CBMIR: SHAPE-BASED IMAGE RETRIEVAL USING CANNY EDGE DETECTION AND K-MEANS CLUSTERING ALGORITHMS FOR MEDICAL IMAGES

B.Ramamurthy[1],

1. PhD Research Scholar,
Department of Computer and Information Sciences
PSG College of Technology, Coimbatore 641004

K.R.Chandran[2]

2. Professor,
Department of Computer and Information Sciences
PSG College of Technology, Coimbatore 641004

ABSTRACT

The accumulation of large collections of digital images has created the need for efficient and intelligent schemes for classifying and retrieval of images. In this work, "CBMIR: Shape-Based Image Retrieval Using Canny Edge Detection and K-Means Clustering Algorithms for Medical Images", has been developed to retrieve the medical images from huge volume of medical databases. This requires the preprocessing, feature extraction, classification, retrieval and indexing steps in order to develop an efficient medical image retrieval system. In this work, for preprocessing step, the image segmentation method has been carried out, for feature extraction, basic shape feature has been extracted using canny edge detection algorithm, and for classification, K-means classification algorithm has been used. For retrieval of images, Euclidian distance method values are calculated between query image and database images. The goal of this work is to provide a medical image retrieval system for further use of medical diagnosis purpose in the field of medical domain.

Keywords: CBMIR, Canny Edge Detection Algorithm (CEDA), K-means classification algorithm, Euclidian distance method (EDM)

1. Introduction

Content-Based Image Retrieval (CBIR) refers to image retrieval system that is based on visual properties of image objects rather than textual annotation. Contents of an image can be of various forms like, texture, color, and shape etc. In this work, shape is selected as a primary feature in indexing the image database. CBIR is more robust and makes it easier for image retrieval [1].In CBIR system; the processing steps are getting the input images, extracting the feature of the images, classifying the images and finally storing the images in an image feature database which is available for on-line services in order to retrieval of similar images from the feature database.

CBIR for medical images has become a major necessity with the growing technological advancements. The contents of an image have to be carefully extracted, classified with efficient techniques for easy retrieval. Contents in an image can be of various forms like, texture, color, and shape etc of which shape is regarded as the most efficient metric. Medical images are usually fused, subject to high inconsistency and composed of different minor structures. So there is a necessity for feature extraction and classification of images for easy retrieval. There are various methods have been proposed for medical image retrieval system using shape feature and shape based classification which uses shape based model description and shape similarity measures [2,3].

The main objective of the work is to retrieve the images from huge volume of medical databases with high accuracy by performing feature extraction, classification process. So that the retrieved images are used for various medical diagnosis purpose. This paper organized as follows: The existing methods and its related literature survey presented in section 2, then the proposed method and its related works presented in section 3 and in section 4 implementation and result analysis of the work has been described.

2. Related Work

Medical images play a pivotal role in surgical planning, medical training, and patient diagnoses. In large hospitals thousands of images to be managed every year [4].These images needs to be indexed, classified, and searched for easy retrieval. Since there was lot of work has been done in this area, a comprehensive survey of CBMIR has been conducted and which is presented in the following section.

The Image Retrieval for Medical applications (IRMA) project1 undertaken at the Aachen University of technology[5, 6] 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. The IRMA system lacks the ability for finding particular pathology that may be localized in particular regions within the image.

The Spine Pathology and Image Retrieval System (SPIRS) [7,8] 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. In the SPIRS system, the images in the collection must be homogeneous.

The Image Map [9] is one of the 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 is likely to occur and therefore, the retrieved image will represent an unexpected organ.

The Automatic Search and Selection Engine with Retrieval Tools (ASSERT) [10] A physician-in-the-Loop content-based retrieval system for HRCT image databases which is implemented to show a human-in-the-loop 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.

