International Journal of Research in Advent Technology (IJRAT) Special Issue
E-ISSN: 2321-9637
Available online at www.ijrat.org
*International Conference "*INTELINC 18*", 12th & 13th October 2018*

Statistical Shape Features Based MRI Image Retrieval System Using PCA and K-Means Clustering Algorithm

Dr. N. Kumaran
Department of Computer Science and Engineering
Annamalai University (Deputed to GPTC, Karur)
Chidambaram, India
kumaran81@gmail.com

Dr. R. Arunmozhi
Department of Electical and Electronics Engineering
Annamalai University (Deputed to GPTC, Karur)
Chidambaram, India
Arunau2006@gmail.com

Abstract— The enormous number of medical image attainment and storage systems in medical field raises burden for new retrieval methods. This directs to get back related medical images based on its visual data. This work deals with the concept of shape based MRI image retrieval system using Canny Edge detection. The PCA shape feature selection is used here to remove the difficulties of feature vector formation. Then K-Means algorithm and Euclidean distance measure are applied to retrieve most excellent MRI images for the given input image. The results obtained by this method confirm its best cluster accuracy and retrieval efficiency.

Keywords—Canny edge detection; Euclidean distance; K-Means clustering; PCA; Shape features

I. INTRODUCTION

Content Based Image Retrieval (CBIR) system is a type of framework which retrieves images based on its features such as colors, texture and shape of the image. The Content Based Medical Image Retrieval (CBMIR) methods are medical field explicit search engine for clinical image databases, which identifying and retrieving clinical images according to their visual data [1-3]. The most imperative intent of the CBMIR system is to retrieve the medical images from the vast volume of databases with high accuracy by carry out feature detection, classification and distance measure process. The retrieved images are used in support of various medical diagnostics.

This paper, proposed the MRI scan image retrieval system in various parts of the human body such as the spine, brain, abdomen and knee using Canny Edge (CE) detection [4] based shape features with Principal Component Analysis (PCA) [5] feature selection algorithm, K-Means clustering [6] and Euclidean distance measure [7]. The accuracy, precision and recall rate of this method is very good compared with using all twenty extracted shape features.

The next part of the paper depicts related works of the system. The short discussion on the proposed work and shape feature detection is in sections 3 and 4. Sections 5 and 6 give details about feature selection and K-Means clustering algorithm. Section 7 deals the best image retrieval. Section 8 explains the experimental results. In section 9, the conclusion of the proposed work with future prospects is specified.

II. RELATED WORKS

There are different existing frameworks that afford different techniques and algorithms for content based medical image retrieval. The main intention of all these frameworks is to prove the improvement of results so as to assist the doctors and radiologists in diagnosis of treatments.

In [8], they expressed a medical image retrieval framework with 90% of precision and recall rate. Here medical images were segmented into several sub images using C-mean clustering algorithm and detecting three gray level features using color moments. Then the sub images were changed to binary image and 7 shape features and 4 texture features were

extracted using co-occurrence matrix. Then the genetic algorithm was applied to pick most favorable features.

Peiqiang Zhang et al. [9] proposed a system using co-occurrence matrix to extract texture data and edge histogram to extract shape data of clinical images. Then Euclidean distance measure was used for image retrieval. Results of experimentation showed that the system had a recall rate of ninety percentages and applied to medical image retrieval with suitable effect. Wei Zhang et al. [10] investigated the statistical characteristics of the modal shape description in classifying nonrigid shape deformations due to various pathologies reflected in dental images. Good classification performance based on modal features is achieved.

In [11], authors proposed a novel descriptor which extended edge histogram method of WAN, collectively local data with global shape data, combined edge of entire image with edge density of its sub-images; it is called as Edge Density Histogram Descriptor (EDHD). Then Support Vector Machine (SVM) and Euclidean distance measure are applied for medical image classification and retrieval.

K Trojacanec et al. [12] used the Edge Histogram algorithm and by MPEG-7 standard to MRIs retrieval. Experimental outcomes proved the Edge Histogram Descriptor attained higher precision. They also combined both methods by averaging their output ranks. Examination showed that averaging the output ranks gives high precision. In [13], a medical image retrieval method had been developed using Generic Fourier shape descriptor. The retrieval efficiency of the descriptor was verified on the huge medical image databases and quantified by the typical precision and recall measures. From the retrieval results, they accomplished that the Generic Fourier Descriptor have best retrieval performance for the medical images. In our previous work [14], we used a texture spectrum to get texture features and edge histogram to obtain shape features of MRI images. Then K-Mean clustering algorithm and Manhattan distance measure is applied for MRI image retrieval. This method in this paper enhanced the recall and precision rate for medical image retrieval system.

