

6-2004

# A Directed FCM Approach for Analysis of Stained Tissues

Shahram Rahimi

*Southern Illinois University Carbondale, rahimi@cs.siu.edu*

N. Mogharreban

*Southern Illinois University Carbondale*

A. Krovi

*Southern Illinois University Carbondale*

Follow this and additional works at: [http://opensiuc.lib.siu.edu/cs\\_pubs](http://opensiuc.lib.siu.edu/cs_pubs)

Published in Rahimi, S., Mogharreban, N., & Krovi, A. (2004). A directed FCM approach for analysis of stained tissues. IEEE Annual Meeting of the Fuzzy Information Processing, 2004. NAFIPS '04, 228-233. doi: 10.1109/NAFIPS.2004.1336282 ©2004 IEEE. Personal use of this material is permitted. However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works must be obtained from the IEEE. This material is presented to ensure timely dissemination of scholarly and technical work. Copyright and all rights therein are retained by authors or by other copyright holders. All persons copying this information are expected to adhere to the terms and constraints invoked by each author's copyright. In most cases, these works may not be reposted without the explicit permission of the copyright holder.

---

## Recommended Citation

Rahimi, Shahram, Mogharreban, N. and Krovi, A.. "A Directed FCM Approach for Analysis of Stained Tissues." (Jun 2004).

# A Directed FCM Approach for Analysis of Stained Tissues

S. Rahimi

Department of Computer Science  
Southern Illinois University  
Carbondale, Illinois 62901-4511  
rahimi@cs.siu.edu

N. Mogharreban

Department of Computer Science  
Southern Illinois University  
Carbondale, Illinois 62901-4511  
namdar@cs.siu.edu

A. Krovi

Department of Computer Science  
Southern Illinois University  
Carbondale, Illinois 62901-4511  
akrovi@cs.siu.edu

## ABSTRACT

*The use of digital imagery has increased phenomenally especially in the clinical field. These images are obtained from different modalities such as X-ray and MRI. Digital imaging of the more traditional imagery such as stained tissues has opened up new means of investigation. Hence a need to build a system to analyze the stained tissues and extract the salient information has risen. A Directed FCM Approach for Analysis of Stained Tissues introduces a modified FCM to analyze the tissues. The analysis can be controlled by the user by selecting the number of clusters, size of the clusters and the centers for the clusters. The results of this analysis are reported as the percent of changes in a specific square area.*

**Keywords:** Fuzzy C-mean, Stained Tissue Analysis, Image Analysis.

## 1. INTRODUCTION

Due to the increasing number of diagnostic imaging techniques in the medical field, image analysis has been one of the prominent research areas over the last few years. Types of images used in the medical field are X-ray, magnetic resonance imaging (MRI), single photon emission topography (SPECT), and positron emission topography (PET) [14]. Digital imaging has opened up new means of investigation for traditional images such as stained tissues. Computer analysis of these many images has proven essential. However, since medical personnel might not understand the technical aspects of the image analysis, a proper reporting methodology has to be adapted.

The objective of this work is to analyze the images generated from stained tissue treated with angiogenic agents. Angiogenesis is the process of vascular growth and is associated with the abnormal growth of tissue as in cancer. Angiogenesis, however, is also significant in healing. Regeneration and growth of blood vessels and capillaries are essential in the process of tissue repair. The images utilized in this analysis are tissue images stained specifically to pick up the proteins in the lining of the blood vessels. The images we consider in our analysis are 256 color images in standard formats such as like TIFF. The images were taken with a standard microscope-mounted camera with a standard focal length and magnification (magnification is an important

aspect of the image). Changes in vessel growth must be reported within a specific area. We perform the analysis on the colors of the stained tissues specified by the user. We call this the Directed FCM approach (DFCM) because the user selects the number of colors to be identified, the hue for each color and the size of the cluster. Our Directed FCM Approach for Analysis of Stained Tissue can be broken down into three steps: 1) Applying the DFCM, 2) clustering the image by mapping the neighboring elements, 3) reporting the results.

The standard FCM can be used for clustering, however, the partitioning process is very time consuming. Many algorithms have been proposed for speeding up the FCM. In this paper we are proposing an approach to FCM that is directed by the user's choice.