3. Proposed Model

The proposed CBMIR framework is shown in Fig.1.Initially medical images were taken as input to the system and preprocessing of the images carried out in order to improve the flexibility of the images for further processing of the system

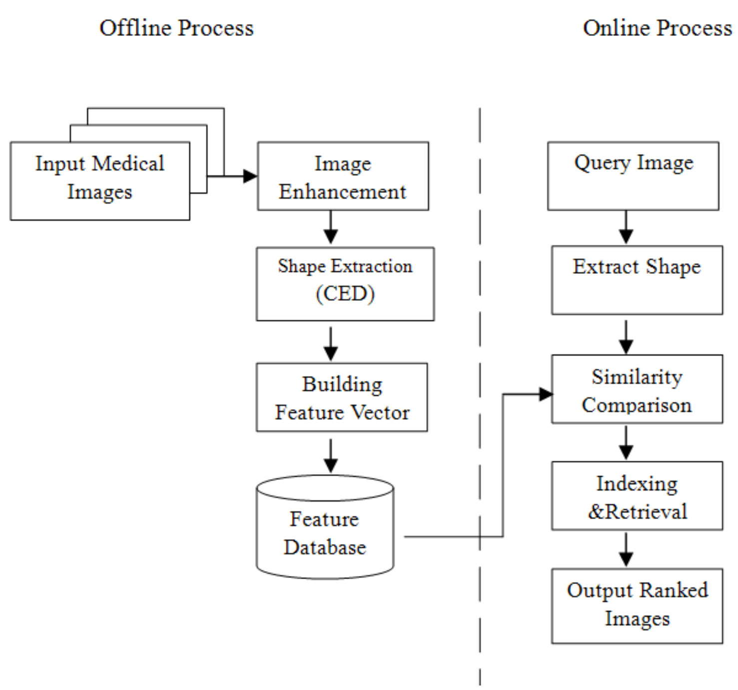

Fig. 1 Proposed CBMIR Framework

In proposed system there are two parts of the process has been carried out namely offline and online. In offline process, Image feature database has been constructed through one of the dominant feature of the image such as shape extraction by using Canny Edge Detection (CED) algorithm. In online process, Graphical User Interface (GUI) for the user interaction has been developed through which user can interact with the system for retrieval of their desired images from the image database. For retrieval process, similarity comparison technique has been carried out between online user query image and offline image feature database. After comparison, resulting images are indexed and retrieved based on their rank.

3.1. Shape

Shape is an important and most powerful feature used for image classification, indexing and retrievals. Shape information extracted using histogram of edge detection. In this work, the edge information in the image is obtained by using the canny edge detection [11].Other techniques for shape feature extraction are elementary descriptor, Fourier descriptor, template matching, Quantized descriptors etc. Canny edge detection outperforms many of the newer algorithms that have been developed in the industry [12].

Algorithm: For detecting edges using Canny Edge Detection Algorithm

Step 1: Smoothing: Smooth the image with a two dimensional Gaussian. In most cases the computation of a two dimensional Gaussian is costly, so it is approximated by two one dimensional Gaussians.

Step 2: Finding Gradients: Take the gradient of the image this shows changes in intensity, which indicates the presence of edges. This actually gives two results, the gradient in the x direction and the gradient in the y direction.

Step 3: Non-maximal suppression: Edges will occur at points where the gradient is at a maximum. The magnitude and direction of the gradient is computed at each pixel.

Step 4: Edge Threshold: The method of threshold used by the Canny Edge Detector is referred to as “hysteresis”. It makes use of both a high threshold and a low threshold.

Step 5: Thinning: Using interpolation to find the pixels where the norms of gradient are local maximum.

The following Figure 2(a) and 2(b) in Fig 2 depicts the results of before and after feature extraction of the work for a brain image.

