations or contractions (scaling) and translations (shifts). They have advantages over traditional Fourier methods in analyzing physical situations where the signal contains discontinuities and sharp spikes.

Calculating wavelet coefficients at every possible scale is a fair amount of work, and it generates an awful lot of data. If the scales and positions are chosen based on powers of two, the so-called dyadic scales and positions, then calculating wavelet coefficients are efficient and just as accurate. This is obtained from discrete wavelet transform (DWT).

The reconstruction of the image is achieved by the inverse discrete wavelet transform (IDWT).Wavelet transform is first performed on each source images, then a fusion decision map is generated based on a set of fusion rules. The fused wavelet coefficient map can be constructed from the wavelet coefficients of the source images according to the fusion decision map. Finally the fused image is obtained by performing the inverse wavelet transform.

When constructing each wavelet coefficient for the fused image. We will have to determine which source image describes this coefficient better. This information will be kept in the fusion decision map. The fusion decision map has the same size as the original image. Each value is the index of the source image which may be more informative on the corresponding wavelet coefficient. Thus, we will actually make decision on each coefficient.

There are two frequently used methods in the previous research. In order to make the decision on one of the coefficients of the fused image, one way is to consider the corresponding coefficients in the source images as illustrated by the red pixels. This is called pixel-based fusion rule. The other way is to consider not only the corresponding coefficients, but also their close neighbors, say a 3x3 or 5x5 windows, as illustrated by the blue and shadowing pixels. This is called window-based fusion rules. This method considered the fact that there usually has high correlation among neighboring pixels.

In our research, we think objects carry the information of interest, each pixel or small neighboring pixels are just one part of an object. Thus, we proposed a region-based fusion scheme. When make the decision on each coefficient, we consider not only the corresponding coefficients and their closing neighborhood, but also the regions the coefficients are in. We think the regions represent the objects of interest.

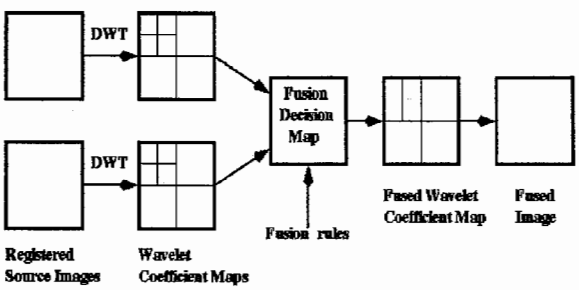


Fig. 1 Block Diagram region-based image fusion Technique.

4. PROPOSED WORK

Neural Network and Fuzzy Logic approach can be used for image fusion. Such a image fusion could belong to a class of image fusion in which case the features could be input and decision could be output. The help of Neuro-fuzzy of fuzzy systems can achieve this. The system can be trained from the input data obtained from the sensors. The basic concept is to associate the given sensory inputs with some decision outputs. After developing the system. another group of input data is used to evaluate the performance of the system.

Following algorithm and .M file for pixel level image fusion using Fuzzy Logic illustrate the process of defining membership functions and rules for the image fusion process using FIS (Fuzzy Inference System) editor of Fuzzy Logic toolbox in Matlab.

4.1 Algorithm

Step 1: Read the images into the variables M1 & M2 respectively. Variables MI and M2 are images in matrix form where each pixel value is in the range from 0-255. Use Gray color map.

Step 2: Apply wavelet decomposition and formspatial decomposition Trees and Convert the images in column form which has C= **zl*sl** entries.

Step 3: Create fuzzy interference system.

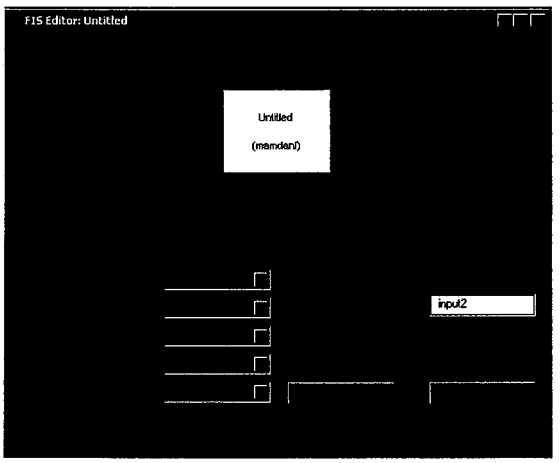


Fig. 2 Fuzzy Interference creation

Step 4: Decide number and type of membership functions for both the input images by tuning the membership functions. Fig 3. Tuning membership functions

Step 5: For num=l to C in steps of one, apply fuzzification using the rules developed above on the corresponding pixel values of the input images which gives a fuzzy set represented by a membership function and results in output image in column format.

Step 6: Convert the column form to matrix form and display the fused image.

4.2 Performance Evaluation

It has been common to evaluate the result of fusion visually. According to visual evaluation, human judgment determines the quality of the image. Some independent and objective observers give grade to corresponding image and the final grade is obtained by taking the average or weighted mean of the individual grades. Obviously this evaluation method has some drawbacks, namely it is not accurate and depends on the observer's experience. For an accurate and truthful assessment of the fusion product some quantitative measures (indicator) is required. Two different measures are used in this project to evaluate the results of fusion process. They are (1) Information Entropy and (2) Root Mean Square Error.

4.2.1 Entropy

One of the quantitative measures in digital image processing is Entropy. Claude Shannon introduced the entropy concept in quantification of information content of messages. Although he used entropy in communication, it can be also employed as a measure and quantify the information content of digital images. A digital image consists of pixels arranged in rows and columns. Each pixel is defined by its position and by its grey scale level. For an image consists of L grey levels, the entropy is defined as:

$$H = -\sum_{i=1}^{L} P(i) \log_2 P(i) \quad \text{..... (1)}$$

where is the probability (here frequency) of each grey scale level. As an example a digital image of type uint8 (unsigned integer 8) has 256 different levels from 0 (black) to 255 (white). It must be noticed that in combined images the number of levels is very large and grey level intensity of each pixel is a decimal, double number. But the equation is still valid to compute the entropy.

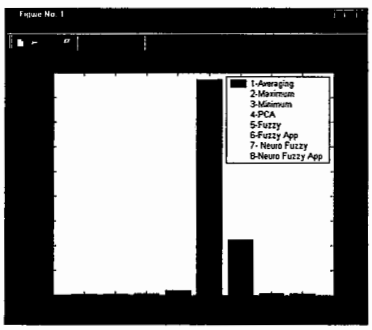


Fig 4. Time Analysis

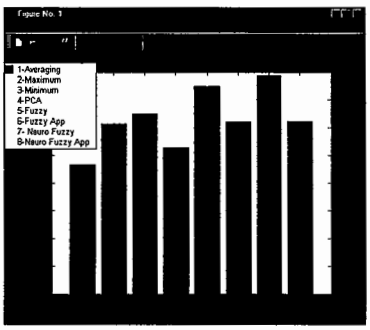


Fig.5 Entropy Analysis

For images with high information content the entropy is large. The larger alternations and changes in an image give larger entropy and the sharp and focused images have more changes than blurred and misfocused images. Hence, the entropy is a measure to assess the quality of different aligned images from the same scene.

4.2.2 Root Mean Square Error

The Root Mean Square Error between the reference image, I and the fused image is defined as: F

$$RMSE = \sqrt{\frac{\sum_{i=1}^{M}\sum_{j=1}^{N}[I(i,j)-F(i,j)]^2}{N.M}} \quad \ldots(2)$$

where and $i\,j$ denotes the spatial position of pixels, M and are the dimensions of the images. N This measure is appropriate for a pair of images containing two objects. First a reference, everywhere-infocus image I is taken. Then two images are provided from this original image. In one image the first object is focused and the second one is blurred. In the other image the first object is blurred and another one is remained focused. The fused image would contain both well-focused objects.

Often the perceptual quality of the resulting fused image is of prime importance. In these circumstances, comparisons of quantitative quality can often be misleading or meaningless. However, a few authors [1, 8, 9] have attempted to generate such measures for applications where their meaning is clearer.

Firstly, a "ground truth" image needs to be created that can be quantitatively compared to the fusion result images. This is produced using a simple cut-and-paste technique, physically taking the "in focus" areas from each image and combining them. The quantitative measure used to compare the cut-and-paste image to each fused image wastaken from [I]

$$\rho = \sqrt{\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}[I_{gt}(i,j)-I_{fd}(i,j)]^2}{N^2}}, \quad \ldots(3)$$

whcrc Igt is thc cut-and-pastc "ground truth" imagc, Ifd is thc fuscd imagc and N is the size of the image. Lower values of _ indicate greater similarity between the images Igt and Ifd and therefore more successful fusion in terms of quantitatively measurable similarity.

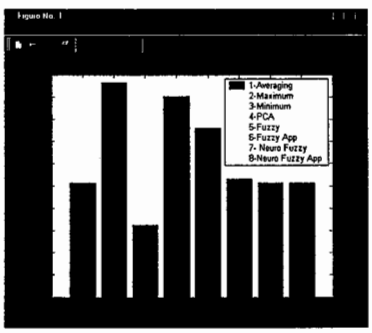


Fig. 6 MSE for Image 1

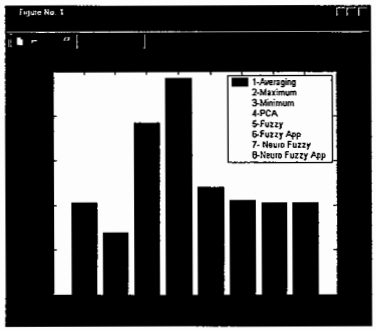


Fig 7. MSE for Image 2

The following table shows the results for the various methods used. The average pixel value method, the pixel based PCA and the DWT methods give poor results relatively to the others as expected.

Different Tech		Entropy	Process Time	Mean Square Error	
				Image1	Image2
Two Max	Averaging	9.3198	0.5150	1.0241e+004	1.0241e+004
	Maximum	12.2813	0.5150	1.9268e+004	6.9486e+003
	Minimum	12.9647	0.5460	6.9486e+003	1.9268e+004
	PCA	10.5563	3.3280	1.8058e+004	2.4251e+003
Two Max Fuzzy		14.9813	174.7040	1.5173e+004	1.2026e+004
Two Max Fuzzy App		12.4181	44.7030	1.0643e+004	1.0581e+004
Two Max Neuro Fuzzy		15.7258	0.8750	1.0241e+004	1.0241e+004
Two Max Neuro Fuzzy App		12.4556	0.6250	1.0241e+004	1.0241e+004

Table 1 Results of various methods

The DT-CWT methods give roughly equivalent results although the New-CWT method gave slightly worse results. The results were however very close and should not be taken as indicative as this is just one experiment and the transforms are producing essentially the same sub-band norms.

The WBV and WA methods performed better than MS with equivalent transforms as expected in most cases. The residual low pass images were fused using simple averaging and the window for the WA and WBV methods were all set to 3×3. The table shows the best results for all filters available for each method.

5. CONCLUSION

In this work, the use of Discrete Wavelet Transform (DWT) ,Fuzzy, Neuro Fuzzy, the fusion of biometric images taken by digital camera were studied. The pixel-level-based fusion mechanism applied to sets of images. All the results obtained by these methods are valid in case of using aligned source images from the same scene.

In order to evaluate the results and compare these methods two quantitative assessment criteria Information Entropy and Root Mean Square Error were employed. In fact if the result of fusion in each level of decomposition is separately evaluated visually and quantitatively in terms of entropy, no considerable differences are observed. Experimental results demonstrated indicate that LPT algorithm reaches its best quality in terms of entropy in lower levels than DWT.

The RMSE values represented in Table show that neither LPT nor DWT has better performance in all levels, although the best result belongs to the LPT method. However the RMSE results compared to quality and entropy of fused images indicate that RMSE can not be used as a proper criterion to evaluate and compare the fusion results. Finally the experiments showed that the LPT approach is implemented faster than DWT. Actually LPT takes less than half the time in comparison with DWT and with regard to approximately similar performance. LPT is preferred in real-time applications.

Fuzzy and Neuro-Fuzzy algorithms have been implemented **to** fuse a variety **of** images. The results **of** fusion process proposed are given in terms of Entropy and Variance.

6. REFERENCES

1. A.K. Jain, Y. Chen and M. Demirkus, "**Pores and Ridges: High Resolution Fingerprint Matching Using Level 3 Features**", **IEEE** Transactions on Pattern Analysis and Machine Intelligence, Vol 29, No 1, January 2007
2. Huiqing Chen Gang Dong, Andawell Corp., Beijing; "**Fingerprint Image Enhancement** by **Diffusion Processes" Image** Processing, 2006 **IEEE**
3. Rebecca Heyer, "Biometrics Technology Review 2008", Defense Science and Technology Organization.
4. Shutao Li, James T. Kwok, Ivor W. Tsang, Yaonan Wang, " **Fusing images with different focuses using support vector machines"** IEEE Transactions on Neural Networks, 15(6):1555- 1561, Nov. 2004.
5. P. J. Burt and R. J. Lolczynski, " **Enhanced image capture through fusion"** In Proc. the 4th Intl. Conf. on Computer Vision, pages 173-182, Berlin, Germany, May 1993.
6. Z. Zhang and R. Blum, **"A categorization of multiscale-decomposition-based image fusion schemes with a performance study for a digital camera application"** Proceedings of the IEEE, pages 1315 -1328, August 1999.

HB LED Color Mixture Traffic Light Solution Based on EZ-Color

Hong Qunhuan Wang Weijian Xiao zhihong
(Department of Electronic & Informatin Engineering, Zhejiang Wanli University,P.R.China.315100)
zwuhqh@126.com adkings7@126.com jhxiaozh@hotmail.com

ABSTRACT

This paper introduces a traffic light design solution using HB LED color mixture,namely an implementation method using a set of traffic lights composed of red, green and blue HB LED to replace ordinary red, yellow and green lights. This scheme realizes HB LED color mixture lighting design on the basis of Cypress EZ-Color controller and codeless embedded design software PSoC Express.

Keywords: HB LED; color mixture; EZ-Color controller; PSoC Express

0. INTR ODUCTION

Some large and medium-sized cities at home and abroad have replaced traditional traffic lights using incandescent lamps as light sources with traffic lights using LED as light sources in quantities in the past two years. Traditional traffic lights are generally composed of incandescent lamp (or metal tungsten halide lamp), reflector and color lens and this kind of traffic lights have many disadvantages, such as false display, short lifetime, high energy consumption and large carbon dioxide emission, etc. The traffic light replaced by HB LED can not only reduce electric energy consumption and sixty percent carbon dioxide emission but also promote climatic melioration.

The single set of traffic lights composed of red, green and blue LED this paper discusses realizes time interval display of red, yellow and green traffic signals by color mixture control and this kind of application scheme can not only increase traffic light performance and environmental protection index but also greatly reduce traffic light construction costs and maintenance funds of urban highway management departments. This scheme is mainly realized by using a programmable controller EZ-Color specially aimed at HB LED lighting color mixture application and its development tool PSoC Express pushed out by Cypress.

1. BASIC PRINCIPLE OF LED COLOR MIXTURE

LED color mixture refers to combining several LED of independent color to form a color spectrum and using a dynamic color mixture system to carry out additive color mixture,thereby producing various colors by 3 to 4 HB LED of different colors. Shown as Fig 1, we can get any color visible in the color spectrum by color mixture of red, green and blue.

Fig 1 Additive Color Mixture Diagram

Fig 2 Chromatic Diagram of CIE 1931

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463G · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.853568

The color mixture of LED is realized by adjusting the brightness of LED and brightness can be controlled by a PWM and a LED driver, PWM mainly provides different current space factors, the lower space factor is, the darker LED brightness is. The light each LED gives off has different wavelength characteristics, we can get various colors visible to human eyes in Chromatic Diagram of CIE 1931[1] shown as Fig 2 through color mixture of LED of different wavelengths or colors.

Although using LED of different colors can get various colors, the color mixture design of HB LED of different types still faces specification code and temperature compensation problems of LED, because the specification code of HB LED directly influences the accuracy of keeping colors of color points and temperature change will result in instability of color points. In addition, temperature change will influence light output and wavelength change of HB LED. And therefore, the color mixture of HB LED brings engineering design great challenges and EZ-Color controller pushed out by Cypress provides a better solution for the color mixture of HB LED.

2. HB LED COLOR MIXTURE SOLUTION OF EZ-COLOR

2.1 Introduction about EZ-Color controller

EZ-Color controller is a type of chip specially used for HB LED lighting color mixture in multi-series programmable soc chips produced by American Cypress Company, shown as Fig 3. This chip integrates peripheral modules [2], such as 8-bit microprocessor,programmable analog line module, programmable digital line module, hardware multiplication accumulator, Flash, SRAM and I2C communication, etc. EZ-Color has three series products, namely CY8CLED04, CY8CLED08 and CY8CLED16, choosing different products can meet different design requirements, for example, CY8CLED16 can realize a LED lighting color mixture system of 16-way LED control channel and it can be also designed into a lighting color mixture system of 32-bit resolution 4-way LED control channel.

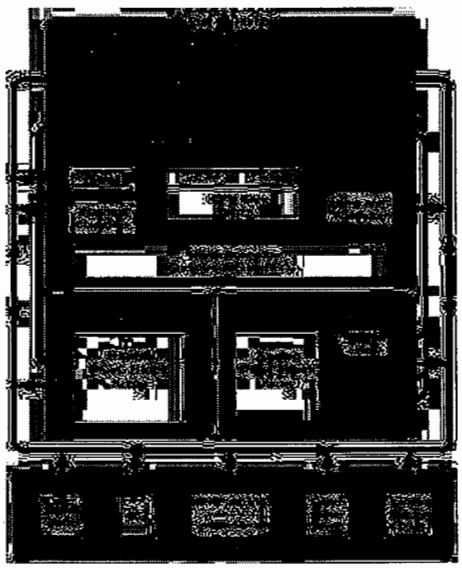


Fig 3 Internal Block Diagram of EZ-Color Controller

LED color mixture scheme of EZ-Color uses (x,y,Y) to represent input, x and y represent color hue and color

saturation, Y represents light flux and the color input request message of EZ-Color is indicated according to the representation method of Chromatic Diagram of CIE 1931. Output is RGB three-color LED whose hardware driver adopts SSDM (random signal strength modulation) line module. SSDM line module is the hardware realization of Cypress PrISM (precise illumination signal strength modulation) technique. PrISM technique compares the value in random counter with that in the signal strength register required by the user, if the value in random counter is less than signal strength value, high level signals are outputted and conversely, low level signals are outputted, thereby producing precise illumination pulse signals whose high and low level widths vary randomly to control the brightness of HB LED. PrISM technique is a LED brightness adjustment technique exclusively owned by Cypress and it can solve electromagnetic interference and low frequency twinkling problems effectively [2].

2.2 Introduction about PsoC Express development tool

PSoC integrated development tool PSoC Express pushed out by Cypress Company is visual embedded system development software [3] which doesn't need assembly language or C language programming, product developers only need to choose an input/output driver according to system requirements, add and define system interfaces, etc, PSoC Express can completely and correctly generate program files and automatically generate design instruction, schematic diagram and device list, etc. After design files are downloaded to PSoC series device through PSoC Programmer and CY3217 MiniProg, product development is finished. And therefore, this development tool can remarkably simplify the design work of embedded system, thereby improving design efficiency and shortening product development cycle.

In order to be convenient for designers to simply and rapidly realize HB LED color mixture design, PSoC Express constructs a color mixture module consisting of red, green and blue LED and this color mixture module has designed the configuration of PSoC internal register, interconnectors between various line modules, LED type code and temperature compensation algorithm software, users only need to select three-color LED color mixture input/output driver and drag and drop it into the design area of PSoC express, all illumination color mixture software will be generated automatically.

In a word, apart from high integration, low cost, good flexibility and easy use, LED color mixture solution of EZ-Color can also solve LED low frequency twinkling and electromagnetic interference problems effectively.

3. DESIGN SCHEME OF HB LED TRAFFIC LIGHT

International universal traffic light is composed of red, yellow and green lights, according to additive color mixture diagram shown as Fig 1, we can know that through additive color mixture of red, green and blue and after being treated by LED light guide plate, yellow light can be got easily. Shown as Fig 4, EZ-Color controls the brightness of HB LED and then light HB LED in different time intervals through power driver, red, yellow and green traffic signals which display alternately and circularly can be realized.

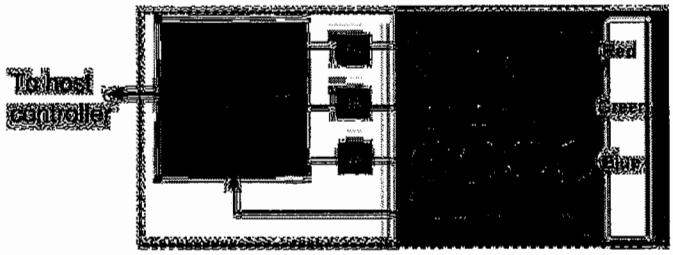


Fig 4 System Diagram of HB LED Controlled by EZ-Color

LED color mixture scheme of EZ-Color uses (x,y,Y) to represent input value, by changing the values of x and y, namely changing the brightness ratio of red light to green light to blue light, various lights can be got. According to chromatic diagram of CIE 1931 shown as Fig 2, set chromatic values of three lights of traffic light as follows.

	Red Light	Yellow Light	Green Light
x	0.75	0.5	0.25
y	0.25	0.5	0.75

For the display of red, yellow and green of traffic lights at urban crossroads, its cycle display time is set according to road traffic, assume the alternate display time of red→yellow→green→yellow→red is as follows: 60s→10s→60s→10s→60s, through the programmable timer in EZ-Color, it can be realized easily by using the loop delay and state machine valuator of PsoC Express.

4. PSoC EXPRESS DESIGN OF HB LED TRAFFIC LIGHT

PSoC Express is a graphic visual design tool, its software is very intelligent, the work that designers need to do is very simple and main steps of its system development are shown as Fig 5 [4].

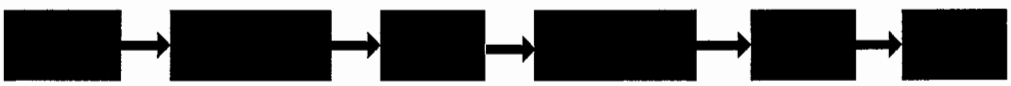


Fig5 Workflow Diagram of PSoC Express System Development

PSoC Express is a tool based on system-level design, it is completely graphical and its operation is simple and understandable. For a control system, designers don't need to know hardware circuit and PsoC chip interior very well. Next, a brief introduction to design method using PSoC Express to develop HB LED color mixture traffic light will be given.

After starting PSoC Express, newly build a project and then establish color mixture drivers of red, yellow and green respectively according to Fig 6 and set parameters of x and y in valuators according to values corresponding to chromatic diagram of CIE 1931. For example, x1Valuator and y1Valuator in the diagram correspond to the driver of red light and parameters are set at 7500 and 2500; x2Valuator and y2Valuator correspond to the driver of yellow light and parameters are set at 5000 and 5000; x3Valuator and y3Valuator correspond to the driver of green light and

parameters are set at 2500 and 7500.

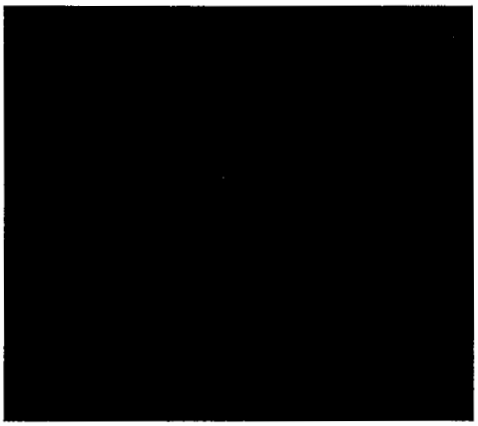


Fig 6 Setting Interface Diagram of Color Mixture Drivers of Red, Yellow and Green

The traffic light displays red, yellow and green lights circularly according to delay time, state diagram can be established through valuator state machine transfer function of PsoC Express and corresponding parameters can be defined according to cycle delay time between signal lights.

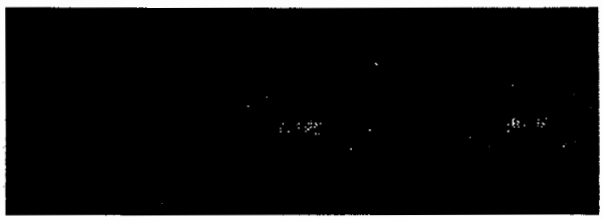


Fig 7 State Diagram of Traffic Light Circle Display

This design adopts a set of red, green and blue HB LED, after color mixture, red, yellow and green traffic signals are outputted respectively, according to time intervals of circle display between red, yellow and green, the three traffic signals control three color mixture drivers respectively to output the current of HB LED and different light colors are realized through the brightness variation of LED. And therefore, color mixture drivers work according to values defined by state transfer function, shown as Fig 8, establish relations between drivers and valuators and define delay time according to specific needs.

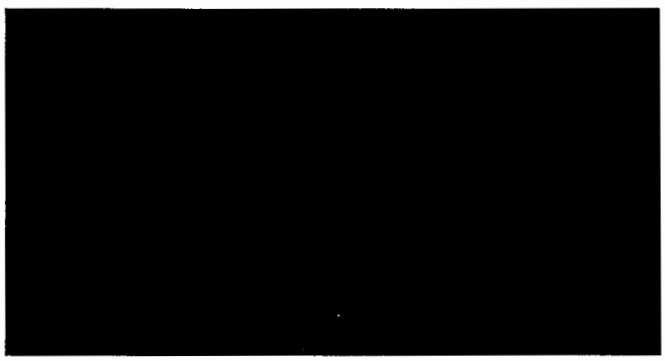


Fig 8 Relational Graph between Color Mixture Drivers and State Function Valuators

5. CONCLUSION

HB LED color mixture lighting technique has been widely used and EZ-Color controller and its development software PsoC Express of Cypress Company provide a great help for designers to develop lighting systems. The design of traffic signal introduced in this paper is only one simple application of EZ-Color, aimed at introducing to vast LED lighting design fans a kind of simple and easy-to-use design method which can be widely used in more LED color mixture lighting designs.

REFERENCES

[1] Xu Fushan, Xu Haisong, Wang Yong, A New Conversion Algorithm between Munsell Color Order System and CIE1931 Standard Colorimetric System [J], Acta Photonica Sinica, 36, 4:650-654.
[2] Ai Xiaohui, HB LED Lighting Color Mixing Solution Based on EZ-Color Controller [J], Electronic Engineering & Product World, 2007, 10:62-66.
[3] Ye Zhaohui, Hua Chengying, Principles and Practical Applications of Programmable System on Chip (PSoC) [M], Beijing: Tsinghua University Press, 2008.
[4] Zhu Mingcheng, Li Xiaobin, PsoC Principle and Application Design [M], Beijing: China Machine Press, 2008.3.

A Study of Interval-Valued Fuzzy Morphology based on the Minimum-Operator

M. Nachtegael[a], P. Sussner[b], T. Mélange[a], E.E. Kerre[a]

[a]Ghent University, Dept. of Applied Mathematics and Computer Science
Fuzziness and Uncertainty Modeling Research Unit
Krijgslaan 281 - S9, 9000 Gent, Belgium
email: Mike.Nachtegael@UGent.be

[b]University of Campinas, Dept. of Applied Mathematics
Campinas, SP 13083 859, Brazil
email: sussner@ime.unicamp.br

ABSTRACT

Pixels of a grayscale image are classically associated with a single grayscale value. However, capturing grayscale images comes along with two kinds of uncertainty: numerical uncertainty (do we measure the actual value of the pixel or just an approximation?) and spatial uncertainty (does the measured pixel correspond to the actual spatial position or has it shifted?). Interval-valued fuzzy set theory provides a framework to model grayscale images of which the captured grayscale values are uncertain. This is realized by associating every pixel with a closed interval of *possible* grayscale values instead of with one *single* value. Based on this image model, a new corresponding morphological framework to process these images (e.g., using dilation and erosion) has been developed. In that way, we are not only able to model the uncertainty that is present during image capturing, but we are also able to process it such that the information regarding the uncertainty is never lost.

In this paper, we study the interval-valued fuzzy morphological model based on the minimum-operator. Properties that are relevant in the context of image processing, as well as some interesting decomposition and construction properties, are discussed. This study gives an insight in the morphological model and will help researchers when they want to apply it in practice.

Keywords: image processing, interval-valued morphology, image acquisition, fuzzy logic

1. INTRODUCTION

"One image says more than a thousand words". This popular phrase, based on the fact that an image can contain an enormous amount of relevant data, expresses the growing importance of images as information carriers in today's world. The scientific and technological innovations and the easy way to capture, develop and publish images has even enhanced this evolution.

Fuzzy set theory[17] – introduced to model uncertain and imprecise information – is one of the mathematical theories that has many applications in image processing. Uncertainty and imprecision are encountered in many image processing applications such as segmentation (which pixels are on an edge?), noise filtering (which pixels are contaminated with noise, and how heavy is the contamination?) or image similarity (how similar are two images?). The theory is also used as a *tool* to construct image processing operators, specifically in the field of mathematical morphology. The basic morphological operators dilation and erosion constitute the fundamentals of this theory[12] and transform an image into another image using a structuring element. As an extension from binary to grayscale morphology, different fuzzy morphological models[3,10,14] have been introduced. These models are based on the observation that grayscale images and fuzzy sets are formally modeled in the same way, and consequently tools from fuzzy set theory can be applied in the context of image processing.

We can also take this a step further, and apply extended fuzzy set theory to *model* the uncertainty that occurs with the image capture itself. In the classical approach, every pixel of a grayscale image is associated with one

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463H · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855669

Figure 1. Different captures of the cameraman image: top = take with cloudy sky, middle = take with sunny sky (causes numerical uncertainty), bottom = take with distortion (causes spatial uncertainty).

Figure 2. Interval-valued representation of the cloudy/sunny/distorted take of the cameraman image: top = lower bound of the measured values, middle = upper bound of the measured values, bottom = representation of the interval width (a high width equals a high uncertainty regarding the measured value). The uncertainty is due to both numerical and spatial uncertainty.

certain grayscale value, indicating the amount of black or white at that specific position in the image. However, due to the circumstances in which images are sometimes captured, the measured values might be uncertain and merely indicate a *likely* value. This uncertainty is an immediate fact if one takes into account that any device will round captured values up or down to the finite set of allowed values. The uncertainty grows if several takes of the same scene reveal slightly different grayscale values for some pixels (due to different recording circumstances) or even slightly shifted spatial positions (due to small distortions). Therefore, it can be useful not to work with grayscale *values* but with grayscale *intervals*, where the interval represents the set to which the actual grayscale value belongs. In this approach, we are actually modeling images as interval-valued fuzzy sets[2,4] and we are led to interval-valued fuzzy mathematical morphology[11,15] (see Figures 1 and 2).

The goal of this paper is to extensively investigate the interval-valued morphological model based on the minimum-operator, one of the basic logical operators in fuzzy set theory. In Section 2 we review the basics regarding interval-valued set theory and interval-valued logical operators. In Section 3 we construct interval-valued morphological models, in particular the model based on the minimum-operator. We present a study of the properties that are relevant in image processing, including decomposition and construction properties that provide a link with binary morphology. We end our paper with concluding remarks in Section 4.

