# Table of Contents

<table>
<thead>
<tr>
<th>Papers</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1150843440</td>
<td></td>
</tr>
<tr>
<td>Jun Zhang and Qieshi Zhang and Jinglu Hu</td>
<td>1--9</td>
</tr>
<tr>
<td>RGB Color Centroids Segmentation (CCS) for Face Detection</td>
<td></td>
</tr>
<tr>
<td>P1150729002</td>
<td></td>
</tr>
<tr>
<td>Imran S. Bajwa and M. Shahid Naweed and M. Nadim Asif and S. Irfan Hyder</td>
<td>11--17</td>
</tr>
<tr>
<td>Feature Based Image Classification by using Principal Component Analysis</td>
<td></td>
</tr>
<tr>
<td>P1150846474</td>
<td></td>
</tr>
<tr>
<td>Jiri Giesl and Karel Vlcek</td>
<td>19--26</td>
</tr>
<tr>
<td>IMAGE ENCRYPTION BASED ON STRANGE ATTRACTOR</td>
<td></td>
</tr>
<tr>
<td>P1150845461</td>
<td></td>
</tr>
<tr>
<td>Preeti Aggarwal and H.K. Sardana and Gagandeep Jindal</td>
<td>27--37</td>
</tr>
<tr>
<td>Content Based Medical Image Retrieval: Theory, Gaps and Future Directions</td>
<td></td>
</tr>
<tr>
<td>P1150850525</td>
<td></td>
</tr>
<tr>
<td>Zhengmao Ye</td>
<td>39--46</td>
</tr>
<tr>
<td>Objective Assessment of Nonlinear Segmentation Approaches to Gray Level Underwater Images</td>
<td></td>
</tr>
</tbody>
</table>
RGB Color Centroids Segmentation (CCS) for Face Detection

Jun Zhang, Qieshi Zhang, and Jinglu Hu

Graduate School of Information, Production and System, Waseda University
Hibikino 2-7, Wakamatsu-ku, Kitakyushu, Japan
J.Zhang@Akane.Waseda.jp, Q.Zhang@Akane.Waseda.jp, and Jinglu@Waseda.jp

Abstract
Nowadays, face detection plays an important role in many application areas such as video surveillance, human computer interface, face recognition and face image database management etc. In face detection applications, face usually form an inconsequential region of images. Consequently, preliminary segmentation of images into regions that contain “non-face” objects and regions that may contain “face” candidates can greatly accelerate the process of human face detection. Color information based methods take a great attention, because colors have obvious character and robust visual cue for detection.

This paper presents a new color thresholding method for detecting and tracking multiple faces in video sequence. The proposed method calculates the color centroids of image in RGB color space and segments the centroids region to get ideal binary image at first. Then analyze the facial features structure character of wait-face region to fix face region. The novel contribution of this paper is creating the color triangle from RGB color space and analyzing the character of centroids region for color segmenting. The experimental results show that the proposed method can achieve ideal thresholding result and it is much better than other color analysis based thresholding methods and can overcome the influence of background conditions, position, scale instance and orientation in images.

Keywords: Face detection, Color image thresholding, Color centroids segmentation.

1. Introduction
In our daily life, more and more application techniques based on biometrics recognition such as fingerprints, iris pattern and face recognition are developed to secure access control. Along with the development of those techniques, computer control plays an important role in making the biometrics recognition more economically feasible in such developments.

Face recognition is a major concerned research direction in this field. In recent years, the face recognition becomes a popular research direction and is applied in many various applications such as financial transactions, monitoring system, credit card verification, ATM access, personal PC access, video surveillance etc.

Face detection and tracking are the key processes of face recognition. In recent surveys on face detection, some techniques such as principal component analysis (PCA), neural networks (NN), support vector machines (SVM), Hough transform (HT), geometrical template matching (GTM), color analysis etc. based methods are used to achieve this application.

PCA based methods [1]-[3] need create Eigen face by many dimension data and training sample data. The NN based methods [4][5] require a large number “face” and “non-face” images to train respectively for getting the network model [6]. SVM based method [7] is a linear classifier and can classify goal region in hyper-plane. HT [8] and GTM [9] based methods are incorporated to detect gray faces in real time applications. Face detection methods based on the representation used reveals that detection algorithms using holistic representations have the advantage of finding small faces or faces in low quality images, while those using the geometrical facial features [10] provide a good solution for detecting faces in different poses. A combination of holistic and feature-based approaches is a promising approach to face detection as well as face recognition.

Human skin color has been used and proven to be an effective feature in the applications of face detection and tracking. Although different people have different skin color, several studies have that the major difference lies largely between their intensity rather than their chrominance [11]-[13]. Several color space have been used to detect pixels as skin region including RGB [14], [15], HSV (or HSI) [16], YCbCr [17]-[19] and YIQ [20]. Color information is an efficient tool for identifying facial areas and specific facial features if the skin color model can be properly adapted for different lighting environments. However, such skin color models are not effective where the spectrum of the light source varies significantly. In
other words, color appearance is often unstable due to changes in both background and foreground lighting. To solve above problems, this paper analyzes the character of color instead of using the existing color space and color channel analysis based methods. This paper proposes a color image thresholding algorithm based on color centroids segmentation (CCS) for face detection and tracking. Experiments have been made on video sequences with multiple faces in different positions, scales and poses, or the faces appear or disappear from sequence. This paper is an extension of our paper [21]. We have added automatic multi-threshold selection and robust detection for color images. Experimental results show that the proposed method can achieve ideal detection results.

1.1 Organization
In Section 2, we introduce how to create the CCS model from created color triangle in detail. Section 3 describes the thresholding algorithm by analyzing the color centroids region. Detection and tracking method will be discussed in detail in Section 4. Section 5 presents the thresholding results of our approach and gives the comparative results with other thresholding methods. And we also give some detection results of various situations and compare them with some reference methods. In Section 6, we summarize this paper and propose some future works.

2. Color Centroids Segmentation Model
This section firstly describes how to transform the three components of 3-D RGB color space to 2-D polar coordinate system for creating color triangle. Then calculates the centroids distribution region of all colors and transforms it to histogram. Finally, analyzes the histogram and gets multi-threshold to segment the centroids region. After these processing, the colors in one image can be divided to 2–7 colors by 2–7 thresholds and the result is better than traditional thresholding methods.

2.1. Color Triangle
In image processing, RGB, YCbCr, HSV, HSI etc. color spaces are widely used. These color spaces use three components to reflect different color e.g. RGB color space consists of $R$, $G$ and $B$ components. This paper transforms the 3-D RGB color space (Figure 1(a)) to 2-D polar system as the color triangle (Figure 1(b)) following equation (1).

\[
\begin{align*}
\begin{bmatrix}
\varphi_R \\
\tau_R
\end{bmatrix} &= \begin{bmatrix}
0 & 0 & 0 \\
1 & 0 & 0
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix} + \begin{bmatrix}
90^\circ \\
0
\end{bmatrix} \\
\begin{bmatrix}
\varphi_G \\
\tau_G
\end{bmatrix} &= \begin{bmatrix}
0 & 0 & 0 \\
0 & 1 & 0
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix} + \begin{bmatrix}
210^\circ \\
0
\end{bmatrix} \\
\begin{bmatrix}
\varphi_B \\
\tau_B
\end{bmatrix} &= \begin{bmatrix}
0 & 0 & 0 \\
0 & 0 & 1
\end{bmatrix} \begin{bmatrix}
R \\
G \\
B
\end{bmatrix} + \begin{bmatrix}
330^\circ \\
0
\end{bmatrix}
\end{align*}
\]

Here $R$, $G$ and $B$ is the component value of RGB color space, ($\varphi_R, \tau_R$), ($\varphi_G, \tau_G$), ($\varphi_B, \tau_B$), is the coordinate of $R$, $G$ and $B$ in polar coordinate system. Color triangle is created by following steps:

**Step1**: create a standard 2-D polar coordinate system;
**Step2**: create three color vectors to reflect $R$, $G$ and $B$, the range of them is $[0, 255]$ and alternation $120^\circ$ reciprocally as follows:

\[
\begin{align*}
R \text{ Component: } \varphi_R &= 90^\circ, \quad \tau_R \in [0, 255] \\
G \text{ Component: } \varphi_G &= 210^\circ, \quad \tau_G \in [0, 255] \\
B \text{ Component: } \varphi_B &= 330^\circ, \quad \tau_B \in [0, 255]
\end{align*}
\]

**Step3**: connect the three apexes.
After above processes, the color triangle can be created as Figure 1(b). For different $R$, $G$ and $B$, the shape of triangle is changeable. For example, Figure 1(c) shows three sets of $R$, $G$ and $B$ their corresponding color triangles. From this example, it can be seen that no matter the $R$, $G$ and $B$ values change the main structure is unmodified.

2.2. Centroids Hexagon
Because the direction of $R$, $G$ and $B$ vectors is fixed and the value range is $[0, 255]$, different combination of $R$, $G$, $B$ represents different color, and the shape of color triangle is different too. The different shape triangles have different centroids and all centroids result in a hexagon region shown in Figure 2. This hexagon is divided to 7 regions: M (Magenta), R (Red), Y (Yellow), G (Green), C (Cyan), B (Blue) and L (Luminance, achromatic) regions. So we may use seven threshold curves as the separating lines for thresholding.

Observe the relationship between color and its corresponding centroid position of color triangle, we find that if the $R$, $G$ and $B$ values are close, no matter small or large, it only reflect the luminance information (weak color information). So the centroids of corresponding color triangles will locate in a circular region (L region). And other six color

\[
[R,G,B] = [150,25,15] \quad [R,G,B] = [25,91,143] \quad [R,G,B] = [218,200,143]
\]

(c) The example of color triangle shape with different color
Figure 1. Color triangle model

regions reflect the color character of $R$, $G$ and $B$ components.
3. CCS Based Color Image Thresholding

3.1 Multiple thresholds

Considering the $L$ region is usually not the goal region and existing method cannot effectively divide white and black region which are noises, cluster them into one region will effectively overcome the influence of white, black and other achromatic regions. Here let $r_i$ as the threshold of $L$ region, $\varphi$ as the angle, the function of threshold curve is:

$$r(\varphi) = T_L(0^\circ < \varphi \leq 360^\circ)$$

(3)

The other six regions which around the $L$ region as following formulas show:

$$
\begin{align*}
M & \text{ Region: } \varphi_M \leq \varphi \leq \varphi_M', \quad r_M > r_L \\
R & \text{ Region: } \varphi_R \leq \varphi \leq \varphi_R', \quad r_R > r_L \\
Y & \text{ Region: } \varphi_Y \leq \varphi \leq \varphi_Y', \quad r_Y > r_L \\
G & \text{ Region: } \varphi_G \leq \varphi \leq \varphi_G', \quad r_G > r_L \\
C & \text{ Region: } \varphi_C \leq \varphi \leq \varphi_C', \quad r_C > r_L \\
B & \text{ Region: } \varphi_B \leq \varphi \leq \varphi_B', \quad r_B > r_L
\end{align*}
$$

(4)

In formula (3) and (4), $r_L$, $\varphi_M$, $\varphi_R$, $\varphi_Y$, $\varphi_G$, $\varphi_C$ and $\varphi_B$ are the thresholds, and the initial value of them is $5^\circ$, $60^\circ$, $120^\circ$, $180^\circ$, $240^\circ$, $300^\circ$ and $360^\circ$.

The seven thresholds can divide all colors in one image into seven clusters and the result is shown in Figure 3(b). But these thresholds cannot always get ideal result for different image scene. So we propose an automatic threshold selection method to get the suitable threshold for different images.

3.2 Automatic Multi-threshold Acquisition

By observing the distribution of color centroids in hexagon as Figure 3(d), we can see that the centroid distribution of different color is different, only when the $R=G=B$, the centroid is same and it is the origin of hexagon. The color information is stronger; the centroid is more far away from origin. For example, $(R, G, B) = (255, 0, 0)$ is the pure red color and its centroid is the up peak point in Figure 2. But threshold curve determination in Figure 3(d) is not easy. To display the distribution character more clearly, we transform the centroids hexagon distribution in Polar coordinate system to histogram distribution in Cartesian coordinate system as shown in Figure 3(e). In the Figure 3(e), horizontal axis is $\varphi$ ($\varphi \in (0^\circ, 360^\circ)$), vertical axis is $r$ ($r \in [0, 85]$) and other six vertical color-lines are threshold curves ($\varphi_M, \varphi_R, \varphi_Y, \varphi_G, \varphi_C$ and $\varphi_B$).

To segment goal region ideally, the thresholds must be calculated accurately. Because the histogram is not smooth, we use 1-D iterative median filter to smooth it for analysis. Through many experiments and observation, we define the adjustment range of threshold curves from left $20^\circ$ to right $20^\circ$ and find the left and right valleys respectively in this range by histogram analysis method. For $r_L$, we define the range from 3 to 15 and calculate the average value of each valley bottom (one color region only calculate one minimum value). After this process, Figure 3(c) can be got.

4. Face Detection and Tracking

4.1. Thresholding

By analyzing the characters introduced above, the thresholding results can be acquired to detect face region in color image.