III. PROPOSED WORK

The block diagram of the proposed CBMIR system is shown in Fig.1 in which different parts of the human body MRI scan images such as the spine, knee, abdomen and brain are used to create the training data set.

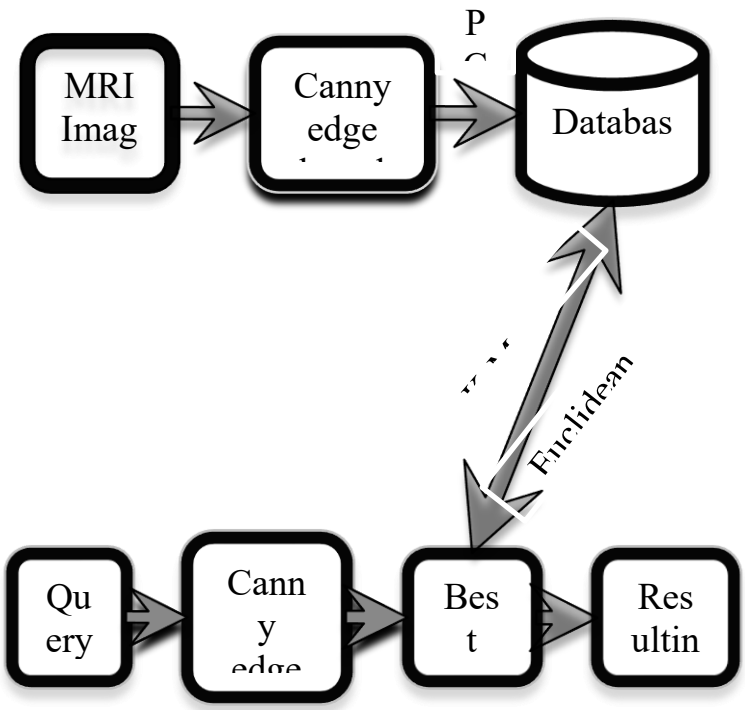

Fig.1. The design of the proposed retrieval system

In this work, we discover edges using Canny Edge detection algorithm to every MRI scan image in the training database. Then twenty shape feature vector values are extracted as feature components and PCA feature selection algorithm is used to create an optimized database. Subsequently, using K-Means clustering the training images are clustered by means of the optimized shape features. When a testing MRI scan image is submitted, the optimized shape features are extracted. After that, for similarity comparison between the query MRI scan image and the clustered MRI scan images, a Euclidean distance function is used. The closest Euclidean distance values for the query image are ranked and the best MRI scan images are retrieved for medical treatment.

IV. SHAPE FEATURE EXTRACTION

The Canny Edge detection algorithm is known as the optimal edge detector. The canny edge detector smoothes the image by eliminate noise. After that it locates the image gradient to emphasize regions with high spatial derivatives. The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum. The gradient array is now further reduced by hysteresis. Hysteresis is used to track along the remaining pixels that have not been suppressed. Hysteresis uses two thresholds and if the magnitude is below the first threshold, it is set to zero. If the magnitude is above the high threshold, it is made an edge. If the magnitude is between the

high and low thresholds, then it is set to zero unless there is a path from this pixel to a pixel with a gradient. The Fig. 2 shows sample canny edge detection for the MRI scan images.