Input image

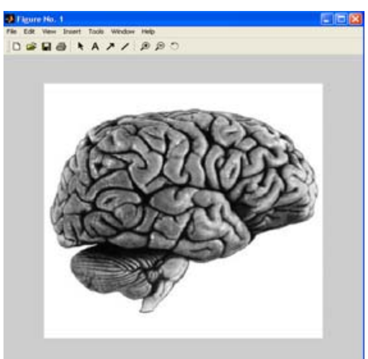

Fig. 2 (a) Brain Image

Output image

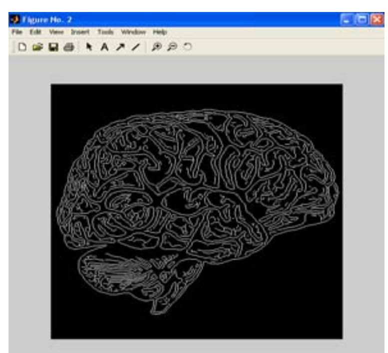

Fig. 2 (b) Edge detected Brain Image

Fig. 2 Sample Edge Detected Brain Image

3.2 Image Classification

Image classification is one of the important steps in image retrieval process because it saves more time while searching the images from huge volume of database. Classification is identification of different regions of the image by which the retrieval efficiency of the system will be improved. A commonly used classification algorithm is the k-means algorithm. In this work, we have used k-means algorithm for identifying different regions of the images in database. So that the performance of the retrieval process has been improved by comparing the classified images with user’s query image. It is an efficient centroid based algorithm which has been widely used in various key areas like micro array dataset [13], high dimensional data sets [14] and especially in image retrieval system [15] etc. The k-means algorithm takes the input parameter k, and partitions a set of n objects into k clusters so that the resulting intra cluster similarity is high but the inter cluster similarity is low. Given a set of observations (x1, x2… xn), where each observation is a d-dimensional real vector, the k-means clustering aims to partition the n observations into k sets (k < n) S = {S1, S2… Sk} so as to minimize the within-cluster sum of squares

$$\underset{s}{\arg\min} \sum_{i=1}^{k} \sum_{xj \in si} \|X_j - \mu_i\|^2$$

Eq. (1)

Where μi is the mean of points in Si.

Algorithm: For K-means clustering Algorithm
The k-means algorithm for partitioning based on the mean value of the objects in the cluster
Input: The number of clusters k and a database containing n objects.
Output: A set of k clusters that minimizes the squared-error criterion.
Method:
Step 1: Enter the number of clusters.
Step 2: Randomly guess k cluster Center locations.
Step 3: Each data point finds out which Center it's closest to.
Step 4: Each Center finds the centroid of the points it owns.
Step 5: Center now moves to the new centroid.
Step 6: Repeat until terminated.

3.3. Similarity Comparison

In this work, for retrieval of the images, similarity comparison technique has been computed. For similarity comparison, we have used Euclidean distance, d using the following equation

$$d = \sqrt{\sum_{i=1}^{N} (F_Q[i] - F_{DB}[i])^2}$$

Eq. (2)

Where FQ[i] is the ith query image feature and FDB [i] is the corresponding feature in the feature vector database. Here N refers to the number of images in the database. [16]

4. EXPERIMENTAL SETUP AND RESULT ANALYSIS

A Intel® Core 2 Duo CPU Workstation with 2GB RAM computer has been used for conducting the experiments. The main browser tool Mozilla Firefox 4.0 Beta1 version was used for developing User Interface components as front end, MATLAB 7.2.0-Image Processing tool Box-Workspace was used as feature database for storage as back end and for image processing work, other MATLAB 7.2.0 utilities were used. For mathematical equations, Math Type tool was also used for writing document. Initially, MATLAB 7.2.0 workspace database with 1000 real time medical images were used for testing the proposed CBMIR system.

4.1. Retrieval Efficiency

For retrieval efficiency, traditional measures namely precision and recall were computed with 1000 real time medical images [17, 18]. Standard formulas have been computed for determining the precision and recall measures.

Precision (P) is the ratio of the relevant images to the total number of images retrieved

$$P=r/n1$$

Eq. (3)

Where,
r-number of relevant images retrieved
n1-total number of images retrieved

Recall(R) is the percentage of relevant images among all possible relevant images

$$R = r / n2 \qquad \text{Eq. (4)}$$

Where,
r-number of relevant images retrieved
n2-total number of relevant images in the database
By randomly selecting some sample query images from the MATLAB-Image Processing tool Box-Workspace Database, the system was tested and the results are shown in the following Table 1.

TABLE 1

PRECISION AND RECALL VALUES IN %

Query Image	Shape	
	Precision	Recall
Image 1	45.0	12.0
Image 2	60.0	52.0
Image 3	15.0	07.0
Image 4	85.0	30.0
Image 5	45.0	20.0

The following Fig 3(a,b,c,d,e, and f) in Fig 3 shows some of the sample images of the work in various steps of Canny Edge Detection Algorithm and Fig 4 shows Sample snapshot of the retrieval system user interface screen.