The number of colors picked up by the user represents the number of clusters for the FCM. The hue picked for each color becomes the center for each cluster. The centers of the clusters remain constant so the FCM need not be repeated iteratively. Instead the size of the cluster is set using a threshold value selected by the user. Each cluster is called a palette. These palettes are used for identifying the stains in the image. FCM assigns unique labels to each palette generated. The stains are identified by mapping the neighboring pixels in the image using the label of the palettes generated by the FCM. A Modified label mapping algorithm is used for this purpose. This is explained in detail in the methodology section.

Scale is the other important factor of the image attribute. The size of the image is in microns. The reported measures need to be in microns to be compared meaningfully by the researchers. The tissues are analyzed before and after angiogenesis and the percentage of color change is reported. It is important that the user makes sure that images of the same magnitude are compared.

This paper proceeds by giving a brief background on image analysis and fuzzy C-mean algorithm. In section three we discuss the methodology applied for clustering. Section four contains the implementation part followed by the conclusion.

## 2. Background

This section gives an over all idea about the image analysis and the Fuzzy C-mean algorithm. Different techniques used for image analysis are explained in detail.

### 2.1. Image analysis

With the increase of available processing power and improvements in digital imaging, image analysis is becoming an important area in research. Images have many features such as color, texture and shape. Color is one of the most important features of an image, since it is not only easy to extract but it also provides a wealth of information about an image. Images are represented by an array of pixels. Color images are usually represented by the distribution of the three color components R (red), G (green) and B (blue). So every pixel in a color image can be quantized to give the RGB values. The other qualities of the image like hue, saturation and intensity can be obtained by applying either linear or nonlinear transformation over the RGB values. [1]

An image  $I$ , can be represented by a two dimensional array of the size  $I_w \times I_h$ , where  $I_w$  represents the width of the image and  $I_h$  represents the height of the image.  $I(i,j)$  represents the pixel at the  $i$ th row and the  $j$ th column. Every pixel has an RGB value ranging from 0-255. Therefore, 24 bits (8 for each color) are required for storing each pixel of an image. In image processing, we not only deal with the two geometric dimensional but also with three dimensional color space (RGB). This adds to the level of complexity for image analysis.

Analysis of images based on color data can be split to a) image quantization and b) image segmentation. There are many techniques for both quantization and segmentation, while there is no single method that is considered acceptable for all types of images. This is because every image has its unique features. For instance, an image may not have representation of all the available colors in the color space. Furthermore a particular color may dominate one image verses another. However, a method developed for analysis of one type of image can be applied and extended to another group of images.

Quantization of an image is a two step process which involves 1) Choosing a small subset of closely related colors (colors that are more similar to each other), and grouping them into color palettes. 2) Mapping all the pixels to the available palettes. Some of the algorithms used for color image quantization are: a) popularity algorithm which uses the densest region in the color distribution of the original image, b) median cut algorithm where the image is divided into smaller regions based on the available colors, and c) the fixed point algorithm [2]. Some of the other popular image quantization algorithms are binary tree palette design, subjectively weighed TSE, erosion based weighting and the LGB algorithm. [3]

Another technique for image analysis is image segmentation. This technique is the process of partitioning an image into clusters or regions in such a way that each region is homogeneous. There are many fuzzy and non fuzzy based

algorithms for image segmentation. Many geographic information systems based algorithms also utilize image segmentation for their analysis. One of the most popular image segmentation algorithm is fuzzy c-mean algorithm. In this algorithm each sample (pixel) is assigned to a cluster based on the cluster membership function. Some of the other traditional image segmentation algorithms are region split and merge (also called quadrant tree), region growing, histogram threshold, edge detection, Marcov random file, Gibbs random file, neural network learning theory [3], low pass filter[3] and fuzzy k-mean clustering[5].

### 2.2. Fuzzy c-mean algorithm

In this application Fuzzy c-mean (FCM) algorithm has been improved. It is used to characterize the image into palettes based on membership values of the pixels. FCM Clustering is an iterative partitioning method that produces optimal c-partitions based on the membership values of the feature vectors [12]. FCM uses the reciprocal of distance to decide the cluster centers.

Consider a set of feature vectors  $Q$ ,  $\{x^{(q)}: q = 1, \dots, Q\}$ . Each feature vector  $x^{(q)} = (x_1^{(q)}, \dots, x_N^{(q)})$  has  $N$  components. The process of clustering is to assign the  $Q$  feature vectors into  $K$  clusters  $\{c^{(k)}: k = 1, \dots, K\}$  usually by the minimum distance assignment principle.