2. INTERVAL-VALUED FUZZY SETS AND LOGICAL OPERATORS

2.1 Interval-valued fuzzy set theory

A fuzzy set A in a universe $\mathcal{U}$ is characterized by an $\mathcal{U} - [0,1]$ mapping, that associates with every element x of $\mathcal{U}$ a membership degree $A(x)$. This can be generalized[6] to the notion of an $\mathcal{L}$-fuzzy set, where $\mathcal{L} = (L, \leq_L)$ represents a complete lattice; a complete lattice is a partially ordered set in which every family of elements has a supremum and infimum. An $\mathcal{L}$-fuzzy set A in $\mathcal{U}$ is characterized by an $\mathcal{U} - L$ mapping; note that the class $\mathcal{L}^{\mathcal{U}}$ of $\mathcal{L}$-fuzzy sets in $\mathcal{U}$ forms a complete lattice if $\mathcal{L}$ is a complete lattice. When $L = [0,1]$, $\mathcal{L}$-fuzzy set theory reduces

to classical fuzzy set theory. An interval-valued fuzzy set (IVFS) corresponds to a mapping A from $\mathcal{U}$ into the class of closed intervals $[\mu_1, \mu_2] \subseteq [0,1]$. Thus, $A(x) = [\mu_1(x), \mu_2(x)]$ for every $x \in \mathcal{U}$. Evidently, if $\mu_1(x) = \mu_2(x)$ for all $x \in \mathcal{U}$ then the interval-valued fuzzy set reduces to a classical fuzzy set.

The class of IVFS can be regarded as $\mathcal{L}^I$-fuzzy sets, with the complete latice $\mathcal{L}^I = (L^I, \leq_{L^I})$ defined by:

$$L^I = \{[x_1, x_2] | [x_1, x_2] \subseteq [0,1]\},$$
$$[x_1, x_2] \leq_{L^I} [y_1, y_2] \Leftrightarrow x_1 \leq y_1 \text{ and } x_2 \leq y_2.$$

Infimum and supremum of a set $C = \{[a_s, z_s] \subseteq [0,1] | s \in I_S \subseteq \mathbb{N}\}$ of closed intervals are given by $\bigwedge C = [\inf_s a_s, \inf_s z_s]$ and $\bigvee C = [\sup_s a_s, \sup_s z_s]$; smallest element $0_{L^I} = [0,0]$, largest element $1_{L^I} = [1,1]$.

Also recall that, for a complete lattice $\mathcal{L} = (L, \leq_L)$, the inclusion of $\mathcal{L}$-fuzzy sets S_1 and S_2 in $\mathcal{U}$ is defined by: $S_1 \subseteq S_2 \Leftrightarrow S_1(u) \leq_L S_2(u)$, for all u in $\mathcal{U}$. The intersection and union of two $\mathcal{L}$-fuzzy sets S_1 and S_2 in $\mathcal{U}$ are defined by: $(S_1 \cap S_2)(u) = S_1(u) \wedge S_2(u)$ and $(S_1 \cup S_2)(u) = S_1(u) \vee S_2(u)$, for all u in $\mathcal{U}$.

2.2 Interval-valued logical operators

As crisp set theory has Boolean logic as underlying logical framework, the extension to fuzzy set theory comes along with the extension of Boolean logic to fuzzy logic. The basic Boolean logical operators (negation, conjunction, implication) were extended to fuzzy logical operators, which respectively model the complement, the intersection, and the inclusion of fuzzy sets. These definitions can be easily generalized to $\mathcal{L}$-fuzzy logical operators which play an important role in $\mathcal{L}$-fuzzy mathematical morphology.[15]

DEFINITION 1 (LOGICAL OPERATORS). *Let $\mathcal{L} = (L, \leq_L)$ be a complete lattice with smallest element 0_L and largest element 1_L.*

- *An $L - L$ mapping $\mathcal{N}$ is called a negator on $\mathcal{L}$ if $\mathcal{N}(0_L) = 1_L$ and $\mathcal{N}(1_L) = 0_L$, and if it is a decreasing mapping; it is called involutive if $(\forall x \in L)\ (\mathcal{N}(\mathcal{N}(x)) = x)$.*
- *An $L^2 - L$ mapping $\mathcal{C}$ is called a conjunctor on $\mathcal{L}$ if $\mathcal{C}(0_L, 0_L) = \mathcal{C}(1_L, 0_L) = \mathcal{C}(0_L, 1_L) = 0_L$, $\mathcal{C}(1_L, 1_L) = 1_L$, and if it has increasing partial mappings. It is called a t-norm (usually denoted as $\mathcal{T}$) if it also is commutative, associative and satisfies $(\forall x \in L)(\mathcal{C}(1_L, x) = x)$.*
- *An $L^2 - L$ mapping $\mathcal{I}$ is called an implicator on $\mathcal{L}$ if $\mathcal{I}(0_L, 0_L) = \mathcal{I}(0_L, 1_L) = \mathcal{I}(1_L, 1_L) = 1_L$, $\mathcal{I}(1_L, 0_L) = 0_L$, and if it has decreasing first and increasing second partial mappings.*

For the complete lattice $([0,1], \leq)$ the standard negator N_s is defined by $N_s(x) = 1 - x$ $(x \in [0,1])$. Popular t-norms on $[0,1]$ are $T_M(x,y) = \min(x,y)$, $T_P(x,y) = xy$ and $T_W(x,y) = \max(0, x + y - 1)$ (Lukasiewicz), popular implicators on $[0,1]$ are $I_{KD}(x,y) = \max(1-x, y)$ (Kleene-Dienes), $I_R(x,y) = 1 - x + xy$ (Reichenbach) and $I_W(x,y) = \min(1, 1 - x + y)$ (Lukasiewicz), always with $x, y \in [0,1]$. Negators, t-norms and implicators are often associated with each other, e.g. it holds that the standard negator N_s is induced by the Kleene-Deenes implicator I_{KD} $(N_s(x) = I_{KD}(x, 0))$ and that the t-norm T_M (the minimum) can be obtained by $T_M(x,y) = N_s(I_{KD}(x, N_s(y)))$, i.e., $\min(x,y) = 1 - I_{KD}(x, 1-y)$. Reversely, it holds that $I_{KD}(x,y) = N_s(T_M(x, N_s(y))) = 1 - \min(x, 1-y)$, so that we can consider I_{KD} as an implicator that is clearly associated with the minimum.

For our purposes we are interested in logical operators on the complete lattice $\mathcal{L}^I = (L^I, \leq_{L^I})$. There are different ways of representing operations on interval-valued fuzzy sets by corresponding operations on classical fuzzy sets, but we only recall those expressions that are useful for this paper.

DEFINITION 2 (REPRESENTABLE LOGICAL OPERATORS).[5] *Let $\mathcal{L}^I = (L^I, \leq_{L^I})$ be a complete lattice, and let $x = [x_1, x_2]$ and $y = [y_1, y_2]$ be closed intervals in $[0,1]$, i.e., $x, y \in L^I$.*

- *Let N be a negator on $[0,1]$, then $\mathcal{N}_N$, defined by $\mathcal{N}_N(x) = [N(x_2), N(x_1)]$ is a negator on $\mathcal{L}^I$, called n-representable negator with representative N.*
- *Let T be a t-norm on $[0,1]$, then $\mathcal{T}_T^r$, defined by $\mathcal{T}_T^r(x,y) = [T(x_1, y_1), T(x_2, y_2)]$ is a t-norm on $\mathcal{L}^I$, called t-representable t-norm with representative T.*

- *Let I be an implicator on $[0,1]$, then $\mathcal{I}_I^r$, defined by $\mathcal{I}_I^r(x,y) = [I(x_2,y_1), I(x_1,y_2)]$ is an implicator on $\mathcal{L}^I$, called i-representable implicator with representative I.*

The origin of the representable operators is quite straightforward: fuzzy logical operators on $[0,1]$ are applied on the lower and upper bounds of the considered intervals. Note that also pseudo-representable operators[5] can be introduced, either by lowering the upper bound (pessimistic approach) or either by lifting the lower bound (optimistic approach) of the resulting intervals.

For example, the standard negator N_s leads to:

$$\mathcal{N}_s(x) = [N_s(x_2), N_s(x_1)] = [1-x_2, 1-x_1].$$

The minimum T_M leads to the following t-representable t-norm:

$$\mathcal{T}_M^r(x,y) = [\min(x_1,y_1), \min(x_2,y_2)],$$

and the Kleene-Dienes implicator I_{KD} leads to the following i-representable implicator:

$$\mathcal{I}_{KD}^r(x,y) = [\max(1-x_2,y_1), \max(1-x_1,y_2)].$$

The latter operators will be considered further in this paper to construct an interval-valued morphological model. Because the operators are all associated with the minimum we will refer to this model as the *model based on the minimum-operator.*

3. INTERVAL-VALUED FUZZY MATHEMATICAL MORPHOLOGY

3.1 Construction

Mathematical morphology was originally introduced for binary images and binary structuring elements, i.e. for objects that can only have values 0 (representing black) and 1 (representing white). In the binary case, images and the structuring elements are represented as crisp subsets of a universe $\mathcal{U}$. The dilation and erosion are defined as follows.

DEFINITION 3 (BINARY DILATION AND EROSION).[12] *Let A, B be crisp subsets of $\mathcal{U}$. The binary dilation $D(A,B)$ and the binary erosion $E(A,B)$ are defined by:*

$$D(A,B) = \{v \in \mathcal{U} | T_v(B) \cap A \neq \emptyset\},$$
$$E(A,B) = \{v \in \mathcal{U} | T_v(B) \subseteq A\},$$

with $T_v(B) = \{u \in \mathcal{U} | u - v \in B\}$ the translation of B by the point v.

The binary dilation and erosion have a very nice geometrical interpretation. For example, the dilation $D(A,B)$ consists of all points v such that the translation $T_v(B)$ of the structuring element has a non-empty intersection with the image A. Consequently, the dilation will typically extend the contours of objects in the image, and fill up small gaps and channels. Similar interpretations can be made for the erosion.

These definitions can be extended to non-binary objects, in particular to interval-valued objects, by fuzzifying the underlying logical framework. This leads to the following definitions.[11,15]

DEFINITION 4 (INTERVAL-VALUED DILATION AND EROSION). *Let $\mathcal{L}^I = (L^I, \leq_{L^I})$ be the complete lattice corresponding to interval-valued fuzzy set theory, let $\mathcal{T}$ be a t-norm on $\mathcal{L}^I$ and let $\mathcal{I}$ be an implicator on $\mathcal{L}^I$. Let A be an image and B be a structuring element, both represented as interval-valued fuzzy sets in $\mathcal{U}$. The interval-valued fuzzy dilation $D_{\mathcal{T}}^I(A,B)$ and the interval-valued fuzzy erosion $E_{\mathcal{I}}^I(A,B)$ of A by B are defined by:*

$$D_{\mathcal{T}}^I(A,B)(v) = \bigvee_{u \in \mathcal{U}} \mathcal{T}(B(u-v), A(u)),$$
$$E_{\mathcal{I}}^I(A,B)(v) = \bigwedge_{u \in \mathcal{U}} \mathcal{I}(B(u-v), A(u)),$$

for all v in $\mathcal{U}$.

In the next subsection, we will discuss the properties of the model based on the minimum-operator, i.e. the model based on the t-norm $\mathcal{T}_M^r$ and implicator $\mathcal{I}_{KD}^r$ on $\mathcal{L}^I = (L^I, \leq_{L^I})$. We will use the notations D_M^I and E_M^I to denote the corresponding interval-valued dilation and erosion.

3.2 Properties of the model based on the minimum-operator

The general theoretical study[11] leads to the following specific properties for the considered interval-valued mathematical morphological model based on minimum-operator. In these properties, A, A_1, A_2 and B, B_1, B_2 represent interval-valued fuzzy sets in the universe $\mathcal{U}$, and $\mathcal{L}^I = (L^I, \leq_{L^I})$ represents the complete lattice associated with interval-valued fuzzy set theory.

PROPERTY 1 (EXPANSIVITY AND RESTRICTIVITY). *If* $B(0_{L^I}) = 1_{L^I}$, *then it holds that:*

$$E_M^I(A,B) \subseteq A \subseteq D_M^I(A,B).$$

PROPERTY 2 (MONOTONICITY). *If* $A_1 \subseteq A_2$, *then:*

$$E_M^I(A_1,B) \subseteq E_M^I(A_2,B) \quad \textit{and} \quad D_M^I(A_1,B) \subseteq D_M^I(A_2,B).$$

If $B_1 \subseteq B_2$, *then:*

$$E_M^I(A,B_1) \supseteq E_M^I(A,B_2) \quad \textit{and} \quad D_M^I(A,B_1) \subseteq D_M^I(A,B_2).$$

PROPERTY 3 (INTERACTION WITH INTERSECTION). *It holds that:*

$$\begin{aligned} D_M^I(A_1 \cap A_2, B) &\subseteq D_M^I(A_1,B) \cap D_M^I(A_2,B), \\ E_M^I(A_1 \cap A_2, B) &= E_M^I(A_1,B) \cap E_M^I(A_2,B). \end{aligned}$$

$$\begin{aligned} D_M^I(A, B_1 \cap B_2) &\subseteq D_M^I(A,B_1) \cap D_M^I(A,B_2), \\ E_M^I(A, B_1 \cap B_2) &\supseteq E_M^I(A,B_1) \cup E_M^I(A,B_2). \end{aligned}$$

PROPERTY 4 (INTERACTION WITH UNION). *It holds that:*

$$\begin{aligned} D_M^I(A_1 \cup A_2, B) &= D_M^I(A_1,B) \cup D_M^I(A_2,B), \\ E_M^I(A_1 \cup A_2, B) &\supseteq E_M^I(A_1,B) \cup E_M^I(A_2,B). \end{aligned}$$

$$\begin{aligned} D_M^I(A, B_1 \cup B_2) &= D_M^I(A,B_1) \cup D_M^I(A,B_2), \\ E_M^I(A, B_1 \cup B_2) &= E_M^I(A,B_1) \cap E_M^I(A,B_2). \end{aligned}$$

From the Properties 3 and 4 it follows that the interval-valued fuzzy dilation D_M^I and erosion E_M^I are actually *algebraic* dilations and erosions[1,7,13] : the dilation interacts with the union of images, the erosion interacts with the intersection of images.

Next, recall that, for a general complete lattice $\mathcal{L} = (L, \leq_L)$ and a given negator $\mathcal{N}$ on $\mathcal{L}$, the $\mathcal{N}$-complement of an $\mathcal{L}$-fuzzy set S in a universe $\mathcal{U}$ is defined by: $(co_{\mathcal{N}} S)(u) = \mathcal{N}(S(u))$, for all u in $\mathcal{U}$. Note that this $\mathcal{N}$-complement can also be regarded as a negator on $\mathcal{L}^{\mathcal{U}}$, induced by the negator $\mathcal{N}$ on $\mathcal{L}$.

PROPERTY 5 (DUALITY). *It holds that:*

$$co_{\mathcal{N}_s} D_M^I(A,B) = E_M^I(co_{\mathcal{N}_s} A, B) \quad \textit{and} \quad D_M^I(co_{\mathcal{N}_s} A, B) = co_{\mathcal{N}_s} E_M^I(A,B).$$

Interpretation of the above properties is straightforward. Expansivity and restrictivity mean that an erosion will enlarge dark objects in the image while the dilation will shrink them or, alternatively, that an erosion will shrink white objects while dilation will enlarge them. Monotonicity w.r.t. images (the first argument) means

that darker (resp. lighter) images will lead to darker (resp. lighter) dilated and eroded images. Monotonicity w.r.t. structuring elements (the second argument) means that darker structuring elements will lead to darker dilated images but to lighter eroded images. From the interaction properties we learn that the erosion of the intersection of two images matches the intersection of the two separate eroded images, and that the dilation of the union of two images matches the union of the two separate dilated images. Also, the dilation and erosion by a union of structuring elements equals the union of the separate dilations, respectively the intersection of the separate erosions. Those equalities can be interesting from a computational point of view. For example, you can decompose a structuring element into basic elements, perform dilation and erosion with those basic elements (fast standard operations), and reconstruct the dilated and eroded images. Finally, the duality property gives a relation between dilation on the one hand and erosion on the other hand, i.e. the dilation can be constructed from the erosion and vice versa.

To conclude the properties we investigate the relation between the interval-valued dilation and erosion and their binary counterparts. Therefore, we first refresh the definitions of the different $[\alpha_1, \alpha_2]$-cuts.

DEFINITION 5 (CUT-SETS OF AN INTERVAL-VALUED FUZZY SET).[16] *For $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$, the weak $[\alpha_1, \alpha_2]$-cut $A^{\alpha_2}_{\alpha_1}$ of A is given by:*

$$A^{\alpha_2}_{\alpha_1} = \{x \in \mathcal{U} | A_1(x) \geq \alpha_1 \text{ and } A_2(x) \geq \alpha_2\} = \{x \in \mathcal{U} | A(x) \geq_{L^I} [\alpha_1, \alpha_2]\}.$$

Let $U_{L^I} = \{[x_1, x_2] \in L^I | x_2 = 1\}$. For $[\alpha_1, \alpha_2] \in L^I \backslash U_{L^I}$, the strict $[\alpha_1, \alpha_2]$-cut $A^{\overline{\alpha_2}}_{\overline{\alpha_1}}$ of A is given by:

$$A^{\overline{\alpha_2}}_{\overline{\alpha_1}} = \{x \in \mathcal{U} | A_1(x) > \alpha_1 \text{ and } A_2(x) > \alpha_2\} = \{x \in \mathcal{U} | A(x) >_{L^I} [\alpha_1, \alpha_2]\}.$$

For $\alpha_1, \alpha_2 \in [0, 1[$, the strict α_1-subcut $A_{\overline{\alpha_1}}$ and the strict α_2-supercut $A^{\overline{\alpha_2}}$ of A are given by:

$$A_{\overline{\alpha_1}} = \{x \in \mathcal{U} | A_1(x) > \alpha_1\} \quad \text{and} \quad A^{\overline{\alpha_2}} = \{x \in \mathcal{U} | A_2(x) > \alpha_2\}.$$

The cases $[\alpha_1, \alpha_2] = 0_{L^I}$ and $[\alpha_1, \alpha_2] \in U_{L^I}$ are excluded for respectively the weak and the strict $[\alpha_1, \alpha_2]$-cut. Since $\{x \in \mathcal{U} | A_1(x) \geq 0 \text{ and } A_2(x) \geq 0\} = \mathcal{U}$ and $\{x \in \mathcal{U} | A_2(x) > 1\} = \emptyset$, these cases do not yield relevant information. For similar reasons, $\alpha_1 = 1$ and $\alpha_2 = 1$ are excluded in the definition of the sub- and supercut.

It turns out that the cuts of the interval-valued dilation reduce to binary dilations (decomposition property) and, reversely, that the interval-valued dilation can also be constructed from its binary counterpart (construction property). One can show[8,9] that D^I_M is the only interval-valued dilation that has these properties.

PROPERTY 6 (DECOMPOSITION OF THE INTERVAL-VALUED DILATION[9]). *Let $\alpha_1 \in [0, 1[$ and $\alpha_2 \in [0, 1[$. It holds that:*

$$\begin{aligned} D^I_M(A, B)_{\overline{\alpha_1}} &= D(A_{\overline{\alpha}_1}, B_{\overline{\alpha_1}}), \\ D^I_M(A, B)^{\overline{\alpha_2}} &= D(A^{\overline{\alpha_2}}, B^{\overline{\alpha_2}}). \end{aligned}$$

Let $[\alpha_1, \alpha_2] \in L^I \backslash U_{L^I}$. It holds that:

$$D^I_M(A, B)^{\overline{\alpha}_2}_{\overline{\alpha}_1} = D(A_{\overline{\alpha}_1}, B_{\overline{\alpha}_1}) \cap D(A^{\overline{\alpha}_2}, B^{\overline{\alpha}_2}).$$

PROPERTY 7 (CONSTRUCTION OF THE INTERVAL-VALUED DILATION[8]). *It holds that:*

$$D^I_M(A, B) = \bigcup_{[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}} [\alpha_1, \alpha_2] D(A^{\alpha_2}_{\alpha_1}, B^{\alpha_2}_{\alpha_1}),$$

with

$$[\alpha_1, \alpha_2] C(x) = \begin{cases} [\alpha_1, \alpha_2] & \text{if } x \in C \\ 0_{L^I} & \text{else} \end{cases}$$

with C a crisp subset of $\mathcal{U}$ and $[\alpha_1, \alpha_2] \in L^I \backslash \{0_{L^I}\}$

Similar decomposition properties hold for the erosion[9] . Regarding the construction property for the dilation, a corresponding interval-valued erosion can then be constructed using the duality property.

4. CONCLUSION AND FUTURE RESEARCH

In this paper we have introduced a new image model (interval-valued image representation), we have introduced the corresponding morphological model (interval-valued fuzzy morphology), and we have specifically investigated the model that is based on the minimum-operator. The interval-valued approach succeeds in modeling the uncertainty regarding grayscale pixel values, and enables us to propagate this uncertainty through morphological operators. The studied model not only has nice properties, it also turns out that the corresponding interval-valued morphological operators can be constructed by or decomposed into their binary counterparts. Future research will focus on a continued effort to gain theoretical insight in the interval-valued models, as well as on practical applications, e.g. edge detection in grayscale images with uncertain values.

REFERENCES

[1] Banon, G. J. F. and Barrera, J., "Decomposition of Mappings between Complete Lattices by Mathematical Morphology, Part 1. General Lattices", Signal Processing 30(3), 299-327 (1993).

[2] Castillo, O. and Melin, P., "Intelligent Systems with Interval Type-2 Fuzzy Logic", International Journal of Innovative Computing, Information and Control 4(4), 771-783 (2008).

[3] De Baets, B., "Fuzzy morphology: a logical approach", Uncertainty Analysis in Engineering and Sciences: Fuzzy Logic, Statistics, and Neural Network Approach (Ayyub B.M. & Gupta M.M., editors), Kluwer Academic Publishers, Boston, 53-67 (1997).

[4] Deschrijver, G. and Kerre, E.E., "On the Relationship Between some Extensions of Fuzzy Set Theory", Fuzzy Sets and Systems 133, 227-235 (2003).

[5] Deschrijver, G. and Cornelis, C., "Representability in Interval-valued Fuzzy Set Theory", International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 15(3), 345-361 (2007).

[6] Goguen, J., "L-fuzzy sets", Journal of Mathematical Analysis and Applications 18, 145-174 (1967).

[7] Heijmans, H., [Morphological Image Operators], Academic Press, New York (1994).

[8] Mélange, T., Nachtegael, M., Sussner, P. and Kerre, E.E., "Construction of Interval-Valued Fuzzy Morphological Operators by Weak Cuts", Proceedings of EUROFUSE 2009 – Eurofuse Workshop on Preference Modeling and Decision Analysis, 227-232 (2009).

[9] Mélange, T., Nachtegael, M., Sussner, P. and Kerre, E.E., "On the Decomposition of Interval-Valued Fuzzy Morphological Operations", Journal of Mathematical Imaging and Vision (accepted for publication).

[10] Nachtegael, M. and Kerre, E.E., "Connections between binary, gray-scale and fuzzy mathematical morphologies", Fuzzy Sets and Systems 124(1), 73-86 (2001).

[11] Nachtegael, M., Sussner, P., Mélange, T. and Kerre, E.E., "Some Aspects of Interval-valued and Intuitionistic Fuzzy Mathematical Morphology", Proceedings of IPCV 2008 – 3rd International Conference on Image Processing, Computer Vision and Pattern Recognition, 538-543 (2008).

[12] Serra, J., [Image analysis and mathematical morphology], Academic Press, London (1982).

[13] Serra, J., [Image Analysis and Mathematical Morphology, Volume 2: Theoretical Advances], Academic Press, New York (1988).

[14] Sussner, P. and Valle, M.,E., "Classification of Fuzzy Mathematical Morphologies Based on Concepts of Inclusion Measure and Duality", Journal of Mathematical Imaging and Vision 32(2), 139-159 (2008).

[15] Sussner, P., Nachtegael, M. and Mélange, T., "L-Fuzzy Mathematical Morphology: an Extension of Interval-Valued and Intuitionistic Fuzzy Mathematical Morphology", Proceedings of NAFIPS 2009 – 28th North American Fuzzy Information Processing Society Annual Conference, 6 pages on CD (2009).

[16] Wang, G., and Li, X., "The applications of interval-valued fuzzy numbers and interval-distribution numbers", Fuzzy Sets and Systems 98, 331-335 (1998).

[17] Zadeh L., "Fuzzy Sets", Information Control 8, 338-353 (1965).

Three-Dimensional Modeling of Plants: A Review

Zhi-yong Bai, Xin-yuan Huang
School of Information Science and Technology, Beijing Forestry University, Beijing, China.
hxy@bjfu.edu.cn

ABSTRACT

The Plant is one of the hot fields in the current virtual reality modeling research, and undoubtedly an important component of factors in the natural scenes. It is difficult to be drawn in terms of shape, so there have be a lot of methods in the research of three-dimensional modeling of plants. This article describes the simulation modeling of plants and related technology research and development situation, summarizes the main research problems, and discusses the future research and application development trends and prospects.

Plants modeling, Three-Dimensional, Virtual Reality

1. Introduction

Plant is one of nature's most common landscapes, with a complex three-dimensional structure. The three-dimensional modeling of plants has been a computer graphics research focus, and a large number of scholars have conducted in-depth research. There are many theories about three-dimensional modeling of plants: the A-system; plants based on fractal technology; plants based on particle systems mold method; the real structure of trees modeling methods; plants growth simulation step by step method and based on the L-system tree growth simulation methods .

Three-dimensional modeling of plants is of great significance in agriculture, forestry, plant science industry, production process management, production optimization modeling and prediction, and urban landscape design, etc. On the other hand, it can be used in the huge potential markets, such as virtual reality, computer animation, games, software, commercial advertising, and educational software and so on.

2. Technologies and Softwares

2.1 Xfrog

Greenworks's Bernd Lintermann 's Xfrog software from Germany's University of Karlsruhe which is based on L-system is run on Windows and SGI, and builds any type of plants. With the basic form and graphics-based modeling method, it uses an interactive method to construct the various plant models. Xfrog defines numbers of small icons including: plant organ, plant structure, function, global variables. Xfrog does not represent the actual growth characteristics of plants, but as a pure graphics sense of plant simulation software it is easy to operate, convenient in three-dimensional viewing and the real effect. Xfrog allows the greatest freedom to the user manipulation of modeling, and has also joined the impact of light and gravity factors and can be exported to 3ds max and Maya.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463I · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855275

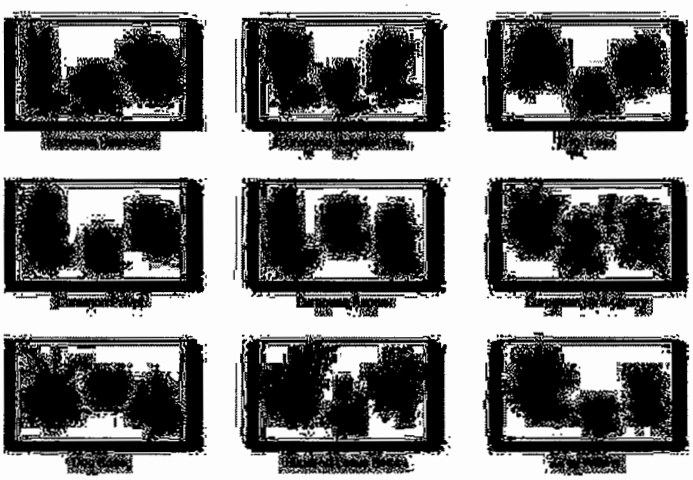


Figure 1: Xfrog Plants Library

2.2 Virtual Plants

Australian manufacturing production Virtual Plants, based on the L-system modeling method, is used to simulate cotton, soybeans, corn, other crops and plant root growth. The main purpose is to predict the impact on plant diseases and insect pests.

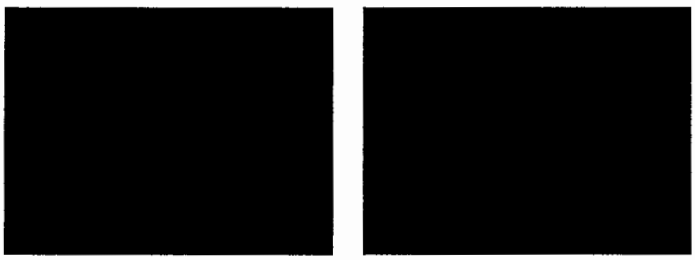


Figure 2: SVS Modeling of Corn

2.3 Virtual Gardening, Virtual BONSAI and Digital Landscapes

Virtual Gardening, Virtual BONSAI and Digital Landscapes from Japan's Iwate University focuses on plant growth characteristics, such as self-pruning, the impact of light on plants, and plant root growth. It provides a model of environmentally sensitive plant growth, root and soil interactions models for simulating the landscape.

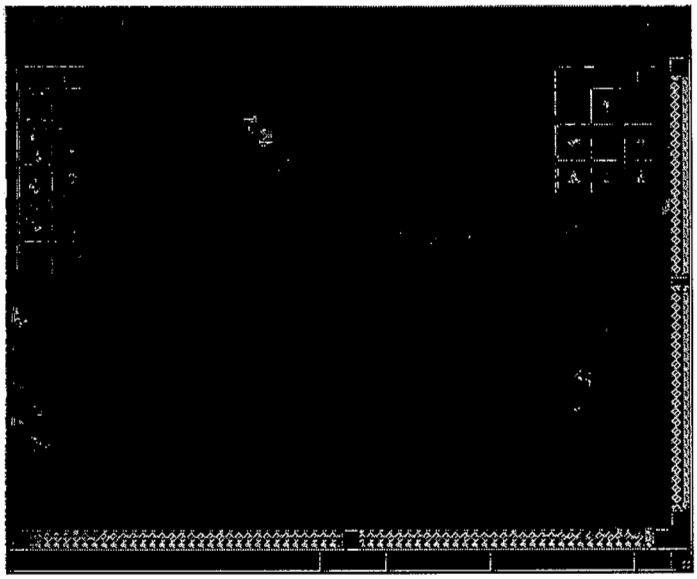


Figure 3: Virtual BONSAI Interface

2.4 SpeedTree

SpeedTree is a powerful modeling and rendering software quickly to help programmers and animation designers to move pretty dynamic three-dimensional leaf patterns to scenes. It is able to produce large areas of forest fine patterns.

SpeedTree has many special effects as well as the optimization technology providing the environment in natural conditions such as wind speed and direction and allow real-time generation of trees with an absolutely realistic swing effect. In addition, SpeedTree engine also optimize the code and is specifically responsible for the scene of "planted" tree of the program. SpeedTree not only create a very real effect of trees and forests, but also as "parts" easily embedded into other rendering engine.

Figure 4: A Tree Based on SpeedTree

2.5 TreeGenerator

TreeGenerator spanning tree is very simple. In accordance with its design idea, a tree is built with only 36 parameters in 5 minutes.

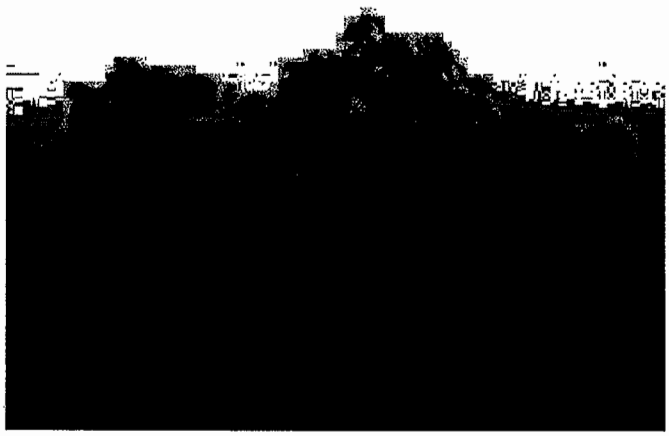


Figure 5: TreeGenerator Modeling

2.5 SVS

United States Department of Agriculture SVS software, using graphics-based approach and a small amount of botanical knowledge, defines a set of control parameters of the shape of the crown. Mainly by the slider control to adjust the various parameters to adjust the slider position of the mouse users it quickly compares the satisfied shape of the crown. SVS is used to create flora plexus and chart a simple form of the distribution of vegetation, and also change the background and the various properties of the ground in a variety of viewing angles.

2.6 AMAP

French Center for International Studies Agricultural Development (CIRAD) developed the use of reference axis technology to produce AMAP (Advanced Modeling of Architecture of Plant) model. AMAP model is applicable to simulation of tall plants, and have successfully constructed a variety of different tropical types of plants growth.