CANNY EDGE DETECTION
Step 1: SMOOTHING
X Derivative:

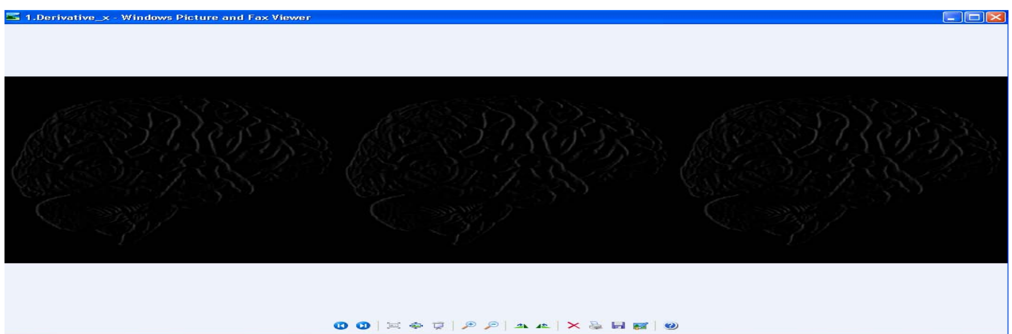

Fig. 3 (a) Sample Image Smoothing for X Derivative

Y Derivative:

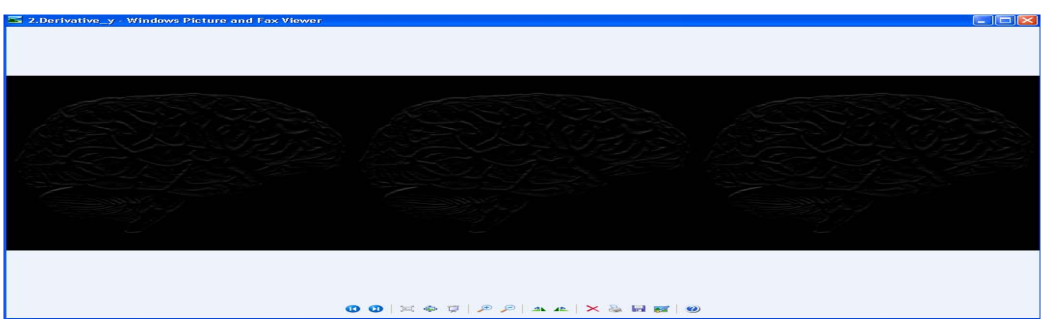

Fig. 3 (b) Sample Image Smoothing for Y Derivative

Step 2: FINDING GRADIENT
Gradient:

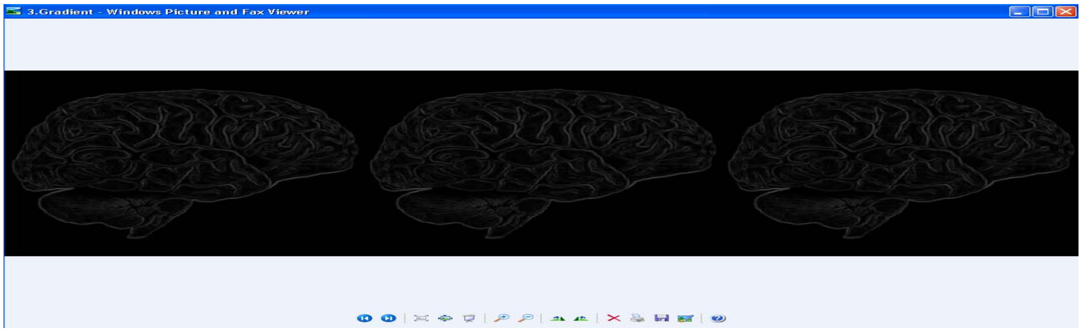

Fig. 3 (c) Sample Image for Finding Gradient

Step 3: NON MAXIMUM SUPRESSION
Non Maximum Suppression:

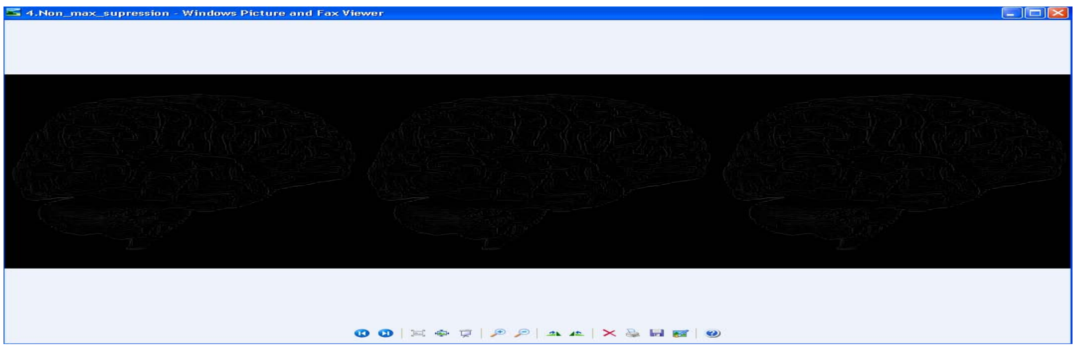

Fig. 3 (d) Sample Image for Non Maximum Suppression

Step 4: THRESHOLDING AND HYSTERESIS:

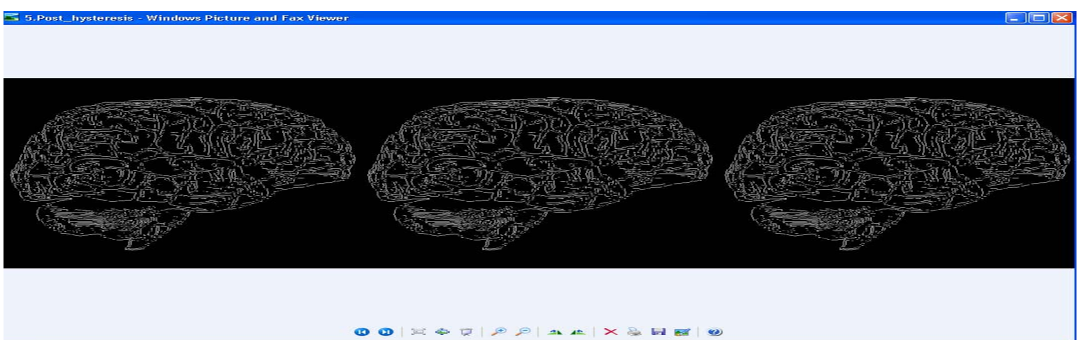

Fig. 3 (e) Sample Image for Thresholding and Hysteresis

Step 5: EDGE DETECTION

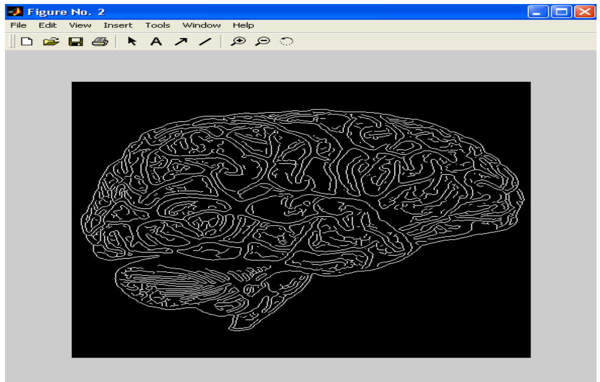

Fig. 3 (f) Sample Edge Detected Image for Canny Edge Detection Algorithm

Fig. 3 Sample Edge Detected Brain Image in various steps in Canny Edge Detection Algorithm

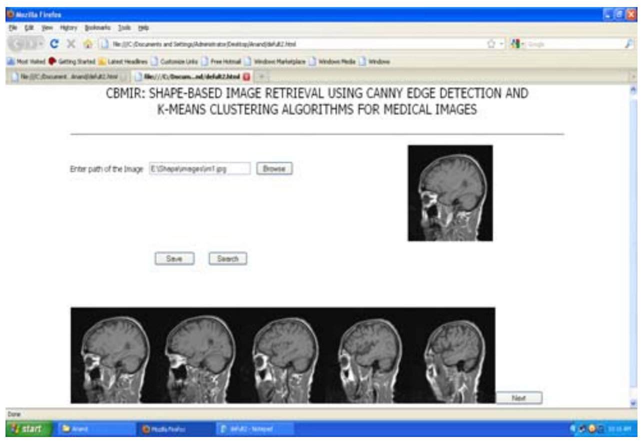

Fig.4 Sample Snapshot of the CBMIR Model Using Shape Feature.

5. CONCLUSION AND FUTURE ENHANCEMENT

In this paper, a novel approach is proposed for medical image retrieval based on shape feature, which uses canny edge detection algorithm for extraction of image shape and K-means algorithm for extraction of different regions of the image in order to improve better matching process between user query image and feature database images. We have shown that Canny Edge Detection and K-means clustering algorithms are quite useful for retrieval of relevant images from image database. Our results indicate that the proposed approach offers significant performance improvements in retrieval of medical images.Further, by fine tuning of shape feature extraction and using other shape feature extraction methods, performanece of the retrieval process can be improved more.