The average weight of a feature vector is,

$$X_n^{(k)} = \sum_{\{q: q \in k\}} w_{qk} x_n^{(q)}$$

Where  $w_{qk}$  is the membership value of the fuzzy cluster  $k$ , which is also called the fuzzy weight. It is generated by the reciprocal of distances and is represented as:

$$w_{qk} = (1/(D_{qk})^2)^{1/(m-1)} / \sum_{(k=1,K)} (1/(D_{qk})^2)^{1/(m-1)},$$

$$m > 1$$

$$D_{qk} = \| X_i - v_i \|$$

$$v_i = \sum_{(k=1,K)} (w_{qk})^m \cdot x^{(q)} / \sum_{(k=1,K)} (w_{qk})^m,$$

where  $v_i$  is the center of the cluster  $i$ .

$$\sum_{(k=1,K)} (w_{qk}) = 1 \text{ for each } q$$

The mean square error is represented by:

$$J(w_{qk}, v_i^{(k)}) = \sum_{(k=1,K)} \sum_{(k=1,K)} (w_{qk}) \| x^{(q)} - v_i^{(k)} \|^2$$

- $m \in (1, \infty)$  is the control parameter of fuzziness or fuzzifier. It helps in reducing the square distance error.

FCM assumes that the number of clusters and the centers of the clusters are known in advance. It then proceeds by generating the membership functions for each feature vector with respect to the clusters. Then new centers for each cluster

are calculated. It then generates the membership functions with the new centers. This process continues until the difference between the old and new membership values is less than the threshold or the condition for the maximum number of iterations is satisfied. Choosing the center of clusters is crucial for the FCM.

### 3. METHODOLOGY

In our approach, we break the process into three steps which include:

- 3.1. Performing a DFCM on an image,
  - (a) Selecting the centers and the threshold value.
  - (b) Assigning a label to each pixel in the image based on their membership values.
- 3.2. Clustering the image by mapping the neighboring elements and
- 3.3. Calculating the percentage of each color stain (based on the clusters).

The image to be analyzed is first quantized to the RGB format and then the above three steps are performed on it in the sequential order.

#### 3.1. Directed FCM

The most important aspect of DFCM is selecting the number of clusters and the centers for each cluster. The images we deal with are the images of tissues and cells. The color values in these images are densely populated only in a few regions and many of the colors in the color space are not present.

Based on the number of colors selected by the user to be identified the user must select the hue for each color from the image. This hue acts as the center of the palette (cluster). Hue is selected by clicking on the color from the image. Choosing the cluster centers is a crucial step for FCM. One more important aspect of FCM is selecting the size of the palette. It is based on the threshold value  $\epsilon$ . A default value is given for  $\epsilon$ , but it can be controlled by the user.

DFCM (*Dynamic FDM*) creates a palette for each color to be identified. The clusters formed by DFCM are called *image dependent dynamic palettes*. It is called dynamic palette because the number of palettes is selected by the user. The center of the palette is selected by clicking on the required color from the image, so it is called *image dependent palette*. The design of the palette when three colors is selected is shown in Fig 1.

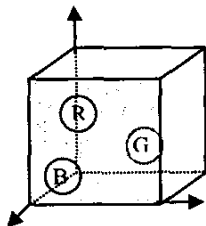


Fig. 1 Clusters in the color space

The big cube represents the total color space in three dimensions. The three spheres inside the large cube represents the palettes generated by the DFCM (each sphere

represents a palette). The center of the palettes (colors selected by the users) represents the center of the sphere and radius of the sphere is represented by the threshold value  $\epsilon$ .

The palettes are labeled based on the color. The palette for the first color is labeled 1, the palette for the second and the third color are labeled 2 and 3 respectively. This labeling helps in mapping the pixels in the image to the palette.

Once the number of clusters and the centers of the clusters have been defined, DFCM generates the membership values of every pixel with respect to each cluster. Then the pixels are mapped to their corresponding clusters by using the cut off value,  $\epsilon$ . This process is explained in detail in Fig 2. The size of the palette can be controlled in the large color space. Since the centers are fixed the FCM need not be repeated iteratively, it can be stopped with a single iteration.

To map the pixels based on the palettes available, a two dimensional array, labels are initialized to zero. The array is of the same size as the image so that every pixel has its corresponding label. The membership values (each cluster has a corresponding membership value) of every pixel in the image are checked with the threshold value. If they fit into any palette the value of the pixel in the label array will be updated with the label of the palette. Otherwise they will have a default label value of 0 which signifies that the pixel doesn't fit into any palette. This is clearly depicted in the figure below. The first figure shows a section from the array of pixels and the second figure shows the corresponding section from array of labels after the mapping is done.