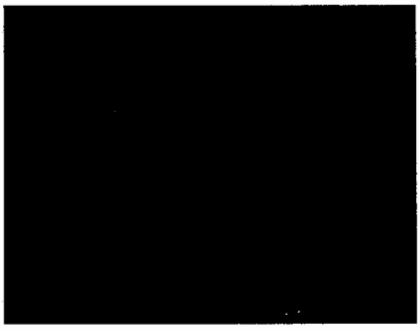


Figure 6: AMAP Modeling

2.7 Virtual Laboratory(Vlab)

Prusinkiewicz in the University of Calgary in Canada used the L-system as the plant morphology description of the framework developed in the SGI workstations to develop Unix-based system virtual plant Laboratory Virtual Laboratory (Vlab) and plant and fractal generator CPFG (plant and fractal generator with continuous parameters).The system enables different types of plants simulation. The defect is the simulation of some higher plants is not ideal, because some plants' L-system structure is more complex and difficult to be extracted.

2.8 Arbaro

It reads in an XML file with the tree parameters and writes out Povray, DXF or Wavefront OBJ files.

Figure 7:Arbaro's Some Desert Trees

3. Summary

In short ,the real three-dimensional structure plant model plays an important role in the simulation in a natural scene, urban landscape design, virtual reality, computer games and applications in agriculture and forestry. The reconstruction of three-dimensional model of the real plants in the computer graphics has a broad application prospects.

References

[1] ZHAO Qin Ping, "survey on virtual reality", Science in China,20-25(2009)

[2] Frederic Boudon, "ractive design of bonsai treemodels", EUROGRAPHICS 22 ,591-599(2004) .

[3] Bedrich Benes, Juan Miguel, "Clustering in virtual plantecosystems", .WSCG 2004., 9-17(2004)

[4] Kui-yu Li, "ased on the depth of image-based rendering algorithm", Chinese Academy of Sciences Master's degree thesis, 211-216

[5] IDV, [peedTreeCAD User Reference] IDV Company,100-200(2008)

Using Ontology for Domain Specific Information Retrieval

[1]H.L.Shashirekha
Department of Studies in Computer Science, Mangalore University, Mangalore, Karnataka, India

[2]S.Murali
PET Research Center, PES College of Engineering, Mandya-571401, Karnataka India

[3]P.Nagabhushan
Department of Studies in Computer Science, University of Mysore, Mysore, Karnataka, India

ABSTRACT

This paper presents a system for retrieving information from a domain specific document collection made up of data rich unnatural language text documents. Instead of conventional keyword based retrieval, our system makes use of domain ontology to retrieve the information from a collection of documents. The system addresses the problem of representing unnatural language text documents and constructing a classifier model that helps in the efficient retrieval of relevant information. Query to this system may be either the key phrases in terms of concepts or a domain specific unnatural language text document. The classifier used in this system can also be used to assign multiple labels to the previously unseen text document belonging to the same domain. An empirical evaluation of the system is conducted on the domain of text documents describing the classified matrimonial advertisements to determine its performance.

Ontology, Information Retrieval, Unnatural Language Text Documents

1 INTRODUCTION

The task of finding relevant information from large domain specific text document collection is a field of active research within the information retrieval community. More and more business, government, and academic institutes need domain-specific search service. Domain-specific Information Retrieval, as a branch of IR technologies which have been extensively used in search engines, focusing on a specific subject domain, can provide substantial support to domain specific search engine system [1, 2]. Retrieving the information effectively and efficiently from a large document collection depends on the proper organization of the text documents. The crucial aspect in achieving efficient retrieval of information is the delivery of minimal irrelevant information while ensuring relevant information is not overlooked. Also, the transfer of irrelevant information which is of no use to the user community consumes unnecessary bandwidth leading to wastage of resources. Traditional solution to the search mechanism employs keyword-based search which is simple and easy to implement. But these approaches relate only identical keywords thereby ignoring the conceptual and semantic similarity, context sensitivity and the dependency of words. The conventional representation of text documents based on bag-of-words also plays a major role in search mechanisms. High dimensionality of text documents is an important problem to be considered in the search mechanisms. These limitations are mainly due to the word centric representation of text documents. Instead of considering words that are present explicitly to represent the text documents, documents can be represented as a bag-of-concepts and the relation between the concepts where a concept is an idea or notion [3]. It's more natural to consider a document as a collection of concepts rather than a simple collection of words. The concepts and relation between the concepts can be modeled as ontology which is a natural way of representing any application domain. The reduced vocabulary of the ontology addresses the high dimensionality of text documents. Ontology as an emerging discipline has shown great potentials to manipulate, manage and explicate information in several domains [4]. It creates an agreed upon vocabulary for the exchange of information accepted by a

[1] hlsrekha@gmail.com, [2] nymurali@yahoo.com, [3] pnagabhushan@hotmail.com

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463J · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855965

group of people or community. In this work we use ontology to represent an application domain where the application domain is described by a set of domain specific unnatural language text documents.

1.1 Unnatural Language Text Documents

The unnatural language text documents contain text just in the form of collection of words or bag-of-words. There may not be any sentences at all or may contain partial sentences and some punctuation symbols but not according to any specific order. Instead they may have a number of identifiable constants such as dates, place, prices and names. Language of these types of documents in which it is written may be similar to telegraphic, cryptic or short message service language. These types of documents do not adhere to the syntax of any of the conventional natural language like English. As a result, the conventional natural language processing techniques cannot be applied. An altogether new set of tools and techniques are to be developed to handle these types of documents. Text documents describing classifieds such as matrimonial advertisements, rentals advertisements and job postings, weather reports, financial statements are examples of these types of documents. Each of these types of documents contains more or less the same information at the conceptual level. Figure-1 illustrates a few examples of texts describing classified matrimonial advertisements.

Tall Medico/ pfly qlfd b'ful h'nly girl for h'some t'tlor Saxena Kayasth BE MBA 29/ 185/ 25k Mgr MNC Mum rsptd NI family. Caste no bar: Send BHP to Box. No MUM 141518C Times of India, Mumbai – 400001

ALLIANCE for Telugu boy SC (Mala), 30/5'7" B.Com Officer (Forex) Nationalised bank, caste no bar. Write with BHP to Box No MUM 141570C Times of India, Mumbai - 1. E: pals2@rediffmail.com.

BRAHMIN KKB BE 30/6' h'some S/W, MNC B'lore 12+lpa seeks tall b'ful prof girl. Send BHP (must). Box No BAN 313496K Times of India, Bangalore-560001 E-mail: northstar@touchtelindia.net

BENGALI MATCH 4 prob EB Basu girl 1976/ 5'6"/ 10Lpa fair, slim B.Arch gold medal from well plcd, Pg profl. Email: rbsb@rediffmail.com Box No. MUM 147497C Times of India, Mumbai - 400001.

KAYASTHA BEAUTIFUL well Qlfd. wrkg . girl for IIT Kanpur M.TEch Ex. BARC Scientist wrkg. MNC Chennai 28/ 6'/ 5 Lpa, fair & h'some. Caste no bar. Write Box No. LUC 516938C Times of India, Lucknow-226001.

Figure-1. Examples of Classified Matrimonial Advertisement Text Documents

As the words in these types of documents act as the values of the attributes of some unknown schema, these words cannot be considered as keywords and hence documents cannot be represented using vector space model. As a result we use domain ontology to represent the unnatural language text documents. In this work, we use domain ontology to conceptualize the document collection and construct a classifier hierarchy that helps in retrieving the required information. The rest of the paper is organized as follows: Section-2 gives a brief overview of the work carried out in the related field. Our approach to domain specific information retrieval is explained in section-3. Section-4 deals with the experimental results, illustration and analysis. Finally the paper concludes in section-5.

2 RELATED WORK

Ontology based approach is a promising way for information retrieval applications. The ontology based information retrieval model proposed by Castells P. et al. [5] is based on an adaptation of the classic vector-space model, including an annotation weighting algorithm and a ranking algorithm. However, when ontology is used to model an application domain, these similarity measures cannot be used. The similarity measures in addition to addressing conceptual similarity, should also give importance to the hierarchical nature of ontology. In the information retrieval model proposed by Truong Q.D. et al. [6] documents, queries and indexing terms are viewed as vertices of a bipartite graph where edges from a document or a query – first node type go to an indexing term – second node type. Edges reflect the link that exists between documents or queries on the one hand and indexing terms on the other hand.

Most of the text classifiers are either binary or multi-class. It is observed that very less attention is given for multi-label classifiers. An overview of multi-label classification and a comparative experimental result of certain multi-label classification approaches are presented by Tsoumakas G. et al. [7]. Most of the text classification methods are usually based on flat model that ignores the hierarchical structure, treating each category as a separate class. These flat classifiers although are computationally simple, fail to exploit the information inherent in the structural relationship

between the categories. Recently some approaches for hierarchical text classification are introduced to overcome the restrictions of flat models [8, 9]. Usage of ontology also addresses the issues related to multi-label classification and hierarchical classification. Ontology based classifiers are used to classify news documents [10], web pages [11], Chinese news clippings [12] and so on.

3 OUR APPROACH

The key aspects to be considered in retrieving information from domain specific collection include representation of text documents, constructing a classifier hierarchy and finally techniques to retrieve the required information. The prime requirement in our approach to retrieve the required information is a domain ontology which models the application domain in the form of concepts and relations between the concepts. This domain ontology is constructed by extracting the knowledge from the domain specific text documents as explained in our earlier work [13]. The advantage of using ontology is that users can build dynamic contexts of interests for categorization and while using the same ontology and same document corpora, documents can be classified in a variety of ways according to different sets of interesting features.

3.1 Representation of Text Documents

The first step in retrieving information from document collection is to transform the text documents which typically are strings of characters into a representation suitable for retrieving task. The limitations of conventional vector space model based on bag-of-words provide scope for alternate representation of documents. Ontology based approaches which give importance to conceptual and semantic similarity is used to represent the documents. In our approach domain ontology is used to map the text documents to a set of concepts and populate a knowledgebase of the documents which participate in the information retrieval task. As each record in this knowledgebase acts as an equivalent representation for the corresponding document, each document in the collection is represented by a knowledgebase corresponding to that document.

3.2 Construction of a Classifier Hierarchy

For a given document collection, a vector space model of concepts is constructed where each vector of concepts represents a document. This vector space model is used to construct a classification hierarchy. The higher level categories of the hierarchy are defined based on the concepts of the ontology. As a result, the number of categories/classes is fixed subject to the maximum number of concepts in the ontology. The lower level categories/classes are defined based on the instances of the concepts which appear as the child node of the categories representing the concepts. Text documents finally appear as the leaf nodes in this hierarchy. By applying conceptual similarity as explained in our earlier work in [14], documents having identical concepts will be categorized under respective classes. This results in a hierarchical multi label text classifier model as shown in Figure-2.

The hierarchy can be constructed based on all the concepts and all the instances of the concepts or based on a selective set of concepts and their instances. Depending on the relevance of the concept, concepts can be considered as 'strong' concept or 'weak' concept. A 'strong' concept is a relevant concept which is of much significance in describing the domain. However, some concepts, even though describe a domain, may not be of much significance. Such concepts are 'weak' concepts which are to be selected and ignored. An example of a hierarchy constructed using the concepts "age" and "caste" along with their instances is illustrated in Figure-3. Each document is assigned a unique id which represents the document as shown in Figure-3. A much detailed classification hierarchy including the concepts "age and "caste" is illustrated in Figure-4. The documents in this figure are represented by their respective id.

The hierarchy can be used to obtain all the category labels of a document. For example, as shown in Figure-3 the document with id '0' is a member of the categories '27', 'age', 'brahmin' and 'caste'. While 'age' and 'caste' are higher level categories, '27' and 'brahmin' are lower level categories which are instances of the categories 'age' and 'caste' respectively. Moving from lower level category to a higher level category illustrates the concept of specialization to generalization. In this hierarchy the leaf nodes represent the documents and the nodes at the higher level represent the categories. A simple bottom-up traversal of the hierarchy gives all the labels of a document.

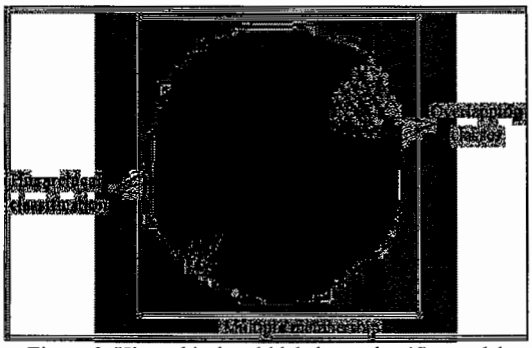


Figure-2. Hierarchical multi label text classifier model

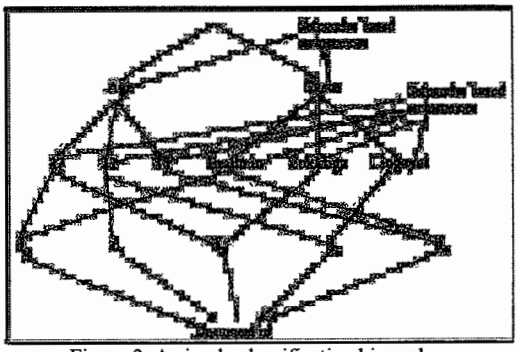


Figure-3. A simple classification hierarchy

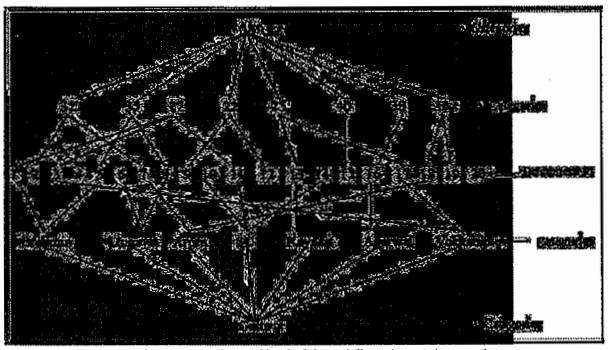


Figure-4. Detailed Classification hierarchy

3.3 Retrieving the Required Information

In addition to obtaining all the labels of a document, the documents that match the user's requirements can also be retrieved. Information from the document collection can be retrieved based on a query, where a query may be a single concept, a combination of concepts or any document belonging to the same domain. The retrieval process is simple if the query is based on concepts. Given a concept, all documents that are indexed under the category corresponding to that concept are retrieved. For example, if the query is for the documents including the concept "profession", then all the documents that are indexed under the concept "profession" will be retrieved. A combination of concepts can also be used to retrieve the required documents. The Boolean operators are used for the combination of queries. For example, if A and B are the concepts, then the Boolean operators defining the combination are:

'and' retrieves a set of documents that have the membership of both the concepts A and B
'or' retrieves a set of documents that have membership of either a concept A or concept B or both
'not' retrieves a set of documents that do not have the membership of a particular concept A

If the query is a document, then retrieving the required information is to search for similar documents in the collection. In this case a query document will be mapped to the concepts based on the underlying ontology and the concepts in combination are used to search the documents in the collection.

In our approach the search is conceptual which is based on the concepts rather than the conventional search which is based on keywords. The advantages of our approach is

- automated method for organizing the documents which helps in effective and efficient retrieval of documents
- reduces the search space of related documents making searches more efficient
- documents are allowed to participate in more than one category or have more than one label which reflects the true nature of the classification domain. For example,

- In medical diagnosis, a document related to a patient suffering from diabetes and high blood pressure at the same time can be grouped into two classes viz. 'diabetes and 'blood pressure'
- newspaper article related to an accident met by a sports person may be a member of class 'sports' and also of class 'accidents'.

4 EXPERIMENTAL RESULTS

The developed domain specific information retrieval system is tested on the domain of unnatural language text documents describing the classified matrimonial advertisements. A set of documents describing the classified matrimonial advertisements are used to construct domain ontology and the classifier hierarchy. Further a set of documents are selected at random to test the efficacy of the proposed system. The retrieval system is tested by considering the concepts, combination of concepts as well as domain specific text documents as queries.

4.1 Results of Concept Extraction

As the retrieval task crucially depends on the extraction of concepts using the domain ontology, the performance of the system depends on how efficiently the concepts are extracted from the ontology. The efficiency of extracting the concepts from the text documents is illustrated in Table-1 in terms of precision and recall. The formulation of precision and recall metrics is given below:

Precision = C / (C + I), Recall = C/N, where N – total number of concepts, C – number of concepts extracted correctly and I – number of concepts extracted incorrectly.

Table-1. Performance of Concept Extraction

Sl. No	No. of documents in the test set =250					
	Concepts	N	C	I	Recall %	Precision %
1	age	162	138	15	85.19	90.2
2	height	165	144	18	87.27	88.89
3	salary	66	60	12	90.91	83.33
4	appear	222	213	10	95.95	95.52
5	bhp	65	58	8	89.23	87.88
6	prof	204	183	22	89.71	89.27
8	fstatus	57	51	5	89.47	91.07
9	qual	273	243	18	89.01	93.1
10	cnb	30	25	6	83.33	80.65
11	language	18	12	0	66.67	100
12	citizen	12	8	0	66.67	100
13	caste	171	138	20	80.7	87.34
14	place	87	63	6	72.41	91.3
15	religion	21	9	0	42.86	100
16	divinfo	15	10	3	66.67	76.92
17	mstatus	10	6	2	60	75

4.2 Illustration of Information Retrieval

The task of retrieving the required information from a domain specific document collection is tested on 250 documents describing matrimonial advertisements. However, for the purpose of illustration we have considered 15 documents with only 5 concepts. Table-2 shows the populated knowledgebase for these 15 documents. This knowledgebase includes the concepts (field names) and the instances (field values) that can be used to construct the classification hierarchy that helps in retrieving the documents. It is clear from the knowledgebase that not all concepts may be present in all the documents. In addition, more than one instance of a concept may be present in any document. The set of documents in a collection together defines the hierarchy representing the entire domain. A classification hierarchy constructed for the concepts

"age" and "caste" is shown in Figure-3 and Figure-4. Table-3 illustrates the documents listed with respect to concepts and their instances (lexical level). The retrieval of documents based on the combination of concepts or categories is illustrated in Table-4.

Table-2. Populated Knowledgebase

Docid	Age	Caste	Qualification	Appearance	Salary
1	27	Brahmin	-	good-looking	20000
2	32	-	MCA	smart	40000
3	27	Vokkaliga	BSc	smart	30000
4	25	Vokkaliga	M.Com	beautiful	-
5	35	Brahmin	BE	-	50000
6	-	Lingayat	B.Com	handsome	-
7	32	Arora	MBBS	-	50000
8	35	Lingayat	-	fair	40000
9	22	-	MCA	beautiful	-
10	-	SC	MBA	good-looking	40000
11	27	SC	M.Sc	-	20000
12	30	Arora	M.Sc	very_beautiful	-
13	25	kayasat	MBA	smart	-
14	-	SC	MBBS	-	20000
15	38	Agarwal	BE	good-looking	30000

Table-3. Classification at conceptual and lexical level

Category label	Document id	
	Classification at conceptual level	Classification at lexical level
age	1, 2, 3, 5, 7, 8, 9, 11, 12, 13, 15, 16, 17, 18, 19	22(9, 12), 25(13, 17), 27(1, 3, 11), 30(18, 19), 32(2,7), 35(5, 8), 38(16), 40(15)
caste	1, 3, 5, 6, 7, 8, 10, 11, 13, 14, 16, 17, 18	Agarwal(16), Arora(7, 18), Brahmin(1, 5), kayasat(13), lingayat(6, 8), SC(10, 11, 14), Vokkaliga(3, 17)
qualification	2, 3, 4, 5, 6, 7, 10, 11, 13, 14, 16, 17, 18, 20	B.Com(6), BE(5, 16), BSc(3,4), M.Com(17), M.Sc(11, 18), MBA(10, 13), MBBS(7, 14), MCA(2, 9)
appearance	1, 2, 3, 4, 6, 8, 9, 10, 12, 13, 15, 16, 17, 19, 20	Beautiful(9, 17), Fair(8, 20), good_looking(1, 10,16), handsome(6, 12,15), smart(2, 3, 4, 13, 19)
salary	1, 2, 3, 4, 5, 7, 8, 10, 11, 12, 14, 15, 16, 19, 20	10000(12, 19), 20000(1, 11, 14, 20),30000(3, 15, 16), 40000(2, 3, 8, 10), 50000(5, 7)

Table-4 Retrieval based on the combination of concepts/categories

Combination of categories	Document id
age 'or' caste	1, 2, 3, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19
age 'and' caste	1, 3, 5, 7, 8, 11, 13, 16, 17, 18
'not' appear	4, 6, 10, 14, 20
'not' caste	2, 4, 9,12, 15, 19, 20
caste 'or' qualification	1, 2, 3, 4, 5, 6, 7, 8, 10, 11, 13, 16, 17, 18, 20
caste 'and' qualification	3, 5, 6, 7, 10, 11, 13, 14. 16, 17, 18

It can be observed that the system retrieves only those documents that are indexed by the concepts in the classifier hierarchy. Further, retrieval is subjected to only those concepts in the classification hierarchy. In case of retrieval based on a document as a query, a query document is mapped only to the concepts of the underlying domain ontology.

5. CONCLUSION

This paper describes the task of retrieving the required information from the domain specific document collection based on domain ontology. Constructing a classification hierarchy for the purpose of retrieving the information is explained. As the crucial aspect of this model is the extraction of concepts, the performance measure in terms of extracting the

concepts from the text documents is presented. The approach developed is illustrated on the domain on texts describing the classified matrimonial advertisements.

ACKNOWLEDGEMENT

The authors would like to acknowledge the financial support provided by University Grants Commission, New Delhi, India, vide No. 34-523/2008(SR) to carry out this part of the research work.

REFERENCES

[1] Sharma S, "Information Retrieval in Domain Specific Search Engine with Machine Learning Approaches", In *Proceedings of world academy of Science, Engineering and Technology,* Vol. 32, pp. 142-145, AUGUST 2008.

[2] Kang K, Lin K., Zhou C. and Guo F., "Domain-Specific Information Retrieval Based on Improved Language Model", *In Proceedings of Fourth International Conference on Fuzzy Systems and Knowledge Discovery* (FSKD), 2007.

[3] Stanley Loh, Leandro K. W and Palazzo D. "Concept-based Knowledge discovery in Texts Extracted from the Web", In *Proceedings* of *ACM SIGKDD Explorations*, 2000.

[4] Gruber T., "Toward Principles for the Design of Ontologies Used for Knowledge Sharing", *In International Journal Human-Computer Studies,* Vol. 43, Issues 5-6, pp.907-928, November 1995.

[5] Castells, Fernandez M. and Vallet D., "An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval", *In IEEE Transactions on knowledge and data engineering*, Vol. 19, No. 2, pp.261-272, February 2007.

[6] Truong Q.D., Dkaki T., Mothe J. and Charrel P.J, "Information retrieval model based on graph comparison", *JADT 2008, 9es Journées internationales d'Analyse statistique des Données Textuelles,* Textual Data Statistical Analysis, 2008.

[7] Tsoumakas G. and Katakis I., "Multi-Label Classification: An Overview", *In International Journal of Data Warehousing and Mining, David Taniar (Ed.)*, Idea Group Publishing, 3(3), pp.1-13, 2007.

[8] Freitas and A.C.P.L.F. de Carvalho, "A Tutorial on Hierarchical Classification with Applications in Bioinformatics", *In Taniar, D., ed. Research and Trends in Data Mining Technologies and Applications. Advances in Data Warehousing and Mining*, IGI Publishing, USA, pp.175-208, 2007.

[9] Sum, Lim E. and Ng W., "Hierarchical Text Classification Methods and Their Specification, In Cooperative Internet Computing", *Edited by Alvin T. S. Chan, Stephen C. F. Chan, H. V. Leong, and Vincent T. Y. Ng, Kluwer Academic Publishers*, pp.236-256., Boston, March 2003.

[10] Janik, K. Kochut. "Training-less Ontology-based Text Categorization", *In Proceedings of Workshop on Exploiting Semantic Annotations in Information Retrieval (ESAIR 2008) at the 30th European Conference on Information Retrieval (ECIR'08)*, Glasgow, Scotland, 30 March 2008.

[11] Song M.H., Lim S.Y., Kang D.J. and Lee S.J., "Automatic Classification of Web Pages based on the Concept of Domain Ontology", *In Proceedings of the 12th Asia-Pacific Software Engineering Conference, APSEC*, pp.645-651, 2005.

[12] Wu S.H., Tsai T.H. and Hsu W.L., "Text categorization using automatically acquired domain ontology", *In Proceedings of the 6th International workshop on Information retrieval with Asian languages*, Vol. 11, pp.138-145 Sapporo, Japan, 2003.

[13] Shashirekha H.L., Murali S. and Nagabhushan P., "Constructing Domain Ontologies for Unstructured Text Documents: A Bootstrapping Approach", *In Proceedings of 3rd International Conference on Information Processing ICIP'09*, pp. 3-12, Bangalore, August 2009.

[14] Shashirekha H.L., Murali S. and Nagabhushan P., "Ontology based similarity measure for Text Documents", *In Proceedings of International Conference on Signal & Image Processing, ICSIP 2009*, pp. 602-606, Mysore August 2009.

Genetic Algorithm and Evolvable Hardware for Adaptive Filtration and Analysis using Texture, Color and Boundary

Vandana.V
&
Soumya Raja

Electronics and Communication Engineering Department
Sri Venkateswara College of Engineering
Pennalur , Sriperumbudur , India
vandana90@gmail.com
soumya.s14@gmail.com

ABSTRACT

Image de-noising is usually required to be performed before further processing like segmentation, object recognition and texture analysis. This paper gives a novel method of combining the filtration process and the post-processing techniques. For this purpose , evolvable hardware is combined with Genetic Algorithm to offer potential cost efficiency together with the flexibility of an adaptive system, producing a high-speed non-linear adaptive median filter . The spatial domain filtering techniques discussed are various forms of median filtering from which the 'FITTEST' algorithm is chosen depending on user requirements . Adaptive median filter removes the impulse & salt and pepper noise efficiently while retaining the edges and other detailed features. Further , after filtration , an object in the image is represented on a scale of categories and recognition algorithms are used to find the most detailed category according to information extracted from the image. The categorization is based on the color and texture content of the image. Border detection is also done by classifying the pixels as homogeneous and heterogeneous by comparison with the neighboring pixels . These techniques can be effectively implemented to process aerial photographs and to detect oil spills in the sea and properties of the ocean useful for fishery and navigation .

Keyword list: Impulse noise, Salt and pepper, adaptive window size, genetic , texture content , segmentation , evolutionary , evolvable.

1. INTRODUCTION

Image Processing is any form of signal processing for which the input is an image the output is either an image or a set of characteristics related to the image. Images might be corrupted with noises which have to be removed in order to make the image useful for interpretation. Methods to remove impulse noise are presented in the following sections .

Impulse noise is caused by malfunctioning pixels in camera sensors, faulty memory locations in hardware, or transmission in a noisy channel. Two common types of impulse noise are the salt-and-pepper noise and the random valued noise. For images corrupted by salt-and-pepper noise, the noisy pixels can take only the maximum and the minimum values in the dynamic range. Filters can be used to remove these noises .

Standard Median (SM) filter exploits the rank-order information of the input data to effectively remove impulse noise by substituting the considered pixel with the middle-position element (i.e., median) of the re-ordered input data. Doing this, they preserve edges in the image but may destroy small features in the image.

The Weighted Median (WM) filter uses a set of weighting parameters to control the filtering performance in order to preserve more signal details . Center Weighted Median filter is a special case of the WM filter, where only the center pixel of the filtering window has a weighting factor. Since the filtering should be applied to corrupted pixels only while leaving those uncorrupted ones intact , adaptive median filtering is used , wherein the window size is variable so as to detect the impulse noise and remove it effectively .One of the major problems with the median filtering is that it is relatively expensive and complex to compute. A combination of genetic algorithm with evolvable hardware can be used to compare the input with the target and calculates the fitness function depending on the input and user requirements to choose the best suited filtration process , overcoming the drawbacks mentioned above.

After processing the image , useful interpretations and conclusions can be made by classification of the image through segmentation and categorization . Segmentation is based on border , color and texture . Applications of these techniques are discussed in Section 8.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463K · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856322

2. MEDIAN FILTERS

Median filtering is a non-linear, low-pass filtering method, which is used to remove "speckle" noise from an image. Median filter removes isolated pixels, whether they are bright or dark. Figure i) illustrates an example calculation.

123	125	126	130	140
122	124	126	127	135
118	120	150	125	134
119	115	119	123	133
111	116	110	120	130

Neighbourhood values:

115, 119, 120, 123, 124, 125, 126, 127, 150

Median value: 124

Figure i) Calculation of median value

2.1. Algorithm for Median Filter

1. Store the neighboring pixels in an array. The neighboring pixels can be chosen by any kind of shape, for example a box or a cross. The array is called the window, and it should be odd sized.
2. Sort the window in numerical order
3. Pick the median from the window as the pixels value.
4. Repeat the process for every pixel.

3. ADAPTIVE MEDIAN FILTER

The unique feature in the Adaptive Median Filter is that the size of the window surrounding each pixel is variable. This variation depends on the median of the pixels in the present window. 'Processing' the image basically entails the following: Evaluation of the center pixel. If it is an impulse, then it is replaced with the median value of the pixels in that window and the window size is expanded. Else , the value of the center pixel is retained in the filtered image.

It self-adjusts its transfer function according to an optimizing algorithm. Because of the complexity of the optimizing algorithms, adaptive median filters adapt their performance based on the input signal. For some applications, adaptive coefficients are required since some parameters of the desired processing operation (for instance, the properties of some noise signal) are not known in advance. In these situations it is common to employ an adaptive filter, which uses feedback to refine the values of the filter coefficients and hence its frequency response.

3.1. Implementation of Adaptive Median Filter

The adaptive filter works on a rectangular region **Sxy**. The adaptive median filter changes the size of **Sxy** during the filtering operation depending on certain criteria as listed below. The output of the filter is a single value which the replaces the current pixel value at (x, y), the point on which *Sxy* is centered at the time. The following notation is used :

Zmin = Minimum gray level value in **Sxy**.; **Zmax** = Maximum gray level value in **Sxy** ; **Zmed** = Median of gray levels in **Sxy** ; **Zxy** = gray level at coordinates **(x, y)** ; **Smax** = Maximum allowed size of **Sxy**

The adaptive median filter works in two levels denoted Level A and Level B as follows:

Level A: A1= **Zmed – Zmin and**

A2= **Zmed - Zmax**

If A1 > 0 AND A2 < 0, Go to level B ,

Else increase the window size

If window size <= **Smax** repeat level A ,

Else output Zxy.

Level B: B1 = **Zxy – Zmin and**

B2 = **Zxy – Zmin**

If B1 > 0 And B2 < 0 output Zxy ,

Else output Zmed.

3.2 Purpose of Adaptive algorithm

To remove 'Salt and Pepper' noise

To smoothen any non impulsive noise

To reduce excessive distortions such as too much thinning or thickening of object boundaries.

4 .EVOLVABLE HARDWARE

Evolvable hardware(EHW) is based on the idea of combining reconfigurable devices with evolutionary algorithms such as Genetic Algorithms. The basic concept in EHW is to regard the configuration bits for reconfigurable hardware devices as chromosomes for Genetic

Algorithm(GA). By choosing an appropriate fitness function for the given task, GA can autonomously find the best hardware configuration in terms of chromosomes i.e. configuration bits. The algorithm for evolving circuits on a reconfigurable fabric is shown in figure ii.