4.1.1. CCS for Thresholding

The CCS can solve the problem of existing methods based on color information, because it can overcome the influences of color and luminance. The proposed method only calculates the direction of color and clusters the dark and light region into one cluster which will result in removing some noises. By analyzing the many sample face region color, we find that it always distributes in range $50^\circ$-$150^\circ$, shown in Figure 4. So we only need to calculate $\varphi_M, \varphi_R$ and $r_L$ for time saving. The pre-face region:

$$r(\varphi) = r_{\text{face}} \quad (\varphi \in [\varphi_M, \varphi_R], r \in [r_L, 85])$$

(5)

Then the method described in Section III is used to select thresholds $\varphi_M, \varphi_R$ and $r_L$ to get the threshold curves for thresholding. Other thresholds keep invariant as initial values without calculation. By this way, the binary image can be got as Figure 5(b). From the result we can see that sky, background, white scarf and red cloth region are clustered to black...
and only the face color similar regions are clustered to white. Through this processing, many noise can be removed, especially the excessive bright, dark and different color regions. But because some color of background and cloth is close to the face region color, it is clustered to pre-face region.

4.1.2. Nonlinear Thresholding for Correction

Despite the CCS thresholding can get better result, but it cannot remove some noises caused by dark color regions. To denoise them, this paper uses the nonlinear thresholding method to correct the binary image acquired by CCS. Considering the gray values of the dark color is lower, apply the nonlinear transform described as

\[ f_{\text{nonlinear}}(x, y) = \ln(1 + 255f_{\text{original}}(x, y)) \]

where, \( k \) is the number of cluster, here \( k = 2 \). From Figure 5(c), it can be seen that the background of binary image with white color and scarf region has been clustered to same value with the face region by nonlinear thresholding, so it is hard to separate the face. But this binary image can overcome some shortages of CCS, for example the dark color regions. Sometimes, the centroids of some bright regions are also in the goal region, but in fact they are noise regions. So use formula (6) to correct image processed by the CCS based method with "and operation".

\[ f_{\text{final}}(x, y) = f_{\text{CCS}}(x, y) \cap f_{\text{nonlinear}}(x, y) \]

It will result in ideal result. Figure 5(d) is the corrected binary image.

4.2. Pre-face Region Decision

Once the binary image is got, median filter is used to denoise as Figure 6(b) show. The white region is the wait-decision region, it maybe include face, hand, skin region and other close color regions (white region on Figure 6(b)). Here all wait-decision regions are analysed in a selection process and some of them are accepted by aspect ratio and size.

**Accepted by size:** Calculate the average area \( S_{\text{ave}} \) of all wait-decision regions, delete the regions whose area is larger than 50% or smaller than 0.25% average area. If the area of face \( S_{\text{face}} \in [0.8S_{\text{ave}}, 1.2S_{\text{ave}}] \), it will be accepted as Figure 6(c).

**Accepted by aspect ratio:**

\[ C = 4\pi S / L^2 \]

where \( C \) aspect ratio, \( L \) is the perimeter of boundary and \( S \) is the area of wait-decision region. If \( C \in [1, 1.7] \), it will be accepted as Figure 6(d).

4.3. Face Region Tracking

After face region is fixed, use a circle to draw it by six steps:

**Step1:** Use the binary face region as the mask to detect the face region from original image.

**Step2:** Because of the eyes and mouth region is usually darker than face skin, divide the face region to 9 sub-blocks. Then maximum entropy method is used to thresholding them respectively, shown in Figure 7.

**Step3:** Remove noise point by median filter and find all pre-eye and pre-mouth regions in the inscribed circle of face region.

**Step4:** Fix eyes and mouth region by aspect ratio and occupancy as follows:

Aspect ratio: between 0.2 and 1.7.

Occupancy: between 0.5% and 4% of face region.

**Step5:** Calculate area centroids respectively.

**Step6:** Draw a circumcircle (blue circle in Figure 7) of triangle which is created by the three centroids (red points in Figure 7). Then use concentric circle multiplied by 1.5 to mark face (green circle in Figure 7).
5. Experimental Results

All the experiments are implemented using Matlab 7.0 on a Celeron M 1.73GHz platform with 2G memory.

5.1. The Result of Proposed Method

5.1.1. Thresholding results and comparison

Figure 8 shows an example with complex background under outdoor situation, (a) is the original image; (b) is the thresholding result by proposed method; (c)–(f) are different results by traditional thresholding methods. Figure 9 shows the different thresholding results for R channel gray image of RGB color space. Figure 9, Figure 10, Figure 11, Figure 12 and Figure 13 show the different thresholding results for C channel gray image of CMYK color space, V channel of HSV color space, L channel of Lab color space and Y channel of YIQ color space. Compare those thresholding results; we can easily conclude that the proposed method is better than the others.

5.1.2. Face detection results

Figure 14 shows an indoor image which contains more than one person under situation. From Figure 14(b), we can see that it is easy to segment the goal color region, the (c) is binary image and green circles of (a) are detected result.

5.2. Comparison with reference [22]

Figure 17 shows a sample image of reference [22], (c)–(f) are the results by [22]. It is easy to see that Figure 17(e)(f) are bad thresholding results. (d) is better but include some noise. Both (b) and (c) are acceptable results and (b) is better than (c) because (c) lost a small region under left eye. So (b) is the best result in the five.

6. Conclusions and Future Works

We have presented a novel color thresholding method to segment the color image for temporal face detection and tracking. Experimental results demonstrate that the proposed method can detect and track faces under various conditions effectively. First, calculate the centroids distribution of all colors in one image. Then use the binary image acquired by nonlinear thresholding to correct the binary image acquired by CCS method. Finally, close operation and filter are used to get the ideal binary image. In addition, the face can be marked correctly following our proposed tracking method.

Experimental results show an excellent performance of the proposed method for different environments such as the change of room, lighting, complex background and color.

In the future, we need to complete the following items to improve the performance of our method further:

- Overcome the change of pose and view point.
- Use motion analysis for effective prediction.
- Integrate the detection and tracking information to make a face model for real-time recognition.
Figure 8. Sample image thresholding with different methods

Figure 9. R channel of RGB color space thresholding with different methods

Figure 10. C channel of CMKY color space thresholding with different methods

Figure 11. V channel of H color space thresholding with different methods
Figure 12. L channel of Lab color space thresholding with different methods

Figure 13. Y channel of YIQ color space thresholding with different methods

Figure 14. Detection with indoor condition

Figure 15. Detection with outdoor condition
7. References


International Conference on Automatic Face and Gesture Recognition, pp. 88-93, 1996.


Biographies

Jun Zhang received the B.E. degree in Automation and the M.E. degree in Pattern Recognition and Intelligence Systems from Xi'an University of Technology, China in 2001 and 2005 respectively, and then studies the Doctoral course in Graduate School of Information, Production and Systems, Waseda University, Japan. She was an instructor of the Department of Information and Control Technology of Xi'an Institute of Post & Telecom from 2005 to 2007. Her research interest includes image recognize, detection and analysis. She is a member of Chinese Institute of Electrons and a student member of IEEE, IEICE, IIEJ and JAMIT.

Qieshi Zhang received the B.E. degree in Automation from Xi'an University of Technology, China in 2004, and Master degree in Information, Production and Systems Engineering from Waseda University, Japan in 2009. Now studies the Doctoral course in Graduate School of Information, Production and Systems, Waseda University. Japan. He was an associate of the Department of Mechanical Electronically Technology in Xi'an Siyuan University, China, from 2004 to 2006. His research interests include image compression, detection and recognize. He is a member of Chinese Institute of Electrons and a student member of IEEE, IEICE and IIEJ.

Jinglu Hu received the M.S. degree in Electronic Engineering from Zhongshan University, China in 1986, and the Ph.D degree in Computer Science and Engineering from Kyushu Institute of Technology, Japan in 1997. From 1986 to 1993, he worked in Zhongshan University, where he was a Research Associate and the Lecturer. From 1997 to March 2003, he was a Research Associate in Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan. From April 2003 to March 2008, he was an Associate Processor in Graduate School of Information, Production and Systems, Waseda University. Since April 2008, he has been a Processor in Graduate School of Information, Production and Systems, Waseda University. Dr. Hu is a member of IEEE, SICE, IEEJ and IEICE.
Feature Based Image Classification by using Principal Component Analysis

Imran S. Bajwa¹, M. Shahid Naweed ¹, M. Nadim Asif¹, S. Irfan Hyder²

1 Department of Computer Science & IT, The Islamia University of Bahawalpur, Pakistan
2 PAF-KIET, Karachi, Pakistan.

imransbjawa@yahoo.com, shahid_naweed@hotmail.com, nasif@softresearch.org, hyder@pafkiet.edu.pk

Abstract

Classification of different types of cloud images is the primary issue used to forecast precipitation and other weather constituents. A PCA based classification system has been presented in this paper to classify the different types of single-layered and multi-layered clouds. Principal Component Analysis (PCA) provides enhanced accuracy in features based image identification and classification as compared to other techniques. PCA is a feature based classification technique that is characteristically used for image recognition. PCA is based on principal features of an image and these features discreetly represent an image. The used approach in this research uses the principal features of an image to identify different cloud image types with better accuracy. A classifier system has also been designed to exhibit this enhancement. The designed system reads features of gray-level images to create an image space. This image space is used for classification of images. In testing phase, a new cloud image is classified by comparing it with the specified image space using the PCA algorithm.

Keywords: Feature identification, Multi-layered cloud types recognition, principal components, eigenvectors, weather prediction

1. Introduction

Satellites are major source of data for forecasting weather applications. Weather forecasting applications require better processing speed and also accurate analysis of data (images). Here analysis is comprised of a number of steps: image pre-processing, image enhancement, image identification, image classification, etc. Various image processing and pattern matching and recognition techniques and methodologies are used to analyze image information. The used technologies can be differentiated into two halves: statistical and non-statistical approaches. Some statistical methodologies like FDA [4], RBFNN [1] and SVM [12] are incorporated for image analysis. In non-statistical techniques, Neural Networks is an often-used approach [3, 13] for image processing. Existing used methodologies have some issues: require more training time, exploit more processing time and have limited accuracy of about 70% [11]. This level of accuracy often degrades classification of clouds, and hence the accuracy of rain and other weather predictions is reduced [15]. We presented PCA approach for identification and analysis of single layered cloud-types [19]. PCA presented comparatively better results then previously used techniques. In this research, another variation of PCA, KPCA is also used for image identification.

1.1 PCA vs KPCA

Principal Component Analysis (PCA) [1] extracts principal features of an image. These features are integrated in a single module or class [6]. These features can principally differentiate among various input images. This technique produces results in fast and relatively more accurate manner [7].

On the other hand, KPCA [18] is Kernel Principal Component Analysis (Kernel PCA) is another variation of PCA. KPCA is a method of non-linear feature extraction, closely related to methods applied in Support Vector Machines (SVMs) [19]. PCA has ability to identify relatively fewer “features” or components that as a whole represent the full object state and hence are appropriately termed “Principal Components”. Thus, principal components extracted by PCA implicitly represent all the features. However, these abstracted features may or may not include a specific feature [5]. Better accuracy in cloud classification means accurate categorization of
clouds according to high, mid and low levels. These high, mid and low-level clouds are further classified in their particular sub classes illustrated in Section 3.3.

1.2 Principal Features
Principal features in PCA are represented by Eigenvectors. The Eigenvectors are defined to be a related set of spatial characteristics [6] of an image that a computer uses to identify and recognize a specific cloud type. Eigenvectors of the covariance matrix is computed from the training set of images. These eigenvectors represent the principal components of the training images [7]. These eigenvectors are often ortho-normal to each other. In the context of clouds classification, these eigenvectors would form the cloud space. They may not correspond directly to any cloud feature like height, width and density. Cloud Detection consists in locating a cloud in complex scenery, by locating and cutting it out. Some methods search elliptical and polygonal forms [2], others seek the texture and color of the clouds [3] and still others seek the patterns and boundaries of the cloud. When the eigenvectors of these features are displayed, they look like a ghostly cloud. They can be thought of as a set of features that together characterize the variation between cloud images.

In the following section of this paper, related work review and used methodology has been presented. The used algorithm has also been elaborated later in the section. Next section presents the experiments and the results with analysis. Last section of the paper describes the conclusion of the research work.

2 Literature Review
Cloud classification is the major research area in meteorology. Many researchers have done work in this area to classify different types of clouds. Multi-layered clouds tend to cover large areas and are indicated on a ground-based or satellite picture by an area of uniform brightness. On the other hand, single-layered clouds are usually formed by air being heated from below.

S. C. Ou and Y. Takano [13] designed a system for remote sensing of cirrus cloud parameters using advanced very-high resolution radiometer. Bankert, R. L. [2] classified clouds into one of ten output classes using a probabilistic neural network (PNN) applying on advanced very high resolution radiometer data of 16 pixel × 16 pixel sample areas. Bryan A. Baum [5] presented a fuzzy logic classification (FLC) methodology is to discriminate between clear sky and clouds in a 32 × 32 pixel array and to discriminate between single-layered and multilayered clouds within the sample. Cloud-top height was determined using ATSR data by Turner, P.J. [17]. The work estimated more accurate cloud top height from satellite images for weather forecasting and analyzing climate changes and they used Statistical Methods. This article presents a facial expressions classification experiment using neural networks.