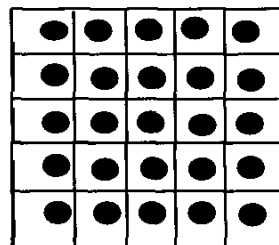


Fig. 2(a) Array of Pixels

1	1	0	2	0
0	1	1	0	0
3	3	3	0	0
0	3	3	1	1
2	0	0	1	1

Fig. 2(b) Array of labels after mapping

Each black dot in Fig.2(a) represents a pixel (before DFCM). Fig.2(b) shows the labels array after the DFCM. The value 1 represents the first palette, 2 represents the

second, 3 represents the third and 0 specifies that the pixel doesn't belong to any palette.

We call this a DFCM because the number of clusters is decided by the user (based on the number of colors to be identified). The centers of the clusters and the size of the clusters are also controlled by the user. The user controls most of the parameters for the DFCM.

### 3.2. Mapping the neighbors

The DFCM assigns a label to every pixel in the image based on the palette. The next step is to group the pixels into clusters based on the labels. There are many fuzzy clustering and non fuzzy clustering algorithms which could perform the image segmentation process. We use a modified label mapping algorithm for clustering the image based on the labels.

The algorithm can be explained as follows:

1. Initialize a 2-Dimensional array, mapped-labels with unique values (in ascending order). The size of the array is the same as that of the image.
2. The mapped-labels array is labeled in such a way that it holds the minimum value of itself or its neighbors if the corresponding elements in the label array are connected.
3. The elements in the labels array are said to be connected if there is a path of contiguous link between them.
4. The contiguous link is formed if it shares the same label value with any of its neighbors.
5. Steps 2, 3 and 4 are repeated until a consistent mapped-labels array is formed.

The only modification that is applied to this algorithm is that instead of comparing value of the mapped-labels to all its eight neighbors, the comparison is done with respect to the immediate neighbors as depicted in Fig 3.

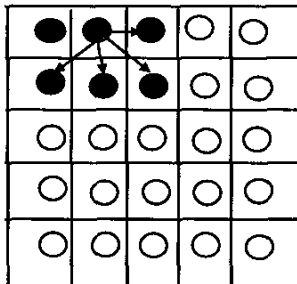


Fig. 3 Mapping the labels.

If the element is the first element in that row it is compared with three of its adjacent neighbors. For all the other elements the comparison is done as shown in the figure above. The comparison can be done for (N-1) rows and (N-1) columns, where N is the size of the array. The clusters are generated after a finite number of iterations. Fig 4 represents the image after the mapping is completed. Each cluster represents a stain and color of the stain is attained from the labels array. This method rules the scope of over segmentation and under segmentation. And the computational cost is still optimal.

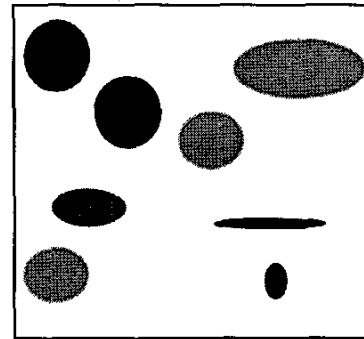


Fig. 4 Image after applying FCM

### 3.3. Calculating the size of the stain

The final step is calculating the number of stains (each cluster is considered a stain) for each color and the size of each stain. A vector can be used to hold the size of the stains. The combination of mapped-label and label arrays gives the color of the stain. Since we know the magnification and the size of the image the result can be displayed in what ever scale is required (such as nanometers or microns) by referring to some common conversion tables. The clustering of the pixels into homogeneous regions can be enhanced using some other fuzzy algorithms like FCFM [8], 3-D labeling, low pass filter and so on. [4]

## 4. IMPLEMENTATION

This section describes the implementation of our design. The image analysis software is implemented on java platform. Java has several advantages such as platform independence, user interaction and a high degree of computing. One more reason for choosing java as the platform is the ease with which it can be extended to a web version. A user interface has been designed to allow data input and for displaying the results in a meaningful fashion.