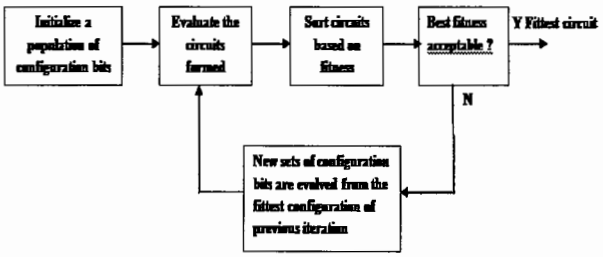


Figure ii) Algorithm for evolving circuits on reconfigurable hardware

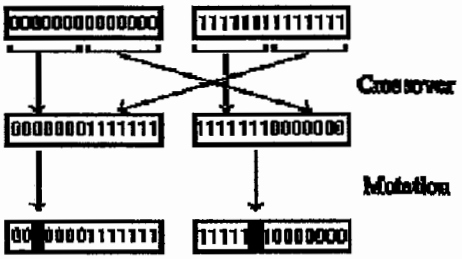


Figure iii) Crossover and Mutation

5. GENETIC ALGORITHM

Genetic Algorithm (GA) determines how the hardware structure should be reconfigured whenever a new hardware structure is needed for better performance. It is an evolutionary algorithm that uses techniques such as inheritance, mutation, selection and crossover. (Figure iii). The canonical GA has a population of chromosomes that are usually represented by the strings of binary characters (0 and 1s). At an initial state, the chromosomes in the population are generated at random, and are processed by many operations like selection, crossover and mutation (called genetic operations) i.e., in each generation the fitness of every individual in the population is evaluated , indicating how well the chromosome performs for the given problem. According to the fitness values, the selection determines which chromosome can survive in the next generation. The crossover chooses some pairs of chromosomes, and exchanges the sub-strings randomly. Finally, the mutation randomly picks some positions in the chromosome and flips their value. A typical genetic algorithm requires a genetic representation and a fitness function to evaluate the solution domain

Evolvable hardware is based on the concept of utilizing this robust capability of GA by reducing its competitional cost .

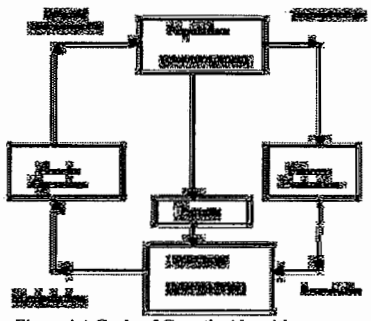


Figure iv) Cycle of Genetic Algorithm

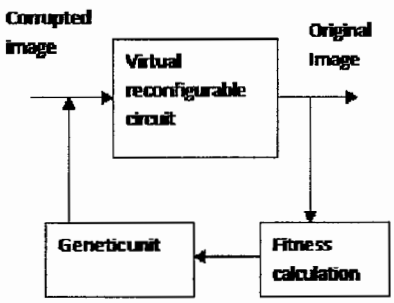


Figure v) Representation of Fitness calculation

6. DIGITAL IMAGE FILTER DESIGN USING GENETIC ALGORITHM

The digital image filter contains a virtual reconfigurable circuit and a genetic unit , the input for which is the corrupted image . Comparison of the output image with the original image yields the fitness value, according to which the most desirable architecture of the reconfigurable circuit is chosen .This is shown in figure v .

6.1 Algorithm

a) Read the corrupted image and the original image. Convert this to a bitmap file and store in the input buffer.
b) Generate an initial population of size n with each chromosome length l and for each chromosome in the population/selection:
   b.1. Choose N*N window and identify the pixels with impulse noise.
   b.2. Apply the required filtering technique using appropriate architecture. Repeat this for the entire image.
c) Select parent chromosome from the most efficient technique.
d) Apply crossover n mutation on the selected chromosomes to get the new child. Replace the old population.
e) Repeat from step three till the comparator gives the desired filtered image with minimum distortion.

7. PROCESSING OF AERIAL IMAGES

Aerial photography is the taking of photographs of the ground from an elevated position (platforms for cameras include aircrafts and rockets). Cameras may be hand held or mounted, and photographs may be taken by a photographer, triggered remotely or triggered automatically. Since
the condition in which an aerial image is captured is unknown and variable , the image can be filtered using any one of the (fittest) filtering techniques chosen by the reconfigurable circuit (GA) depending on the input and target image defined by the user .

8.1.Image Classification

Image classification is classified into two : a)Segmentation b) Categorization

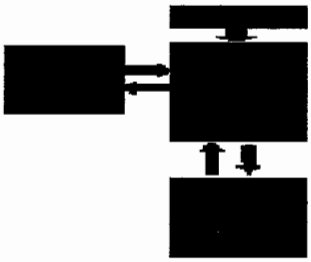


Figure vi) Segmentation

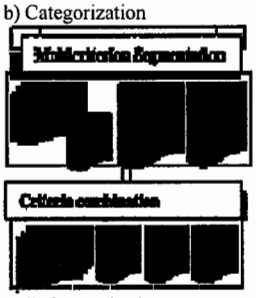


Figure vii) Categorization

7.1.1 Segmentation task

The three criteria of segmentation are : a)Borders b)Color c)Texture

7.1.1.1.Segmentation by the detection of borders:

If a region border is not known but regions have been defined in the image, borders can be uniquely detected. Assume that the image with regions is either binary or that regions have been labeled. An **inner** region border is a subset of the region An **outer** border is not a subset of the region.

7.1.1.1.1Algorithm for inner boundary tracing

1. Search the image from top left until a pixel of a new region is found ; this pixel P0 then has the minimum column value of all pixels of that region having the minimum row value . Pixel P0 is the starting pixel of that region border . Define a variable dir which stores the direction of the previous move along the border from the previous border element to the current border element . Assign
   a) dir=0 if the border is detected in 4-connectivity (Figure viii a)
   b) dir=7 if the border is detected in 8-connectivity (Figure viii b)
2. Search the 3*3 neighborhood of the current pixel in an anti-clockwise direction , beginning the neighbourhood search in the pixel positioned in the direction
   a) (dir+3) mod 4 (figure viii c)
   b) (dir+7)mod 8 if dir is even (figure viii d)
   c) (dir +6)mod 8 if dir is odd (figure viii e)

The first pixel found with the same value as the current pixel is a new boundary element Pn. Update the dir value .

3. If the current boundary element Pn is equal to the second border element P1 , and if the previous border element Pn-1 is equal to P0 , stop . Otherwise repeat step(2). The detected inner border is represented by pixels P0... Pn-2.

If the goal is to detect an outer region border, the given algorithm may still be used based on 4-connectivity.

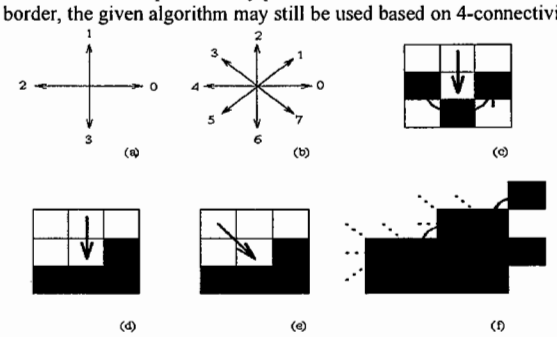


Figure viii) Inner boundary tracing : a)Direction notation , 4-connectivity , b) 8-connectivity , c) pixel neighborhood search sequence in 4-connectivity , d),e) search sequence in 8-connectivity f) Boundary tracing in 8-connectivity (dashed lines show pixels tested during border tracing)

7.1.1.2.Segmentation by the detection of color:
Color is a very important cue in extracting information from images. Color histograms are used in content-based retrieval systems . The difficulty is , for example , in distinguishing between a field of orange flowers and a tiger , because it lacks information about how the color is distributed spatially . It is important to group color in localized regions and to fuse color with textural properties . We treat the hue-saturation-value (HSV) color space as a cone : for a given point (h,s,v) , h and sv are the angular and radial coordinates of the point on a disk of radius v at height v are black , regardless of their h and s values . The cone representation maps all such points to the apex of the cone, so they are close to one another . The Cartesian coordinates of the points in the cone can be used to find color differences .

7.1.1.3.Segmentation by the detection of texture:
Texture analysis refers to the characterization of regions in an image by their texture content. It attempts to quantify intuitive qualities described by terms such as rough, smooth, silky, or bumpy as a function of the spatial variation in pixel intensities. In this sense, the roughness or bumpiness refers to variations in the intensity values, or gray levels . While color is a point property , texture is a local-neighborhood property . Various parameters like local range of an image , local standard deviation of an image, the local entropy of a grayscale image can be used .

7.1.1.3.1. Algorithm for detection of texture
1. NxN matrix is chosen
2. a) For local range deviation the centre pixel is replaced with the deviation between the first and the last pixel.
   b) For local standard deviation the centre pixel is replaced with the standard deviation between the first and the last pixel. Entropy of the neighboring pixel is calculated .
3. Based on the pixel value and that of the neighboring value , the texture is analyzed. For example if the deviation is zero then the surface is assumed to be smooth. This is a comparative study.

7.1.2. Categorization task
In this phase the system has to supply object prototypes corresponding to zones. The only characteristics allowing the selection of a list of the prototypes are the texture, the color, and the shape.
In this example, the entry is constituted by the recognizable color and texture zones as shown.
Zone1: Uniform color blue (selected objects: sky, sea);
Zone2: Brown color and texture of bare ground (selected objects: mountain);
Zone3: Green color and forest texture.


Figure x) Prototype bases


Figure xi) Segmented zones


Figure xii) Resulting zones

8. APPLICATIONS

8.1.Glacier and ice-berg detection
This is based on the segmentation using boundary detection. The inner boundary specifies the region within the image. Hence inner boundary detection is used .This can be of great help to the navigators.

8.2.Detection of Oil spills in the sea
This is based on segmentation using texture analysis. The difference in the densities of oil and water provide a variation in the texture of the components. Hence by analyzing the pixel values , the textural deviation can be calculated. Hence oil spills can be spotted in the sea. This helps to decipher the concentration of aquatic life and other marine properties.

8.3 Contour Detection
The various contours like forest, dry land, mountains etc can be detected by a combination of texture, boundary and color segmentation. The image is compared with the data to identify the object. (As shown in figures x , xi , xii). Genetic algorithm is used for this. The best suited result is chosen from the population. The image is filtered using the evolvable hardware and genetic algorithm. Hence it is highly efficient.

Figure xiii) Glacier detection : AB-Ablation zone , AC- accumulation zone , T- Tributary Glacier

Figure xiv) Oil spill Detection (White pixel=oil)

9. RESULTS

The adaptive median filter is designed to remove impulsive noise from images. Therefore, our algorithm's performance was first tested with basic salt and pepper noise with a noise density of 0.25.

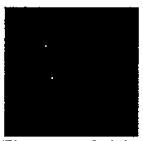


Figure xv) Original image

The corrupted image

Result of SM filter

Result of AM filter

10. CONCLUSION

The work has been presented as a novel approach to digital image filtering. This adaptive median filter successfully removes impulse noise and preserves the edges. Furthermore in order to improve the flexibility, an attempt has been made to merge the above technique with genetic algorithm. This algorithm is implemented in an evolvable hardware which changes the task requirements through the ability to reconfigure its own structure. Hence the type of filtering that gives the best result for different kinds of images can be obtained automatically by the system. The above technique can be used in real-time applications to process aerial images by choosing the most appropriate method of filtering using genetic algorithm and to recognize objects in the image by using texture analysis through segmentation and categorization on the basis of texture , color and border . This is useful to detect oil spills in the sea and to detect glaciers and ice bergs that are retreating .

11. REFERENCES

[1] Zdenek Vasicek, Lukas Sekanina, Novel Hardware Implementation of Adaptive Median Filters 978-1-4244-2277-7/08/ ©2008 IEEE.

[2] V.V. Bapeswara Rao and K. Sankara Rao, A New Algorithm for Real-Time Median Filtering, IEEE Transactions on Acoustics, Speech, Processing VOL ASSP-34. NO. 6, December 1986.

[3] N.C. Gallagher and G.L. Wise, "A Theoretical Analysis of the Properties of Median Filters," IEEE Transactions on Acoustic, Speech and Signal Processing,

[4]Thomas S. Huang, Two-Dimensional Digital Signal Processing II: Transforms and Median Filters, Springer-Verlag New York, Inc., Secaucus, NJ, 1981

[5] M. Karaman, L. Onural, and A. Atalar, "Design and Implementation of a General Purpose Median Filter in VLSI," in VLSI Signal Processing III, IEEE Press, 1988.

[6]. G.R. Arce and P.J. Warter, "A Median Filter Architecture Suitable for VLSI Implementation," in Proceedings of the 23rd Annual Allerton Conference of Communication Control Computing , Oct. 1984, pp. 172-181.

New method for image denoising using Nonsubsampled WBCT

Li Min[1] Wang Ting[2] Wang Cheng-biao[1] Fu Zhi-qiang[1] Li Wei-qing[1]

[1]School of Engineering and Technology, China University of Geosciences, Beijing 100083, China

[2]Shan Dong Machinery Group Co. Ltd, Zibo 255201, China

ABSTRACT

A novel denoising method based on the nonsubsampled wavelet-based contourlet transform (N-WBCT) was proposed in this paper. It employs wavelet transform for multi-scale decomposition, and nonsubsampled directional filter banks (NSDFB) for directional decomposition. N-WBCT has not only the properties of multi-resolution and multi-direction, but also the property of translation invariance, which is useful in eliminating the Gibbs phenomenon. The experiment results showed that this algorithm can get higher PSNR and the better visual.

Keywords: Image denoising; Contourlet transform; Wavelet-based contourlet transform (WBCT); Nonsubsampled directional filter banks

1 INTRODUCTION

Image denoising is the basic problem in image processing. The existing image denoising methods cannot capture the geometric information of images and tend to amplify noises when they are applied to noisy images since they cannot distinguish noises from weak edges. Recently, various wavelet-like Multiscale Geometric Analysis (MGA) methods to effectively solve this problem. Among them, contourlet transform is one of the most important approaches, it extracts the geometric information of images, which can be used to distinguish noises from weak edges[1][2]. Since it can capture significant information from a nature image, contourlet become a promising technique for image denoising.

To further exploit the potential of contourlet, the WBCT algorithm was first introduced by Eslami and Radha at 2004. It is a non-redundant version of the contourlet transform, and appropriately used for image denoising. It applied the directional filter banks (DFB) to all the detail subbands of wavelets in a similar way that one constructs contourlet. The main difference is that wavelets instead of the Laplacian pyramids employed in contourlet.

The main disadvantage of the WBCT is the occurrence of artifacts that are caused by setting some transform coefficients to zero for denoising. In this paper, a new method was proposed to solve this problem, nonsubsampled WBCT. First, we employ wavelet as the subband multiresolution decomposition, then apply the NSDFB. We also

[1] Li Min, postgraduate student in School of Engineering and Technology, China University of Geosciences, Email: lilinlimin@163.com

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463L · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855572

compare this method with other traditional method. The result show that our proposed scheme is capable of retaining textures and fine details.

2 NONSUBSAMPLED WBCT

Sec. 2.1 Contourlet Transform and WBCT

Contourlet transform proposed by M. N. Do and Vetterli was a "true" two dimensional transform. Compared to wavelet, contourlet transform has the ability of image multi-scal and direction decomposition using the anisotropic basis, so it is better to capture the curve singularity. The multi-scale analysis and the directional analysis is separate in the contourlet transform. Firstly uses the Laplacian Pyramid transform (LP) to do multi-scale decomposition, secondly uses the DFB to filter the high frequency from each Laplacian Pyramid channel[2].

Both the LP and the DFB use perfect-reconstruction filters, so the contourlet transform is perfect reconstructable. But the contourlet transform has a redundancy ratio that is less than 4/3, which comes from the LP. As a remedy for this problem, the WBCT algorithm was proposed. It brings sparser representation and can capture the intrinsic geometrical structure of images, which is widely applied in image denoising and attracted great attention from the image processing community.

Similar to the contourlet transform, the WBCT consists of two filter band stages[4]. The first stage adopts wavelet to implement subband decomposition to replace Laplace pyramid decomposition, thus the redundancy caused by the Laplace pyramid decomposition can be eliminated. At each level in the wavelet transform, we obtain three highpass bands corresponding to the LH, HL and HH bands. The second stage of the WBCT uses the DFB, which provides angular decomposition at the same directions. Figure 1and Figure 2 illustrate a schematic plot of the WBCT using 3 dyadic wavelet levels and 8 directions at the finest level. The directional decomposition is overlaid the wavelet subbands. The wavelet transform and the DFB have no redundancy ratio, so WBCT is a non-redundant transform contrast to the contourlet transform[5][6][8].

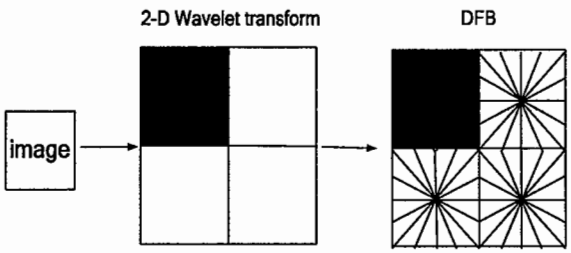


Figure 1. A framework of the WBCT

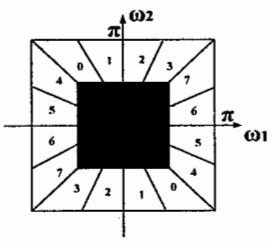


Figure 2. DFB with 8 frequency bands

Sec. 2.2 Nonsubsampled WBCT and NSDFB

Due to downsampling and upsampling, DFB applied to the WBCT are lack of translation invariance, and the lack of

translation invariance causes pseudo-Gibbs phenomena around singularities. However, translation invariance is desirable in image analysis applications such as edge detection, contour characterization, and image denoising. Hence, for image analysis and primitive extraction purposes, the N-WBCT is proposed.

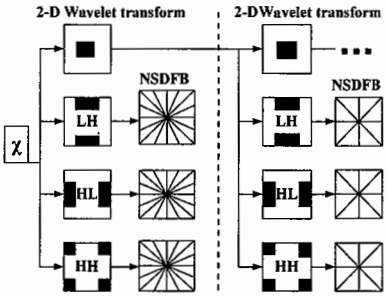


Figure 3. The implementation of N-WBCT

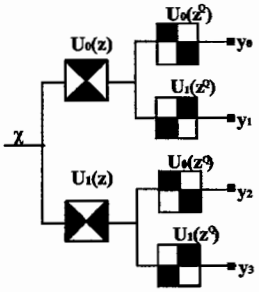


Figure 4. Frequency decomposition of four-channal NSDFB

To achieve the translation invariance, the N-WBCT built wavelets and NSDFB, as shown in Figure 3. Taking advantage of the wavelet transform that has efficient nonlinear approximation property, the important features of directionality and translation invariance are added by applying NSDFB to all subbands of the wavelet. The artifacts usually introduced by DFB are eliminated.

The NSDFB is a translation invariance version of the critically sampled DFB in the N-WBCT, which is obtained by simply switching off each of the downsamplers and the upsamplers in the DFB tree structure and upsampling the filters accordingly. This results in a tree composed of two-channel nonsubsampled filter banks. All filters in the tree structure are obtained from a single nonsubsampled filter bank with fan filter. An L-level tree structured decomposition leads to 2^L subbands with wedge-shaped frequency partition. To obtain finer directional decomposition, NSDFB is iterated repeatedly. For next level, all filters are upsampled by a quincunx matrix[7].

Figure 4 illustrates a four channel decomposition. In the second level, the upsampled fan filters $U_i(Z^Q)$, i=0 ,1 have checker-board frequency support, and when combined with the filters in the first level give the four directional frequency decomposition. The building block of the nonsubsampled pyramid is a two channel nonsubsampled filter bank. And it has no downsampling or upsampling. Hence it is translation invariance. The perfect reconstruction condition is given as

$$H_0(Z^Q)G_0(Z^Q)+ H_1(Z^Q)G_1(Z^Q)=1 \quad (1)$$

3 THE PROPOSED DENOISING SCHEME

Figure 5 shows the flow graph of N-WBCT denoising scheme. The algorithm can be described as follows:

Step 1: Apply wavelet transform to the noising image.

Step 2: In each scale, apply NSDFB to the every directional subband of wavelet decomposition

Step 3: Using soft and hard thresholding to denoise the N-WBCT.

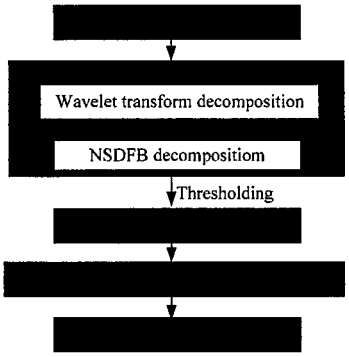


Figure 5. Flow graph of the Nonsubsampled WBCT denoising scheme

4 EXPERIMENT RESULTS AND DISCUSSION

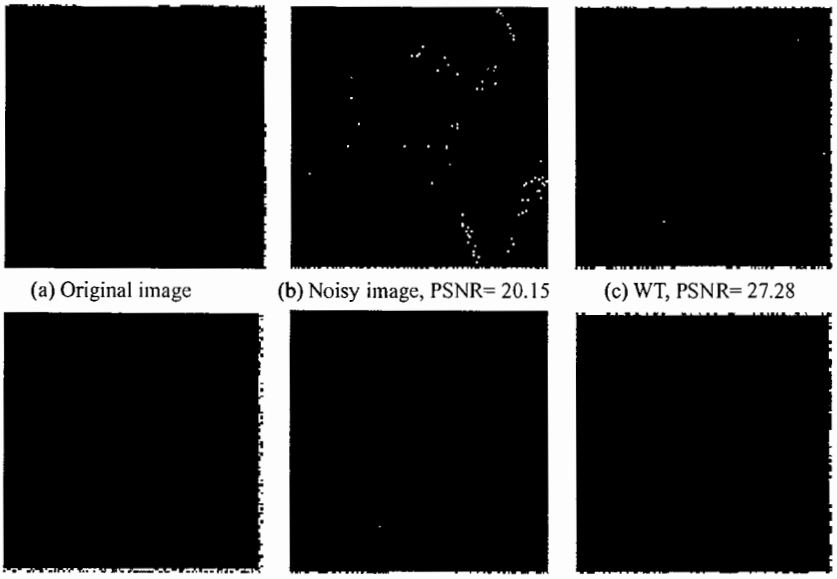
To test our proposed algorithm ,we selected the standard "lena" and "cameraman" images ,ues four approaches for our experiments: wavelet (WT), contourlet (CT), WBCT, and N-WBCT. The images show in Figure 6and Figure 7. The Table1 showed the peak signal-to-noise ratio (PSNR) results that we proposed approach in this article is superior to others.

Table 1. The comparison of PSNR in dB for denoising performance

		PSNR/dB				
Image	σ	Noisy image	WT	CT	WBCT	N-WBCT
Lena	20	22.05	28.39	28.61	27.98	29.97
	25	20.15	27.28	27.63	27.31	28.83
	30	18.57	26.28	26.81	25.17	27.79
	35	17.23	25.49	26.05	22.82	27.18
	40	16.06	24.69	25.59	20.63	26.53
cameraman	20	22.11	25.89	25.13	24.89	27.36
	25	20.14	24.73	24.22	24.27	26.52
	30	18.53	23.78	23.47	22.94	25.81
	35	17.23	23.08	22.89	21.34	25.15
	40	16.07	22.31	22.45	19.77	24.44

The reconstructed image of wavelet algorithm has a high PSNR but it blurred image and bad in capture the detail

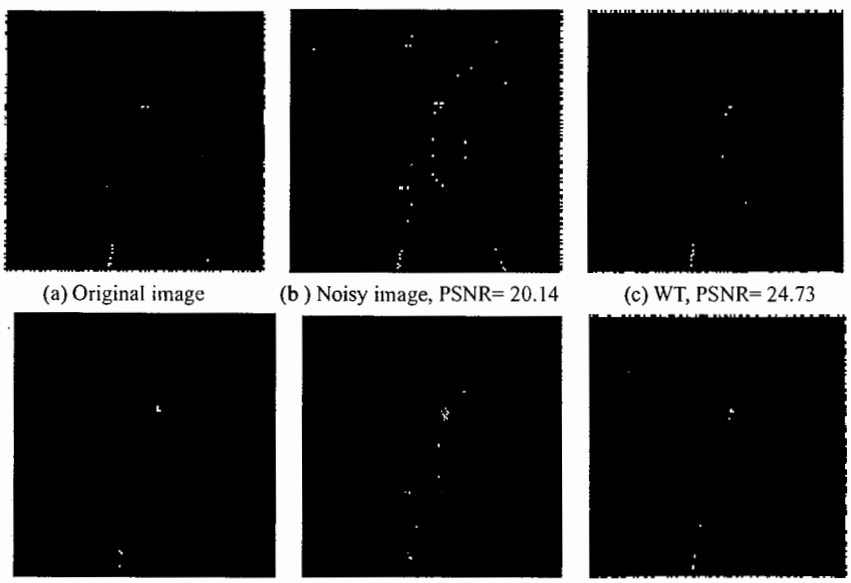
of image. From the contourlet in image denoising we find that there some additive noise and shift value on reconstructed image, it causes the low PSNR results and bad in vision. The WBCT algorithm causes pseudo-Gibbs phenomena on reconstructed image. The N-WBCT is a fully translation invariance, multi-direction, and multi-scale, so reconstructed image of N-WBCT algorithm is obviously better than the WT, CT and WBCT algorithm.



(a) Original image (b) Noisy image, PSNR= 20.15 (c) WT, PSNR= 27.28

(d) CT, PSNR= 27.63 (e) WBCT, PSNR= 27.31 (d) N-WBCT,PSNR=28.83

Figure 6. Denoising experiment with the lena image corrupted with a gaussian noise of σ=25



(a) Original image (b) Noisy image, PSNR= 20.14 (c) WT, PSNR= 24.73

(d) CT, PSNR= 24.22 (e) WBCT, PSNR= 24.27 (d) N-WBCT, PSNR=26.52

Figure 7. Denoising experiment with the cameraman image corrupted with a gaussian noise of σ=25

5 CONCLUSION

In this paper, we present a new image denoising method based on the N-WBCT which is an improvement on WBCT. Our experiment results clearly demonstrated the capability of the proposed scheme in image denoising can effectively deal with images with smooth contours and maximally suppress the Pseudo-Gibbs phenomenon.

REFERENCES

[1] Do M. N., and Vetterli M., "The contourlet transform: an efficient directional multiresolution image representation", IEEE Trans. Image Processing, IEEE, Washington D.C.,2005,14 (12), pp.2091-2106.

[2] M. N. Do and M. Vetterli, "Contourlets: A directional multi-resolution image representation", Proc. of IEEE intl. Conf. on

Image Processing, Rochester, September 2002.

[3] D. Donoho. "Denoising by soft-thresholding". IEEE.Trans. Info. Theory, 43:613-627, 1995.

[4] R. Eslami and H. Radha, "Wavelet-based contourlet transform and its application to image coding," in proc. of IEEE International Conference on Image Processing, Oct. 2004.

[5] R. Eslami and H. Radha, Wavelet-based Contourlet Packet Image Coding, Conference on Information Science and Systems, The Johns Hopkins University, Mar 16-18 2005.

[6]R. H. Bamberger and M. J. T. Smith, "A filter bank for the directional decomposition of images: theory and design," *Signal Processing, IEEE Transactions on*, vol. 40, pp. 882-893, 1992.

[7] J. P. Zhou, Arthur L. Cunha, and Minh N. Do.,"Nonsubsampled contourlet transform: construction and application in enhancement", IEEE ICIP. 2005, pp. 469-472.724

[8] Duncan D Y Po, and Do M N, "Directional multi-scale modeling of images using the contourlet transform", IEEE Trans. Vol. 15, 2006,pp. 1610-1620.

Symmetry Based Fast Marching Method for Icosahedral Virus Segmentation

Guihua shan, Jun Liu, Liang Ye, Xuebin Chi

Computer Network Information Center, Chinese Academy of Sciences

Graduate University of Chinese Academy of Sciences

ABSTRACT

Segmentation of icosahedral virus density map from cryo-electron microscope (CryoEM) is a challenging task because virus structure is complex and density map is at low resolution. Fast marching method is widely used in segmentation, in which seed selection is essential for correct segmentation results. However, the selection of an appropriate seed is difficult. In this paper, we present the method of selecting the seed in fast marching algorithm by making use of the shape symmetry to improve the fast marching method for icosahedral virus segmentation. Based on the feature of icosahedron, we compute and get its symmetry axes inside the density map. With these symmetry axes, we specify the initial seeds with the local maxima value along symmetry axes. Further, the new data structures are presented, which can effectively reduce the memory cost when implement the fast marching algorithm. Experimental results show that the approach can obtain segmentation results of the density maps fast and accurately.

symmetry, icosahedron, segmentation, fast marching method

1 INTRODUCTION

The functions of macromolecular machines that are mainly composed by proteins are dependent on their structures. Viruses are the simplest life unit in the world. 95% of human viruses and most plant and animal viruses have icosahedral shaped shells [1], which have very well symmetry features. These shells come in a variety of sizes depending upon the viruses which are generally made up of pentagons, trimers and hexagons .

CryoEM is a very important method for studying the structures of macromolecular machines. Three main steps are involved in this technology. First is picture collection from virus by CryoEM. The second is 3D density map reconstruction from these pictures. The last step is data analysis for reconstructed density map by visualization, segmentation and structure-fitting methods. The virus is made up of multiple components. For understanding the underlying structure of the complex virus, the segmentation of the density map into more manageable substructures is indispensable. However, segmentation is considered as one of the hardest open problems in CryoEM [3].

Currently, such segmentation is typically carried out by hand. The popular visualization tools for analyzing density map only support manual segmentation, such as Chimera and Spider. This results in inconveniency that the users have to decompose the density map by tracing the elements of interest with a graphical interface. It is tedious and time-consuming. Moreover, it segments the mesh of a specific iso-surface, not 3D volumetric data. When iso-value changed, user has to restart a new segmentation.

To resolve this issue, we propose symmetry based fast marching method for icosahedral virus segmentation, where we make use of the shape symmetry to select the seed to improve the traditional fast marching method, which is able to automatically segment the whole density data into primary subunits. Our approach includes four steps. Firstly, we compute the location of all symmetry axes inside density map based on the feature of icosahedra. Secondly, we make use of the symmetry information to automatically select the seeds for multi-front growing. Thirdly, all front grow simultaneously. Finally, the front stops growing when it's narrow band become empty.

The rest of this paper is organized as follows. In Section 2, we summarize the related work on seed selection for fast marching method for segmentation. In section 3, we detect the symmetry axes inside the density map. Then we propose the symmetry based fast marching method in section 4. Section 5 describes the new data structures for implement.

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463M · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855737

Experimental results on the virus density maps are also presented in this section. Finally we conclude this paper in Section 6.

2 RELATED WORK

Segmentation is one of the hardest fields in image processing. Many people have contributed to this topic and have obtained a lot of advances [2] [4-7]. Level set, developed by Sethian[10][11], is one of the widely used methods. Fast marching method is a simple version of the level set method, which based on the theory of evolution. Front evolutes from a seed point in the shape of interest. The seed selection is essential for correct segmentation results. However, the selection of an appropriate seed is difficult. Nabil et.al.[14] proposed a visual attention guided seed selection method. A multicue saliency-based visual attention model is built for detects the most salient parts of a given scene. And then using those detected regions as seeds for segmentation. Zhu et. al [12] present a seed selection method based on the assumption that the real edge of the object is on boundary of the over-segmentation results. The image first is over-segmented by watershed transform into very small regions, and then let the center point of these small regions as start seeds for fast marching methods. Yu et al. [8] describe a modified fast marching method, in which critical points are obtained by gradient vector filter and the seeds are selected from these critical points.

3 SYMMETRY DETECTION OF DENSITY MAP

Icosahedron has twelve global 5-fold rotational symmetry axes, which site on six lines, and twenty or thirty local three-fold symmetry axes. Considered rectangular Cartesian coordinate system , usually the center of the macromolecular density map is the origin and one of these 5-fold symmetry axes is coincident with the Z-axis。 Mark V_1 the normal vector of this symmetry axis, then $V_1 = (0\ 0\ 1)$. There are five 5-fold symmetry axes evenly distributed around V_1, and form θ angle with V_1. The other symmetry on the same line with V_1 is $V_{12} = (0\ 0\ -1)$. The left five 5-fold symmetry axes are evenly distributed around V_{12} and form θ angle with V_{12}. Let us denote the angle of X-axis by φ_i and the projection of V_i on the XY plane. We get $\varphi_i = (i-2)\times 2\pi/5$ because V_i $(i = 2,\cdots,6)$ are evenly distributed around V_1. The relationship is shown in Fig1.