Su Hongtao presented a classification system [1] that was based on attributes extraction from human face images using the principal component analysis (PCA) technique. Well-framed images were used in order to simplify the face detection on the image. Two different models of neural networks were applied as classifiers: back-propagation and RBF networks. For classification of single-layered cloud types another system [19] was presented that was based on PCA. The system had better accuracy in classification results. Now the same project is extended to classify multi-layered clouds as well.

3. Designed System’s Methodology
The developed classifier system discriminates the single-layered cloud types. It carries out classification in five modules: image acquisition, detection, extraction of related attributes, comparison of these attributes and finally classification. Following are some major phases of the designed system.

a. Image Acquisition - This module helps to acquire new image. The image can be acquired through different sources e.g. digital camera. Images for testing and training phases are converted to 256-bit Gray color image. Images are also scaled to 50 x 50 ratio. This ratio can vary from 30 x 30 to 50 x 50.

b. Detection of Cloud Fragments - This module detects the presence of the cloud fragments in the images. In this module it is specified that rather the images contain the clouds or not.

c. Extraction of Attributes - This module identifies the various patterns in data. Cloud attributes are extracted from images using Principal Component Analysis algorithm.

d. Image Comparison - This module compares the principal features of the test image with the imagespace of already given images in training set. After matching, it infers that rather image is recognized or not.

e. Cloud Type Classification - This module finally detects and classifies the cloud type. Images are classified to their respective types on the basis of the matching inferences provided by previous module. First image is classified as single-layered or multilayered cloud image and then afterwards the sub type is identified. Figure 2.1 represents the described architecture of the designed system.

4. Feature Extraction using PCA
The system presented in this work exemplifies the concept of Eigenvectors. These eigenvectors are a small group of characteristics extracted by the designed classifier system using PCA. PCA is a two-phase algorithm [19].

• Training Phase
• Recognition Phase
4.1. Training Phase
Training phase constructs an image-space, called a cloud space, which is later required for classification in testing phase. In training phase, the classifier system is trained by using sample data input. If it is required, output pattern can be enhanced and improvised by retraining the system by more refined and conspicuous data. Training is performed using n images in the following 6 steps:

**STEP 1**
Each sample image is converted into a row vector. A row vector can be constructed by concatenating each row with first row in sequence. As in fig-2 a \( m \times n \) matrix is converted into a single row \( 1 \times mn \) vector \( X \).

A training image of 50 x 50 pixels is taken and converted into a 1-row vector of length 2500 x 1. The procedure has been explained in figure 2.0.

**STEP 2**
The row vector matrix is constructed by combining together the row vectors of \( n \) cloud images. \( X \) is a row vector of a sample image \( i \), where \( i = 1 \ldots n \).

1-row vectors of 10 images are combined to make a 2-D array of 10 x 2500

**STEP 3**
A mean cloud vector \( \Psi \) of \( n \) row vectors is calculated to extract required principal features.

\[
\Psi = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

A mean vector (1 x 2500) is taken of 2-D array of 10 x 2500.

**STEP 4**
A new matrix \( \Phi \) is constructed by subtracting mean cloud vector \( \psi \) from each cloud image \( X \) of the training set.

\[
\Phi_i = X_i - \psi
\]

Mean vector (1 x 2500) is subtracted from matrix of step 2 to make a new 2-D array of 10 x 2500

**STEP 5**
A data covariance matrix \( C \) is calculated by multiplying matrix \( \Phi \) with its transpose matrix \( \Phi^t \).

\[
C = \Phi^t \times \Phi
\]

Covariance matrix of step 4 is calculated using Matlab 6.5

**STEP 6**
A number of highest valued eigenvectors are then picked to make an image space from the resultant covariance matrix \( C \).

In this experiment, 20 highest values of each image are taken to result a final matrix that is used for recognition

* (Step5,6 were performed using Matlab 5.6)

4.2. Recognition Phase
In testing phase, each new image is analyzed and its principal features are located. Then these principal features are compared with the principal features of image-space. If some match is found there, then the image is classified according to the previously defined rules. Recognition or testing phase is performed in the following two steps.

**STEP 1**
A new cloud image is categorized by calculating projection on image-space by

\[
\Omega = U_i \times (Z - \psi)
\]

Where \( U_i \) is image-space and \( Z \) is the new Image

**STEP 2**
If threshold \( \Omega \) matches with one of the thresholds in image space then cloud recognition occurs and the particular cloud type is specified.

\[
\Phi_i = \frac{1}{K} \max (\Omega_i - \Omega_j) \text{ where } (i, j = 1, \ldots n)
\]

If threshold of any recognition image matches with the training work space, recognition occurs

**STEP 3**
A set of rules are defined to classify the corresponding type of the matched image. Each training image has attached related information in coded form e.g. main type of cloud, sub-type of cloud, etc. This information is coded at the time of training.

The cloud image is classified into specific type according to the specified rules.

5. Experiments and Analysis

A series of experiments were done using the developed classified system to evaluate its accuracy. Experiments were performed using following steps:

1. Data Collection
2. Normalizing Cloud Images
3. Define Classes
4. Cloud Type Classification
5. Evaluate Accuracy
6. Comparison with other Technologies

Data Collection

Image data of cloud’s different types was obtained for training purposes. This data is available from different sources. Ground-based cloud images have been used in this experiment. These images of the general and sub-cloud types are available at different websites of world’s major weather forecasting organizations [13], [14], [15].

Overlapping sets of training images and testing images were used for the experiment. Global daytime cloud images are used in development and implementation aspects of a principal component analysis classification system. The designed image classifier system is used to find the presence of clouds and classification of single layer clouds in cloud images.

Define Classes

An efficient and effective image classifier system often consists of a defined set of classes. These precisely defined classes are well separated by a set of features that are typically derived from the multi-dimensional radiometric image data. The selection of classes is often influenced by desired application and classes may be complicated. In this research, there are two major classes.

1. Single-layered Clouds
2. Multi-layered Clouds

Single-layered cloud images are further classified into following sub-categories.

1. Clear sky
2. Low-level clouds
   i)- cumulus
   ii)- Stratocumulus
   iii)- Stratus
3. Mid-level clouds
   i)- Altocumulus
   ii)- Nimbostratus
   iii)- Altostratus
4. High-level clouds
   i)- Cirrus
   ii)- Cirrostratus
   iii)- Cirrocumulus

Normalizing Cloud Images

Images for testing and training phase are of 256-bit Gray color image and are overlapping. If the acquired image is not in specified bitmap format then it is converted into required format. The system obtains the image in the form of BMP of JPEG format.

Acquired Image was of size 50 x 50 pixels for processing in the designed system. But this ratio can be tuned from 30 x 30 to 50 x 50. This module gets the image in integer or short co-ordinate i.e. perform scaling at 50 x 50 scale.

Cloud Type Classification

Two types of satellite images have been used, first as training image and other images with clouds for testing. Comparing the individual pixel values within 50 x 50 array with a clear sky images depicts cloud fragments in a sample image. Often the array of 32 x 32 array is used in conventional image recognition applications. As the greater number of pixels can immensely affect the memory usage so array of smaller range is preferred. But this procedure also affects the overall image processing accuracy. But PCA handles images so conveniently that an array of greater range may be used to get still higher accuracy.

If the cloud matches with the existing collected data then the program will display as match is found. It displays cloud’s general type as low, mid or high. Program has also the capacity of prescribing the cloud’s sub type and also describing the properties of each cloud sub type.

Details of Experiments

A software system was designed to practically perform the experiments and check the affectivity of the PCA algorithm. The code of designed system was written in Visual C language. First phase converts normal 24-bit image into 8-bit gray-scale image.

![Cumulus – Normal Coloured Image](50 x 50 Gray scale (8-bit) bmp image)

Figure 3.1: Conversion to 24-bit to 8-bit gray-scale image

Following are the some examples of the single-layered cloud images that have been used for the training or recognition purpose.
Figure 3.2: 8-bit Converted images for training

First window is load training that is used for image acquisition. First of all a set of images that would be use in training are loaded in system, so that their principal features can be studied. Training can take place with out this step but this procedure make the process of training faster and causes minimum problems.

Window shown in figure 3.3 performs the preliminary steps involved in training. At this step actual training of the system is performed by extracting different principal features of the image. Then it is informed to the system that to which cloud type this image belongs and then the further sub-cloud type is also specified manually. This phase need an adequate training. The sufficiency of training would be specified by the variance of data. Higher the data variance in the sample images, higher the training would be needed.

A data file was created on the basis of first four steps of the training phase. Step 5 and step 6 were implemented in MATLAB 6.5 software. MATLAB comprehensively handles the complexity of steps 5 and 6. After receiving the processed file from MATLAB, the designed software system completes the remaining training procedure. Figure 3.4 shows working of image recognition module that performs actual matching of the images. If the match occurs then the inferences are sent to the next module for further classification. Here the co-variance matrix is generated and finally the highest valued eigenvectors are selected.

Third and final window of the designed system presents actual classification of the cloud images. This testing is performed on the basis of training performed and features specified by previous phase. This class simple compares the principal features of new testing image with the existing training image’s features. On the basis of the various existing features of the cloud images the classifier software system displays the main type and sub type of that cloud.

Evaluate Accuracy

To test the accuracy of the designed system images of different type were used. 20 Clear sky images were used for testing and all images were successfully categorized. 36 images of each single-layered and
multi-layered cloud types were used and showed results with high accuracy. A matrix of results of testing images is shown below.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Clear sky</th>
<th>Single-layered</th>
<th>Multi-layered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear sky</td>
<td>20</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Single-layered</td>
<td>0</td>
<td>33</td>
<td>3</td>
</tr>
<tr>
<td>Multi-layered</td>
<td>0</td>
<td>5</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 3.1 - Testing results of different cloud type images

A matrix representing classification accuracy test (%) for cloud free and single-layered and multi-layered cloud types is constructed. Overall classification accuracy for single-layered clouds is determined by dividing number of correctly classified samples by the total number of samples. An accuracy test (%) table is shown here.

<table>
<thead>
<tr>
<th>Classes</th>
<th>Clear sky</th>
<th>Single-layered</th>
<th>Multi-layered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clear sky</td>
<td>100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Single-layered</td>
<td>0</td>
<td>91.6</td>
<td>2</td>
</tr>
<tr>
<td>Multi-layered</td>
<td>0</td>
<td>4</td>
<td>86.11</td>
</tr>
</tbody>
</table>

Table 3.2 - Average Accuracy = 88.85%

Comparison with other Techniques
Various classification techniques and algorithms are used for image classification. Every technique provide with respective accuracy level. The derived results using principal component analysis are compared with the results of other technologies used for cloud classification. Different technologies provide with different level accuracies. Results show that PCA, relative to other statistical techniques, is more accurate [table. 3]. Other statistical techniques include Fuzzy Logic based systems that give 84% accuracy [5] but Fuzzy systems itself are dependent on the appropriateness of the initial categories defined i.e. much effort is needed for domain knowledge and efficiency issues. Neural networks demand intense domain knowledge and intuition for representation otherwise suffer from divergent training sessions and inaccurate results [11].

Comparison with other Techniques

<table>
<thead>
<tr>
<th>Technology Name</th>
<th>Accuracy Per.</th>
<th>Error Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA (Principal Component Analysis)</td>
<td>88.85%</td>
<td>1.2%</td>
</tr>
<tr>
<td>NMF (Non-Negative Matrix Factoriz.) [8]</td>
<td>69.94%</td>
<td>8.1%</td>
</tr>
<tr>
<td>BPNN (Back Propagation NN) [9]</td>
<td>71.80%</td>
<td></td>
</tr>
<tr>
<td>RBFNN (Radial Basis Function NN) [1]</td>
<td>73.20%</td>
<td>7.3%</td>
</tr>
<tr>
<td>SVM (Super Vector Machine) [12]</td>
<td>84.11%</td>
<td></td>
</tr>
<tr>
<td>Fisher Discriminant Analysis [4]</td>
<td>64.00%</td>
<td></td>
</tr>
<tr>
<td>FLNN (Fuzzy Logic Neural Networks) [5]</td>
<td>81.00%</td>
<td>3.4%</td>
</tr>
<tr>
<td>Wavelet Transforms [6]</td>
<td>78.30%</td>
<td>3.9%</td>
</tr>
<tr>
<td>Probabilistic Neural Networks [2]</td>
<td>86.01%</td>
<td></td>
</tr>
<tr>
<td>K-SOM (K-Self Organizing Maps) NN [7]</td>
<td>80.00%</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.3 - Accuracy comparison in different techniques

PCA is the image classification technique, which provides higher accuracy up to 90%. Statistics show that PCA based image classifier system is a better classifier than other used techniques. There is a comparison of PCA with other techniques is also given below.

6. Conclusion
PCA is an efficient identifier in terms of time and provides better accuracy in cloud image recognition. A PCA-based system provides high speed processing with relatively better accuracy. PCA also easily handles a large amount of data due to its capability of reducing data dimensionality and complexity. PCA algorithm provides a more accurate cloud classification that infers better and concise forecasting of rain. Probably, the more long-term weather forecasting is also possible.