6. REFERENCES

[1] V.S.V.S. Murthy, E.Vamsidhar, J.N.V.R. Swarup Kumar, P.Sankara Rao, “Content Based Image Retrieval using Hierarchical and K-Means Clustering Techniques”, International Journal of Engineering Science and Technology, Vol 2(3), pp. 209-212, 2010.

[2] Roberto Paredes, Daniel Keysers, Thomas M. Lehman, Berthold Wein, Hermann Ney, and Enrique Vidal, “Classification of Medical Images Using Local Representations”, Bildverarbeitung fur die Medizin , pp. 171-174, 2002.

[3] Wei Zhang, Sven Dickinson, Stanley Sclaroff , Jacob Feldman, Stanley Dunn, “Shape- Based Indexing in a Medical Image Database”, Biomedical Image Analysis, pp-221 – 230, 26-27 Jun 1998.

[4] Muller, H., N.Michoux, D.Bandon and A.Geissbuhler, “A review of content based image retrieval systems in medical applications-clinical benefits and future directions”, International Journal of Medical Information., pp: 73:13, 2004.

[5] Thies C, Guld MO, Fischer B, Lehmann TM, “Content- based queries on the CasImage database with in the IRMA framework”, Lec Notes in Comp Sci 2005; 3491:781-92.

[6] Antani S, Long LR, Thoma GR, “Content-based Image retrieval for large biomedical image archives. In: Proc 11th World Cong Medical Informatics: 2004. p. 829-33.

[7] Long LR, Antani SK, Thoma GR, “Image Informatics at a national research center, CompMed Imaging &graphics 2005; 29:171-93.

[8] Thoma GR, Long LR, Antani SK, Biomedical Imaging research and development: knowledge from images in the medical enterprise. Technical Report Lister Hill National Ctr for Biomedical Communications, US National Library of Medicine, NIH 2006;LHNCBC-TR-2006-002.

[9] Petrakis, Euripidies G. M and C. Faloutsos, "ImageMap: An Image indexing Method Based on Spatial Similarity", IEEE Trans on Knowledge and Data Eng., 14(5):979-987, 2002.

[10] Chi-Ren Shyu, Carla E. Brodley, Avinash C. Kak, Akio Kosaka,"ASSERT: A Physician-in-the-Loop Content-Based Retrieval System for HRCT Image Databases", Computer Vision and Image Understanding, Vol. 75, Nos. 1/2, July/August, pp. 111-132, 1999.

[11] Canny, J., "A computational approach to edge detection", IEEE Trans on Pattern Analysis and Machine Intelligence, 8:679-698, 1986.

[12] Ehsan Nadernejad, Sara Sharifzadeh, Hamid Hassanpour, "Edge Detection Techniques: Evaluations and Comparisons", Applied Mathematical Sciences, no. 31, pp.1507 – 1520, Vol. 2, 2008.

[13] Bernard Chen, Phang C. Tai, R. Harrison1 and Yi Pan1, "Novel Hybrid Hierarchical-K-means Clustering Method (H-K-means) for Microarray Analysis", Computational Systems Bioinformatics Conference, 2005, Workshops and Poster Abstracts. IEEE, pp105 – 108, 8-11 Aug. 2005.

[14] Madjid Khalilian, Norwati Mustapha, MD Nasir Suliman, MD Ali Mamat, "A Novel K-Means Based Clustering Algorithm for High Dimensional Data Sets", Proceedings of the International MultiConference of Engineers and Computer Scientists 2010 Vol I, IMECS 2010, March 17-19, 2010.

[15] Hong Liu , and Xiaohong Yu, "Application Research of k-means Clustering Algorithm in Image Retrieval System", Proceedings of the Second Symposium International Computer Science and Computational Technology(ISCSCT '09), Huangshan, P.R. China, pp.274-277, 26-28, Dec. 2009.

[16] S. Nandagopalan, Dr. B. S. Adiga, and N. Deepak "A Universal Model for Content-Based Image Retrieval" World Academy of Science, Engineering and Technology 46 2008

[17] Mohammed Eisa and Ibrahim Elhenawy and A. E. Elalfi and Hans Burkhardt, "Image Retrieval based on Invariant Features and Histogram Refinement", ICGST International Journal on Graphics, Vision and Image Processing, March 2006, pp. 7-11.

[18] M. Flickner, et al, "Query by Image and Video Content: The QBIC System", IEEE Computer, 28(9): 23-32, 1995.