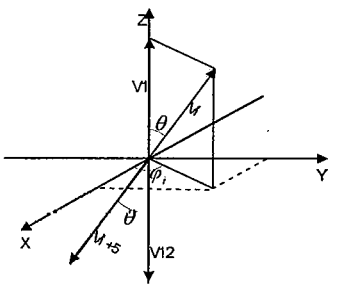


Fig.1 the distribution of the global five-fold symmetry

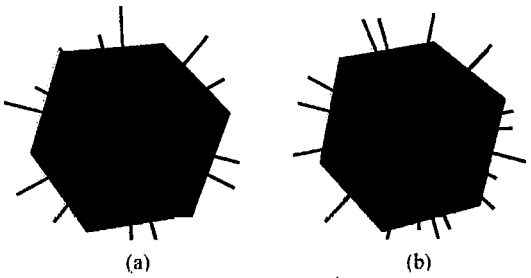


Fig.2 two cases of the distribution of the local three-fold symmetry: a)symmetry axis on center of triangles,b) symmetry axis on center of edges

Then, V_i satisfy

$V_i = (\sin\theta\cos\varphi_i, \sin\theta\sin\varphi_i, \cos\theta)$,

Since V_{i+5} and V_i are on the same line, we can derive

$V_{i+5} = -V_i$

$= (-\sin\theta\cos\varphi_i, -\sin\theta\sin\varphi_i, -\cos\theta)$

Due to $\langle V_1, V_2 \rangle = \langle V_3, V_4 \rangle$

We obtain

$\cos\theta = \sin^2\theta\cos(2\pi/5) + \cos^2\theta$

Solve the above equation, we get

$\cos\theta = \sqrt{5}/5$, $\sin\theta = 2\sqrt{5}/5$

So we obtain all the global 5-fold symmetry axes. According the definition of icosahedrons and Euler's theorem, we derive its 20 surfaces and 30 edges. Let us denote 20 surfaces by S_i $(i = 1,\cdots,20)$, and 30 edges by L_j $(j = 1,\cdots,30)$ is organized as following:

$S_1=(V_1,V_2,V_3)$, $S_2=(V_1,V_3,V_4)$,
$S_3=(V_1,V_4,V_5)$, $S_4=(V_1,V_5,V_6)$,
$S_5=(V_1,V_6,V_2)$, $S_6=(V_{12},V_7,V_8)$,
$S_7=(V_{12},V_8,V_9)$, $S_8=(V_{12},V_{10},V_{11})$,
$S_9=(V_{12},V_{10},V_{11})$, $S_{10}=(V_{12},V_{11},V_7)$,
$S_{11}=(V_2,V_3,V_{10})$, $S_{12}=(V_3,V_4,V_{11})$,
$S_{13}=(V_4,V_5,V_7)$, $S_{14}=(V_5,V_6,V_8)$,
$S_{15}=(V_6,V_7,V_9)$, $S_{16}=(V_7,V_8,V_5)$,
$S_{17}=(V_8,V_9,V_6)$, $S_{18}=(V_9,V_{10},V_2)$,
$S_{19}=(V_{10},V_{11},V_3)$, $S_{20}=(V_{11},V_7,V_4)$,
$L_1=(V_1,V_2)$, $L_2=(V_1,V_3)$, $L_3=(V_1,V_4)$,
$L_4=(V_1,V_5)$, $L_5=(V_1,V_6)$, $L_6=(V_2,V_3)$,
$L_7=(V_3,V_4)$, $L_8=(V_4,V_5)$, $L_9=(V_5,V_6)$,
$L_{10}=(V_6,V_2)$, $L_{11}=(V_{12},V_7)$, $L_{12}=(V_{12},V_8)$,
$L_{13}=(V_{12},V_9)$, $L_{14}=(V_{12},V_{10})$, $L_{15}=(V_{12},V_{11})$,
$L_{16}=(V_{11},V_7)$, $L_{17}=(V_2,V_3)$, $L_{18}=(V_3,V_4)$,
$L_{19}=(V_3,V_{10})$, $L_{20}=(V_3,V_{11})$, $L_{21}=(V_4,V_{11})$,
$L_{22}=(V_4,V_{12})$, $L_{23}=(V_5,V_7)$, $L_{24}=(V_5,V_8)$,
$L_{25}=(V_6,V_8)$, $L_{26}=(V_6,V_9)$, $L_{27}=(V_7,V_8)$,
$L_{28}=(V_8,V_9)$, $L_{29}=(V_9,V_{10})$, $L_{30}=(V_{10},V_{11})$,

As for local three-fold, there are two different cases:
1) There are twenty 3-fold symmetry axes which lie on the center of the twenty triangles of icosahedra(see Fig.2(a));
2) There are thirty 3-fold symmetry axes which lie on the center of the thirty edges of icosahedra (see Fig.2(b)).
In case 1, let us denote the i th triangle by $S_i=(V_{i1},V_{i2},V_{i3}), i=1,\cdots,20$ then $C_i=(V_{i1}+V_{i2},V_{i3})/3$ is the local 3-fold symmetry axis.
In case 2, let us denote the i th edge by $L_i-(V_{i1},V_{i2}), i-1,\cdots,30$, then $C_i=(V_{i1}+V_{i2})/2$ is the local 3-fold symmetry axis.
From the above computing, we finally obtain all 5-fold symmetry and local 3-symmetry axes.

4 SYMMETRY BASED FAST MARCHING METHOD

We have obtained symmetry axes in section 3.For each axis of symmetry, compute the local maxima and select the maxima of all the local maxima along this axis as seed for fast marching method, since the maxima density capture the most important features of a given map. Each seed corresponds to a different segmentation object which is five-fold or three-fold.

Let all fast marching fronts grow simultaneously. This may cause problems when each of the front stop and what happened if the point that is the minimum in its narrowband is already be occupied by the other front. In the fast marching methods, the moving direction is decided by the time value T[14] of the Narrowband. If the point with the minimal time value is already be occupied by other front, it will be deleted from its narrowband. So this front will never go to that way. The more of its neighbors are reached by the others, the less of directions it can move in. When all points in its neighborhood are occupied by others, its close set will be empty, and then it will automatically stop moving.

Algorithm. Symmetry based fast marching methods:

Initialize:

For each symmetry Axis:

1. *Alive points:* let alive point be the grid point (i, j, k) with max value of local maxima along this axis, which is seed for initial front. Set T (i, j, k) =0;
2. *Narrowband points:* Let Narrowband points be the set of alive points' 6-voxel neighborhood. Set $T(i,j,k)=1/F(i,j,k)$, $F(i,j,k)=e^{-\alpha|\nabla G_\sigma * I(i,j,k)|}$,where G_σ is a Gaussian filter and α is a positive constant.
3. *Far point*: Let far points be all other data. Set T (i, j, k) = Time_Max .

Loop step:

For each symmetry Axis:

1. If the Narrowband is empty, quit loop. If the Narrowband is not empty, select point (i,j,k) with the smallest value for T among Narrowband points

2. Judge whether point (i, j, k) is alive point in the other Narrowband. If true, delete it from Narrowband. Go back to step 1, else, continue.
3. Tag point (i, j, k) as alive point, and delete it from Narrowband.
4. Tag as Narrowband points all neighbors of point(i, j, k) , which are not Alive. If the neighbor is in Far set, remove it from that list and add to Narrowband.
5. Recompute the value T of Narrowband points.

The algorithm has two parts: initialization and loop. In the initialization part, all points are classified to three statuses: alive, narrowband and far,where the seeeds are selected as alive points. In the loop part, all fronts are marching according to five steps. Compared with traditional fast marching algorithm, there are three main differences in our algorithm. Firstly, we specify the initial seeds with the local maxima value along symmetry axes. Secondly, the algorithm works for multi-objects related to the symmetry axes, while traditional method is designed for one object segmentation. Thirdly, the automatic stop condition is defined in the loop step1.

5 IMPLEMENTATION AND APPLICATION

In this section, firstly we propose the new data structures for effective reducing the memory cost. Then we apply symmetry detect method and the segmentation method to two virus density maps.

When each front is directly initialized and go forwards according to traditional fast marching algorithm, it will need a matrix D for recording its own time map over the whole grid, a matrix S for marking all points' status and a min-heap-data structure for save narrowband to quickly obtain the minimum of time value. Let m is the number of objects which the virus will be decomposed into. And n is the size of the virus. The memory cost is 4×n×m bytes. When m is large (for example m=42), the memory cost is huge. To solve this problem, we design several novel data structures to largely reduce the memory consume. The new data structures for D, S and min-heap go as follows:

```
struct nodestucture {
        Int index; //No.index front
        float time; //time map
        struct nodestucture * nextnode;
};
nodestucture * D[n]
```

The min heap data structure is also changed similar to the time map:

```
struct heapstructure{
                int index;
                fibheap_el * node;
                struct heapstructure * nextnode;
        };
Heapstructure *heap_pool[n];
```

To record segmentation status of all fronts, we only use one matrix S to save the alive information for each front. The vale i in S at point (x,y,z) means this point is alive point of ith front and no information saved for narrowband and far status. Memory cost is decreased from 4×n×m to 4×n.

If each member of D has m nodes, the memory cost is the same with the traditional algorithm. But in icosahedron, each five-fold object has 5 local three-fold neighbours, and each three-fold object has 6 five-fold neighbours. Berger B, et.al.[1] and Baker [15] proposed more detail about the virus structure in icosahedron. None member of D and min-heap have more then 3 nodes. Moreover only the common edges have members with multi-nodes. Common edges are very small parts of the whole icosahedron, and most of members in D have only one node. So we are able to largely reduce the memory cost.

With these new data structures, we modified the initialization step described in Algorithm 1. We only record the alive points for each front on the same matrix. narrowband points are record in min-heap. Time map D is initialized as following:

```
For each symmetry axis,
  if (D.at(i,j,k)==NULL){
  D.at(i,j,k)= new nodestucture();
  D.at(i,j,k)->distance=0:
  D.at(i,j,k)->index=index;
  D.at(i,j,k)->nextnode=NULL;}
```

We present the segmentation results on two reconstructed density map : Polio Virus (PLV) and hepatitis B virus (HBV) from CryoEM.

PLV is of size 150*150*150 with T=4 ,which has 12 five-fold subunits and 20 six-fold subunits. The pentons and trimers are the primary subunits. To segment out these members will do great help to understand the virus structures. The

segmentation results are presented in Fig.3. Fig.3. (a) shows original isosurface view with the detected symmetry axes before segmentation. Fig.3.(b) shows one snapshot when the fronts are growing simultaneously from the seeds distributed over the virus based symmetry information; Fig.3.(c) shows that PLV is decomposed into 32 subunits-pentons and trimers which are all marked in different colors. Fig.3.(d) shows one of the segmented pentons. Fig.3.(e) shows one of the segmented trimers.

The HBV is a DNA virus and a member of the hepadna family of viruses. The size is 160*160*160. Its T value is equal to 3. It is typical icosahedra shape with 12 five-fold subunits and 30 six-fold subunits. The pentons and hexons are the primary members of this virus. Fig.4 shows the segmentation results. Different segmentation objects are marked with different colors. Fig.4.(a) show the detected symmetry axes; Fig.4.(b) shows one snapshot when all fronts are growing simultaneously from the seeds distributed over the virus based symmetry information; Fig.4.(c) shows HBV is decomposed into 42 subunits: pentons and hexons, which are all marked with different colors; Fig.4.(d) shows one of the segmented pentons. Fig.4.(e) shows one of the segmented hexons .

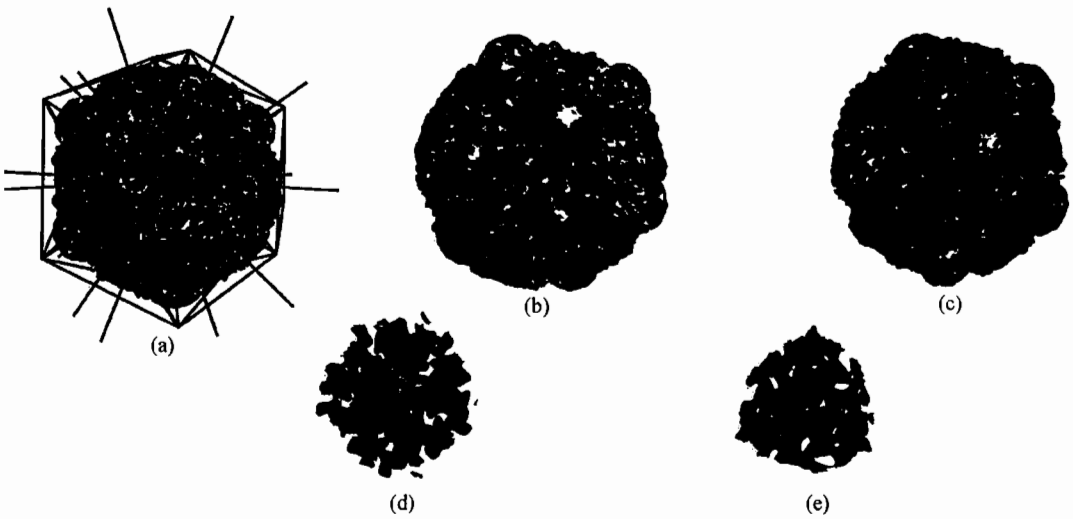


Figure3. Polio Virus segmentation results, different segmented objects marked in different colors. (a).The original isosurface view with detected symmetry axes before segmentation; (b) A snapshot of the 32 fronts are growing simultaneously from seeds distributed over the virus based symmetry information; (c) The virus is segmented into 32 subunits: pentons and trimers; (d) One of the segmented pentons; (e) One of the segmented trimer.

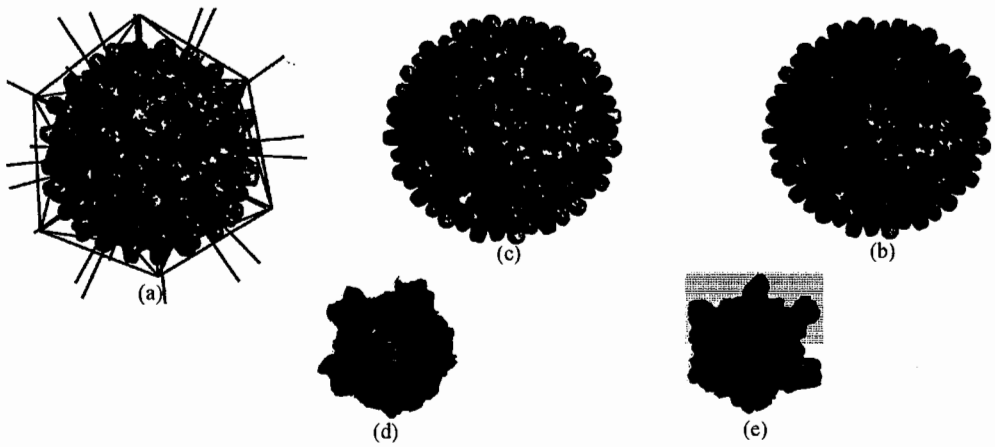


Figure4. HBV segmentation results, different segmented objects are marked with different colors. (a).Original iso-surface view with detected symmetry axes before segmentation; (b) A snapshot of the 42 fronts are growing simultaneously from seeds which distributed over the virus based symmetry information (c) The virus is decomposed into 42 subunits: pentons and hexons; (d) One of the segmented pentons; (e) One of the segmented hexons.

6 CONCLUSION

We propose symmetry based fast marching method for icosahedral virus segmentation. It makes use of symmetry feature inside the data and biological domain knowledge to improve the fast marching method, which is able to automatically segment the whole density data into primary subunits. Seeds are selected by local maxima value along symmetry axis. Each seed start a fast marching front. Those front grow simultaneously and will stop when their narrowband become empty. For effective implementation, new data structures are designed which can effectively decrease the memory cost. Experimental results show that the efficiency of the method.

7 ACKNOWLEDGE

The work was supported by National Science Fund grants No.60673064 and Knowledge Innovation Project of The Chinese Academy of Sciences (No.KGGX1-YW-13). We are grateful to Prof. Fei Sun and his research group from Institute of Biophysics Chinese Academy of Sciences for helpful discussions and the reconstructed CryoEM density maps of HBV.

REFERENCES

[1] Berger B, Shor PW, Tucker-Kellogg L, King J "Local rule-based theory of virus shell assembly." Proc. Natl. Acad. Sci. USA Vol. 91, 7732-7736, (1994).

[2] K. Haris, S.N. Efstratiadis, N. Maglaveras, and A.K. Katsaggelos. Hybrid image segmentation using watersheds and fast region merging. IEEE Trans. on Image Processing, 7(12),1684-1699 (1998.)

[3] M. L. Baker, T. Ju, W. Chiu, Identification of Secondary Structure Elements in Intermediate Resolution Density Maps, Structure, 15(1):7-19, (2007).

[4] G. Khne, J. Weickert, O. Schuster, and S. Richter. A tensor-driven active contour model for moving object segmentation. In Proceedings of IEEE International Conference on Image Processing, 73-76 (2001)

[5] Masouri, A- R., Sirivong, B., Konrad, J. Multiple motion segmentation with level sets. Proceedings of the SPIE, Vol 3974. 584-595 (2000).

[6] Niels Volkmann, A novel three-dimensional variant of the watershed transform for segmentation of electron density maps, Journal of structural Biology vol.138,123-129(2002).

[7] A. Frangakis and R. Hegerl, Segmentation of two and three dimensional data from electron microscopy using eigenvector analysis *Journal of Structural Biology*, vol. 138,No. 1-2, pp. 105–113, 2002

[8] Zeyun Yu, Chandrajit Bajaj, Automatic ultra-stucture segmentation of reconstructed CryoEM Maps of Icosahedral Viruses, IEEE Transactions on Image Processing: Special issue on molecular and cellular bioimaging, 2005.

[9] Mahinda Pathegama, Edge-end pixel extraction for edge-based image segmentation, transactions on Engineering, computing and technology, vol.2, 213-216 (2004).

[10] Sethian, J.A., Level Set Methods: Evolving Interfaces in Geometry Fluid Mechanics, Computer Vision and Materials Sciences. Cambridge University Press, 1999

[11] Sethian, J.A., "A fast marching level set method for monotonically advancing fronts". Proceedings of the Natural Academy of Sciences, 1591–1595, (1996).

[12] Zhu Juping, Tian Jie, jinyao Ge Xingfei, "Modified fast marching and level set method for medical image segmentation". Journal of X-Ray Science and Technology 11 ,193–204(2003)

[13] T. S. Baker, N. H. Olson, and S. D. Fuller, "Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs," Microbiology and Molecular Biology Reviews, vol. 63(1999), No. 4, pp. 862–922

[14] Nabil Ouerhani , Neculai Archip , Heinz Hügli and Pierre-Jean Erard, "Visual attention guided seed selection for color image segmentation", proceedings of the 9th international conference on computer analysis of images and patterns, 630-637 (2001).

Satellite Image Compression Using Wavelet

Alb. Joko Santoso, F. Soesianto, B. Yudi Dwiandiyanto
University of Atma Jaya Yogyakarta
albjoko@mail.uajy.ac.id

ABSTRACT

Image data is a combination of information and redundancies, the information is part of the data be protected because it contains the meaning and designation data. Meanwhile, the redundancies are part of data that can be reduced, compressed, or eliminated. Problems that arise are related to the nature of image data that spends a lot of memory. In this paper will compare 31 wavelet function by looking at its impact on PSNR, compression ratio, and bits per pixel (bpp) and the influence of decomposition level of PSNR and compression ratio.

Based on testing performed, Haar wavelet has the advantage that is obtained PSNR is relatively higher compared with other wavelets. Compression ratio is relatively better than other types of wavelets. Bits per pixel is relatively better than other types of wavelet.

Keyword: Wavelet, Compression, Satellite Image

1. INTRODUCTION

Image is data in picture forming that can represent those data. Image data is combination with information and redundancy, a part of information is a part of data that is existence because having meaning and aiming data. Otherwise a part of redundant is a part of data than can be reduction, compressed, and turned away.

Problems can be happened with image data characters that spend too much memory. Memories that can be spent by those pictures reduce capability saving of data image. In telecommunication image data transmission need transmission channel that have big bandwidth. In Computer and Internet world, Compressing file is used in all necessity, if we want to make data refill, It is not necessary to copy all original file but wit compressing that file so saving capacity will be smaller. If anytime the data will be needed, we must return to original file.

Specific aim this research is getting a wavelet function that appropriate to compress satellite image so this advantage is to save saving space, saving time of central processing unit and if we use the computer, data that have been compressed is not need long time so it can save delivery time.

2. DEFINITION OF WAVELET

Wavelet is a mathematic function that divide data to be several component that have different frequency, It learn every component with appropriate resolution for every scale [1]. Wavelet is a wave forming that has limited duration with zero mean value. Successful application with wavelet is data image compressing, water making, edge detection, radar system, finger print code. There are 3 reasons for using wavelet transformation to refreshing image namely:

1. Wavelet transformations have unconditional basis character. It have meaning that coefficient transformation result will be reduced quickly from approach coefficient to detail coefficient or transformation result have small or zero value, So image character can be represented by a small part of coefficients transformation result or a sign can be represented effectively with using a small part of coefficients transformation result.
2. Wavelet transformation have sign location characteristic, so it can be separated irregulars sign component at specific space and frequency well
3. Wavelet transformations have quick transformation process, so it is caused computation time will be short and It is appropriate to apply in digital computer.

Wavelet transformation is explanation of a sign or image that use wavelet function with different location and size also it is counted with pyramid algorithm. In Image processing, digital Image are formed Discrete Wavelet Transform. Wavelet is a base, wavelet base is come from a scaling function. Scaling function has character namely it can be arranged from several copying than have been dilation, translated, scaled. This function is reduced from

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463N · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855734

dilatation equation, as basic from wavelet theory. From scaling function can be formed first wavelet equation (mother wavelet) namely:

$$\psi_{ab}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)$$

From this mother wavelet can be formed the next wavelet (ψ^1, ψ^2, etc) with dilating, moving mother wavelet. Based on scaling function, mother wavelet, has different name examples:

- Wavelet Haar has scaling function with coefficient $c_0 = c_1 = 1$.
- Wavelet daubechies with 4 coefficients (Db4) have scaling function with coefficient $c_0 = (1+\sqrt{3})/4$, $c_1=(3+\sqrt{3})/4$, $c_2= (3-\sqrt{3})/4$, $c_3 = (1-\sqrt{3})/4$
- Wavelet B–Spline cubic has scaling function with coefficient $c_0 = 1/8$, $c_1 = 4/8$, $c_2 = 6/8$, $c_3 = 4/8$, $c_4 = 1/8$.

3 . DEFINITION OF COMPRESSION

The development of information technology that allows large amounts of data accumulated. For example, nationwide department store sales save each using a POS (Point of Sales). Database sales data can reach several GB per day for a supermarket network is. Therefore there is need for compression, which can reduce the storage space limitations can be speed up access times. The nature of the data compression there are two, that is, without a lossless, and lossy. Data compression without loss is used to reduce the size of the data by providing the exact appearance of the original. For compression without losing accuracy of the results of compression rate is 100% of the original data. Data compression is lost, there is a difference compression results with the original data but still within the specified tolerances. In this study, the nature of the data compression is used without a loss, because the compressed image database to produce 100% accuracy level.

4. IMAGE COMPRESSION USING WAVELET

Destination image compression is to reduce the storage capacity without losing image quality significantly. A characteristic of an image is a close correlation between a pixel with other pixels. Image data compression can be done with wavelets transform. Stollnitz [2] says that one of the characteristics of wavelet is the infrequency. In fact, many coefficients in the representation of wavelet whose value is zero or very small. This nature provides an opportunity to perform image data compression. The main properties of wavelet brief silence in image compression is the minimum distortion in the image of disappearances carried out reconstructs although the coefficients are near zero transform. Whereas wavelet transform the image will result in many subfields images that have very small magnitudes. Threshold determination of non-negative, subfields elements of the image is very small value can be zero which can lead to very rare matrix. The existence of the matrix is very rarely make it easier to be transmitted and stored, even a reconstructed image with threshold (quantization) can provide acceptable results in a visual eye. Currently wavelet applications are received much attention in the world of research, one of which is to analyze the image. Signal analysis techniques as discrete 2-dimensional, such as images, wavelet signals into signals decomposes average, details of vertical, horizontal and diagonal at some desired level. Or, the original signal wavelet becomes decomposes signals in some frequency bands (called multi-resolution analysis). Analysis can be done with the Discrete Wavelet Transform [3] or the standard decomposition techniques and non-standard with the Haar wavelet ([4], and [5]. Signature image generated by wavelet derived from the wavelet coefficients at a certain level (eg 3, 4 or 5) and can be sized much smaller than the original image.

5. SATELLITE IMAGE

Satellite imagery is the data that uses satellite imagery as property. The satellite uses sensors to record the condition or description of the earth's surface. Generally applied in activities related to monitoring of natural resources in the earth's surface (there are even a few satellites that can record up to below the earth's surface), the study of land and environmental changes, and other applications that involve human activities on the earth's surface. The advantages of the technology, especially in this decade are the ability to record a wide range of areas and levels of resolution in the recording a very high objects. The data generated from satellite imagery and then lowered to the thematic data and stored in a database for use in various applications.

6. RESEARCH METHODOLOGY

Proposed research is designed as a review of research literature and experiments. Literature research looking for some wavelet functions wavelets which already exist and image processing theory, and writing source code. While study conducted experiments to test the wavelet function is suitable or appropriate for the compression of satellite images and tests multiple satellite images.

The processes of research are:

1. Literature research with wavelet, image database, image processing, compressing techniques.
2. Experiment of several wavelet function. It will be used in process of image satellite compressing. Function wavelet will be used to experiment are:
    a. Haar
    b. Daubechies (db2, db3, db4, db5)
    c. Coiflets (coif1, coif2, coif3, coif4, coif5)
    d. Symlets (sym2, sym3, sym4, sym5, sym6, sym7, sym8)
    e. Biorthogonal (bior1.3, bior 1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8)

   The result data are:
    a. Ratio compression
    b. MSE
    c. PNSR
    d. Bit per pixel (Bpp)
3. Analyzing
    a. wavelet type vs compression ratio
    b. wavelet type vs PNSR
    c. wavelet type vs bit per pixel (bpp)
4. Conclusion of advantage and disadvantaged every wavelet function.

7. RESULT AND DISCUSSION

This study using satellite images of 24-bit color with 512x512 size obtained from satellite data. The image is tested in this study can be seen in Figure 2.

Figure 2. Satellite imagery is used for testing (a) bigmap.bmp, (b) 22.bmp, (c) awan1.bmp

It is used several wavelet function In this research namely wavelet Haar, db2, db3, db4, db5, coif1, coif2, coif3, coif4, coif5, sym2, sym3, sym4, sym5, sym6, sym7, sym8, bior1.3, bior1.5, bior2.2, bior2.4, bior2.6, bior2.8, bior3.1, bior3.3, bior3.5, bior3.7, bior3.9, bior4.4, bior5.5, bior6.8. This wavelet then tried to use to compress the test image. The results of compression of each wavelet are then compared by using several parameters, ie PSNR (Peak Signal to Noise Ratio), MSE (Mean Square Error), bpp (bits per pixel), and the compression ratio. In this paper will be presented some results of the prototype program developed using Matlab program assistance, including the following:

1. Compression and decompression functions
2. Function calculation PSNR, MSE, bpp, and the compression ratio

Also to be presented also the compression test results using several types of wavelets and its impact on parameter values PSNR (Peak Signal to Noise Ratio), MSE (Mean Square Error), bpp (bits per pixel), and compression ratio. Here is the output of the program.

Figure 3. Image compression and Decompressing Awan 1

Figure 3 is an example of image compression and decompression tests. The image displayed on the image is compressed using wavelet biorthogonal 6.8, with the number of decomposition level 4. Based on statistical calculations, obtained MSE values are small enough to show that the image reconstruction results similar to the image that will be compressed

7.1. The Influence of Wavelet Type of PSNR

PNSR is one of parameter that can be used to make quantification image qualities. Parameter of PNSR is always used as resembling level between reconstruction image and original image. The Bigger PNSR will make better qualities image.

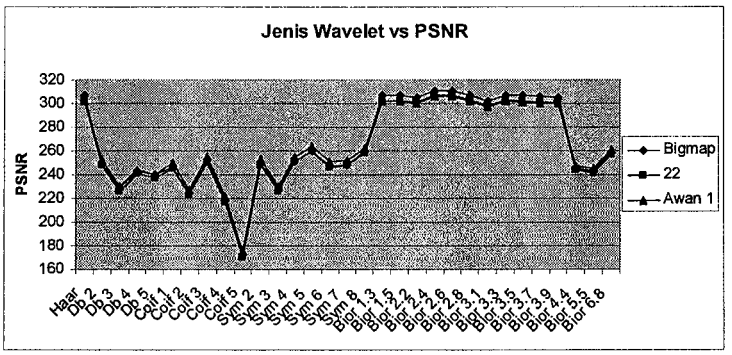


Figure 4. Graphic Relationship Between Wavelet type of PSNR

From the graphs can be seen that the sequence of the 3 types of wavelets have the highest PSNR is
a. For bigmap.bmp satellite imagery are Bior 2.4, Bior 2.6, and Haar.
b. For 22.bmp and awan1.bmp satellite images are Bior 2.6, Bior 2.4, and Haar.
Wavelet Biorthogonal 2.4 more suitable for color image which is almost similar, whereas bior 2.6 is more suitable for color image is more varied. While Haar wavelet is more adaptable to several types of test images.

7.2. The Influence of Wavelet Type of Compression Ratio

Compression ratio is used to measure the ability of data compression, namely by comparing the size of the original image with the size of the compressed image. The greater the compression ratio means the better the wavelet function.

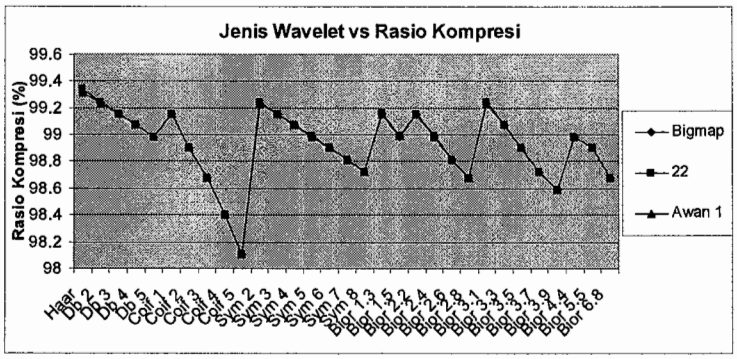


Figure 5. Relationship Between Graph Types of Wavelet Compression Ratio

From the graph can be seen that the sequence of 3 types of wavelet have the highest compression ratio are Haar, Bior 3.1, then Symlet 2 or Db 2.

7.3. The Influence of Wavelet Type of The Bit Rate or Bit per Pixel (bpp)

Parameter bit rate is a mathematical parameter which can quantify the performance of compression. More specifically the bit rate is a parameter used to compare elements of nonzero quantized image with elements not zero in the original image. The lower the bit rate the better is the ability of wavelet as image compressor. Parameters so that the bit rate can be used to test the performance of wavelet compression.