References


[16] http://www.nottingham.ac.uk/meteosat/


Author’s Biography

Imran Sarwar Bajwa is Assistant Professor in the Islamia University of Bahawalpur. He has been teaching and doing research in various universities in Pakistan and University of Coimbra, Portugal.

Prof. Dr. Irfan Hyder is president of PAF-KIET University in Pakistan. He has done his PhD from UT-Austin.
IMAGE ENCRYPTION BASED ON STRANGE ATTRACTOR

Jiri Giesl, Karel Vlcek
Department of Applied Informatics, Tomas Bata University, Nad Stranemi 4511, Zlin 760 05, Czech Republic
jgiesl@fai.utb.cz
http://www.fai.utb.cz

Abstract
Chaotic systems are extremely sensitive to initial conditions and this feature can be very helpful in the field of cryptography. Various encryption schemes use chaotic systems for encryption key generation and this key is then used for pixel permutation and pixel diffusion. But chaotic systems and their maps can be used directly for encryption purpose. Proposed scheme is based on extended map of Clifford strange attractor, where each dimension has specific role in the encryption process. Two dimensions are used for pixel permutation and the third dimension is used for pixel diffusion. The theoretical and simulation results prove many properties of this scheme such as large key space and high security.

Keywords: image encryption, chaos encryption, strange attractors, chaotic maps, image security analysis

1. Introduction
Image is a multimedia signal providing the most information to the man. For this reason the question appears which way can be signal secured against unauthorized reading e.g. in the medicine or military fields. Position permutation and diffusion of the pixels belongs to basic methods of image encryption. Their combination leads to better security against known attacks and is very often used for practical usage. However, these methods remains openness for various encryption algorithms and that is why the knowledge of chaotic systems can be used. These systems are extremely sensitive to initial conditions and so they are suitable candidates in the field of cryptography. Many papers were written on this theme for that reason.

Chaos-based image encryption is discussed in detail in [1]. Most of the papers have described process of generation of the time series based on a chaotic map. These series are used for creation of the binary sequence as an encryption key and pixel of plain image are then rearranged and XOR operated with this key. For example in [2] three logistic maps are used in key stream generator and this improved the linear complexity of key stream. Each paper proposed various type of key generator or improvements of chaotic encryptions in terms of security and speed [3,12,13] but only a few of them show different way of encryption, such as using hyper-chaotic system for confusing the relationship between the plain-image and the cipher-image [5], Lorentz system for key-stream generation [6] or S-box algebraic operations [10,11]. When other methods such as image encryption in wavelet domain proposed in [9] are used with chaos-based encryption scheme, we should expect interesting results. The results for the audio signals are presented in our previous research work in [15], where the wavelet coefficients were modified and the audio signal becomes inaudible.

Proposed scheme uses formula of strange attractor for encryption purposes. It does not create any encryption key but uses attractor map for pixel permutation and diffusion directly. Parameters of attractor map play the role of encryption keys here and key-space is very large due to their non-integer character.

The remainder of this paper is organized as follows. Section 2 defines strange attractor and its main properties. Section 3 presents the proposed encryption algorithm. Some security analyses are given in Section 4, performance is evaluated in Section 5 and finally Section 6 concludes this paper.

2. Strange attractors
Strange attractors are complicated sets with a fractal structure to which chaotic dynamical systems evolve after a long enough time. These attractors can be generated in several ways – the most commonly used are quadratic map (1) and trigonometric map (2) where parameters \( a, b, c, d, e, f, g, h, i, j, k, l \) define each strange attractor.

\[
x_{n+1} = a + b \cdot x_n + c \cdot x_n^2 + d \cdot x_n y_n + e \cdot y_n + f \cdot y_n^2
\]

\[
y_{n+1} = g + h \cdot x_n + i \cdot x_n^2 + j \cdot x_n y_n + k \cdot y_n + l \cdot y_n^2
\]

(1)

\[
x_{n+1} = \sin(b \cdot y_n) + c \cdot \cos(d \cdot x_n)
\]

\[
y_{n+1} = e \cdot \sin(f \cdot y_n) + g \cdot \cos(h \cdot x_n)
\]

(2)
Strange attractor can reveal after several iterations of map (1) or (2). When the strange attractor is represented geometrically it is obvious that fixed points are locally unstable but the system is globally stable.

Attractor is chaotic when Lyapunov exponent for that map is positive. Two dimensional chaotic maps have not only a single Lyapunov exponent. It has a positive one corresponding to the direction of expansion, and a negative one corresponding to the direction of contraction. The signature of chaos is that at least one of these exponents is positive and the magnitude of the negative exponent has to be greater than the positive one [16].

For a map $x_{n+1} = f(x)$ a small deviation $\delta x_0$ of coordinate $x_0$ leads to a small change in $x_1$.

$$\delta x_1 = \delta x_0 \cdot f'(x_0)$$ (3)

For $n$ iterations

$$\delta x_n = \delta x_0 \cdot \prod_{i=0}^{n-1} f'(x_i)$$ (4)

Then the Lyapunov exponent is determined as

$$\Lambda = \lim_{n \to \infty} L_n$$ (5)

where

$$L_n = \frac{1}{n} \log \left| \frac{\delta x_n}{\delta x_0} \right| = \frac{1}{n} \sum_{i=0}^{n-1} \log |f'(x_i)|$$ (6)

Deviation $|\delta x_n|$ grows with increasing of $n$ for a chaotic orbit and this leads to positive Lyapunov exponent $\Lambda > 0$.

Strange (chaotic) attractors are associated with motion which is unpredictable. If we attempt to predict motion of a chaotic system then even the small deviation in the initial conditions will be amplified exponentially over time and will rapidly destroy the accuracy of our prediction. Eventually, all we will be able to say is that the motion lies somewhere on the chaotic attractor in phase-space, but exactly where it lies on the attractor at given time will be unknown to us. These properties of chaotic system, extreme sensitivity to initial conditions and unpredictability, can be very helpful for the encryption purposes.

Strange attractors themselves are markedly patterned, often having elegant, fixed geometric structures, despite the fact that the trajectories moving within them appear unpredictable. The strange attractor's geometric shape is the order underlying the apparent chaos.

### 3. Image encryption scheme

In proposed encryption scheme the Clifford attractor is used. It belongs to trigonometric strange attractors and is described by formula (7). This formula is very similar to formula (2) but more reduced regarding number of parameters and their utilization. Figure 1 shows example of its geometrical representation with predefined parameters.

$$x_{n+1} = \sin(a \cdot y_n) + c \cdot \cos(a \cdot x_n)$$

$$y_{n+1} = \sin(b \cdot y_n) + d \cdot \cos(b \cdot x_n)$$ (7)

Figure 1. Example of Clifford attractor

Formula (7) can be used for pixel permutation easily. Coordinates of pixels can be put in formula (7) and new coordinates will appear after a few iterations. But encryption with pure position permutation only is not sufficient in terms of security. It can be decrypted easily e.g. by system of fuzzy ergodic matrices [4]. Because modern encryption schemes use also modification of pixel values, one more axis must be added to formula (7). This 3D extension creates chaotic map (8) which can be used for encryption of coordinates and value of each pixel.

$$x_{n+1} = \sin(a \cdot y_n) + c \cdot \cos(a \cdot x_n)$$

$$y_{n+1} = \sin(b \cdot y_n) + d \cdot \cos(b \cdot x_n)$$

$$z_{n+1} = \sin(e \cdot z_n) + f \cdot \cos(e \cdot y_n)$$ (8)

When coordinates and values of pixels are put in (8), new coordinates and values will be gained after several iterations and quantization. Encryption scheme is resistant against reverse iteration of chaotic system due to that quantization.

Define $P(x_0, y_0)$ as the pixel value at coordinates $(x_0, y_0)$ and $x_k, y_k, z_k$ as values of chaotic map in $k$-th iteration. These values were created when
\{x_0, y_0, P(x_0, y_0)\} were putted into (8). Original pixel at coordinates \((x_0, y_0)\) is then swapped with the pixel at coordinates \((x_k, y_k)\). Also pixel value at coordinates \((x_k, y_k)\) is modified by XOR operation of value \(z_k\) according to formula (9) and then XOR operated with the value of the previous processed pixel \(Q\).

This process must be done for every pixel in image and can be done \(m\) times.

\[ P(x_0, y_0) \leftrightarrow P(x_k, y_k) \oplus z_k \oplus Q \]  
\(9\)

Figure 2 shows flowchart of this encryption scheme. Main parameters (key) are used for iterative process of Clifford system. Pixel of image is used as initial value of Clifford system. New positions and modification value is gained after \(k\) iterations and quantization. These positions are then used for pixel permutation and the modification value is XOR operated with original pixel value and value of previous pixel. Encrypted pixel is gained this way.

4. Security Analysis

Every image encryption scheme should match the security requirements mentioned in [7,8]. This part analyses proposed encryption scheme in terms of security.

4.1. Distribution of pixels

The encryption scheme mentioned above was experimentally tested on “Lena” and “Cameraman” images of 256x256 pixels size. Figure 3 and Figure 5 show original images and histograms of their pixel values distribution. Figure 4 and Figure 6 show images after encryption process and their pixel values distribution. Encryption process returns noisy images. Their histograms are very close to uniform distribution, significantly different from that of the original image and contain no statistical resemblance to the original image. This aspect makes the encryption scheme resistant against the known-plaintext attack.

4.2. Information Entropy

Illegibility and indeterminateness are the main goals of image encryption. This indeterminateness can be reflected by one of the most commonly used information-theoretical measure – the entropy.

The entropy \(H\) of a symbol source \(S\) can be calculated as:

\[ H(S) = \sum_{i=1}^{N} P(s_i) \cdot \log \frac{1}{P(s_i)} \]  
\(10\)

where \(P(s_i)\) represents the probability of symbol \(s_i\) and \(\log\) represents the base 2 logarithm.

Entropy of an image is maximal when all values of pixels are distributed equally. Requested effect of encryption process is to get maximal entropy. Table 1 shows entropies of plain images and their encrypted forms which are very close to maximum entropy. This means a high permutation and diffusion is achieved after encryption process. Encryption system has a robust performance varying the different complexity of the original image and therefore this system is secured upon the entropy attack.

<table>
<thead>
<tr>
<th>Image</th>
<th>Entropy of plain image</th>
<th>Entropy of encrypted image</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grey scale</td>
<td>8</td>
<td>7,996</td>
</tr>
<tr>
<td>Black color</td>
<td>0</td>
<td>7,996</td>
</tr>
<tr>
<td>Lena</td>
<td>7,201</td>
<td>7,997</td>
</tr>
<tr>
<td>Cameraman</td>
<td>7,117</td>
<td>7,996</td>
</tr>
</tbody>
</table>

4.3. Cross-correlation of images and adjacent pixels

Cross-correlation is a standard method of estimating the degree to which two series are correlated. Consider two series \(x_i\) and \(y_i\) where \(i = 1, 2, \ldots, N\) and \(E(x)\), \(E(y)\) are the means of the corresponding series according to (11).

\[ E(x) = \frac{1}{N} \cdot \sum_{i=1}^{N} x_i \]  
\(11\)
Figure 3. (a) original “Lena” image, (b) histogram of original “Lena” image

Figure 4. (a) encrypted “Lena” image, (b) histogram of encrypted “Lena” image

Figure 5. (a) original “Cameraman” image, (b) histogram of original “Cameraman” image

Figure 6. (a) encrypted “Cameraman” image, (b) histogram of encrypted “Cameraman” image
Table 2: Correlation coefficients of original/encrypted image

<table>
<thead>
<tr>
<th>Direction of adjacent pixels</th>
<th>Original image</th>
<th>Encrypted image by the proposed scheme</th>
<th>Encrypted image by the [1] scheme</th>
<th>Encrypted image by the [5] scheme</th>
<th>Encrypted image by the [14] scheme</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal</td>
<td>0.942755</td>
<td>0.001522</td>
<td>0.005776</td>
<td>-0.014200</td>
<td>0.030800</td>
</tr>
<tr>
<td>Vertical</td>
<td>0.970970</td>
<td>0.001866</td>
<td>0.028434</td>
<td>-0.007400</td>
<td>0.030400</td>
</tr>
<tr>
<td>Diagonal</td>
<td>0.920081</td>
<td>-0.003903</td>
<td>0.020662</td>
<td>-0.018300</td>
<td>0.031700</td>
</tr>
</tbody>
</table>

The cross correlation $r$ at delay $d$ is defined as:

$$r(d) = \frac{\sum_{i}(x_i - E(x)) \cdot (y_{i-d} - E(y))}{\sqrt{\sum_{i}(x_i - E(x))^2} \cdot \sqrt{\sum_{i}(y_{i-d} - E(y))^2}}$$  \hspace{1cm} (12)

Cross-correlation can be used as a measure of similarity of two images that differ only by a shift. Cross-correlation was computed from delay $d = 0$ to $d = 255$. The denominator in the expression (12) normalizes the correlation coefficients such that $-1 \leq r(d) \leq 1$. This bound indicates maximum correlation and $r(d) = 0$ indicates no correlation. A high negative correlation indicates a high correlation in the case of inversion of one of the series. [17]

Figure 7 shows cross-correlation of original and encrypted image. It is obvious that correlation value does not exceed 0.008. That means very low correlation and very low similarity of images and their pixels.