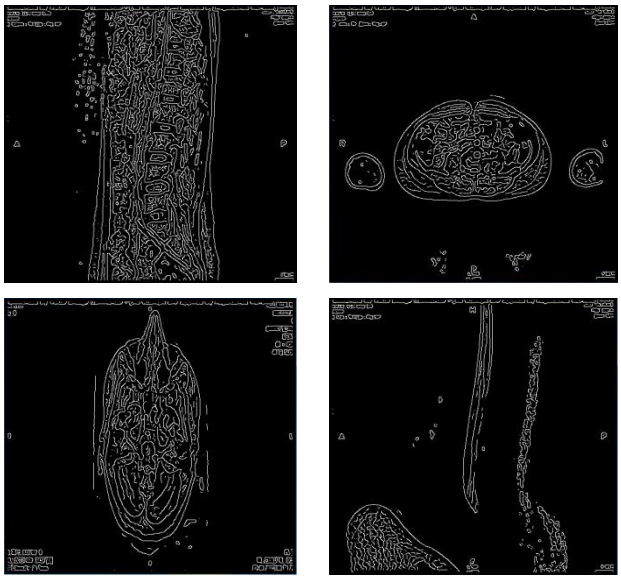

Fig.2. Sample canny edge detected MRI scan images

After detecting the edges the twenty statistical shape features like area, mean, standard deviation, centroid based on X and Y coordinates, center of mass based on X and Y coordinates, perimeter, circularity, Feret's diameter (min & max), Feret's angle and starting coordinates, integrated density, median, skewness, kurtosis, aspect ratio and roundness are extracted. Fig. 3 shows the output screen for shape feature extraction of a knee MRI scan image.

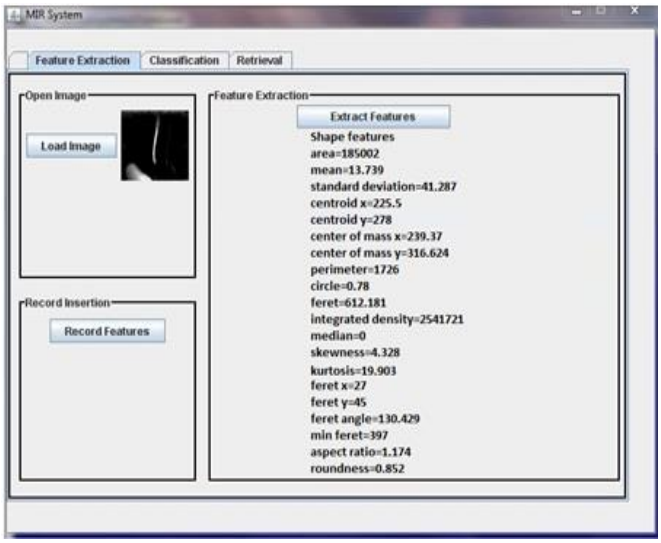

Fig.3. Output screen for shape features extraction

V. FEATURE SELECTION

After extracting twenty shape features, the database is normalized using the Z-Transform [15] - [17] and rescales the feature values. The normalization $Z = \frac{(X-\mu)}{\sigma}$. Where, x is a shape feature value, μ and σ is the mean and stranded deviation of the shape feature value. Then, we have used Principal Component Analysis as a feature selection algorithm. PCA is useful when we have obtained features on large number of attributes and believe that there is some redundancy in those features. In our case, redundancy means that some of the features are correlated with one another, possibly because they are measuring the same construct.

We have selected eight best features (i.e) area, standard deviation, Feret's diameter (min & max), Feret's starting coordinates (X & Y), Feret's angle and perimeter for MRI scan image retrieval using PCA with variance as 0.95. So, instead of using twenty shape features we are using only eight shape features for best MRI image retrieval.

VI. K-MEANS CLUSTERING

K-Means clustering algorithm is the method of partitioning a cluster of data into a least number of groups. In this work, the ten categories of normal and abnormal training images are given in a 8-D metric space and determine a partition of the images into maximum twenty clusters, such that the images in a cluster are more similar cases to each other than two images in different clusters.

We initialize 20 clusters by randomly picking one image to represent each cluster. Each of the remaining images is assigned to a cluster and the clustering condition is applied to calculate the cluster average. This average value is used as the new centroid and each and every image is reassigning to the cluster that it is very much similar to. This carry on until there is no modification in the centroids.

VII. IMAGE RETRIEVAL

The Euclidean Distance (ED) is calculated between the query image and the clustered images. If x_i and yi are 2D feature vectors of the clustered training images and query image respectively then the distance measure is defined as,

$$ED_{(x,y)} = \sqrt{\sum_{i=1}^{d}(x_i - y_i)^2}$$

The calculated distances are sorted in increasing order and display the first 10 images as the best similar MRI scan images for medical treatment. The sample output screen for MRI brain image retrieval is shown in Fig.4.