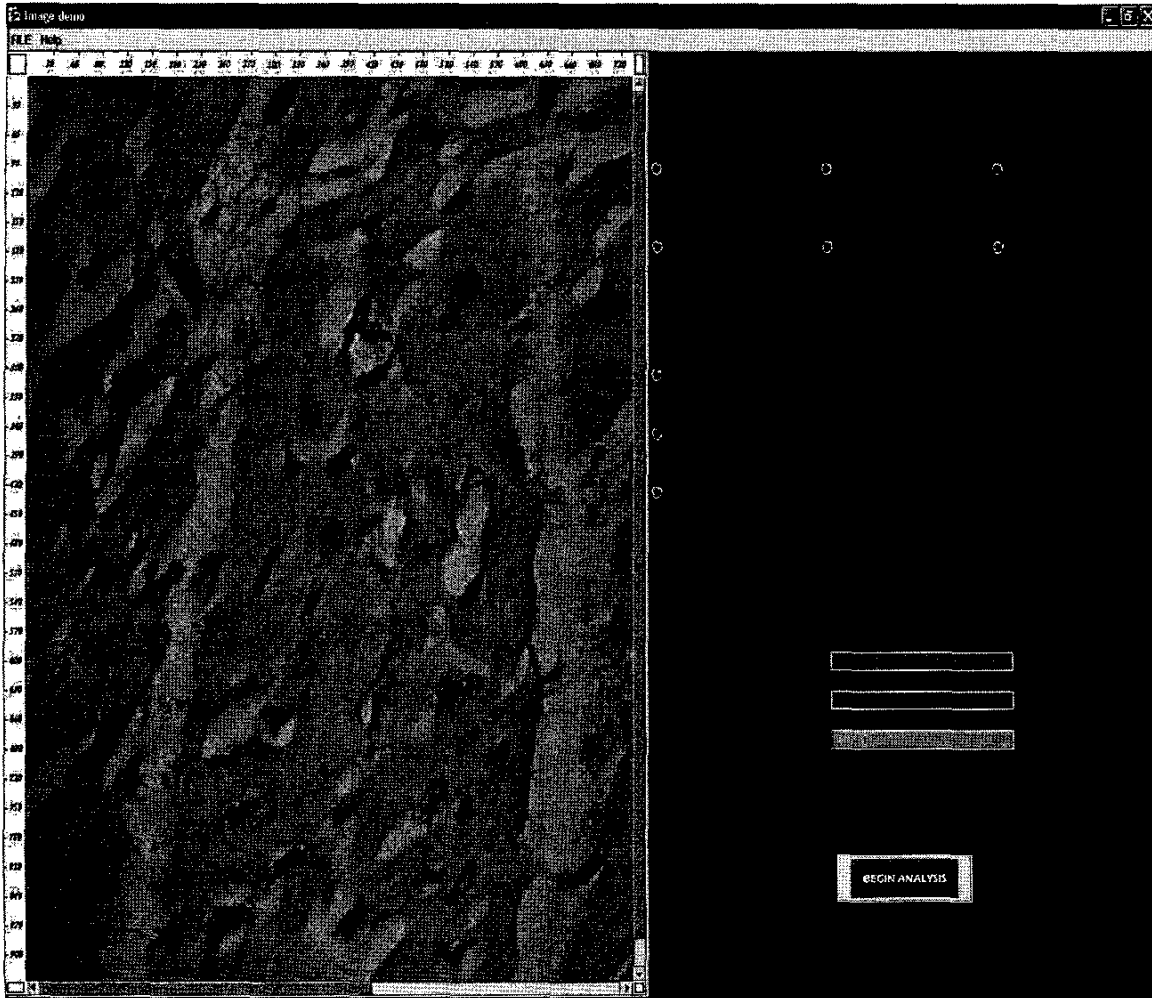


Fig. 5 GUI Design

The application has a menu with options like file and help. The image to be analyzed can be selected by browsing through the directories using the file> open option. The image is displayed on the left panel. Then the magnification is selected using the radio buttons on the top right hand side of the application. Magnification is important because the size of the image changes according to the magnification. The image is then quantized into the RGB format and stored in a two dimensional pixel array. The result is displayed based on the magnification. This helps the expert who is analyzing the image to know what image magnification he is working with.

Based on the number of colors a user decides to pick, boxes appear in the next panel. Then the user clicks on the colors to be identified from the image. Each color appears in its corresponding box. These colors form the centers of the clusters in our DFCM algorithm.