![Cross correlation](image1)

Figure 7. Cross-correlation of original image and encrypted image

In general, adjacent pixels of the most plain-images are highly correlated. One of the requirements of an effective image encryption process is to generate encrypted image with low correlation of adjacent pixels. Correlation between two horizontally, vertically and diagonally adjacent pixels of original and encrypted image was analyzed. Each pair of adjacent pixels of original image was chosen and correlation coefficient was computed according (12). Then the same process was done for the encrypted image. Correlation coefficients in different directions of adjacent pixels are listed in Table 2.

As can be seen, every encryption scheme can effectively de-correlate adjacent pixels in the original image. The correlation coefficients are the smallest when the proposed encryption scheme is used. Thus, proposed scheme has the best permutation and diffusion properties.

4.4. Key sensitivity

Two set of keys (parameters) of chaotic attractor (8) are defined for the encryption process to test key sensitivity:

$$\{a = -1.85, b = 1.48, c = -1.55, d = -1.87, e = -4.32, f = 0.63\} \hspace{1cm} (13)$$

$$\{a = -1.85, b = 1.4800001, c = -1.55, d = -1.87, e = -4.32, f = 0.63\} \hspace{1cm} (14)$$

Set of keys (13) and (14) are very similar, only one parameter is different with minimal divergence.

First test of key sensitivity is based on encryption of “Lena” image by keys (13) and encryption of “Lena” image by keys (14). Cross-correlation of their encrypted forms was then computed. Figure 8 shows cross-correlation of images encrypted by these two different set of keys.

![Cross correlation](image2)

Figure 8. Cross-correlation of two encrypted images

As a result, the image encrypted by keys (13) has 99.61% difference from the image encrypted by keys (14) in terms of pixel gray-scale values.

Second test of key sensitivity is based on the encryption by keys (13) and decryption by keys (14). Figure 9 shows illegible image and uniform-like distribution of pixels in the case of incorrect decryption. Image can not be reconstructed to approximate form even when using keys with minimal divergence.

![Cross correlation](image3)

Figure 9. Illegible image and uniform-like distribution of pixels
4.5. Plain-image sensitivity

In differential attacks, some plain-images (plaintext) with little difference may be selected to analyze the according difference between their encrypted forms. In this way, an opponent may be able to find out a relationship between plain-image and its encrypted form. Thus, much difference between encrypted forms is expected in order to keep high security. Plain-image sensitivity is the most analyzed by two measures: NPCR and UACI.

Number of pixel change rate (NPCR) means the number of pixels changed in the encrypted form when only one pixel value is changed in plain-image. The larger the NPCR is, the higher sensitivity the plain-image has and the more difficult the system’s security against differential attack is.

Let two encrypted images of the same size be $A$ and $B$, where $B$ has only one pixel difference. NPCR is then defined as:

$$\text{NPCR} = \frac{\sum_{i,j} D(i,j)}{W \times H} \times 100\%$$  \hspace{1cm} (15)$$

where $W$ and $H$ are the width and the height of images and $D(i,j)$ is bipolar array defined as:

$$D(i,j) = \begin{cases} 0, & A(i,j) = B(i,j) \\ 1, & A(i,j) \neq B(i,j) \end{cases}$$ \hspace{1cm} (16)$$

Expression $A(i,j)$ or $B(i,j)$ means the pixel in the $i$-th row and $j$-th column of that image.

Unified average changing intensity (UACI) means changing intensity of the corresponding pixels of the plain-image and encrypted image. The larger the UACI is, the more resistant to the differential attack the encryption scheme is.

The UACI is defined by:

$$\text{UACI} = \frac{1}{W \times H} \sum_{i,j} \left[ \frac{|A(i,j) - B(i,j)|}{255} \right] \times 100\%$$ \hspace{1cm} (17)$$

In our plain-image sensitivity test we changed gray-scale value of pixel at position $(2,3)$ from 159 to 160. Figure 10 and Figure 11 show that during the fifth encryption round the system has already reached the satisfying encryption performances and kept stable at all rounds. Table 3 shows the real values of NCPR and UACI in each encryption round.
Table 3: NPCR and UACI

<table>
<thead>
<tr>
<th>Encryption rounds m</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>25</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>NPCR (%)</td>
<td>40.87</td>
<td>78.12</td>
<td>91.85</td>
<td>96.81</td>
<td>99.29</td>
<td>99.51</td>
<td>99.58</td>
<td>99.57</td>
<td>99.55</td>
<td>99.61</td>
<td>99.66</td>
<td></td>
</tr>
<tr>
<td>UACI (%)</td>
<td>4.28</td>
<td>22.69</td>
<td>29.60</td>
<td>32.06</td>
<td>33.16</td>
<td>33.23</td>
<td>33.44</td>
<td>33.53</td>
<td>33.41</td>
<td>33.57</td>
<td>33.49</td>
<td></td>
</tr>
</tbody>
</table>

It is obvious that this encryption scheme is resistant against differential attacks after fifth encryption round. Generally, with the increase of encryption rounds, the influence of one-pixel change is increased. Hence, it is reasonable to increase encryption rounds to achieve higher security but at the expense of processing speed.

4.6. Key space

Good encryption scheme must be resistant against any brute-force attacks, so the key space must be too large. Total precision of common PC processor is 16 decimal digits therefore number of different combinations of one parameter is \(10^{16}\) and it corresponds approximately to \(2^{53}\) size key space. Six attractor parameters are used in the proposed scheme; hence the key space is enlarged to \(2^{318}\). Also number of iterations \(k\) of Clifford system (8) and number of encryption rounds \(m\) can be considered as keys. Thus, key space of proposed scheme is large enough to make the brute-force attack infeasible. Table 4 shows comparison of key spaces of various encryption schemes. Proposed encryption scheme has the largest key space.

Table 4: Key space comparison

<table>
<thead>
<tr>
<th>Encryption scheme</th>
<th>Key space</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>(2^{318})</td>
</tr>
<tr>
<td>[1]</td>
<td>(2^{128})</td>
</tr>
<tr>
<td>[2]</td>
<td>(2^{138})</td>
</tr>
<tr>
<td>[5]</td>
<td>(2^{232})</td>
</tr>
<tr>
<td>[15]</td>
<td>(2^{256})</td>
</tr>
</tbody>
</table>

5. Performance evaluation

The above-described image encryption algorithm is not so effective in terms of speed and this property is the main weakness of this algorithm because of its constraints in real-time applications. Encryption speed is the most influenced by number of iterations \(k\) of Clifford system and number of encryption rounds \(m\). Security of encrypted image is directly proportional to height of parameter values and to time needed for encryption. Table 5 shows average elapsed time in seconds with different encryption conditions at 8-bit image of \(256 \times 256\) pixels size. The computer used in this test is an AMD Athlon, 1.50-GHz CPU with 512-MB memory and 80-GB hard disk capacity.

Simulation shows that the peak speed can reach up to 820kB/s on 1.50-GHz AMD Athlon computer. Parameters \(k\) and \(m\) should be set to 5 for the sufficient security. Speed is then much lower only about 30kB/s. Thus, this encryption scheme is not suitable for the real-time applications but can be used for archiving purposes.

6. Conclusion

Proposed encryption scheme uses Clifford strange attractor and its maps for image encryption purposes. This scheme does not create any encryption key but uses attractor map and its parameters for pixel permutation and diffusion directly. Coordinates and value of each pixel becomes initial condition for appropriate dimension of attractor map. New coordinates and value of pixel is created after several iterations. Encrypted image with proper security properties will be gained this way. Some security analyses are presented to demonstrate that this encryption scheme has very good permutation and diffusion properties and is sufficient for practical usage in terms of security. But this high security is achieved at the expense of processing speed. Thus, proposed scheme can be used for archiving purposes well.

7. References


Biography

Jiri Giesl received his Masters Degree in Information Technologies from Tomas Bata University in Zlin, Czech Republic in 2007. He is PhD candidate in the Department of Applied Informatics currently. His areas of interest are Digital signal processing, Encoding and Compression, Wavelet theory, Chaos & Fractals and their application for signal analysis and synthesis.

Karel Vlcek received his Masters Degree from University of Technology in Brno, Czech Republic in 1971, Doctoral Thesis (1989) and Assoc. Prof. Defence (1993) from Czech Technical University in Prague, and Full Professorship from VSB Technical University of Ostrava, Czech Republic in 2003. His professional carrier starts in the company TESLA Roznov (1973), as a development specialist in the group of “Testing of integrated circuits”, and continues in the company TESLA Valasske Mezirici (1982), in the working group of “Application of Microprocessors”. His research activities at the universities are Digital Signal Processing, Theory of Information and Coding, Design of Electronic Circuits with support VHDL, Telecommunications, and Biomedical Engineering. Currently he is at the Department of Applied Informatics of Faculty of Applied Informatics, Tomas Bata University in Zlin, Czech Republic.
Content Based Medical Image Retrieval: Theory, Gaps and Future Directions

Preeti Aggarwal¹, H.K. Sardana², Gagandeep Jindal³
¹Dept. of Computer Sci. & Engg, UIET, Panjab University
²CSIO, Sector 20, Chandigarh, ³Landran Engg. College, Mohali
pree_agg2002@yahoo.com¹, hk_sardana@hotmail.com², gagandeep_pec@yahoo.com³

Abstract
Content-based image retrieval (CBIR) has been one of the most vivid research areas in the field of computer vision, and substantial progress has been made over the last years. CBIR has a potential for making a strong impact in diagnostics, research, and education. CBIR is a promising technology to enrich the core functionality of picture archiving and communication systems (PACS). Even in modern PACS that are based on the Digital Imaging and Communications in Medicine (DICOM) standard, image data is addressed by alphanumeric indexes such as patient name and examination date. The main purpose of this paper is to disseminate the knowledge of the CBIR approach to the applications of medical image retrieval and to attract greater interest from various research communities to rapidly advance research in this field. The semantic gap divides the high-level scene analysis of humans from the low-level pixel analysis of computers. In this paper, we suggest a more systematic and comprehensive view on the concept of gaps in Content based medical image retrieval (CBMIR) research. Also, several research directions for improving the retrieval quality based on the experiences from other closely related research fields are given in the paper. Possible clinical benefits from the use of content-based access methods are described as well as promising fields of applications.

Keywords: Content based medical image retrieval (CBMIR), National Health and Nutrition Examination Survey (NHANES), National Library of Medicine (NLM), Pathology bearing region (PBR), Picture archiving and communication systems (PACS), Region of interest (ROI).

1. Introduction
Content-based image retrieval (CBIR), a technique that uses visual contents to search images from large-scale image databases according to users' interests, has been an active and fast advancing research area since the 1990s [1]. During the past decade, remarkable progress has been made in both theoretical research and system development. Figure 1 shows the general model used in CBIR.

Figure 1: Diagram of content based image retrieval system

However, many challenging research problems continue to attract researchers from multiple disciplines. Before introducing the fundamental theory of content-based retrieval, we will take a brief look at its development. The application potential of image database management techniques has attracted the attention of researchers [2, 3, 4, 5]. Early techniques were not generally based on visual features but on the textual annotation of images. In other words, images were first annotated with text and then searched using a text-based approach from traditional database management systems. Comprehensive surveys of early text-based image retrieval methods can be found in [6, 7]. Text-based image retrieval uses traditional database techniques to manage images.

However, since automatically generating descriptive texts for a wide spectrum of images is not feasible, most text-based image retrieval systems require manual annotation of images. Obviously, annotating images manually is a cumbersome and expensive task for large image databases, and is often subjective,
context-sensitive and incomplete. As a result, it is difficult for the traditional text-based methods to support a variety of task-dependent queries.

A. Content based medical image retrieval

In the medical field, images, and especially digital images, are produced in ever increasing quantities and used for diagnostics and therapy. The Radiology Department of the University Hospital of Geneva alone produced more than 12,000 images a day in 2002. The cardiology is currently the second largest producer of digital images, especially with videos of cardiac catheterization (~1800 exams per year containing 1800 images each).

With DICOM [8, 57] a standard for image communication has been set and patient information can be stored with the actual image(s), although still a few problems prevail with respect to the standardization. In several articles, content-based access to medical images for supporting clinical decision-making has been proposed that would ease the management of clinical data and scenarios for the integration of content-based access methods into Picture Archiving and Communication Systems (PACS) have been created.