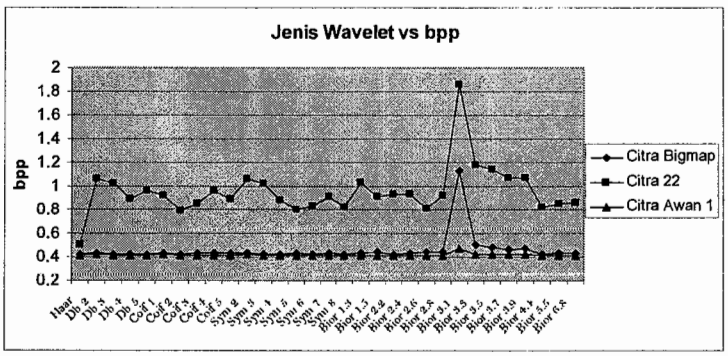


Figure 6. Graphic Relationship Between Wavelet type of bits per pixel

From the graphs can be seen that the type of wavelet which has the lowest bpp is Haar. To image Bigmap.bmp and awan1.bmp bit per pixel (bpp) images tended to same except for the type of wavelet Bior 3.1 there is a significant difference. As for his image 22.bmp tends bit per pixel (bpp) fickle, and significant changes occurred in the types of wavelet Bior 3.1.

7.4. Influence Decomposition Level

Affect the decomposition level PSNR and compression ratio, taking into account both aspects can be determined at what level is chosen as the test image. Determining the level of decomposition is very important because it will affect the computing process, the greater the level the greater the computing process. But keep in mind these two parameters are PSNR and compression ratio can be determined so that the optimal level of decomposition. In this paper the decomposition level is used 4 the optimal for color satellite image measuring 512 x 512.

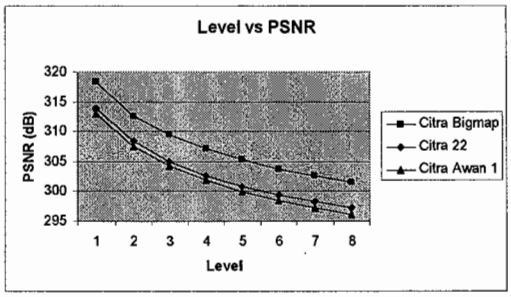


Figure 7. Graphic Relationship Between Decomposition Level of PSNR

From the graphs can be seen that the greater the level of decomposition the smaller PSNR value, this means that the value of error between the original image and the image reconstruction greater.

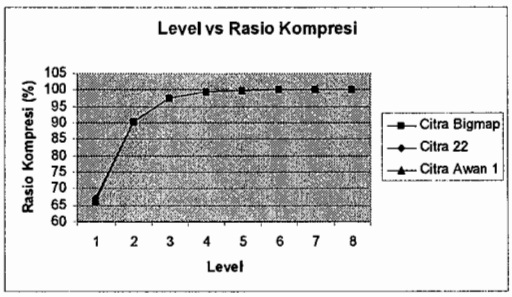


Figure 8. Graphic Relationship Between Decomposition Level of Compression Ratio

From the graphs can be seen that the greater the level of decomposition, the greater the value of the compression ratio.

8. CONCLUSION

Based on testing performed, Haar wavelet has the advantages in terms of:
a. PSNR obtained relatively higher compared with other wavelet
b. Compression ratio is relatively better compared with other types of wavelet
c. Bits per pixel (bpp) is relatively better compared with other types of wavelet
Based on the testing also found that biorthogonal wavelets have similar characteristics with the Haar wavelet, which has a better PSNR, bpp is good, and the compression ratio is better.

REFERENCE

[1] Mubarak, Riyad, 2003, PEMAMPATAN DATA CITRA DENGAN MENGGUNAKAN TRANSFORM GELOMBANG-SINGKAT, UGM, Yogyakarta.
[2] Stollnitz, E.J, DeRose, T.D., dan Salesin, D.H., 1996, WAVELETS FOR COMPUTER GRAPHICS: THEORY AND APPLICATIONS, Morgan Kaufman Publisher, USA, San Fransisco.
[3] Mallat, S, 1999, A WAVELET TOUR OF SIGNAL PROCESSING, Academic Press, USA
[4] Chakrabarti,K., Garofalakis, M., Rastogi, R., Shim, K., 2000, APPROXIMATE QUERY PROCESSING USING WAVELET, Proceedings of the 26th VLDB Coference, Cairo, Egypt.
[5] Natsev, A., Rastogi, R., Shim, K., 1999, WALRUS: A SIMILARITY RETRIEVAL ALGORITHM FOR IMAGE DATABASES, Duke University and Bell Laboratories, USA

KNOWLEDGE BASE IMAGE CLASSIFICATION USING P-TREES

Dr. M.Seetha, *G. Ravi
Professor, Dept. of CSE, GNITS, Hyderabad-8, A.P., E-mail: smaddala2000@yahoo.com
*Asst. Prof, Dept. of IT, CBIT, Hyderabad, A.P, E-mail: gottipatiravim.tech@gmail.com

ABSTRACT

Image Classification is the process of assigning classes to the pixels in remote sensed images and important for GIS applications, since the classified image is much easier to incorporate than the original unclassified image. To resolve misclassification in traditional parametric classifier like Maximum Likelihood Classifier, the neural network classifier is implemented using back propagation algorithm. The extra spectral and spatial knowledge acquired from the ancillary information is required to improve the accuracy and remove the spectral confusion. To build knowledge base automatically, this paper explores a non-parametric decision tree classifier to extract knowledge from the spatial data in the form of classification rules. A new method is proposed using a data structure called Peano Count Tree (P-tree) for decision tree classification. The Peano Count Tree is a spatial data organization that provides a lossless compressed representation of a spatial data set and facilitates efficient classification than other data mining techniques. The accuracy is assessed using the parameters overall accuracy, User's accuracy and Producer's accuracy for image classification methods of Maximum Likelihood Classification, neural network classification using back propagation, Knowledge Base Classification, Post classification and P-tree Classifier. The results reveal that the knowledge extracted from decision tree classifier and P-tree data structure from proposed approach remove the problem of spectral confusion to a greater extent. It is ascertained that the P-tree classifier surpasses the other classification techniques.

Keyword list: Maximum Likelihood Classification, bit Sequential (bSQ), Knowledge Base Classification, Peano Count Tree

1. INTRODUCTION

Remotely sensed image analysis is a challenging task and is accomplished by digital image classification. Traditionally used pixel based classification methods are based on conventional statistical techniques and do not resolve inter-class confusion. As a result, in recent years, alternative strategies have been proposed, particularly the use of artificial neural networks, decision trees, methods derived from fuzzy set theory, and incorporation of secondary information such as texture, context and terrain features [14]. The untrained classes will only display membership to trained classes, which can introduce a significant bias to classification accuracy [4]. The high resolution image classification is presented based on fuzzy rules by the means of descriptors such as: form, texture and relations between objects and sub-objects [1]. It was shown that the KFLANN was capable of fast and accurate assimilation of data [13].

Image classification based on unsupervised image segmentation algorithm and its combination with MPEG-7 low-level descriptors and a Bayes classifier [8]. Image Processing and Analysis is very important for GIS work, since the classified image is much easier to incorporate than the original unclassified image [2]. Digital image classification [12] is the process of sorting all the pixels in an image into a finite number of individual. The approach on different urban zones within the city of Istanbul compared to a digitized layer based on visual classification [6].

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463O · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856324

Artificial neural networks can handle non-convex decisions. To remove such spectral confusion extra spectral and spatial knowledge is required. Classification done using such knowledge is known as knowledge base classification. But such classification needs strong knowledge base, which sometimes become drawback of this process because of the knowledge acquisition process [5].

Decision tree was generated using See5 decision tree software [9]. The main advantage of See5 is that it can convert a decision tree into classification rules. Thematic Mapper (TM) satellite images are in BSQ format [15]. The Peano Count Tree (P-tree) structure is used to build the classifier. P-trees represent spatial data bit-by-bit in a recursive quadrant-by-quadrant arrangement. Each new component in a spatial data stream is converted to P-trees and then added to the training set as soon as possible [10]. P-trees can be generated quite quickly and can be viewed as a "data mining ready" and lossless format for storing spatial or any relational data. The training dataset for the decision tree is different as it is non-parametric classifier [7]. In P-tree, the first step changes the data format to bSQ format. Subsequently, the P-tree was constructed by applying the normal decision tree algorithm [16].

There is a dearth on the classification methods using knowledge based classification using P-tree and accuracy assessment using IRS 1D LISS III image using all the classification techniques. The implementation of P-tree data structure for classification is mainly to improve the accuracy of classification. Hence, this paper emphasizes on the accuracy assessment of image classification using p-tree data structure and a comparative analysis of MLC (Maximum Likelihood Classification), Neural network classification using back propagation algorithm, KBC Knowledge Base Classification and PTC (P-tree Classifier).

2. BACK PROPAGATION ALGORITHM

The BP algorithm is used in layered feed-forward Artificial Neural Network (ANN). The network receives inputs by neurons in the input layer, and the output of the network is given by the neurons on an output layer. There may be one or more intermediate hidden layers. BP algorithm uses supervised learning, which means that the algorithm is provided with the inputs and outputs that are to be computed by the network and then the error (difference between actual and expected results) is calculated. The idea of the BP algorithm is to reduce this error, until the ANN learns the training data. The training begins with random weights, and the goal is to adjust them so that the error will be minimal.

3. KNOWLEDGE BASE CLASSIFICATION USING DECISION TREE

Decision tree is a non-parametric classifier and therefore it doesn't take into account the statistical distribution of the training data. It is a set of records written in a text editor. Each training pixel in itself is a record and the various spectral and ancillary layers serve as different attributes of the record. The following subsection explains about decision tree classifier and the steps for preparation of the training dataset for the decision tree classifier [3].

3.1. Data Preparation for Decision Tree

Total five layers were taken from satellite image. Three layers RED, GREEN, and BLUE from IRS 1D LISS-3 image, which were at 23.5 m resolution and two layers of DEM and slope were

considered. To investigate the affect of the size of training set on the accuracy of the classifier, different training samples of different sizes were prepared (for example 50, 100, 150 samples etc). A training set of 225 pixels with 25 pixels from all the five classes were taken first and then subsequently equal number of pixels from all the five classes were increased.

3.2 Decision Trees to Rules

To simplify a decision tree convert them into rules, which are easier to understand and to implement. Every path from the root to a leaf is converted to an initial rule by regarding all the test conditions appearing in the path as the conjunctive rule antecedents while regarding the class label held by the leaf as the rule consequence [11]. Finally, a default rule is created for dealing with instances that are not covered by any of the generated rules.

3.3 Knowledge Base Classification

Two approaches were followed to use the extracted classification rules for the classification.

Classification using Extracted Knowledge:
In first approach, classification rules were used directly using knowledge base classifier to classify the image.

Post-Classification Sorting:
In second approach, post-classification sorting method was used to reclassify the pixels, which were misclassified during maximum likelihood classification.

4. CLASSIFICATION BY DECISION TREE INDUCTION USING P-TREES

A Decision Tree is a flowchart-like structure in which each node denotes a test on an attribute. Each branch represents an outcome of the test and the leaf nodes represent classes or class distributions. The basic algorithm for inducing a decision tree from the learning or training sample set is as follows:

- Initially the decision tree is a single node representing the entire training set.
- If all samples are in the same class, this node becomes a leaf and is labeled with that class label.
- A branch is created for each value of the test attribute and samples are partitioned accordingly.
- The algorithm advances recursively to form the decision tree for the sub-sample set at each partition. Once an attribute has been used, it is not considered in descendent nodes.
- The algorithm stops when all samples for a given node belong to the same class or when there are no remaining attributes (or some other stopping condition).

5. RESULTS AND DISCUSSIONS

The LISS III images are classified for five classes namely Water, Agricultural field (AF), Greenary field (GF), Urban and Open area for various classification methods (Tables 1-5). The user's accuracy, producer's accuracy and overall accuracy for all the classified images were compared (Figures 4 to 9). It reveals that P-Tree classifier gave better accurate results than other methods.

Table 1: User's accuracy of LISS-3 Classified Image1

	Water	AF	GF	Urban	Open area
MLC	32.14	100	66.67	100	84.62
NNC	79.33	70.15	85.55	73.23	70.33
KBC	78.57	72.72	90	71.42	71.42
PCS	72.72	88.88	81.81	75	72.72
PTC	83.33	90	90.9	75	77.77

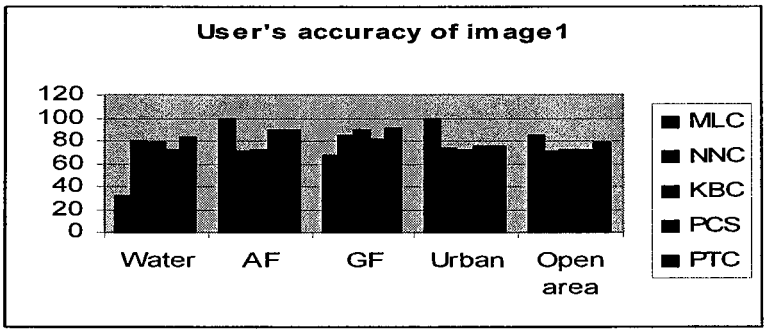


Figure 4: User's Accuracy for Different classifiers of image1

Table 2: Producer's accuracy of LISS-3 Classified Image1

	Water	AF	GF	Urban	Open area
MLC	70	66.67	25	40	50
NNC	89.86	79.71	73.33	63.13	85.13
KBC	91.66	80	75	50	83.33
PCS	80	80	75	75	90
PTC	83.33	90	83.33	75	87.5

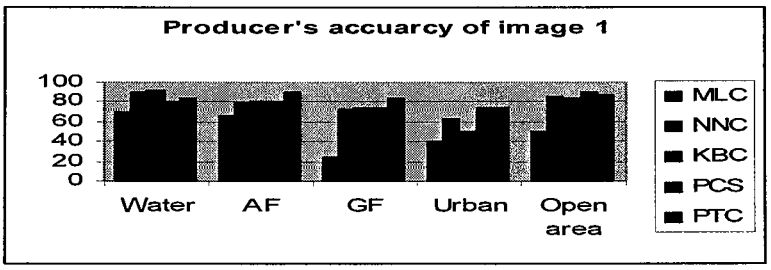


Figure 5: Producer's Accuracy for Different classifiers of image1

Table 3: User's accuracy of LISS-3 Classified Image2

	Water	AF	GF	Urban	Open area
MLC	58.82	57.14	77.28	42.87	50
NNC	74.6	61.33	76.4	79.12	74.54
KBC	73.33	62.5	70	80	75
PCS	70	80.88	80	60	75
PTC	100	81.81	83.33	87.5	81.81

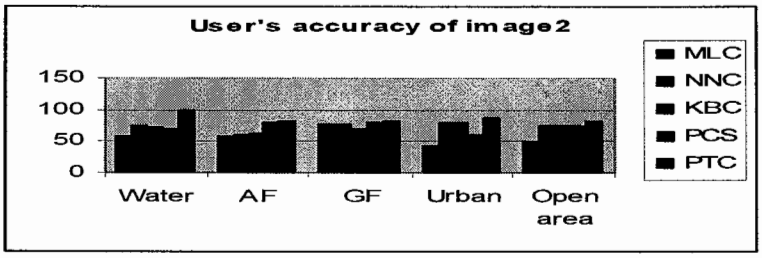


Figure 6: User's Accuracy for Different classifiers of image2

Table 4: Producer's accuracy of LISS-3 Classified Image2

	Water	AF	GF	Urban	Open area
MLC	80	44.44	87.5	40	50
NNC	82.13	74.12	72.23	70.15	89.12
KBC	84.61	55.56	70	66.67	88.88
PCS	63.33	80	75	85.71	90
PTC	85.71	81.81	83.33	100	83.33

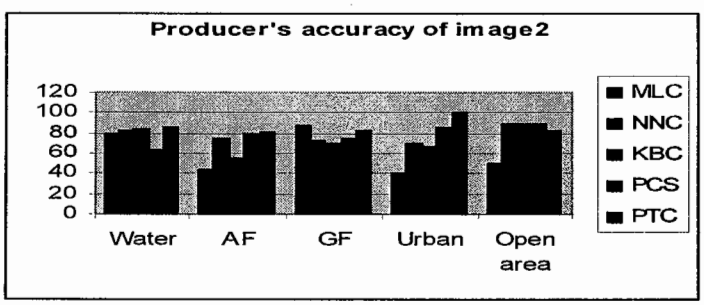


Figure 7: Producer's Accuracy for Different classifiers of image2

Table 5: Overall Accuracy of LISS-3 Classified images

	MLC	NNC	KBC	PCS	PTC
Image1	56	78	76	78	84
Image2	54	77	74	78	86
Image3	45	68	68	72	79
Image4	55	76	78	80	89

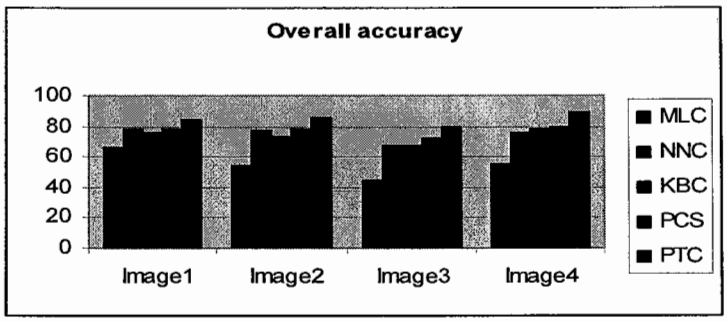


Figure 8: Overall Accuracy for Different Classifiers of different images

The overall accuracy was compared for the four classified images using different methods. Maximum likelihood classification (MLC) has the lowest Overall accuracy and P-Tree classifier has the highest Overall Accuracy. The comparative analysis reveals that P-Tree classifiers provide better solutions. MLC has given the accuracy of 56% for image1 and 54% for image2. The MLC classified image and original images when given as input to the KBC, produced higher accuracy than conventional MLC. Original images using KBC gave accuracy of 76% for image1 and 74% for image2. MLC using KBC has given 78% accuracy for both images. The accuracy of P-Tree was 84% for image1 and 86% for image2. It was observed that the P-Tree classifier provided superior results to other methods of classification.

6. CONCLUSIONS

Image classification was an important area of data mining. The accuracy assessment of the image classification using Maximum Likelihood Classification, Knowledge Base Classification and P-Tree classifier was emphasized in the study. This is due to misclassification of different pixels. The image classification using the extracted knowledge in the form of classification rules using a decision tree approach has obtained the accuracy of 76% and 74%. The post classification sorting has obtained the accuracy of 78% for both images. This is due to the improvement by addition of ancillary information into classification process. It was observed that the accuracy of image classification using P-Tree Classifier is 82% and 84%. This is due to the improvement in the loss less compressed representation of the spatial data set and fast classification of measurements. The results demonstrated that P-Tree classifier surpasses other classification methods. In future, the training data set can be increased by using hyper spectral images or by adding more ancillary layers such as aspect, soil maps, geology maps etc. The accuracy of the classification can be improved by segmenting the image and extracting improved edges.

References

1. Alzir Felippe, B. Antunes, Christel Lingnau, Jorge Centeno Da Silva, "Object-Oriented Analysis and semantic network For High Resolution Image Classification", 1UFPR- Universidade Federal do Paraná, Brazil, 2003.
2. E Hall, Image Processing and Analysis, 2000.
3. Evans, F. "An investigation into the use of Maximum Likelihood Classifiers, Decision Trees, Neural Networks and Conditional Probabilistic Networks for Mapping and Predicting Salinity", Msc thesis 1998. WWW site: http://www.cmis.csiro.au/Fiona.Evans/papers/pdf/evansmsc.pdf

4. Foody, G., "Estimation of sub-pixel land cover composition in the presence of untrained classes", Computers and Geoscienes 26: PP 469-478, 2000.
5. Han, Kamber, M. Data Mining: Concepts and Techniques, Morgan Kaufmann Publishers 2001.
6. Hannes Taubenbock, Thomas Esch, Achim Roth, Humboldt-Universitat zu BerlinEARSeL, Workshop of the SIG Urban Remote Sensing, An urban classification approach based on an object-oriented analysis of high resolution satellite imagery for a spatial structuring within urban areas, German Remote Sensing Data Center (DFD), German Aerospace Center (DLR), Wessling, 2-3 March 2006, Germany.
7. Pal, M. & P.M.Mather, Decision Tree Based Classification of Remotely Sensed data, 2001, WWW site: http://www.crisp.nus.edu.sg/~acrs2001/pdf/046PAL.PDF
8. Perera, Serazi, Perrizo Bayesian Classification on Spatial Data Streams Using P-Trees, 2003.
9. Quinlan, J. R., See5 Manual, 1997, WWW site: http://www.rulequest.com/see5-info.html.
10. Qiang Ding, William Perrizo, Victor Shi, Integrating Query Processing and Data Mining in Relational DBMSs using P-Tree,-2003
11. Qin Ding, Maleq Khan, Amalendu Roy and William Perrizo, The P-tree Algebra, 2003.
12. Rafael c. Gonzalez,Richard e.Woods, "Digital Image Processing using MATLAB" third edition, 2005.
13. Symeon Papadopoulos, Vasileios Mezaris, Ioannis Kompatsiaris, Michael G. Strintzis, "A Region-based approach to conceptual image classification", University of Thessaloniki, Thessaloniki 54006, Greece, 2005.
14. Tso, B. and Mather, P.M, "Classification Methods for Remotely Sensed Data", Taylor and Francis Inc, 2001.
15. William Perrizo, Qin Ding, Qiang Ding and Amalendu Roy, "Deriving High Confidence Rules from Spatial Data using Peano Count Trees", Springer-Verlag, LNCS 2118, July 2001.
16. William Perrizo, "Peano Count Tree Technology", Technical Report NDSU-CSOR-TR-01-1, 2001.

Spatial data clustering using an improved Evolutionary Algorithm

Yiping Tang[1], Wenxing Long[2], Chuan Hu[2]

1. School of Environment and Resource, Southwest University of Science and Technology, Mianyang 621002, PR China
2. School of Transportation and Municipal Engineering, Sichuan College of Architectural Technology, Deyang 618000, PR China

ABSTRACT

Considering the difficulties for traditional methods in clustering analysis of spatial data, in this paper, a novel spatial data clustering method based on an improved evolutionary algorithm is proposed. It effectively solved the two main problems puzzling many researchers, i.e., 1) difficulty in coping with the local optimum, and 2) sensibility to the center selections of the initial clustering. Empirical evaluation of our method indicates that it has better performance, compared with the other methods in literature.

Keywords: spatial data; evolutionary algorithm; clustering

1. INTRODUCTION

In the last few years, information technology and data acquisition technology gained a great development, making the data stored in database increase at an unprecedented speed. Therefore, it increasingly become a problem of difficult to obtain valuable information or knowledge from the huge data sea.

Data mining [1] is the process of extracting patterns from data. As more data are gathered, with the amount of data doubling every three years, data mining is becoming an increasingly important tool to transform these data into information. It is commonly used in a wide range of profiling practices, such as marketing, surveillance, fraud detection and scientific discovery.

Spatial data mining [1] in particular is the discovery of interesting relationships and characteristics that may exist implicitly in spatial databases. Because of the huge amount of spatial data that my be obtained from satellite images, medical equipments, video cameras, etc., it is costly and often unrealistic for users to examine spatial data in detail. Spatial data mining aims to automate such a knowledge discovery process. Thus, it plays an important role in 1) extracting interesting spatial patterns and features; 2) capturing intrinsic relationships between spatial and non-spatial data; 3) presenting data regularity concisely and at higher conceptual levels; and 4) helping to recognize spatial databases to accommodate data semantics, as well as to achieve better performance. There exist many methods to do the spatial data mining work, such as.

In this paper, we only focus on spatial data clustering. Spatial data clustering [2] is the process of grouping a set of objects

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463P · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855584

into classes or clusters so that objects within a cluster have high similarity in comparison to one another, but are dissimilar to objects in other clusters. As a branch of statistics, cluster analysis has been studied extensively for many years, focusing mainly on distance-based cluster analysis. However, the existing methods always have two main defects, i.e., they have difficulty in coping with the local optimum, and they are very sensitive to the center selections of the initial clustering [3].

To solve the defects methods in clustering analysis of spatial data, we firstly proposed an improved evolutionary algorithm. It improved the traditional evolutionary algorithm in the following two aspects, individual initialization and crossover method. Then we used this improved evolutionary algorithm to solve the spatial data clustering problem. The results of empirical evaluation indicates that our method has a better performance, compared with other methods in literature.

The paper is organized as follows. Section 2 gives a brief introduction to traditional evolutionary algorithms, and their defects. Section 3 discusses the approach proposed in this paper in detail. Section 4 is the simulation experiment. In Section 5 we make a brief conclusion.

2. TRADITIONAL EVOLUTIONARY ALGORITHMS

In this section, we will give a brief introduction to the tradition evolutionary algorithm and their defects in solving problems.

2.1 A brief introduction to traditional EAs

Evolutionary Algorithm (EA) [4] [5] is a heuristic that mimics the evolution of natural species in searching for the optimal solution to a problem. It is one of the population based, parallel problem solver, which is based on Darwin's theory of the survival of the fittest and Mendel's theory of heredity. When solving a problem, a group of individuals (called a population) will be generated first, and use strategies like crossover, mutation and selection to lead the population to a better direction, and finally to get global optimum. Traditionally, it mainly consists of Genetic Algorithm (GA), Evolutionary Strategy (ES), Evolutionary Programming (EP), and Genetic Programming (GP) [5]. Nowadays, EAs have been widely used to solve a large number of complex problems, for their merits such as parallelism, self-organizing, adaptation and self-learning intelligence.

2.2 Defects of traditional EAs

As is talked in Ref. [6], traditional EAs always have three main problems:

Premature convergence: They always lead to premature convergence, especially when the SGA is used to solve complex optimization problems.

Their computational quantity is large and their precision is not high: If we do not use some strategies, it will always take a long time to obtain optimized solutions, and the precision of the solutions is not high.

They lack a proper stopping criterion, and it is difficult to select operation parameters.

3. THE APPROACH

Just as we talked above, if we want to use EA to solve spatial data clustering problem, we should improve them. In the

field of EAs, researchers always improve the performance of EAs by improving the evolutionary operators, which are seen as one of the key factors that influence the performance of the algorithms [7]. And taking into consideration the spatial clustering problems, in this section, a novel method for Spatial Data Clustering based on an Improved EA is proposed. Hereafter, we call this method SDC-IEA. Since SDC-IEA is based on the traditional EAs, only modifying some key operators. So the main framework of SDC-IEA is very similar to that in traditional EAs. The following sections will elaborated some the new key operators of SDC-IEA one by one.

3.1 Individual encoding and population initialization

EAs start from a group of individuals, which always is called population. So to use EAs to solve the spatial data clustering problems, the first problem we should solve is how to use the individual to represent the solution of spatial data clustering problem, i.e., how to encode the individuals. In SDC-IEA, we use the a vector with the form as {1,2,3,4} to denote the solution of each spatial clustering result. The number of dimensions denote the number of data samples should be processed by SDC-IEA, 4 in the above vector. And the number in dimension denotes which clustering does the *i-th* (from left to right) data sample belongs to. In SDC-IEA, we use functions *srand()* and *rand()* together to generate *popsize* individuals, which together form the initial population.

3.2 Crossover operator

In literature, researchers always use the typical crossover operation, such as one-point crossover and two-point crossover [5]. Since different cluster identifiers in two individuals may mean the same meaning. For example, suppose there are two individual (2,5,5,5,2) and (6,7,7,7,6). We may note that cluster identifier 2 and 6, 5 and 7 are all share the same meaning respectively. So here we use a special crossover operator. First we sort the individuals according to fitness value from large to small, and store the top 0.1popsize ones. It can ensure the good individuals will not be destroyed by crossover operation. Then, we choose two individuals from the left *0.9poposize* individuals, which is denoted as *individual_1* and *individual_2*. The crossover operation used in this paper then can be described as follows: first randomly pick one dimension dim_i in *individual_1*, then travel through the *individual_1*, detect other dimensions with the same value as dim_i and let the value of the same dimensions in *individual_2* be the same as dim_i. To make it clear, we use the scenario shown in figure 1 as a running example. Where the highlight dimensions are the points that the crossover operation occurred. And whether to accept or reject the newly generated individual is determined by whether its fitness is larger or smaller than the original one.

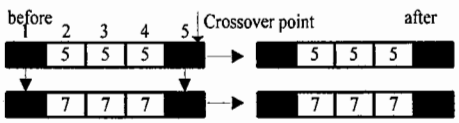


Figure 1. crossover operation

3.3 Mutation operator

In this paper, the one-point mutation operation [5] will be introduced. i.e., choose one dimension of an individual, and let the value of it be other existing cluster identifier. The process can be described as follows, where the highlight dimension is the point the mutation operation occurred. And whether to accept or reject the newly generated individual is determined by whether its fitness is larger or smaller than the original one. See figure 2.

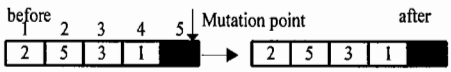


Figure 2. mutation operation

3.4 Fitness calculation

SDC-IEA is also a distance-based cluster analysis method as many previous work do. Fitness calculation is a key step in EAs, for the fitness of individuals always determine whether this individual will be kept in the next generation. So how to design the fitness function is of vital importance for the efficiency of SDC-IEA. In SDC-IEA, we use the following formula to calculate the fitness of each individual.

$$Fitness = \frac{1}{\sum_{i=1}^{k} \sum_{d_j \in C_i} | d_j - \frac{1}{|C_i|} \sum_{d_j \in C_i} d_j |^2} \tag{1}$$

Where d_j is a vector denoting the data samples with many dimensions; C_i is denotes the *i-th* cluster; and the subtraction and summation operations all operate on the corresponding dimension of the vector d_j. i.e., suppose d_1=(1,2,3,4), d_2=(5,6,7,8), then d_1+d_2=(1+5,2+6,3+7,4+8)=(6,8,10,12), d_1-d_2=(1-5,2-6,3-7,4-8)=(-4,-4,-4,-4).

3.5 The procedure of SDC-IEA

Step 1: Initialize the population with *popsize* individuals $Population_{popsize}=\{x_1,x_2,...,x_{popsize}\}$ at random, and the maximum generation *MaxGen*; set the current generation *gen*:=0.

Step 2: Calculate the fitness of each individual in the population according to formula (1), then store the top *0.1popsize* individuals. After that, randomly select two individual from the left *0.9popsize* individuals, and do crossover operation according to that proposed above.

Step 3: Do the mutation operation proposed above and *gen*=*gen*+1.

Step 4: If *gen*<=*MaxGen*, goto step 2; else output the solution.

4. NUMERICAL EXPERIMENTS

There is lack of an open dataset that can be used to validate the efficiency of the spatial data clustering methods. We use the data set widely used in literature to validate our method, which are chosen from Refs. [8] [9] and shown in Table 1 and 2.

Table 1 Data set

data	data	data	data
d_1(5,3)	d_2(1,0)	d_3(3,5)	d_4(0,1)
d_5(2,7)	d_6(7,2)	d_7(3,3)	d_8(3,4)
d_9(4,3)	d_{10}(6,4)	D_{11}(4,6)	$d1_2$(1,6)

Table 2 Data set

data	data	data	data
d_1(1,0)	d_2(0,1)	d_3(1,1)	d_4(2,1)

$d_5(1,2)$	$d_6(2,2)$	$d_7(3,1)$	$d_8(1,3)$
$d_9(4,3)$	$d_{10}(5,4)$	$d_{11}(4,6)$	$d_{12}(1,5)$

Table 3 shows the parameter setting of SDC-IEA when solving the three complex function optimization problems.

Table 3 Parameter Setting

Parameter	value	Parameter	value
popsize	100	*MaxGen*	10,000

Table 4 shows the comparison data with two other methods. Where sensitivity means whether the method is sensitive to the center selection of the initial clustering. It is clear that the results of the three methods all obtain the same results when doing the two simple examples. But our method has a better performance efficiency. The time it consumed is least.