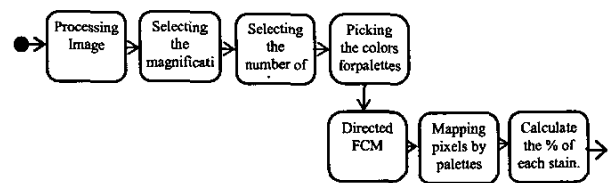


Fig. 6 Design for the Directed FCM approach

When the color selection is done and the threshold has been set the analysis can begin. All the pixels in the image are mapped based on the colors and the threshold. Then they are finally grouped into clusters (individual stains) as explained in the earlier section. Every region  $C_i$  (where  $0 < i < n$ , and  $n$  is the total number of regions) has the label of the color palette to which it belongs and the total count of pixels in that region. The percentage of color in each stain is calculated. Then the conversion to the required units is done and the result is displayed. The result can also be saved into a text file.

Figure 5, explains the methods using the UML diagram. Each block in the UML diagram specifies the functions and the sequence in which they are performed.

## 5. CONCLUSION

In this paper an efficient directed Fuzzy c-means approach for analysis of stained tissues has been proposed. A special merit of this work is that the parameters of directed fuzzy c-mean scheme used, namely, number of clusters, hue for each color and the lower threshold value for the size of the cluster, can be controlled by the user. This system is being implemented using java as the language and the output of the system is a set of values indicating the number of stains for each color in the system and the size of the individual stains. Another salient feature of the system is that it can be used to analyze other images as well; vis-à-vis; satellite images, thermal images, etc. The performance evaluation of separate modules of this system looks promising; however, a complete evaluation of the system will be performed upon completion of the implementation.

## 6. References

- [1]. Nikhil Pal and Sankar k Pal, "A Review on Image Segmentation Techniques," *Pattern Recognition*, vol. 26, No. 9, pp. 1277-1294, February 1993.
- [2]. Paul s. Heckbert, "Color Image Quantization for Frame Buffer Display," in *SIGGRAPH '82*, pp. 297-307, July 1982.
- [3]. Michael T. Orchard, "Color Quantization of Images," *IEEE Transactions on Signal Processing*, vol. 39, NO. 12, December 1991.
- [4]. David x. Zhong, "Color Space Analysis and Color Image Segmentation," *ACM International Conference*, pp.47-49, 2000.
- [5]. Vitorino Ramos, Fernando Muge, "Map Segmentation by Color Cube Genetic k-Mean Clustering," *4<sup>th</sup> European Conference, EC DL 2000*, pp. 319-323, June 2003.
- [6]. Dogan Ozwmir and Lale Akarun, "A Fuzzy Algorithm for Quantization of Images," *Pattern Recognization*, vol. 35, pp. 1785-1791, 2002.
- [7]. Alan Wee-chung Liew, Wing Hong Lau and Shu Hung Leung, "Segmentation of Color Lip Images by Spatial Fuzzy Clustering," *IEEE Transactions on Fuzzy Systems*, vol. 11, No4, pp. 542-549, August 2003.
- [8]. Liyan Zhang, "Professional Paper: Comparison of Fuzzy c-means Algorithm and New Fuzzy Clustering and Fuzzy Merging Algorithm," University of Nevada, Deaprtment of Computer Science, Technical Report, May 2001.
- [9]. Krishna k. Chintalapudi and Moshe Kam, "The Credibilistic Fuzzy C Mean Clustering Algorithm," *IEEE SMC'98*, pp. 2034-2040, 1998.
- [10]. Chuai – aree, Lurisinsap, Preaphon and Siripant, " Fuzzy c-Mean: A Stastical Feature Classification of Text and Image Segmentation Methods," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 9, No. 6, pp. 661-671, 2001.
- [11]. M.A Egan, M. Krishnamoorthy and K. Rajan, " A Java Simulation Tool for Fuzzy Clustering," *Concurrency: Practice and Experience: Special Issue on Java for Science and Engineering9*, pp.1327-1322, 1997.
- [12]. Uthai Sangthongpraow, Punya Thitimajshima and Y. Rangsanseri, " Modified Fuzzy C-mean for Satellite Image Segmentation," *AARC ,ACRS 1999, Image processing*, <http://www.gisdevelopment.net/aars/acrs/1999/ts11/ts11466pf.htm>.
- [13]. Abhishek Krovi and Shahram Rahimi, "A Distributed Approach to Content Based Image Retrieval," *The 2003 International Conference on Parallel and Distributed Processing Techniques and Applications (PDPTA'03), Las Vegas, Nevada*, pp. 458-463, 2003.
- [14]. F.Masulli, M.Artuso, P. Bogus, A. Schenone, "Fuzzy Clustering Methods for the Segmentation of Multivariate Medical Images," *Computrized Medical Image and Graphics*, vol. 20, pp. 119-129, 1996.