B. Challenges in Medical Image Retrieval

Before the emergence of content-based retrieval, medical images were annotated with text, allowing the images to be accessed by text-based searching. Through textual description, medical images can be managed based on the classification of imaging modalities, regions, and orientation. This hierarchical structure allows users to easily navigate and browse the database. Searching is mainly carried out through standard Boolean queries. However, with the emergence of massive image databases, the traditional text based search suffers from the following limitations:

- Manual annotations require too much time and are expensive to implement. As the number of images in a database grows, the difficulty in finding desired images increases. Muller, Michous, Bandon, and Geissbuhler (2004a) reported that the University Hospital of Geneva produced approximately 12,000 medical images per day. It is not feasible to manually annotate all attributes of the image content for this number of images.
- Manual annotations fail to deal with the discrepancy of subjective perception. The phrase, “an image says more than a thousand words,” implies that the textual description is not sufficient for depicting subjective perception. Typically, a medical image usually contains several objects, which convey specific information. Nevertheless, different interpretations for a pathological area can be made by different radiologists. To capture all knowledge, concepts, thoughts, and feelings for the content of any images is almost impossible.
- The contents of medical images are difficult to be concretely described in words. For example, irregular organic shapes cannot easily be expressed in textual form, but people may expect to search for images with similar contents based on the examples they provide
  - Low resolution and strong noise are two common characteristics in most medical images see Figure 2. With these characteristics, medical images cannot be precisely segmented and extracted for the visual content of their features. In addition, medical images obtained from different scanning devices may display different features, though some approaches to image correction and normalization have been proposed (Buhler, Just, Will, Kotzerke, & van den Hoff, 2004);
  - DICOM images are available in various modalities even for a same type of image like X-rays, MRI, CT and many i.e. if one need image of spine then we can have X-ray as well as MRI image for spine. Then for the researchers it is very difficult which one to choose.

The remainder of the paper is organized as follows: Section (2) focuses on existing CBIR systems, Section (3) role of shape in CBMIR, section (4) segmentation methods of medical images, section (5) comparison techniques used, section (6) other retrieval techniques used in CBMIR, section (7) built-in tools in CBMIR, section (8) various gaps in CBMIR, section (9) applications of CBMIR in various departments and section (10) conclusion and future work.

2. Existing Content-based image retrieval systems

Although early systems existed already in the beginning of the 1980s [9], the majority would recall systems such as IBM’s QBIC (Query by Image Content) as the start of content-based image retrieval [10, 11]. Another commercial system for image [12]
and video [13] retrieval is Virage2 that has well known commercial customers such as CNN.

Using higher level information, such as segmented parts of the image for queries, was introduced by the Blobworld4 system [14, 15]. PicHunter [16] on the other hand is an image browser that helps the user to find an exact image in the database by showing to the user images on screen that maximizes the information gain in each feedback step. A system that is available free of charge is the GNU Image Finding Tool (GIFT) [17, 18]. Most of these systems have a very similar architecture for browsing and archiving/indexing images comprising tools for the extraction of visual features, for the storage and efficient retrieval of these features, for distance measurements or similarity calculation and a type of Graphical User Interface (GUI). This general system setup is shown in Figure 3.

In contrast, the Spine Pathology and Image Retrieval System (SPIRS) [22,23,24] at the U.S. National Library of Medicine provides localized vertebral shape-based CBIR methods for pathologically sensitive retrieval of digitized spine x-rays and associated person metadata that come from the Second U.S. National Health and Nutrition Examination Survey (NHANES II). In the SPIRS system, the images in the collection must be homogeneous, i.e., a single type imaging the same anatomy in the same view, e.g., vertebral pathology expressed in spine x-ray images in the sagittal plane.

MIMS[25] first takes into account the complexity observed when describing semantic content of images and, second, the graphical aspect of certain objects. This existence creates critical problems of subjectivity, although such an approach is made as general as possible.

ImageMap [26] is so far the only existing medical image retrieval that considers how to handle multiple organs of interest. However, it works based on spatial similarity. Consequently, a problem caused by user subjectivity is likely to occur, and, therefore, the retrieved image will represent an unexpected organ.

ASSERT[27] (Automatic Search and Selection Engine with Retrieval Tools) A physician-in-the-Loop content-based retrieval system for HRCT image databases which is implemented to show a human-in-the-loop (a physician-in-the-loop, more specifically) approach in which the human delineates the pathology bearing regions (PBR) and a set of anatomical landmarks in the image when the image is entered into the database.

WebMIRS[28] In the WebMIRS system, the user manipulates GUI tools to create a query such as, “Search for all records for people over the age of 60 who reported chronic back pain. Return the age, race, sex, and age at pain onset for these people.” In response, the system returns a display of values for these four fields for all matching records, plus a

Figure 3: Principal component of all CBIR systems

A. Existing Content based medical image retrieval systems

Generally all the CBMIR have the following principle components Figure 4. In current picture archiving and communication systems (PACS), retrieval of image information is done using limited text keywords in special fields in the image header (e.g., patient identifier).

Since these keywords do not capture the richness of features depicted in the image itself, content-based image retrieval (CBIR) has received significant attention in the literature as a promising technique to facilitate improved image management in PACS systems [19,20].

The Image Retrieval for Medical Applications (IRMA) Project1 undertaken at the Aachen University of Technology (RWTH) [20,21] aims to provide visually rich image management through CBIR techniques applied to medical images using intensity distribution and texture measures taken globally over the entire image. This approach permits queries on a heterogeneous image collection and helps identify images that are similar with respect to global features, e.g. All chest x-rays in the AP (anterior-posterior) view. The IRMA system lacks the ability for finding particular pathology that may be localized in particular regions within the image.

Figure 4: General structure of CBMIR system
display of the associated x-ray images. SPIRS and WebMIRS are almost the same systems as both are developed for C-spine and L-spine image retrieval by Communication Engineering Branch (CEB) and National Health and Nutrition Examination Survey (NHANES). Various other CBMIR systems are available like FC, CasImage and many more but all have their advantages and disadvantages. So there is a need of absolute error free, efficient and automatic CBMIR system which can really helpful in medical stream.

3. Role of shape feature in CBMIR

Medical images have many features which can have used to make feature vector for CBMIR like color, texture and shape. But the most important and valuable feature is shape as most of the medical images gray scale images so color plays no role. Texture can also be considered but generally due to noise in medical images exact texture cannot be evaluated. So it is observed that only shape features appear promising for indexing the images, since the images are gray scale and offer very little in terms of texture for the anatomy of interest. The many shape representation and similarity techniques found in the literature. These techniques may be grouped under the following categories:

- **Shape geometry based methods.** These use shape properties (Ang et al., 1995) such as area, perimeter, convexity, elongation, orientation, etc.
- **Invariant moments.** Several forms of invariant moments are seen in the literature such as Hu invariant moments (Hu, 1962), generalized complex moments (Kim and Kim, 1997), affine moments, and Zernike moments (Eakins et al., 2001; Ip et al., 1997). Multi-stage modification using invariant moments has yielded very good results (Jain and Vailaya, 1998).
- **Polygon approximation methods.** Methods that remove small variations and less significant features and then represent the curve in tangent space (Arkin et al., 1991; Latecki and Laka’mper, 2001, 2002). Matching is done using the turn angle function.
- **Deformable shape based methods.** Methods that employ elastic deformation of templates (Jain et al., 1996). These are artificial, closed contours/surfaces able to expand or contract over time, within an image, and conform to specific image features. Active contours and active surfaces are included in this category. Energy-minimization models [29] and front-propagation methods based on level-sets [30] are common types. Propagation is toward a local optimum.
- **Appearance Models:** An active appearance model (AAM) is an extension of an ASM where shape plus intensity of an object, referred to as an image patch, are integrated into a statistical model [31]. A brief survey describing a number of AAM variants can be found in [32].
- **Fourier transform based methods.** Representing the cumulative shape boundary as a function of its normalized length (Zahn and Roskie, 1972). Shapes or contour points have also been described in the frequency domain (Gonzalez and Woods, 2002; Sonka et al., 1999).

These shape representation methods exhibit and retain different shape characteristics. Determining the suitability of an algorithm for a CBMIR system application can only be done after an evaluation of the shape methods on the particular shapes that populate the database. As mentioned in section 2.1, SPIRS had used deformable models and appearance models for the segmentation and extraction of pathology bearing region (PBR), which was vertebra in that case.

4. Segmentation methods of medical images

Segmentation is the process of identifying and classifying data found in a digitally sampled representation. Typically the sampled representation is an image acquired from such medical instrumentation as CT or MRI scanners. Fully automated segmentation of images into objects itself is an unsolved problem. Even in specialized domains, fully automated segmentation causes many problems and is often not easy to realize. Semi-automated segmentation in which little user interaction is there can make CBMIR more interactive and efficient.

In image retrieval, several systems attempt to perform an automatic segmentation of the images in the collection for feature extraction [33, 34]. After segmentation, the resulting segments can be described by shape features that commonly exist, including those with invariances with respect to shifts, rotations and scaling [35, 36].

In section 3 various segmentation methods are listed like AAM, ASM, ACS and many more. Based on the requirement and current dataset the corresponding method can be selected.

MATITK [37] is the insight segmentation tool by kitware especially for the segmentation of 3D medical images in MATLAB (MATLAB+ITK). Basically ITK (insight toolkit) is an open-source software toolkit for performing registration and segmentation. Registration is the task of aligning or developing correspondences between data. For example, in the medical environment, a CT scan may be aligned with a MRI scan in order to combine the information contained in both. ITK is implemented in C++. Because ITK is an open-source project,
developers from around the world can use, debug, maintain, and extend the software. Canny operator and GVF snake model can also be used for the segmentation of medical images [38]. Canny operator has preferable anti-noise ability. However the edge based on Canny operator is not consecutive. GVF Snake model is used widely in image segmentation. But there are problems in convergence processing to boundaries of some medical image because of noise. See Figure 5, shows its result.

A. Region based image retrieval
Most CBMIR systems use region based image retrieval segmentation as medical systems are mostly concerns with particular organs only. So first the pathology bearing regions (PBR) are extracted, where PBR are the regions affected with a disease to which the user is more concerned. The three main tasks of region based image retrieval system are:

- Image segmentation to obtain objects/regions. Images are segmented by grouping pixels with similar descriptions to form objects/regions.
- Object clustering for faster retrieval.
- Similarity distance computation between the image query and database. Figure 6 provides an overview of the system architecture developed in this study

In this figure each database image is converted into segmented objects i.e. all the objects in the image are extracted by finding the shape features and indexing each image and similarly the query image’s objects are also extracted and compared with database segmented objects. Then similar objects are retrieved. By finding the region-of-interest (ROI), retrieval time gets reduced as only particular region is used instead of the whole image for comparison.

5. Comparison techniques used
Medical systems often use measurement systems such as the easily understandable Euclidean vector space model [11, 39] for measuring distances between a query image (represented by its features) and possible results representing all images as feature vectors in an n-dimensional vector space. This is done, although metrics have been shown to not correspond well to human visual perception (Tversky [40]). Several other distance measures do exist for the vector space model such as the city-block distance, the Mahalanobis distance [11] or a simple histogram intersection [41]. Still, the use of high-dimensional feature spaces has shown to cause problems and great care needs to be taken with the choice of distance measurement to be chosen in order to retrieve meaningful results [42, 43].

These problems with a similarity definition in high-dimensional feature spaces are also known as the curse of dimensionality and have also been discussed in the domain of medical imaging [44]. A formal definition of vector-space, probabilistic and Boolean models for information retrieval is attempted in [45].

A. Evaluation methods of CBMIR
In the general image retrieval domain it is difficult to compare any two retrieval systems. For medical image retrieval systems, the evaluation issue is almost non-existent. Still, there are several articles on the evaluation of imaging systems in medicine [46] or on general evaluation of clinical systems and the problems with it.
The most common evaluation measures used in CBIR are precision and recall (see Eq. 1), usually presented as a precision vs. recall graph (PR graph). Researchers are familiar with PR graphs and can extract information from them without interpretation problems.

\[
\text{precision} = \frac{\text{no. relevant items retrieved}}{\text{no. items retrieved}} \quad (1)
\]

\[
\text{recall} = \frac{\text{no. relevant items retrieved}}{\text{no. relevant items}} \quad (2)
\]

In medical statistics commonly used measurements are sensitivity and specificity defined as follows:

\[
\text{sensitivity} = \frac{\text{pos. items classified as pos.}}{\text{all positive items}} \quad (3)
\]

\[
\text{specificity} = \frac{\text{neg. items classified as neg.}}{\text{all negative items}} \quad (4)
\]

Systems that use sensitivity and specificity include [47, 48]. Various problems in evaluating the CBIR or CBMIR are listed in [46]. These papers show that although we are towards the development of CBMIR systems still many problems are there in evaluating these systems. The need for standardization is clear, since several measures are slight variations of the same definition as Eq-1 and Eq-3.

6. Other retrieval techniques in CBMIR

There are a large number of other important techniques to improve the performance of retrieval systems. One of the most prominent techniques is relevance feedback that is well known from text retrieval [49]. This technique has proven to be important for image retrieval as well [50-53] because often unexpected or unwanted images show up in the result of a similarity query. Very rare CBMIR systems have used this technique. But now more emphasis is given on this. Often, the performance of a retrieval system with feedback is regarded as being even more important than without as only with feedback the user's subjectivity can seriously be taken into account. An overview of interaction techniques in image retrieval is given in [54].

Other techniques from the artificial intelligence community are also used for image retrieval such as long-term learning from user behavior based on data mining in usage log files [55] using the well-known market basket analysis. Use of DICOM header for retrieval is the current area of interest by the researchers [55]. DICOM header contains lots of important information of patient like his name, age, sex, image modality like CT/X-ray/MRI and many more which can be saved in the database directly as text and can be further used for retrieval purposes [56].