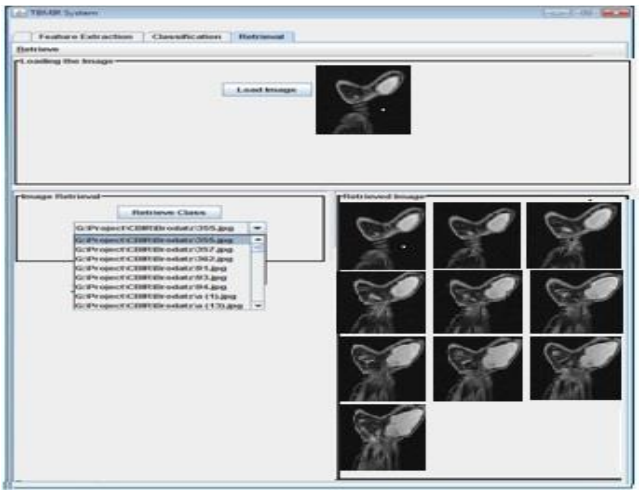

Fig.4. Best Retrieved Spine MRI Scan images

VIII. EXPERIMENTS AND RESULTS

This technique is implemented using JAVA programming language and Ms-Access as the database. In this work, we used around 1850 MRI scan images as a training data set such as 900 spine, 450 brain, 350 abdomen and 250 knee images and 150 MRI scan images as a testing data set such as 50 spine, 50 brain, 25 abdomen and 25 knee images with the size of 256 x 256 as an image database.

The nature of clustering of the system is to cluster the normal and abnormal human body MRI scan images using 1850 training data set. The K-Means clustering gave an examination accuracy of 81.2% while using 20 statistical shape features. The proposed PCA based feature selection algorithm with 8 shape features gave the examination accuracy of 92.4%. The effectiveness of the K-Means clustering algorithm between Canny Edge based shape features and Canny Edge based shape features with PCA given in Table I.

TABLE I. Comparison of Accuracy, Sensitivity & Specificity

Methods	Accuracy %	Sensitivity %	Specificity %
CE based shape features	81.2	75.9	84.7
CE based shape features + PCA	92.4	88	96.4

Fig.5 shows the empirical receiver operating characteristic curves in support of various cutoff points with a false positive rate on the X-axis and true positive rate on the Y-axis.

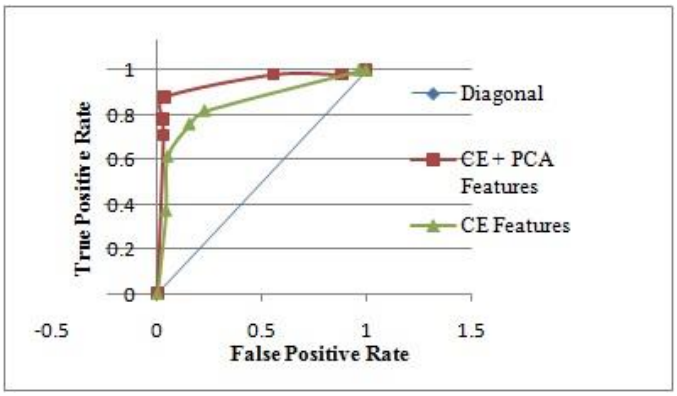

Fig.5. Empirical ROC curves

The effectiveness of the proposed method can be measured by recall and precision, which are often referred together since they measure the different aspects of the system performance. Recall is defined as the relation between the total number of retrieved relevant images and the number of relevant images in the database. Precision dealings the retrieval accuracy and is defined as the relation among the number of retrieved relevant images and the number of total retrieved images. The average of precision and recall graph is given in Fig. 6.

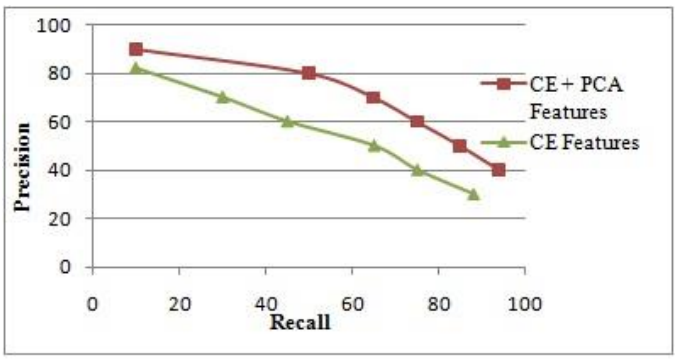

Fig.6. Precision and Recall graph

IX. CONCLUSION

In this paper, we have proposed an efficient MRI scan image retrieval system using Canny Edge and PCA based optimized statistical shape features. The experimental results demonstrate that the proposed method has the best accuracy, precision and recall rate than usual statistical shape features based MRI scan image retrieval methods. We have planned to extend our work with all types of human body scan images.