Table 4 Comparisons Data

	Method	Result	Sensitivity	Time/s
Data set 1	Ref. [8]	C1: 2/4 C2: 3/5/11/12 C3: 1/6/10 C4: 7/8/9	No	3.500
	Ref. [9]		No	0.922
	SDC-IEA		No	0.854
Data set 2	Ref. [8]	C1: 8/12 C2: 4/5/7/12 C3: 1/2/3/5 C4: 9/10/11	No	4.297
	Ref. [9]		No	1.469
	SDC-IEA		No	0.978

5. CONCLUSIONS

In this paper, we have presented a novel method for solving the spatial data clustering problem. It first proposed an improved evolutionary algorithm by modifying the traditional evolutionary algorithms. It develops a new operator concerning crossover operation. Empirical evaluation of our method indicates that it has better performance, compared with the other methods in literature

REFERENCES

[1] Ng RT and Han JW, "Efficient and effective clustering methods for spatial data mining," Proc. of the 20th VLDB Conf., 144-155 (1994).

[2] Han J, Kamber M and Tung AKH, "Spatial clustering methods in data mining: a survey," In: H. Miller and J. Han, Editors, Geographic Data Mining and Knowledge Discovery, Taylor & Francis, London (2001).

[3] DASU T. Exploratory data mining and data cleaning. Wiley, New York (2003).

[4] Yao X and Yong X, "Recent Advance in Evolutionary Computation," Computer Science & Technology, 21, 1-18 (2006).

[5] Pan Z, Kang L and Chen Y, Evolutionary Computation. Tsinghua University Press, Beijin (1998). (In Chinese)

[6] Li K, Pan W, Chen Z, et al, "A Bionics Evolutionary Algorithm Based on Average Vector Deviation," Proc. of the 4th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD 07), 386-390 (2007).

[7] Baeck T, Hoffme F, Schwefel H, "A Survey of Evolution Strategies,". Proc. of the 4th International Conference on Genetic Algorithms and Their Application. CA, Morgan Kanfmann (1991).

[8] Lan X, Xu H, Pan W, et al. Spatial Data Clustering Method based on Improved Evolutionary Algorithm. Computer Engineering, 34(22) (2008). (In Chinese)

[9] Huang X, Ji B, Pan W, et al. Novel Particle Dynamical Evolutionary Algorithm for Spatial Data Clustering, Application Research of Computer, 26(3) 860-863 (2009). (In Chinese)

Voice Conversion Using Dynamic Features for High Quality Transformation

Wei Wang[1], Zhen Yang
Institute of Signal Processing and Transmission
Nanjing University of Posts and Telecommunications, Nanjing, China

ABSTRACT

A novel voice morphing method is proposed to make the speech of the source speaker sound like the voice uttered by a target speaker. This method is based on the Gaussian Mixture Model (GMM). However, the traditional GMM has the over-smoothed phenomenon and may get discontinuity of the converted speech due to the inaccuracy of the extracted feature information. In order to overcome it, we consider the dynamic spectral features between frames. The conversion function is also modified to deal with the discontinuities. The Speech Transformation and Representation using Adaptive Interpolation of weiGHTed spectrogram (STRAIGHT) algorithm is adopted for the analysis and synthesis process. Objective and perceptual experiments show that the quality of the speech converted by our proposed method is significantly improved compared with the traditional GMM method.

Voice conversion; dynamic features; modified GMM

1. INTRODUCTION

Voice conversion, also named as voice transformation and voice morphing, is a technique for modifying a source speaker's speech to sound as if it was spoken by some designated target speaker. It is a potential method that can flexibly synthesize all kinds of speech with a small training corpus.

Generally, speech production is often represented by a source-filter model. Both parts of this model contribute to forming the speaker individuality. The vocal tract, such as formant positions and bandwidths, are related to the filter, while pitch, duration allocation and speech rate are features mainly related to the source. Most current voice conversion systems focus on spectral conversion and pitch modification. A considerable amount of effort has been dedicated to the problem in the last two decades. Abe applied Vector Quantization (VQ) mapping method [1] to the parameters of the source and target speakers and got an acceptable performance. Valbret presented the Linear Multivariate Regression (LMR) [2] and Dynamic Frequency Warping (DFW) [3]. Narendranath proposed BP artificial neural networks [4] to convert formants. However, these methods typically lead to discontinuities in the transformed speech and the conversion approach suffers from a lack of robustness as well as degraded quality. Hence, the statistical methods have been employed in recent years. Stylianou provided the GMM based voice conversion method [5] for modern systems, in each Gauss component, a linear function is built between source and target speakers using Minimum Mean Square Error (MMSE) method. But the traditional GMM conversion tends to generate overly smoothed utterances [6] and a number of

[1] Email: Y070803@njupt.edu.cn and web site: http://202.119.236.102.102/

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463Q · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.855168

artifacts in consecutive converted frames [7]. In this paper, we employ an efficient modified GMM method considering dynamic features for spectrum conversion under STRAIGHT analysis-synthesis framework [8] to overcome these problems.

The remainder of this paper is organized as follows: section 2 provides an overview of our voice conversion system, section 3 describes the details of the modified GMM conversion method. In section 4, the objective and perceptual experimental results are conducted to evaluate the performance of the proposed method while conclusions are presented in section 5.

2. THE FRAMEWORK OF THE VOICE CONVERSION SYSTEM

The conversion process consists of two phases: training and transforming. In the training phase, we use STRAIGHT model to extract feature vectors, i.e. the smoothed spectrums and pitch tracks, from both of the source and target speakers' voice. We use Linear Spectral Frequency (LSF) to represent feature parameters due to the dimension of spectrum is too large to convert conveniently [9]. The Dynamic Time Warping (DTW) is applied to align LSF and corresponding dynamic features which are used to search for morphing rules by joint GMM. In the transforming phase, we adopt a modified GMM conversion rules into the LSFs, details followed by section 3, while the average pitch value is computed to keep the quality of the synthesized speech in this system. The flow chart of our conversion system is shown in Figure 1.

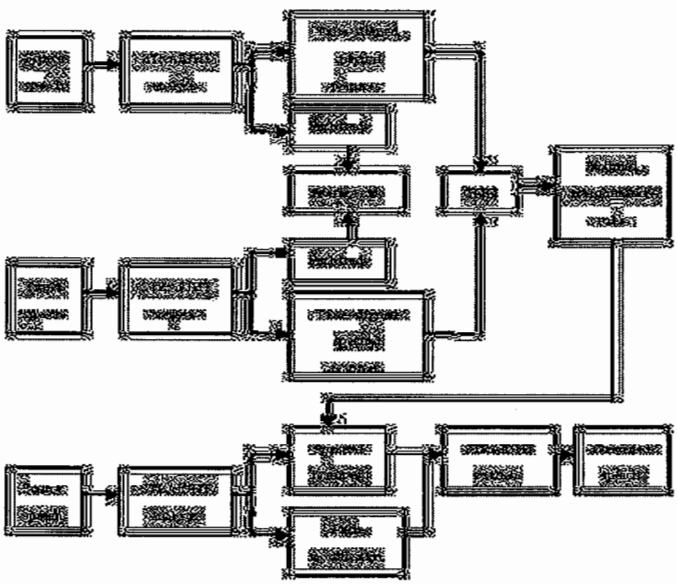


Figure 1. Flow chart of the proposed system

3. IMPROVED GMM BASED VOICE CONVERSION USING DYNAMIC FEATURES

3.1 Conventional GMM-based Voice Conversion

We assume the source speech is represented by an n-frame time series $X = x_1, x_2, \cdots, x_n$, where x_i is a d

dimensional feature vector for the ith frame, i.e. $x_i = [x_1, x_2, \cdots, x_d]^T$, and the target speech be represented by an m-frame time series $Y = y_1, y_2, \cdots, y_m$, where $y_j = [y_1, y_2, \cdots, y_d]^T$. The joint probability density of source and target feature vectors is modeled by GMM as follows[10]:

$$P(z_t|\lambda) = \sum_{i=1}^{M} \omega_i N(z_t; \mu_i, \Sigma_i) \quad (1)$$

Where z_t is a joint vector of time aligned feature pairs $[x_t^T, y_t^T]^T$. ω_i is the prior probabilities of z_t, given the component i, it satisfies $\sum_{i=1}^{M} \omega_i = 1$. $N(z_t; \mu_i, \Sigma_i)$ denotes the $2d$ dimension normal distribution with the mean vector μ_i and the covariance Σ_i. Both of them are given by equation (2)

$$\mu_i = \begin{bmatrix} \mu_i^X \\ \mu_i^Y \end{bmatrix} \Sigma_i = \begin{bmatrix} \Sigma_i^{XX} \Sigma_i^{XY} \\ \Sigma_i^{YX} \Sigma_i^{YY} \end{bmatrix} \quad (2)$$

The parameters $(\omega_i, \mu_i, \Sigma_i)$ for equation (1) can be estimated using the well known Expectation Maximization (EM) algorithm. The transformation function is chosen to be locally linear function between source and target speakers in each GMM component.

$$F(x) = \sum_{i=1}^{M} P_i(x)\left(\mu_i^Y + \Sigma_i^{YX}(\Sigma_i^{XX})^{-1}(x - \mu_i^X)\right) \quad (3)$$

$$p_i(x) = \frac{\omega_i N(x; \mu_i^X, \Sigma_i^{XX})}{\sum_{j=1}^{M} \omega_j N(x; \mu_j^X, \Sigma_j^{XX})} \quad (4)$$

$p_i(x)$ is the probability of x belonging to the ith component.

3.2 Improved GMM Method Using Dynamic Features

For the traditional GMM conversion method, the main drawback is that the converted features obtained from converted spectrum are overly smoothed, so called statistical smoothed phenomena which makes the reconstructed speech unclear.

Chen yining[11] has pointed out that the first component $\sum_{i=1}^{M} P_i(x)\mu_i^Y$ in equation (3) is a mean item through which we obtain the basic shape of the target speech, and the second component $\sum_{i=1}^{M} P_i(x)\Sigma_i^{YX}(\Sigma_i^{XX})^{-1}(x-\mu_i^X)$ is a correlation item which reveals the relationship between the two speakers. The transformation details are obtained by the covariance matrix $\Sigma_i^{YX}(\Sigma_i^{XX})^{-1}$ and the offset vector $x-\mu_i^X$. But most of covariance elements have small values. Our experiment shows that the largest element is 0.73 while about 95% are smaller than 0.1. This means the converted features have lost too many details from the source speaker and this can be explained for the quality degradation of the converted speech.

To maintain enough transformation details, we consider dynamic features Δx and Δy from both the source and target speakers in the training phase. Here $\Delta x_t^i = x_t^i - x_t^{i-1}$ which represents the difference between adjacent parameters in $t-th$ frame. Accordingly $\Delta y_t^i = y_t^i - y_t^{i-1}$. Then we get the $4d$ dimensional feature vectors $z_t = [x_t^T, y_t^T, \Delta x_t^T, \Delta y_t^T]^T$ for training. Equation (4) is modified as follows:

$$\mu_i = \begin{bmatrix} \mu_i^X \\ \mu_i^Y \\ \mu_i^{\Delta X} \\ \mu_i^{\Delta Y} \end{bmatrix}, \quad \Sigma_i = \begin{bmatrix} \Sigma_i^{XX} \Sigma_i^{XY} \Sigma_i^{X\Delta X} \Sigma_i^{X\Delta Y} \\ \Sigma_i^{YX} \Sigma_i^{YY} \Sigma_i^{Y\Delta X} \Sigma_i^{Y\Delta Y} \\ \Sigma_i^{\Delta XX} \Sigma_i^{\Delta XY} \Sigma_i^{\Delta X\Delta X} \Sigma_i^{\Delta X\Delta Y} \\ \Sigma_i^{\Delta YX} \Sigma_i^{\Delta YY} \Sigma_i^{\Delta Y\Delta X} \Sigma_i^{\Delta Y\Delta Y} \end{bmatrix} \qquad (5)$$

On the other hand, Figure 2 shows that the LSFs extracted from the source speech are similar with the corresponding ones extracted from the target speech. Hence we directly use x instead of the first item in equation (3) to avoid the discontinuities caused by the soft classification. The dynamic factors are also considered in the second item. Finally, the modified transformation function is given below.

$$F(x) = x + \sum_{i=1}^{M} P_i(x)\left(\mu_i^Y + \mu_i^{\Delta Y} - \mu_i^X - \mu_i^{\Delta X}\right) \qquad (6)$$

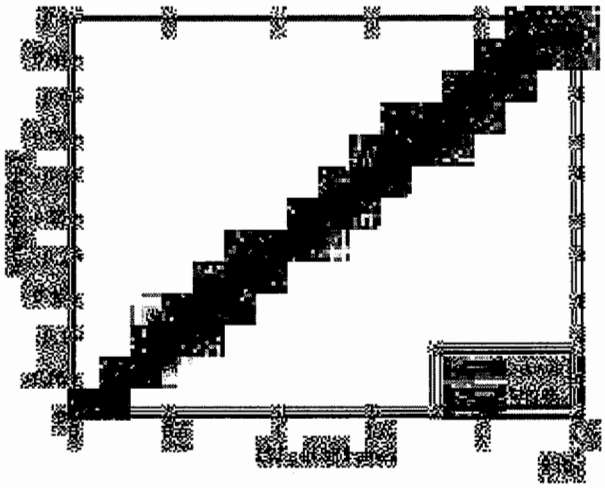


Figure 2. Comparison of LSFs from source and target speakers

4. EXPERIMENTS AND RESULTS

The speech corpus for this study consists of 250 parallel utterances of Mandarin Chinese spoken by two males and two females. These data are sampled at 16 kHz and quantized for 16 bit per sample in a quiet environment. 200 utterances had been used for training and the remaining for the test. We use 18-th order LSF as spectral parameters and set the number of GMM components to be 36.

Objective and subjective tests are conducted respectively, and the female to female, female to male, male to female and male to male voice conversions are performed in each test.

4.1 Objective Evaluation

A log distortion measure in time domain is used to evaluate the objective performance of the voice conversion system, which is defined as

$$SNR = 10\log 10\frac{\sum s(t)^2}{\sum (s(t)-s_c(t))^2} \qquad (7)$$

Where $s(t)$ and $s_c(t)$ denote target and converted voice separately. TABLE I shows the comparison results of the distortion of the traditional and proposed systems.

Table I. Results of the Objective Test

System	Traditional	Dynamic GMM
SNR(dB)	2.5083	3.6722

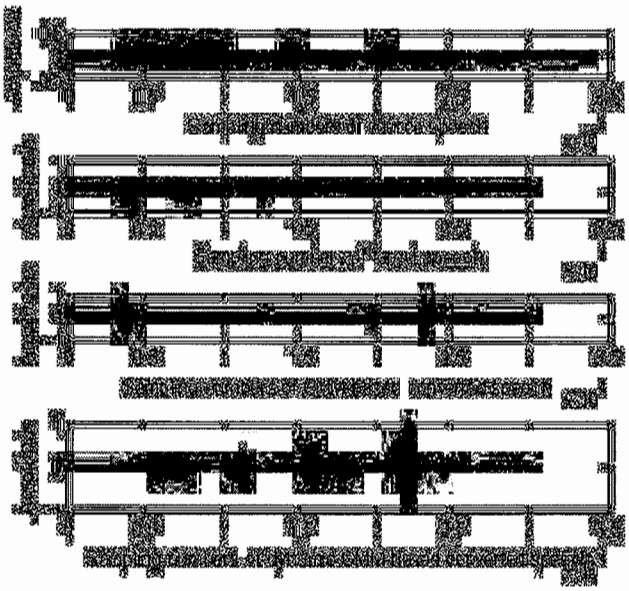


Figure 3. Comparison of four kinds of speech

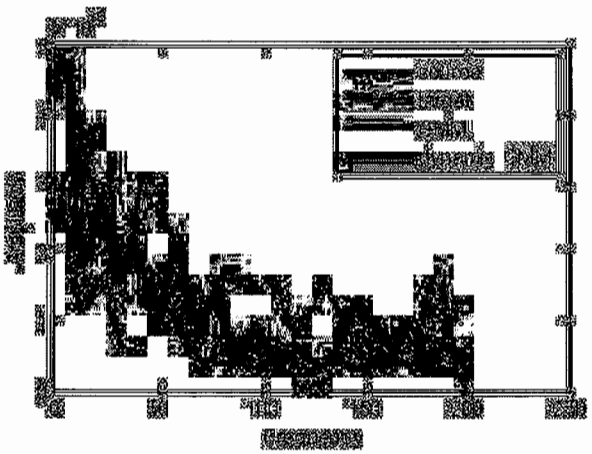


Figure 4. Comparison of corresponding spectrum

The comparison of the speech in time domain from the source and target speakers, and from the traditional GMM-based and dynamic GMM-based conversion systems is given in Figure 3, while Figure 4 shows the comparison of spectral envelope. We can see that according to the proposed system, the converted speech is less noisy and clearer than the traditional converted one. In addition, the corresponding spectrum maintains more details.

4.2 Subjective Evaluation

ABX performance test was performed to evaluate the speaker identification. In this test, 6 listeners were asked to select A (source speaker) or B (target speaker) as being most similar to X (converted speech). TABLE II gives the results

of the correct percentage of the test.

Table II. Results of ABX Test

System	Correct percentage(%)
Traditional	35.6
Dynamic GMM	78.6

Furthermore, they were asked to assess the overall sound quality of the converted speech on an MOS scale between 1 (very bad) and 5 (very good). TABLE III shows the results.

Table III. Results of MOS Test

System	Traditional	Dynamic GMM
MOS	2.4	3.9

5. CONCLUSION

This paper presents a novel method to extend the GMM voice conversion system. In order to overcome the over smoothed problem, dynamic features of spectral parameters in each frame are considered. In addition, the traditional transformation function is modified to alleviate the discontinuities between frames. Objective and subjective experimental results show that our conversion system achieves better performance than the standard GMM method.

Although the quality of the converted speech is enhanced, there is still distortion remained in pitch transformation which makes the performance of speaker identification unsatisfactory, and future work will focus on it.

6. ACKNOWLEDGMENT

This work is mainly supported by the National High Technology Research and Development Program of China (863 program) (No.2006AA010102)

REFERENCES

[1] M.Abe, S.Nakanura, K.Shikano and H.Kuwabara (1988). "Voice conversion through vector quantization".Proc. ICASSP, pp.655-658.

[2] Valbret H., Moulines E., Tubach J.P.,Voice Transformation Using PSOLA Technique, ICASSP, March 1992, 1:145-148

[3] Valbret H., Moulines E., Tubach J.P.,Voice Transformation Using PSOLA Technique, Speech Communication, 1992, 11(7):175-187

[4] Narendranath, M., Murthy, H.A., Rajendran, S., and Yegnanarayana, B., "Transformation of formants for voice conversion using artificial neural networks", Speech Communication, vol. 16, 1995, pp. 207-216.

[5] Y.Stylianou, O.Cappe and E.Moulines (1995). "Statistical methods for voice quality transformation".Proc.EUROSPEECH, pp.447-450.

[6] H. Ye and S. Young, "High quality voice morphing" ICASSP, 2004, pp. I9–I12.

[7] Zhihua Jian, Zhen Yang, "Voice conversion using canonical correlation analysis based on Gaussian mixture model", ICSANP , 2007

[8] H. Kawahara, I. Masuda Katsuse, and A. de Cheveigné,"Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F extraction: Possible role of a repetitive structure in sounds," Speech Commun., vol. 27, no. 3–4, pp. 187–207, 1999.

[9] T.Takigi and H. Kuwabara, "Acoustic parameters of voice individuality and voice quality control by analysis/synthesis method", Speech Communication, Vol.10,pp. 491-495,1991.

[10] Y.Stylianou, O.Cappe, and E.Moulines, "Continuous probabilistic transform for voice conversion", IEEE Trans. Speech Audio Process., vol. 6, no. 2, pp. 131–142, Mar. 1998.

[11] Y.Chen, M.Chu, E.Chang, J.Liu and R.Liu (2003). "Voice conversion with smoothed GMM and MAP adaptation". Proc. EUROSPEECH, pp.2413-2416.

Decoding of QOSTBC Concatenates RS Code Using Parallel Interference Cancellation

Zhenghang Yan1, 2, Yilong Lu1, Maode Ma1 and Yuhang Yang2
1 School of Electrical, Electrical Engineering Nanyang Technological University Nanyang Avenue, 639798, Singapore Email:.eylu@ntu.edu.sg
2 Department of Electronic Engineering Shanghai Jiao Tong University Shanghai, China, 200240
Email:yanzhenghang@sjtu.org

ABSTRACT

Comparing with orthogonal space time block code (OSTBC), quasi orthogonal space time block code (QOSTBC) can achieve high transmission rate with partial diversity. In this paper, we present a QOSTBC concatenated Reed–Solomon (RS) error correction code structure. At the receiver, pairwise detection and error correction are first implemented. The decoded data are regrouped. Parallel interference cancellation (PIC) and dual orthogonal space time block code (OSTBC) maximum likelihood decoding are deployed to the regrouped data. The pure concatenated scheme is shown to have higher diversity order and have better error performance at high signal-to-noise ratio (SNR) scenario than both QOSTBC and OSTBC schemes. The PIC and dual OSTBC decoding algorithm can further obtain more than 1.3 dB gains than the pure concatenated scheme at 10^{-6} bit error probability.

Keywords: OSTBC, QOSTBC, RS Code, Parallel Interference Cancellation

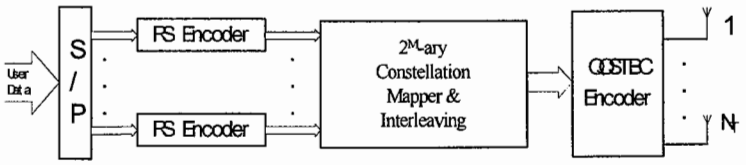
1 Introduction

Quasi-orthogonal space-time block codes (QOSTBC) have recently become an attractive topic because it can achieve high transmission rate compared to orthogonal space-time block code (OSTBC) by scarifying partial diversity [1]. The outage performance upper bound is derived for QOSTBC in [2]. The maximum-likelihood (ML) decoder works with pairwise symbols for QOSTBC, thus its decoding complexity increases exponentially. A low complexity sphere decoding algorithm was put forward to decouple pairwise symbols decoding into single symbol decoding in [3]. J. Kim et. al proposed another efficient decoding algorithm based on iterative interference cancellation [4]. Besides some low complexity decoding algorithms [5, 6], many works have been done to achieve higher transmission rate or higher diversity. An improved QOSTBC design can guarantee both full diversity and fast ML decoding by choosing half of the symbols from a signal constellation set and the other half of them from a rotated constellation set [7, 8]. Using channel correct code and parallel interference cancellation (PIC) is another approach to achieve high throughput and good performance with low complexity detection and decoding algorithm [9]. However, the PIC algorithm in [9] needs enough receive antennas at receiver like in Bell Labs Layered Space-Time (BLAST) [10] and Generalized Layered Space-Time (GLST) architectures [11].

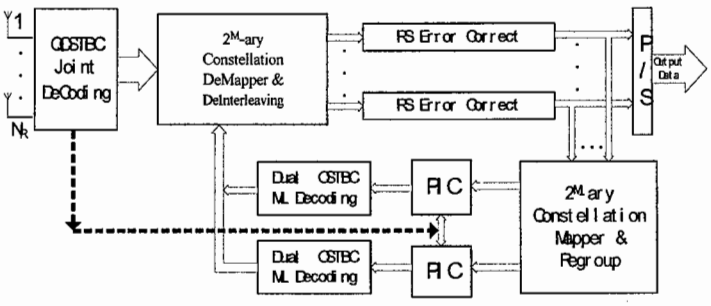
In this letter, we propose a new PIC different from [9]. There is no limits for receive antennas number. Forward error correction (FEC) code is introduced at transmitter. At receiver, the pairwise joint detection is first used in our new method. After the error correction process by FEC, we reconstruct the transmitting constellation symbols and divide them into two groups. When one group of signals are removed from the received signals, the system is equivalent to dual OSTBC systems. Then, linear ML decoding can be processed for the dual OSTBC systems. Full diversity gain and FEC code gain can be simultaneously obtained. Reed–Solomon (RS) error correction codes [12] are used as FEC codes in this paper. RS codes were put forward by Irving Reed and Gus Solomon in the journal of the society for industrial and applied mathematics [13]. These codes have great power and utility, and are today found in a wide variety of commercial applications such as CD, DVD and WiMAX. By using different constellation sizes and different RS code rate, the system can allow for a flexible choice of transmission rates. Numerical results show that the concatenated theme can achieve higher diversity order than both QOSTBC and OSTBC for the same transmission rate and total transmission power constraint. At high signal noise ratio (SNR) scenario, the bit error probability (BEP) performance of the concatenated

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463R · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856327

scheme is far better than that of QOSTBC and OSTBC. The proposed PIC and dual OSTBC decoding algorithm can achieve more than 1.3 dB gains than the concatenated scheme without using PIC algorithms at high SNR scenarios, too.



(a) Transmitter: QOSTBC concatenated with RS code structure



(b) Receiver: PIC and dual OSTBC decoding

Fig 1. Transmitter and receiver structures

The organization of the paper is as follows. Section 2 provides a QOSTBC concatenated with RS error correction code scheme. Section 3 derives the PIC and dual OSTBC ML decoding algorithm for the proposed scheme. Simulation results are presented in section 4 and conclusions are summarized in section 5.

2 The Concatenated QOSTBC Scheme

We focus on the widely used QOSTBC scheme in [1] for systems with four transmit antennas and n_R receive antennas. Analysis for other QOSTBCs and any transmit/receive antennas is similar. Quasi-stationary flat Rayleigh fading channel is assumed and ideal channel state information (CSI) is available at the receiver. The proposed concatenated QOSTBC structure is shown in Fig. 1a. At transmitter, considering 2^M-ary constellations, the input data stream is demultiplexed into M data substreams in series/parallel (S/P) convertor. Each substream is encoded using a RS code respectively. The i^{th} bits from all M RS codewords collectively select the i^{th} 2^M-ary constellation point s_i. Every $4L$ constellation symbols compose one frame. After all $4L$ constellations symbols have been decided in a frame, we take the l^{th}, the $(l+L)^{th}$, the $(l+2L)^{th}$ and the $(l+3L)^{th}$ symbols $(l=1,\ldots,L)$ as a group and denote them as $x_1(l)$, $x_2(l)$, $x_3(l)$, $x_4(l)$. Then, the QOSTBC codewords transmitted at the l^{th} time slot are expressed as [1]

$$\mathbf{A}=\begin{pmatrix}\mathbf{A}_{12} & \mathbf{A}_{34}\\ -\mathbf{A}_{34}^{*} & \mathbf{A}_{12}^{*}\end{pmatrix}=\begin{pmatrix}x_1 & -x_2^{*} & -x_3^{*} & x_4\\ x_2 & x_1^{*} & -x_4^{*} & -x_3\\ x_3 & -x_4^{*} & x_1^{*} & -x_2\\ x_4 & x_3^{*} & x_2^{*} & x_1\end{pmatrix}^{T} \tag{1}$$

where * is conjugate operation and

$$\mathbf{A}_{mn}=\begin{pmatrix}x_m & x_n\\ -x_n^{*} & x_m^{*}\end{pmatrix} \tag{2}$$

The index l is ignored for simplify. Let the total transmitted energy across all n_T transmit antennas be 1 for each time slot. The received signal at the n_R receive antennas is

$$\mathbf{Y} = \mathbf{H}\mathbf{A}^T + \mathbf{N} \tag{3}$$

where received signal $\mathbf{Y}$ is an $N_R \times 4$ matrix. The element h_{ij} of the $N_R \times N_T$ channel matrix $\mathbf{H}$ is the total channel gain from the j^{th} transmit antenna to the i^{th} receive antenna. $\mathbf{N}$ is an $N_R \times 4$ noise matrix whose entries are i.i.d complex Gaussian noise with mean 0 and variance σ_n^2, and independent over time slots.

3 Detection and Decoding

At the receiver, QOSTBC pairwise ML joint detection is first applied to the received signals as shown in Fig. 1b. By minimizing $f_{14}(x_1, x_4)$ the transmitted symbols x_1 and x_4 can be decoded pairwise and the transmitted symbols x_2 and x_3 can be decoded pairwise by minimizing $f_{23}(x_2, x_3)$ as [1]. The simplified ML decoding algorithms in [3, 4] can be used in place of the algorithm in [1] in order to reduce computation complexity. After joint detection, 2^M-ary constellation de-mapping, de-interleaving and RS error correction are operated to the decoded data in a frame. After error correction is completed, we take interleaving and 2^M-ary constellation mapping process to the corrected data frame in the same way as the process in the transmitting process and get x_1^l, x_2^l, x_3^l and x_4^l, where superscript l denotes the l^{th} iterative. The symbols x_3^l and x_4^l are input to a PIC module and the symbols x_1^l and x_2^l are input to another PIC module. At the first PIC module, we remove x_3^l and x_4^l from the received signals by

$$\mathbf{Y}_{12}^l = \mathbf{Y} - \mathbf{H}\begin{pmatrix} 0 & 0 & -x_3^{l*} & x_4^l \\ 0 & 0 & -x_4^{l*} & -x_3^l \\ x_3^l & -x_4^{l*} & 0 & 0 \\ x_4^l & x_3^{l*} & 0 & 0 \end{pmatrix} \tag{4}$$

The subscript of $\mathbf{Y}_{12}^l$ indicates that the signals $\mathbf{Y}_{12}^l$ are contributed by x_1 and x_2. The superscript of $\mathbf{Y}_{12}^l$ indicates iterative number. Assuming x_3^l and x_4^l are successfully decoded ($x_3 = x_3^l$ and $x_4 = x_4^l$), then

$$\mathbf{Y}_{12}^l = \mathbf{H}\begin{pmatrix} x_1 & -x_2^* & 0 & 0 \\ x_2 & x_1^* & 0 & 0 \\ 0 & 0 & x_1^* & -x_2 \\ 0 & 0 & x_2^* & x_1 \end{pmatrix} + \mathbf{N} \tag{5-1}$$

or

$$\mathbf{Y}_{12}^l = \left[\begin{bmatrix} \mathbf{h}_1 & \mathbf{h}_2 \end{bmatrix} \begin{bmatrix} x_1 & -x_2^* \\ x_2 & x_1^* \end{bmatrix} \quad \begin{bmatrix} \mathbf{h}_3 & \mathbf{h}_4 \end{bmatrix} \begin{bmatrix} x_1^* & -x_2 \\ x_2^* & x_1 \end{bmatrix} \right] + \mathbf{N} \tag{5-2}$$

where $\mathbf{h}_i$ is the i^{th} column vector of channel matrix $\mathbf{H}$. Obviously, there are two independent OSTBC received signals in (5-2). We derive the decision metric

$$\left\| \mathbf{Y}_{12}^l - \mathbf{H}\begin{pmatrix} s_1 & -s_2^* & 0 & 0 \\ s_2 & s_1^* & 0 & 0 \\ 0 & 0 & s_1^* & -s_2 \\ 0 & 0 & s_2^* & s_1 \end{pmatrix} \right\|_F^2 \tag{6}$$

where $\|.\|_F^2$ is squared Frobenius norm, s_1 and s_2 are all possible transmitting constellation symbols. Minimizing the decision metric results in a ML decoding. We expand the above metric and get two independent parts, one of which is only a function of s_1 and the other one is only a function of s_2. Thus, after some manipulation x_1 can be detected by minimizing the decision metric

$$\left| \sum_{j=1}^{N_R} \left(y_{12}^l(j,3) h_{j3}^* + y_{12}^{l*}(j,4) h_{j4} \right) - s_1^* \right|^2 + \left| \sum_{j=1}^{N_R} \left(y_{12}^l(j,1) h_{j1}^* + y_{12}^{l*}(j,2) h_{j2} \right) - s_1 \right|^2 + \left(-2 + \sum_{j=1}^{N_R} \sum_{i=1}^{4} |h_{ji}|^2 \right) |s_1|^2 \tag{7}$$

and x_2 can be detected by minimizing the decision metric

$$\left|\sum_{j=1}^{N_R}\left(y_{12}^{1}(j,3)h_{j4}^{*}-y_{12}^{1*}(j,4)h_{j3}\right)-s_2^{*}\right|^2+ \tag{8}$$
$$\left|\sum_{j=1}^{N_R}\left(y_{12}^{1}(j,1)h_{j2}^{*}-y_{12}^{1*}(j,2)h_{j1}\right)-s_2\right|^2+\left(-2+\sum_{j=1}^{N_R}\sum_{i=1}^{4}\left|h_{ji}\right|^2\right)\left|s_2\right|^2$$

where $y_{12}^{1}(j,i)$ is the j^{th} row, the i^{th} column element of $\mathbf{Y}_{12}^{1}$. It is shown that x_1 and x_2 can be solely detected with linear ML decoding algorithm respectively. Full diversity is achieved by using ML decoding algorithm. Comparing with half diversity and pairwise decoding algorithm, the proposed PIC and decoding algorithm can improve diversity gain in the price of not much computation complexity. Similarly, x_3 and x_4 can be detected in the same way in another PIC and dual OSTBC ML decoding branch.