7. Built-in tools for CBMIR

For ad-hoc retrieval, two visual runs were submitted: MedGIFT, the GNU Image Finding Tool (GIFT) for visual retrieval (http://www.dim.hcuge.ch/medgift), and easyIR for text [58]. No textual retrieval was attempted, resulting in lower scores than those using text retrieval but the best visual run. For medical retrieval, visual retrieval was performed with several configurations of Gabor filters and grey level/color quantization as well as combinations of text and visual features. No machine learning was performed, so results were surprisingly good and only topped by systems with optimized learning strategies.

AAMtools, An Active Appearance Modeling Toolbox by George Papandreou [59] is a MATLAB-based toolbox for building Active Appearance Models (AAMs) and fitting them to still and moving images. Active appearance modeling is a successful and popular computer vision technique for deformable object appearance modeling and has found important applications in both image synthesis and analysis problems, most notably for modeling human face images in still images and videos. This tool was developed in June'2008 and still to come in market.

8. Various gaps in CBMIR

CBMIR is an active research area however; it has not made significant inroads as CBMIR applications incorporated into routine clinical medicine or medical research. The cause is often attributed without sufficient analytical reasoning to the inability of these applications in overcoming the “semantic gap”. The semantic gap divides the high-level scene analysis of humans from the low-level pixel analysis of computers. In general, two gaps have been identified in CBIR techniques: (i) there is the semantic gap [60, 61] between the low-level features that are automatically extracted by machine and the high-level concepts of human vision and image understanding; and (ii) sensory gap defined by Smeulders et al between the object in the world and the information in a (computational) description derived from a recording of that scene [61]. However, in our view, other gaps [62, 63] hinder the use of CBIR. techniques in daily routine of medical image management. For instance, there is a gap between the publication of system approaches or technological concepts and their prototypical realization and implementation. As another example, there is a gap if 3D image data is represented by signatures that are based on 2D slices of the data. Some gaps are listed as illustrated in Figure 7.

Researchers are trying to bridge all these gaps [64] to get highly efficient CBMIR system. Lots of other limitations in CBIR line curse of dimensionality and lack of scalability [65] shows that researchers have to concentrate more on this area to make CBIR a highly efficient system.
Various medical departments are in need of CBMIR systems as in routine many similar cases comes for diagnosis. From our research we came to know that following departments are using CBMIR systems: Dermatology, Cytological, Histopathology, Cardiology, Radiology and many more there as CBMIR plays a very important role in saving time for diagnosis.

10. Conclusion and Future Directions
CBMIR is not basically developed to replace a doctor but just help them in diagnostic aid so that their time can be saved. The integration into PACS is an essential step for the clinical use of retrieval systems. PACS solutions currently allow search by patient and study characteristics and are mainly a storage place for images.

Of course, evaluation of the retrieval quality is an extremely important topic as well. Standardized evaluation is needed to identify promising techniques. Retrieval systems need to be compared to identify good techniques. This can advance the field much more than any single technique developed so far. User interaction and relevance feedback are two other techniques that need to be integrated more into retrieval systems as this can help to lead to much better results. Adding relevance feedback can also improve CBMIR performance. There are very few systems which use this concept like QBIC etc. but if relevance feedback is integrated with medical systems can improve the retrieval rate.

It needs to become easy to integrate these new functionalities into other existing applications such as HIS (Hospital Information System)/RIS (Radiology Information System)/PACS or other medical image management or viewing software.

Marcelo Costa Oliveira [66] has proposed a new method of using grid computing can also improve CBMIR performance. Combinations of modalities such as Photon Emission Tomography/Computer Tomography (PET/CT) scanners or the use of image fusion techniques also create multi-dimensional data that needs to be analyzed and retrieved. Integration of already existing CBMIR systems like SPIRS and IRMA are combined to improve the overall performance [67]. Similarly other existing CBMIR systems should also be combined according to the needs and for better results. The development of open toolboxes is another important factor for successful applications. This will help to reduce the number of applications developed and will make it possible to spend more time on the important tasks of integration and development of new methods and system optimizations. Researchers are also diverting towards the use of Genetic Algorithms which will definitely improve the retrieval performance [68]. MATLAB® has usually been tagged a high-level language with a lot of flexibility but inherently slow and memory-consuming, just meant for fast development of algorithms or one-shot applications.
but not for production environment. CBMIR systems which are developed in MATLAB are very slow in response. Methods should be developed to increase the speed of processing [69]. The possibility of fully automating the segmentation and indexing remains a more problematic and distant goal. Selection of efficient and fully automatic segmentation method will always be the first need of CBMIR system. A fully automatic segmentation method ITK was used for lung segmentation [70] and it gave very good results but still ITK is under observation by some researchers as it is highly dependent upon the dataset. Lots of semiautomatic segmentation methods we have studied but no doubt fully segmentation will increase the overall performance. Accuracy of CBMIR system should be good so that the physicians were able to double the diagnostic accuracy.

References


[37] Vincent Chu, Ghassan Hamarneh,Simon Fraser , ”MATITK: Extending MATLAB with ITK”.


[54] H. Muller, D. M. Squire, T. Pun, Learning from user behavior in image retrieval: Application of the market basket analysis, International Journal of Computer Vision (Special Issue on Content-Based Image Retrieval).


[56] A. Grace Selvarani and Dr. S. Annadurai, “MEDICAL IMAGE RETRIEVAL BY COMBINING LOW LEVEL FEATURES AND DICOM FEATURES”, International Conference on Computational Intelligence and Multimedia Applications 2007, 0-7695-3050-8/07 $25.00 © 2007 IEEE DOI 10.1109/ICCIMA.2007.336


[66] Marcelo Costa Oliveira, “Grid computing to make viable the content based medical image retrieval through the image registration techniques”, ACM New York, USA.


[69] M Bister, CS Yap, KH Ng, CH Tok, “Increasing the speed of medical image processing in MatLab”, Biomedical Imaging and Intervention Journal.

[70] Joris Heuberger, Antoine Geissbuhler, Henning Muller, “Lung CT segmentation for image retrieval using the Insight Toolkit (ITK)”.

Biography

**Preeti Aggarwal** is with Panjab University, Chandigarh, India. She is working as Lecturer in the Deptt. of Computer Sci. & Engg. She has done B.E (CSE) in 2000 and M.E (CSE &IT) from PEC, Chandigarh in 2006. Presently she is pursuing her research in the area of CBMIR. Email: pree_aggr2002@yahoo.com

**Gagandeep Jindal** is with Landran College, Mohali and presently working as Asst. Professor in the Deptt. of Computer Sci. & Engg. He has done his M.E (CSE&IT) from PEC, Chandigarh in 2005. Presently he is working in the area of CBIR. Email: gaganpec@yahoo.com

**Dr. H.K. Sardana** is Scientist-F at CSIO, Chandigarh and also Dy. Director of the institution. His area of research is pattern recognition and image/signal processing. Email: hk_sardana@hotmail.com
Objective Assessment of Nonlinear Segmentation Approaches to Gray Level Underwater Images

Zhengmao Ye
College of Engineering, Southern University, Baton Rouge, USA
zhengmaoye@engr.subr.edu

Abstract
Automatic target recognition is a challenging task with a wide variety of potential applications in the industrial, military and medical fields. Region segmentation is a crucial step towards automatic segmentation of images. Under some severe conditions of improper illumination and unexpected disturbances, the blurring images make it more difficult for target recognition, which results in the necessity of segmentation. For instance, the underwater images are generally captured under water dispersing and atmospheric fluctuation. Segmentation is thus needed to clarify feature ambiguity against stochastic disturbances. Region segmentation splits images into regions based on similarity measures, such as pixel intensities, locations and textures or combinations. It categorizes an image into separate parts, which correlates with objects involved. Both K-means segmentation and watershed segmentation can be applied. Segmentation by K-means clustering refers to grouping similar data points into the clusters. It requires that the number of clusters be specified whose distance metrics should be defined to quantify orientation closeness of objects. The winner-take-all algorithm can thus be selected to update the cluster centers. It has the capability of simplifying computation and accelerating convergence. Another typical methodology is watershed segmentation. It is based on the gradient magnitude of images, which can classify diverse objects automatically, where watershed lines separate catchment basins. The erosion and dilation operations are essential procedures involved in watershed segmentation. Also to avoid over segmentation, the foreground and background markers should be selected accordingly. To evaluate the actual role of nonlinear image region segmentation, quantitative statistical measures have been proposed, such as the gray level energy, discrete entropy, relative entropy, mutual information and information redundancy. The assessment measures will further quantify the impact from image segmentation. The objective assessment approach has the potential to solve other image processing issues.

Keywords: K-Means Segmentation, Watershed Segmentation, Gray Level Energy, Discrete Entropy, Relative Entropy, Mutual Information, Information Redundancy

1. Introduction
Image segmentation is a major step for automated object recognition systems. In many cases, image processing is affected by illumination conditions, random noises and environmental disturbances due to atmospheric pressure or temperature fluctuation. The quality of underwater images is directly affected by water medium, atmosphere medium, pressure and temperature. This emphasizes the necessity of image segmentation, which divides an image into parts that have strong correlations with objects to reflect the actual information collected from the real world. The watershed segmentation approach starts from either low or high altitude up to high or low altitude. The gradient gray levels of image data are similar to altitudes in the topographic surface. High region edges correspond to watersheds and low gradient interiors correspond to catchment basins, while catchment basins represent the segmented regions, which can be determined by values of gray levels. Watershed transformation can be accelerated by updating fast watershed algorithms. Segmentation on a basis of watershed approaches is focused on specified local small size images out of image databases. On the other hand, to minimize the medium dispersing, K-means clustering is critical for image processing. It is used to accumulate pixels with similarities together to form a set of coherent image layers [7-8]. For K-means clustering, it is hard to achieve optimization unless certain control algorithms such as the nearest neighbor rule or winner-take-all scheme are applied [5-6]. To implement image segmentation for all possible small and large scale image processing, nonlinear K-means clustering is presented for image segmentation also. To quantify the impact of image segmentation, the quantitative measures of the gray level energy, discrete entropy, relative entropy, mutual information, information redundancy, uncertainty coefficient are introduced [1-14].

From literatures [15-20], entropy of a fuzzy set is viewed as a global measure of the fuzziness of the fuzzy set while the energy of a fuzzy set is viewed as a local measure, where a tradeoff exists between the entropy of fuzzy sets and the information energy. In other cases, the relative entropy is employed to be the fitness function of genetic algorithms in optimal segmentation approaches.
The fuzzy integral based region merging algorithm has been presented to deal with issues of over-segmentation due to watershed transformation. To reduce the number of regions subsequently, the image fusion process has been applied to the regions in a recursive way according to the maximum fuzzy integral.

The organization of this article is as below. Based on the underwater images at the ocean floor, firstly, nonlinear K-means segmentation via the optimal schemes will be applied to analyze the effects of clusters being generated; secondly, watershed segmentation using morphological analysis techniques will be applied to investigate detailed information of the source and resulting images. At last, objective assessment of the outcomes from two nonlinear segmentation techniques has been proposed in terms of quantitative measures from different points of view. As a result, both qualitative and quantitative analysis of deep underwater images will be conducted in this research, which can also be extended to other advanced signal and image processing methodologies.

2. Segmentation Via K-Means Clustering

Image segmentation requires that the number of clusters be specified for partitioning. Centers of each cluster are defined, which represent the mean values of all data points in those clusters. The distance metric should be determined to quantify the relative distances of objects. Euclidean and Mahalanobis distances are two major types of distance metrics. Computation of the distance metrics is based on the spatial gray level histograms of images. The Mahalanobis metric distance is applied, which is formulated as (1), where $X_A$ is the cluster center of any layer $X_A$, $s$ is a data point, $d$ is the Mahalanobis distance, $K_A^{-1}$ is the inverse of the covariance matrix.

$$d = (s - X_A)^T K_A^{-1} (s - X_A)$$

K-means clustering assigns each object a space location, which classifies data sets through K numbers of clusters ($K=3$ in the example). It selects cluster centers and points cluster allocations to minimize errors. Optimal statistical algorithms can be applied for classification, which are categorized as threshold based, region based, edge based or surface based. The distances of any specific data point to several cluster centers should be compared for decision making. Winner-take-all competitive learning (2-3) has been introduced to K-means clustering algorithms so that only one cluster center is updated for each individual input. Images will be decomposed into three recognized physical entities. Clustering depends on the partition of images into a set of layers. The winner-take-all learning network classifies input vectors into one of K specified numbers of categories according to clusters detected in the training dataset. All points are allocated to the closest cluster. Learning is performed in an unsupervised mode. Each cluster center has an associated weight that is listed as $w$’s. The winner is defined as one whose cluster center is closest to the inputs. Thus, this mechanism allows for competition among all input responses, but solely one output is active each time. The unit that finally wins the competition is the winner-take-all cluster, so the best cluster center is computed eventually.

$$w_{ij} = \min(w_{ij}) \quad \text{for } j = 1, 2, 3 \quad \text{and } i = 1, 2, 3$$

$$w_{ij} + w_{i2} + w_{i3} = 1 \quad \text{for } i = 1, 2, 3$$

To minimize the distortion, K-means clustering iterates between the cluster point labeling and center reassigning.
Point labeling groups the data points belong to the same cluster. After completing each iteration, the points are reassigned to the winning center. Center reassigning recalculates centers for all clusters. The winner-take-all algorithm is implemented until the local convergence has been achieved. In Figures 1-4, the source image near the Hawaii Maui Island and all three clusters are shown.