REFERENCES

[1] H. Muller, N. Michoux, D. Bandon & A. Geissbuhler, "A review of content-based image retrieval systems in medical applications: clinical benefits and future directions", International journal of Medical Informatics, vol. 73, 1-23, 2004.

[2] X. S. Zhou, S. Zillner, M. Moeller & *et. al.*, "Semantics and CBIR: A Medical Imaging Perspective", ACM Conf. on Content-based Image and Video Retrieval, 571-580, 2008.

[3] Ceyhun Burak Akgül, Daniel L. Rubin, Sandy Napel, Christopher F. Beaulieu & *et. al.*, "Content-Based Image Retrieval in Radiology: Current Status and Future Directions", Journal of Digital Imaging, vol. 24(2), 208-222, 2011.

[4] B. Ramamurthy & K. R. Chandran "CBMIR: Shape-Based Image Retrieval Using Canny Edge Detection and K-Means Clustering Algorithms for Medical Images", International Journal of Engineering Science and Technology, Vol. 3(3), 1870-1877, 2011.

[5] Usha Sinha, Hooshang Kangarloo, "Principal Component Analysis for Content-based Image Retrieval", RadioGraphics, vol. 22, 1271–1289, 2002.

[6] Gobert N. Lee & Hiroshi Fujita, "K-means clustering for classifying unlabelled MRI data", IEEE Conference of the Australian Pattern Recognition Society on Digital Image Computing Techniques and Applications, pp. 92-98, 2007.

[7] Swarnambiga Ayyachamy,Vasuki S., & Manivannan, "Distance measures for medical image retrieval, International Journal of Imaging Systems and Technology", vol. 23(1), 9–21, 2013.

[8] Hong Shao, Wen-cbeng Cui & Hong Zhao, "Medical Image Retrieval Based on Visual Contents and Text Information", IEEE International Conf. on Systems, Man and Cybemeties, pp. 1098-1103, 2004.

[9] Peiqiang Zhang & Hongguang Zhu, "Medical Image Retrieval Based on Co-Occurrence Martix and Edge Histogram", IEEE Conf. on Multimedia Technology, pp. 5434-5437, 2011.

[10] Wei Zhang, Sven Dickinson, Stanley Sclaroff, Jacob Feldman and Stanley Dunn, "Shape – Based Indexing in a Medical Image Database", Biomedical Image Analysis, pp. 221-230, 1998.

[11] Fu Li-dong & Zhang Yi-fei, "Medical Image Retrieval and Classification Based on Morphological Shape Feature" IEEE International Conf. on Intelligent Networks and Intelligent Systems, pp. 116-119, 2010.

[12] K Trojacanec, G Mirceva & D Davcev, "Application of Edge Histogram Descriptor and Region Shape Descriptor to MRIs", ICT Innovations, Springer, 2010.

[13] A. Grace Selvarani & S. Annadurai, "Content Based Image Retrieval For Medical Images Using Generic Fourier Descriptor", Journal of Computational Intelligence in Bioinformatics, vol. 1(1), pp. 65-72, 2008.

[14] N. Kumaran, Dr. R. Bhavani & E. Elamathi, "MRI Image Retrieval based on Texture Spectrum and Edge Histogram Features", IEEE International conference on Communication and Signal Processing, pp. 1059-1063, 2013.

[15] John Collins & Kazunori Okada "A Comparative Study of Similarity Measures for Content-Based Medical Image Retrieval", CLEF Online Working Notes/Labs/Workshop, 2012.

[16] N. Kumaran & Dr. R. Bhavani "PCA Based Feature Selection for MRI Image Retrieval System Using Texture Features", Advances in Intelligent Systems and Computing, Springer Link, Vol.324, pp. 109-117, 2015.

[17] Chua Sook Ling, Wan Mimi Diyana Wan Zaki, Aini Hussain, W. Siti Halimatul Munirah W. Ahmad and Erica Yee Hing "Shape based image retrieval system for MRI spine", International Conference on Electrical Engineering and Informatics (ICEEI), pp.1-6, 2017.