The detected x_1, x_2, x_3 and x_4 are input to 2^M-ary constellation de-mapper and are dealt with the same process after QOSTBC joint detection. The user data can be obtained at the output of parallel/serial convertor. Furthermore, the PIC and dual OSTBC decoding process can be iteratively implemented after x_1, x_2, x_3 and x_4 are all decoded as shown in Fig.1b.

4 Numerical Results

In this section, we provide simulation results for the proposed PIC and dual OSTBC decoding algorithm and compare it with the results for other transmitter/receiver schemes. In all simulations, four transmit antennas and one receive antenna are deployed similarly as in [1]. The total average transmission power keeps constant for all schemes. Fig.2 provides simulation results for the transmission rate of 2 bits/s/Hz. In order to keep the same transmission rate, we used appropriate constellation size for different transmission schemes. QPSK is used for uncoded scheme and rate one QOSTBC. 16-QAM is adopted for the proposed concatenated scheme and rate ½ OSTBC. The optimal rotation angle $\theta=\pi/4$ is adopted for constellation rotation QOSTBC with QPSK modulation [8]. There is no rate ½ RS code, thus we obtain rate ½ by using three 7/15 RS code blocks and one 9/15 RS code block as one transmission unit.

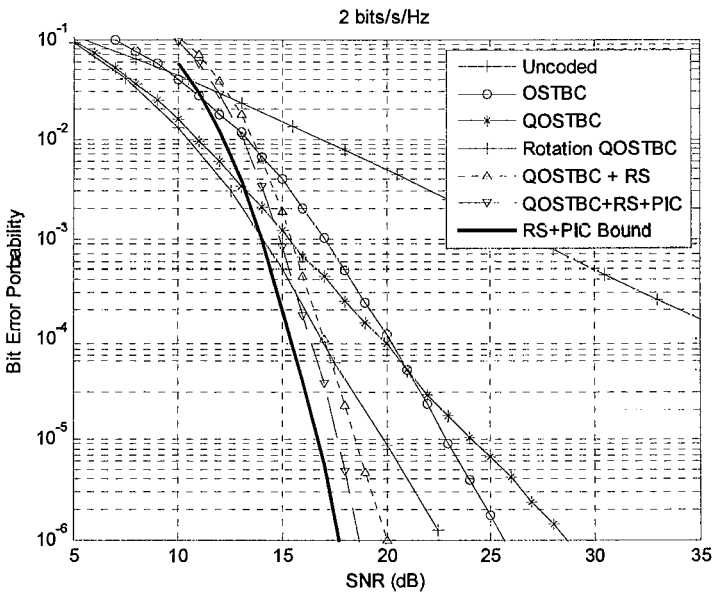


Fig 2. Bit-error probability versus SNR for all transmission schemes at 2 bits/sec/Hz; 4 transmit antennas, 1 receive antenna

The slope of the BEP-SNR curve dictates the degree of diversity. Constellation rotation QOSTBC achieves as full diversity as OSTBC and its error performance is superior to OSTBC and QOSTBC. Fig.2 shows that the scheme using QOSTBC concatenated RS code has higher diversity order than full diversity order because FEC code gains are introduced. At very low SNR, concatenated scheme has bad error performance because large constellation size is adopted. The error performance of the concatenated scheme is better than constellation rotation QOSTBC when SNR>16.0 dB and obtains 2.8 dB gain at 10^{-6} BEP. The proposed PIC and dual OSTBC decoding scheme can achieve rate one, full diversity order and FEC code gain simultaneously. Comparing with pure concatenated scheme, the proposed PIC algorithm only obtains a little gain at very low SNR scenario. The reason lies in that in some frames decoding errors are beyond the correction capability of RS code and more errors are incurred after error correction process on the contrary. The proposed PIC scheme can obtain obvious gain than the pure concatenated scheme except at very low SNR scenario. For example, 1.3 dB gains are obtained at 10^{-6} BEP. The performance upper bound of the proposed PIC scheme is also described in Fig.2. It is assumed that perfect PIC is implemented for performance upper bound. The distance between performance upper bound and BER curve of the proposed PIC scheme is about 0.9 dB. It is introduced by decoding errors at pairwise ML decoding and error propagation in PIC process. Numerical outcome shows that no gain is obtained at the second iterative PIC process and the BEP curve of the second iterative PIC process is ignored in this paper. It is because RS code is a kind of block code and iterative decoding can not improve the performance of block code.

Fig.3 shows the similar outcome at 3 bits/s/Hz. The rate one code and the uncoded system use 8-PSK. The rate ¾ OSTBC and the proposed concatenated scheme use 16-QAM. Since there is no rate ¾ RS code, we obtain rate ¾ by using seven 11/15 RS code blocks and one 13/15 RS code block as one transmission unit. The optimal rotation angle $\theta=\pi/6$ is adopted for QOSTBC with 8PSK modulation. The proposed PIC algorithm can obtain about 1.4 dB gains than the pure concatenated scheme at 10^{-6} BEP. And the distance between the proposed PIC algorithm and the upper bound is about 1.5 dB.

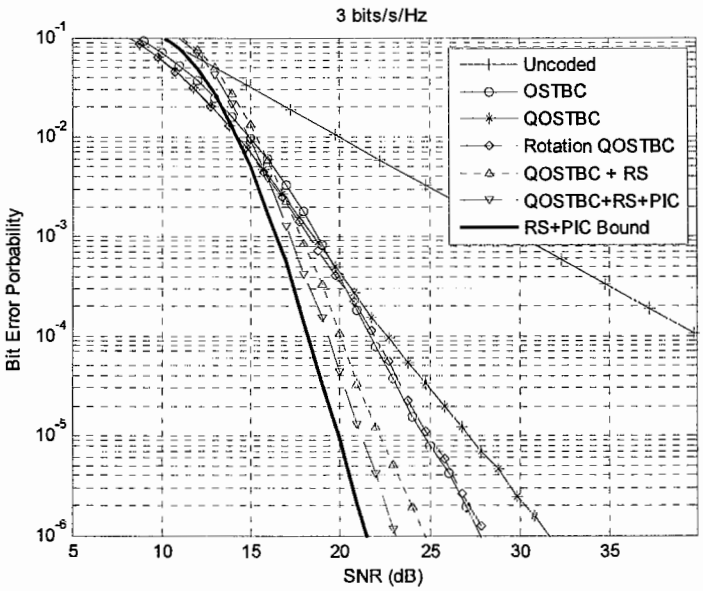


Fig 3. Bit-error probability versus SNR for all transmission schemes at 3 bits/sec/Hz; 4 transmit antennas, 1 receive antenna.

5 Conclusions

We have deployed a QOSTBC concatenated with RS error correction code structure at the transmitter. At the receiver, after pairewise ML detection and error correction, the proposed PIC and dual OSTBC detection are used. In the proposed PIC scheme, rate one is fulfilled for QOSTBC. Full diversity is achieved by dual OSTBC ML decoding after PIC is implemented. Code gain is obtained by using RS error correction code. Thus, full transmitting rate, full diversity and FEC code gain can be obtained simultaneously. The QOSTBC concatenated with RS code structure have better BEP performance than QOSTBC and OSTBC schemes except at very low SNR scenario. The proposed PIC and dual OSTBC decoding algorithm for the concatenation scheme can provide more than 1.3 dB performance gains at 10-6 BEP. Furthermore, the proposed PIC algorithm does not require more receive antennas. It can be used at the tiny terminal only with 1 antenna.

In order to reduce the distance between the performance upper bound and the proposed PIC algorithm, we are going to use constellation rotation QOSTBC in place of traditional QOSTBC in future works.

Acknowledgment

The authors would like appreciate Chinese Scholarship Council for the financial support.

References

[1] Jafarkhani, H., "A quasi-orthogonal space-time block code," IEEE Trans. on Commun., 49(1), 1-4 (2001).

[2] Sezgin, A. and Jorswieck, E. A., "Tight Upper Bound on the Outage Probability of QSTBC," IEEE Commun Letters,10(11), 784-786 (2006).

[3] Peng, A. Y. C., Kim, I.-M., and Yousefi, S., "Low-complexity sphere decoding algorithm for quasi-orthogonal space-time block codes," IEEE Trans. on Commun., 54(3), 377-382 (2006).

[4] Jeongchang, K. and Kyungwhoon, C., "An efficient decoding algorithm for QO-STBCs based on iterative interference cancellation," IEEE Commun. Letters, 12(4), 292-294 (2008).

[5] Azzam, L., and Ayanoglu, E., "Maximum Likelihood Detection of Quasi-Orthogonal Space-Time Block Codes Analysis and Simplification," in IEEE ICC '2008. 3948-3954 (2008).

[6] Chae, C., Jung, T., Hwang, I., "Design of new quasi-orthogonal STBC with minimum decoding complexity for four transmit antennas," IEICE Trans. Commun. E91-B (10) (2008).

[7] Su, W. F., and Xia, X. G., "Signal constellations for quasi-orthogonal space-time block codes with full diversity," IEEE Trans. on Inform. Theory, 50(10), 2331-2347 (2004).

[8] Liang, X. and Liu, H. P., "Optimal rotation angles for quasi-orthogonal space-time codes with PSK modulation," IEEE Commun. Letters, 9(8), 676-678 (2005).

[9] Pau, N. S. J., Taylor, D. P. and Martin, P. A., "Robust high throughput space time block codes using parallel interference cancellation," IEEE Trans. on Wireless Commun., 7, 1603-1613 (2008).

[10] Foschini, G. J., "Layered space-time architecture for wireless communication is a fading environment when using multiple antennas," Bell Labs. Tech. J., 1 (2), (1996).

[11] Tao, M. and Cheng, R.S., "Generalized layered space-time codes for high data rate wireless communications," IEEE Trans. on Wireless Commun., 3(4), 1067–1075 (2004).

[12] Wicker, S. B., [Error Control Systems for Digital Communication and Storage], Upper Saddle River, NJ, Prentice Hall (1995).

[13] Reed, I. S. and Solomon, G., "Polynomial Codes Over Certain Finite Fields," Journal of the Society for Industrial and Applied Math, 8(2), 300-304 (1960).

Color-SIFT model: A robust and an accurate shot boundary detection algorithm

[1]M. Sharmila Kumari

Department of Computer Science and Engineering, P A College of Engineering, Nadupadavu, Mangalore, Karnataka, India

[2]B. H. Shekar

Department of Computer Science, Mangalore University, Mangalore, Karnataka, India

ABSTRACT

In this paper, a new technique called color-SIFT model is devised for shot boundary detection. Unlike scale invariant feature transform model that uses only grayscale information and misses important visual information regarding color, here we have adopted different color planes to extract keypoints which are subsequently used to detect shot boundaries. The basic SIFT model has four stages namely scale-space peak selection, keypoint localization, orientation assignment and keypoint descriptor and all these four stages were employed to extract key descriptors in each color plane. The proposed model works on three different color planes and a fusion has been made to take a decision on number of keypoint matches for shot boundary identification and hence is different from the color global scale invariant feature transform that works on quantized images. In addition, the proposed algorithm possess invariance to linear transformation and robust to occlusion and noisy environment. Experiments have been conducted on the standard TRECVID video database to reveal the performance of the proposed model.

Shot detection; Local descriptors; Video segmentation; Video abstraction.

1 INTRODUCTION

The massive amount of digital media coverage that results in huge video data generates difficulties in identifying and selecting desired information for semantic analysis of video content. The critical issue relies in identification of a continuously recorded sequence of image frames in such a huge data. This massive video data finds its application in digital libraries, security, distance learning, advertising, electronic publishing, broadcasting, interactive TV, video-on-demand entertainment, and so on where it becomes more and more necessary to support users with powerful and easy-to-use tools for searching, browsing and retrieving media information.

Quite a lot of approaches to shot boundary detection were proposed in the literature. The principle methodology of shot boundary detection is to extract one or more features from every n^{th} frame of a video sequence, to compute the difference of features for consecutive frames, and to compare these differences to a given threshold. Each time the threshold is exceeded, a shot boundary is detected. The various approaches differ concerning the used features. The detection of boundaries between video shots provides a basis for almost all of the existing video segmentation and abstraction methods[8]. However, it is quite difficult to give a precise definition of a video shot transition since many factors such as camera motions may change the video content significantly. Usually, a shot is defined to be a sequence of frames that appears to be continuously captured by the same camera [6].

Generally, a shot can encompass camera motions such as pans, tilts, or zooms, and video editing effects such as fades, dissolves, wipes and mattes [3, 8]. In particular, video shot transitions can be categorized into two types: *abrupt/sharp shot transitions*, which is also called *cuts*, where a frame from one shot is followed by a frame from a different shot, and *gradual*

[1] sharmilabp@gmail.com, [2] bhshekar@gmail.com

Second International Conference on Digital Image Processing, edited by Kamaruzaman Jusoff, Yi Xie,
Proc. of SPIE Vol. 7546, 75463S · © 2010 SPIE · CCC code: 0277-786X/10/$18 · doi: 10.1117/12.856328

shot transitions, such as cross dissolves, fade-ins and fade-outs, and various other editing effects. Several models have been proposed by many researchers to address these problems which are fall into one of two domains, either uncompressed or compressed, depending on whether it is applied to raw video stream or compressed video data. As noted in [6], methods working on uncompressed video are, in general, more reliable but require massive storage and computational resources, compared with techniques in the compressed domain. Our research work focus on uncompressed domain and in particular on hard cuts.

It is observed from the literature that the histogram-based approaches have been widely used in shot detection. Beyond the basic histogram comparison algorithm, several researchers have proposed various approaches to improve its performance, such as histogram equalization[8], histogram intersection[15], histogram on group of frames[5] etc. However, experimental results show that approaches which enhance the difference between two frames across a cut may also magnify the difference due to object and camera movements[18].

Another interesting issue is which color space to use when we consider color-based techniques such as color histogram comparison. As we know, the HSV color space reflects human perception of color patterns. In [7] the performance of several color histograms based methods using different color spaces, including RGB, HSV, YIQ, etc. were evaluated. Experimental results showed that HSV performs quite well with regard to classification accuracy and it is one of those that are the least expensive in terms of computational cost of conversion from the RGB color space. Inspired by this work, we addressed the problem of video shot detection in RGB color space.

On the other hand, local descriptors[9, 11, 17] are commonly employed in a number of real-world applications such as object recognition[4, 11] and image retrieval[13] as they can be computed efficiently, are resistant to partial occlusion, and are relatively insensitive to changes in viewpoint. Mikolajczyk and Schmid[12] presented a comparative study of several local descriptors including moment invariants[17], Scale Invariant Feature Transform[11], and cross-correlation of different types of interest points[9, 13]. Their experiments showed that the ranking of accuracy for the different algorithms was relatively insensitive to the method employed to find interest points in the image but was dependent on the representation used to model the image patch around the interest point. The best matching results were obtained using the SIFT descriptor and hence in our work, we focus on this SIFT algorithm and explores alternatives to its descriptor representation. We have used SIFT descriptor to extract local features in our proposed model. However, SIFT works only grayscale information and hence we have considered RGB color spaces to extract local descriptors on each color space and fuse the information available in the three different color spaces to achieve the target.

The remainder of this paper is organized as follows. A review of SIFT descriptor is given in section 2. The proposed model, color-SIFT is given in section 3. Section 4 presents experimental results along with comparative analysis. Conclusion is given in section 5.

2 SIFT DESCRIPTOR: A REVIEW

Scale Invariant Feature Transform (SIFT) is an approach for detecting and extracting local feature descriptors that are reasonably invariant to changes in illumination, image noise, rotation, scaling, and small changes in viewpoint. The SIFT transforms image data into scale-invariant coordinates relative to local features that consists of four major phases: (1) scale-space peak selection; (2) keypoint localization; (3) orientation assignment; (4) keypoint descriptor.

The first stage of SIFT model is scale-space extrema detection, which computes interest points that are invariant to scale and orientation are extracted by searching over all scales and image locations. Interest points for SIFT features correspond to local extrema of difference-of-Gaussian filters at different scales. The first step toward the detection of interest points is the convolution of the image with Gaussian filters at different scales, and the generation of difference-of- Gaussian images from the difference of adjacent blurred images. The convolved images are grouped by octave by having same number of difference-of-Gaussian images per octave with a fixed number of blurred images per octave. An efficient approach to construction of Gaussian pyramid is as follows:

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \qquad \ldots(1)$$

$$D(x,y,\sigma) = L(x,y,k\sigma) - L(x,y,\sigma) \quad \ldots(2)$$

Where I(x,y) is the input image, * is the convolution operator in x and y, σ is a scale factor, k is a constant factor, D is the difference of Gaussian image and

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \quad \ldots(3)$$

The second stage is keypoint localization, where keypoints are identified as local maxima or minima of the DoG images across scales. Each pixel in the DoG images is compared to its 8 neighbors at the same scale, plus the 9 corresponding neighbors at neighboring scales. If the pixel is a local maximum or minimum, it is selected as a candidate keypoint. All the candidate keypoints are localized to sub-pixel accuracy and eliminated if found unstable. An accurate position fix on the key-points located in the previous step has been implemented by fitting a 3D quadratic function to the local sample points:

$$\text{i.e.,} \quad D(x) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D^T}{\partial X^2} X \quad \ldots(4)$$

The third stage is orientation assignment, where the keypoint is assigned an orientation. It involves calculating the gradient vectors in a window around the SIFT feature on the scale at which the feature was detected. To determine the keypoint orientation, a gradient orientation histogram is computed in the neighborhood of the keypoint using the Gaussian image at the closest scale to the keypoint's scale. The contribution of each neighboring pixel is weighted by the gradient magnitude and a Gaussian window with a σ that is 1.5 times the scale of the keypoint. Peaks in the histogram correspond to dominant orientations. A separate keypoint is created for the direction corresponding to the histogram maximum. All the properties of the keypoint are measured relative to the keypoint orientation which provides invariance to rotation.

Finally, keypoint descriptors are computed by having keypoint orientation. The feature descriptor is computed as a set of orientation histograms on 4 x 4 pixel neighborhoods. The orientation histograms are relative to the keypoint orientation and the orientation data comes from the Gaussian image closest in scale to the keypoint's scale. Histograms contain 8 bins each, and each descriptor contains an array of 4 histograms around the keypoint. This leads to a SIFT feature vector of size (4x4) x 8.

3 COLOR-SIFT MODEL: A PROPOSED ALGORITHM

This section presents a novel method of detecting shot boundaries by transforming a given video image containing N number of frames into frames described by key descriptor matrices. The proposed method of shot boundary detection has two stages. In the first stage, each color plane image obtained for each frame of a video image is subjected to feature extraction process, the output of which is fed to similarity computation model among the successive frames that results in three similarity values for each frame. In the second stage, shot detection scheme is employed to identify the shot boundaries (see Fig. 1) based on pre-defined threshold value.

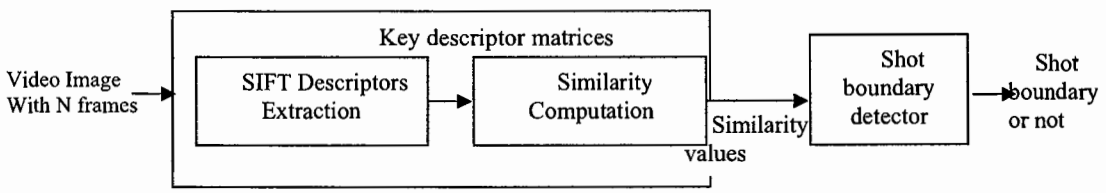


Figure-1. Block Diagram of the proposed method of shot boundary detection.

Let V be the video image containing N number of frames. In the first phase of the proposed model, SIFT descriptors are extracted for each color plane. In our work, we have considered RGB color space and hence three feature descriptors matrices associated with R, G and B color spaces have been obtained for each frame. Subsequently, the similarity between the consecutive frames is obtained. While computing the similarity, we have considered the similarity between the key

descriptor matrices of the corresponding color spaces. This similarity information is fused to obtain a decision on shot detection.

More formally, let ${}^{f1}K_R$, ${}^{f1}K_G$ and ${}^{f1}K_B$ be the three key descriptor matrices obtained using SIFT descriptors associated with frame f_1. Similarly, the key descriptor matrices associated with next consecutive frame f_2 is also computed. Let ${}^{f2}K_R$, ${}^{f2}K_G$ and ${}^{f2}K_B$ be the three key descriptor matrices associated with frame f_2. The number of keypoint matches between these two consecutive frames is obtained using Euclidean distance measure. Let n_R, n_G and n_B be the number of matches found respectively for red, green and blue color planes. The minimum of these matches is the minimum matching count associated with f_2. The percentage of match is obtained by computing:

$$\frac{\min(n_R, n_G, n_B)}{\text{Number of keys}} * 100 \qquad \ldots(5)$$

This percentage of match is stored as the match score associated with frame 2 (f_2). Similarly, the percentage of match score associated with all the frames is obtained.

Once percentage of match score is obtained for all the frames, shot detection is identified based on the set threshold. This threshold value normally lies between [0 and 5]. It is said to be scene change if the frame is having percentage score less than the pre-set threshold.

4 EXPERIMENTAL RESULTS

This section presents the results of the experiments conducted to corroborate the success of the proposed model. We have conducted experimentation on TRECVID video database. We have specifically chosen this video database as these are used by many researchers as a benchmark database to verify the validity of their proposed shot detection models. All experiments are performed on a P-IV 2.99GHz Windows machine with 504 MB of RAM.

Experimentation on 'indi016' video data: This video data is uncompressed to obtain the data in AVI format. We have considered first 1791 frames. The minimum number of matches obtained between consecutive frames is shown in Fig. 2.

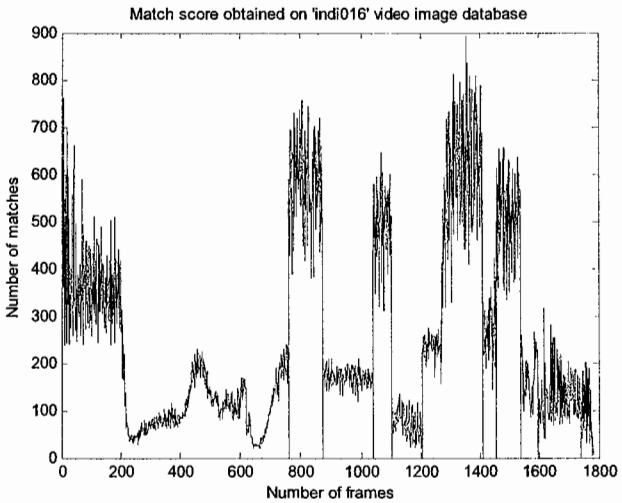


Figure-2. Minimum match score for 'indi016' video image database.

It shall be observed from the above figure that there is a scene change in a frame nearer to value 800. To be specific, the following are the frame numbers associated with scene change. Frame numbers: 762, 872, 1039, 1101, 1205, 1408, 1455, 1537, 1596, 1740, 1778 and 1781.

For illustration purpose, we have presented the following figure that gives an insight into shot change.

Frame Number: 761 Number of key descriptors: Red plane:440 Green plane:401 Blue plane:439	Frame Number: 762 Number of key descriptors: Red plane:396 Green plane:455 Blue plane:407
Number of matches: 0	

Similarly, experimentation has been conducted on 'indi017' database. The shot detection results are shown in fig. 3. The frame numbers that were obtained from the proposed model depicting shot changes are: 114, 190, 248, 491, 624, 659, 760, 912, 1047, 1051, 1063, 1523, 1702 and 1782.

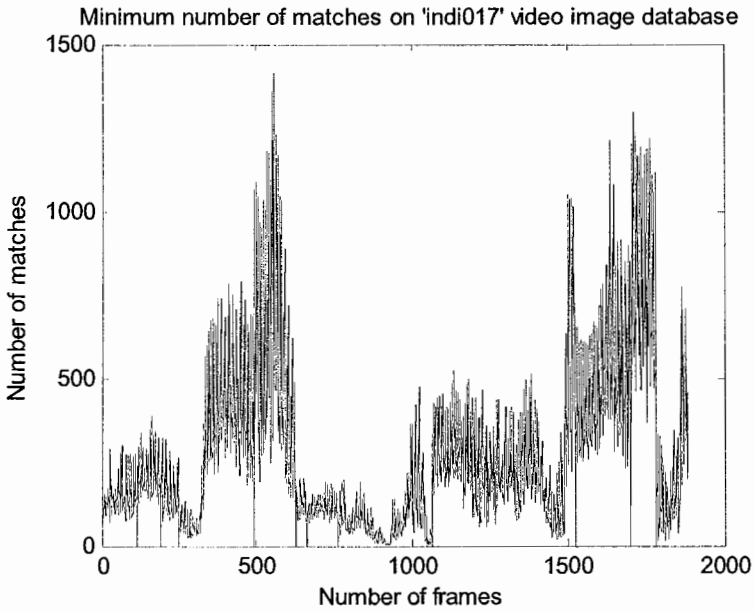


Figure-3. Minimum match score for 'indi017' video image database.

Percentage of match scores for each frame is shown in Fig. 4.

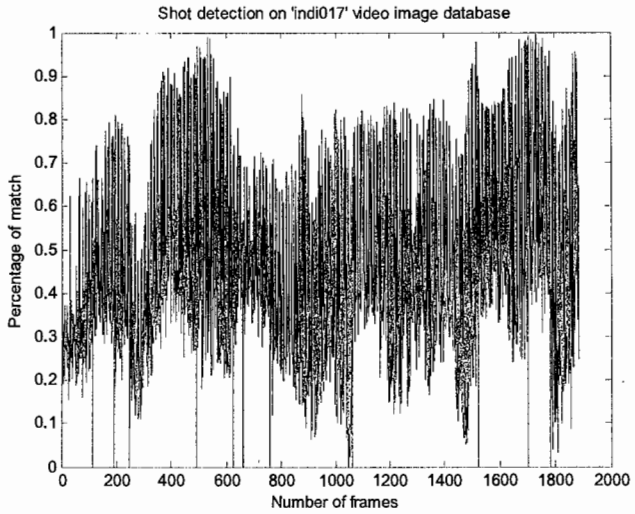


Figure-4. Percentage of match scores for 'indi017' video image database.

Further, to demonstrate the computational efficiency, computing time taken for feature extraction (match score) for a frame is given in Table 1 along with the other models exist in the literature used for shot detection.

Table-1. Comparison of feature extraction time of color-SIFT with other methods.

Shot detection method	**Per frame feature extraction time (sec.)**
Color[10]	0.794
Motion[2]	0.623
Color + Motion	0.820
Temporal slices analysis[14]	0.878
OLFC[16]	0.956
Color-SIFT **(Proposed model)**	0.836

5 CONCLUSION

We have made an attempt to use local descriptor based technique for shot detection. Inspired by the work of many researchers who works on different color spaces and also the significant merits of SIFT descriptors, we have proposed a color-SIFT model that uses color information for accurate detection of shots. Experimental results on standard TRECVID video image database reveal that the proposed model best suits for shot detection purpose. In our future work, we planned to work in HSY color space as it is found to be better than RGB space and modification to SIFT model that suits HSY color space too. In addition, pair-wise clustering technique can also be employed to reduce the number of colors that results in only one color plane which can be subsequently used for shot detection is also our future target.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the funding support provided by University Grants Commission, New Delhi vide No. 36-163/2008(SR) to carry out the part of this project work.

6 REFERENCES

[1]. Aigrain P. and P. Joly. "The Automatic Real-Time Analysis of File Editing and Transition Effects and Its Applications", Computer and Graphics, Vol. 18(1), pp. 93-103 (1994).

[2]. Bruno, E. and D. Pellerin. "Video Shot Detection based on Temporal Linear Prediction of Motion". In Proc. IEEE Intl. Conf. on Multimedia and Exposition, Lausanne, Switzerland, pp. 289-292 (2002).

[3]. Del Bimbo, A. Visual Information Retrieval, Morgan Kaufmann, San Friancisco, CA (1999).

[4]. Fergus, R., P. Perona, and A. Zisserman. Object class recognition by unsupervised scale-invariant learning. In Proceedings of Computer Vision and Pattern Recognition. (2003).

[5]. Ferman, A.M., A. M. Tekalp, and R. Mehrotra, Robust Color Histogram Descriptors for Video Segment Retrieval and Identification, IEEE Transactions on Image Processing, Vol. 11(5), pp. 497- 507 (2002).

[6]. Gargi, U., R. Kasturi, and S. H. Strayer, Performance Characterization of Video-Shot-Change Detection Methods, IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10(1), pp. 1-13 (2000).

[7]. Gargi, U., S. Oswald, D. Kosiba, S. Devadiga, and R. Kasturi, Evaluation of Video Sequence Indexing and Hierarchical Video Indexing, Proceedings of SPIE Conference on Storage and Retrieval in Image and Video Databases, pp. 1522-1530 (1995).

[8]. Hanjalic, A. Shot-Boundary Detection: Unraveled and Resolved? IEEE Transactions on Circuits and Systems for Video Technology, Vol. 12(2), pp. 90-105 (2002).

[9]. Harris, C. and M. Stephens. A combined corner and edge detector, In Alvey Vision Conference, pp. 147–151 (1988).

[10]. Koprinska, I. and S. Carrato. Temporal Video Segmentation, a Survey. Signal Processing, Image Communication-16, pp. 450-477 (2001).

[11]. Lowe, D.G. Object recognition from local scale-invariant features, In Proceedings of International Conference on Computer Vision, pp. 1150–1157 (1999).

[12]. Mikolajczyk, K and C. Schmid. A performance evaluation of local descriptors, In Proceedings of Computer Vision and Pattern Recognition (2003).

[13]. Mikolajczyk, K and C. Schmid. Indexing based on scale invariant interest points. In Proceedings of International Conference on Computer Vision, pp. 525–531 (2001).

[14]. Ngo, C.-W., T.-C. Pong and H.-J. Zhang. On Clustering and Retrieval of Video Shots through Temporal Slices Analysis, IEEE Trans. on Multimedia, 4(4), pp. 446- 458 (2002).

[15]. Swain, M.J., and D. H. Ballard, Color Indexing, International Journal of Computer Vision, Volume 7, Number 1, pp. 11-32 (1991).

[16]. Vadivel, A. Mohan, M. Sural, Shamik Majumdar, A. K. OBJECT Level Frame Comparison for Video Shot Detection, Proceedings of the IEEE Workshop on Motion and Video Computing (WACV/MOTION'05), Vol. 2, pp. 235-240 (2005).

[17]. Van Gool, L., T. Moons, and D. Ungureanu. Affine/photometric invariants for planar intensity patterns. In Proceedings of European Conference on Computer Vision (1996).

[18]. Zhang, H., A. Kankanhalli, and S. Smoliar, Automatic Partitioning of Video, Multimedia Systems, Vol. 1(1), pp. 10-28 (1993).

Author Index

Numbers in the index correspond to the last two digits of the six-digit citation identifier (CID) article numbering system used in Proceedings of SPIE. The first four digits reflect the volume number. Base 36 numbering is employed for the last two digits and indicates the order of articles within the volume. Numbers start with 00, 01, 02, 03, 04, 05, 06, 07, 08, 09, 0A, 0B ... 0Z, followed by 10-1Z, 20-2Z, etc.