3. Watershed Segmentation
In image watershed segmentation, altitude is represented by the gray level of the pixels. All pixels throughout the same catchment basin are connected with the minimum altitude region of the basin. The watershed lines divide individual catchment basins. The high gradient regions correspond to watershed lines and low gradient regions correspond to catchment basins. Monotonic increment or decrement of the threshold is applied to find a path to reach local maxima or minima. In general, the watershed operations consist of the erosion, dilation, opening and closing operations. Dilating is to replace function values at each point by the maximum value and eroding is to replace function values at each point by the minimum value. Thus erosion and dilation are dual operations, while erosion is for shrinking and dilation is for expanding. Opening is the dilation of an eroded function and closing is the erosion of a dilated function. The algebraic difference between the dilation and erosion of an input image is the gradient gray level. The watershed image gradient magnitude is formulated as the difference of unit size dilation and unit size erosion of the input.

\[ \text{grad}(X) = (X \circ B) - (X \bullet B) \]  

(5)

where X is an image to be transformed, B is a structuring element, \( \circ \) and \( \bullet \) denote dilation and erosion. Then the image gradient is expressed as dilation minus erosion.

It is possible that when an image is partitioned into too many regions, it will give rise to over-segmentation, so foreground and background markers can be used to avoid over-segmentation. For the case of opening, downstream watershed segmentation is selected, where dilation is applied after erosion. The source ocean floor coral image at the Hawaii Big Island and its processed image after erosion and dilation are shown in Figures 5-6. The watershed segmentation function is shown in Figure 7. To avoid over-segmentation, foreground and background markers are used, which are shown in Figures 8-9. The image superimposed by the regional maxima is also shown in Figure 10.
4. Objective Assessment
4.1 Probability Distributions Based on Histograms
The histogram is used to display the brightness of the gray level content, showing the occurrence frequency of the pixel counts for each of the 256 intensity levels. The occurrence of the gray level component is described as the co-occurrence matrices of relative frequencies. The probability function of the gray level image is estimated from the percentages of the count at the specific level over the total count across its histogram plot. In Figures 11-14, the probability functions of the ocean floor image and its three clustering are illustrated. In Figures 15-16, probability functions of the Big Island ocean floor image and processed one after the opening operation are shown.
4.2 Gray Level Energy

The gray level energy indicates how the gray levels are distributed. It is formulated as (6), where \( E(x) \) represents the gray level energy with 256 bins and \( p(i) \) refers to the probability distribution functions, which contains the histogram counts. The energy reaches its maximum value of 1 when an image has a constant gray level.

\[
E(x) = \sum_{i=1}^{k} p(i)^2
\]  

(6)

Table 1. Gray Level Energy of Image Segmentation

<table>
<thead>
<tr>
<th>K-Means Segmentation</th>
<th>Energy</th>
<th>Watershed Segmentation</th>
<th>Energy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Image</td>
<td>0.0137</td>
<td>Source Image</td>
<td>0.0122</td>
</tr>
<tr>
<td>Cluster #1</td>
<td>0.2754</td>
<td>Foreground Markers</td>
<td>0.6278</td>
</tr>
<tr>
<td>Cluster #2</td>
<td>0.4516</td>
<td>Background Markers</td>
<td>0.3347</td>
</tr>
<tr>
<td>Cluster #3</td>
<td>0.3058</td>
<td>Erosion and Dilation</td>
<td>0.0132</td>
</tr>
</tbody>
</table>

The larger energy value corresponds to the lower number of gray levels, which means simple. The smaller energy corresponds to the higher number of gray levels, which means complex. From Table 1, in K-means segmentation, each of three clusters is much simpler than the source image. For watershed segmentation, a simpler image is observed after erosion and dilation operations.

4.3 Discrete Entropy

The discrete entropy is the measure of image information content, which is interpreted as the average uncertainty of information source. It is calculated as the summation of the products of the probability of outcome multiplied by the log of the inverse of the outcome probability, taking into considerations of all possible outcomes \( \{1, 2, \ldots, n\} \) in the event \( \{x_1, x_2, \ldots, x_n\} \), where \( n \) is the gray level; \( p(i) \) is the probability distribution, considering all histogram counts. It is formulated as (7).

\[
H(x) = \sum_{i=1}^{k} p(i) \log_2 \frac{1}{p(i)} = -\sum_{i=1}^{k} p(i) \log_2 p(i)
\]  

(7)

For image processing, the discrete entropy is a measure how many bits needed for coding the image data, which is a statistical measure of randomness. The maximal entropy occurs when all potential outcomes are equal. When the outcome is certainty, the minimal entropy occurs which is equal to zero. The discrete entropy represents average amount of information conveyed from each individual image. In Table 2, results have shown that the entropy of the clusters is much smaller than that of the source image, which are much simpler. It is the fact that when the image pixels are distributed among more gray levels, the entropy value will increase. For watershed segmentation, it is also indicated from Table 2 that the processed image after the erosion and dilation operations has a relatively smaller entropy value than that of the source image, which means that the image being processed via the opening operation has been simplified.

Table 2. Discrete Entropy of Image Segmentation

<table>
<thead>
<tr>
<th>K-Means Segmentation</th>
<th>Entropy</th>
<th>Watershed Segmentation</th>
<th>Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Image</td>
<td>6.4079</td>
<td>Source Image</td>
<td>6.5238</td>
</tr>
<tr>
<td>Cluster #1</td>
<td>3.7137</td>
<td>Foreground Markers</td>
<td>1.7682</td>
</tr>
<tr>
<td>Cluster #2</td>
<td>2.7598</td>
<td>Background Markers</td>
<td>2.7020</td>
</tr>
<tr>
<td>Cluster #3</td>
<td>3.4220</td>
<td>Erosion and Dilation</td>
<td>6.0283</td>
</tr>
</tbody>
</table>

Principle of the maximum entropy has also been applied to analyze the potential of image segmentation. Assume mutually exclusive propositions have individual discrete probability distributions. The minimum entropy of an image is equal to zero when one of the distributions is definitely true, which represents the most informative distribution case. On the other hand, when the probability distribution is uniform, the maximum discrete entropy occurs with the discrete entropy value of \( \log_2(n) = 8 \) bits \( (n=256) \). In this case, no proposition is superior to any other available propositions, thus, the least informative distribution occurs. The discrete entropy provides useful numerical measure between zero and \( \log_2(n) \), from the most informative case to the totally uninformative case. All quantities of the discrete entropy in these examples are within a range between 0 and 8, the latter of which is the maximum entropy possible.

4.4 Relative Entropy

Suppose that two discrete probability distributions of the images have the probability functions of \( p \) and \( q \). The relative entropy of \( p \) with respect to \( q \) is then defined as the summation of all possible states of the system, which is formulated as (8).

\[
d = \sum_{i=1}^{k} p(i) \log_2 \frac{p(i)}{q(i)}
\]  

(8)

From Table 3, the relative entropy results for K-means segmentation and watershed segmentation are provided. The smaller values indicate greater similarities between the processed and source images.
4.5 Mutual Information

The notion of the mutual information can be applied as another objective metric. The mutual information acts as a symmetric function, which is formulated in (9).

\[ I(X; Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \]

where \( p(x,y) \) represents the mutual information; \( H(X) \) and \( H(Y) \) are entropy and conditional entropy values. It is interpreted as the information that \( Y \) can tell about \( X \) is equal to the reduction in uncertainty of \( X \) due to the existence of \( Y \). At the same time, it also shows the relationship of the joint and product distributions. The results are shown in Table 4.

<table>
<thead>
<tr>
<th>K-Means Segmentation</th>
<th>Relative Entropy</th>
<th>Watershed Segmentation</th>
<th>Relative Entropy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cluster #1 w.r.t. Source</td>
<td>0.2109</td>
<td>Erosion and Dilation</td>
<td>0.1245</td>
</tr>
<tr>
<td>Cluster #2 w.r.t. Source</td>
<td>0.1606</td>
<td>ForeMarker 6.2756</td>
<td>BackMarker 5.7236</td>
</tr>
<tr>
<td>Cluster #3 w.r.t. Source</td>
<td>0.1120</td>
<td>Enhancing w.r.t. Source</td>
<td>0.2407</td>
</tr>
</tbody>
</table>

From Table 4, in image segmentation, the better match between the source and processed images, the less value of the mutual information. The processed image after the erosion and dilation operations and the enhanced image have shown good similarities with the source image.

4.6 Normalized Mutual Information

The normalized mutual information is a well defined measure covering contents from both discrete entropies and mutual information. It is formulated as (10).

\[ NMI = \frac{ I(X; Y) }{ \sqrt{ H(X) H(Y) } } \]

where \( I(X; Y) \) is the mutual information; \( H(X) \) and \( H(Y) \) are the discrete entropies. Results are shown in Table 5.

Once again, similar to that of the mutual information, the better match between the source and processed images, the smaller the normalized mutual information.

4.7 Information Redundancy

Another symmetric information measure can be used to indicate redundancy in image segmentation. It reaches the minima of zero when all variables are independent. It is formulated as (11) and results are listed in Table 6. The bigger information redundancy, the greater dependence between two images.

\[ R = \frac{ I(X; Y) }{ H(X) + H(Y) } \]

4.8 Uncertainty Coefficient

The uncertainty coefficient presents the uncertainty in the dependent variable. It is based on the entropy principle to ascertain the uncertainty in the dependent variable and to calculate the proportion of uncertainty. It ranges from 0 to 1, which indicates the normalized variants of mutual information, where 0 refers to no reduction in uncertainty of the dependent variable and 1 refers to the complete elimination of uncertainty. The uncertainty coefficient is a non-symmetric function which is defined as (12). The corresponding results for all three clusters of K-means segmentation are shown in Table 7.

\[ C_{XY} = \frac{ I(X; Y) }{ H(Y) } \quad C_{YX} = \frac{ I(Y; X) }{ H(X) } \]

5. Conclusions

Image segmentation is to distinguish objects from images. It classifies each image pixel to a segment according to the similarity in a sense of a specific metric distance. To reduce blurring effects of the underwater images stem from the water and atmospheric media, nonlinear region K-means segmentation has been presented for image segmentation, where the competitive learning winner-take-all rule is applied to update the clustering center with satisfactory results. Meanwhile, another nonlinear region segmentation has also been presented.
segmentation provides the unique morphological analysis of images to determine catchment basins and watershed lines clearly across the entire image range. To avoid over-segmentation, foreground and background marker controlled algorithms are applied with useful outcomes. To evaluate the roles of both segmentation approaches, quantitative metrics are proposed rather than qualitative measures from intuition. Both histograms and probability distributions are calculated to serve as the base functions to assess digital images. Then notions of the gray level energy, discrete entropy, relative entropy, information redundancy, mutual information as well as uncertainty coefficient are applied to imply the effects of two technical segmentation approaches. These methodologies can be easily and broadly expanded to diverse types of image processing for practical implementations.

6. References
**Biography**

Zhengmao Ye received the B.E. degree in thermal engineering from Tianjin University, China; the first M.S. degree in automotive engineering from Tsinghua University, China; the second M.S. and Ph.D degrees in electrical engineering from Wayne State University, Detroit, USA, in 1999 and 2001, respectively, next to his Chinese Ph.D degree being pursued under the Academician Advisor Dr. Shaoxi Shi (Past President of Tianjin University and Chinese Chief Scientist on I.C. Engines) in School of Mechanical Engineering, Tianjin University, China.

He is presently a faculty in the Electrical Engineering Department, Southern University, Baton Rouge, USA, where he is the Founder and Director of the System and Control Laboratory. His current research interests include modeling, control and optimization with diverse applications on electrical, mechanical, automotive and biomedical systems, as well as signal processing and image processing. Dr. Ye is the first cross-disciplinary researcher internationally who has first author publications covering all the leading control proceedings in three most prestigious engineering societies (IEEE, ASME and SAE), specifically, IEEE (CDC, CCA, SMC, ACC, ISIC, FUZZ, IJCNN, CASE, ICCA, SOSE, MSC Congress, WCCI Congress and CCECE Congress), ASME (IMECE Congress, ICES, JRCICE), SAE (USA Congress, EAEC Congress, PFL Congress), and with Sole Authorships in IEEE Transactions and SAE Transactions as well. He was also the academic reviewers for over 80 articles submitted to IEEE, ASME, SAE and some International Journals.

Dr. Ye is the recipient of the Chinese National Fellowship (First Prize) at Tianjin University, the USA Allied Signal Fellowship (First Prize) at Tsinghua University and the Most Outstanding Faculty of Electrical Engineering at Southern University. He was selected for inclusion in Marquis Who’s Who.