ISSN: 0975-8585

Research Journal of Pharmaceutical, Biological and Chemical Sciences

An Efficient Image Recognizing and Data Retrieval Technique for Medical Image Database using Principle Component Analysis

Suvel Rathneswar M[1], and Rosline Nesa Kumari G[2]*

[1]UG Student,Department of CSE,Saveetha School of Engineering,Chennai,Tamilnadu,India.
[2]Professor, Department of CSE,Saveetha School of Engineering,Chennai,Tamilnadu,India.

ABSTRACT

Medical imaging technology has revolutionized health care over the past 30 years,allowing doctors to find disease earlier and improve patient outcomes. Diagnosing and treatment using medical imagesalong with the internet technologyis gaining broader scopes, so there is a need for an image storage and data retrieval technique that can fetch related data of the image. This type of technique provides great assistance for doctors in the field of medical research. This paper presents a content based retrieval technique for medicalimage using Principle component analysis (PCA) The efficient retrieval technique evaluated on functional magnetic resonance(fMRI)images.With various fMRI images the proposed method effectiveness is evaluated.

Keywords: Medical image , Image processing, fMRI,PCA

Corresponding author

ISSN: 0975-8585

INTRODUCTION

In recent years, the medical imaging field is growing up and is generating plenty additional interest in ways and tools, to manage the analysis of medical pictures. To support clinical decision-making, several imaging modalities, such as digital radiography, and ultrasound ,Xray computed tomography (CT),magnetic resonance imaging (MRI) are made.Yet a new type of scan calledFunctional magnetic resonance imaging or functional MRI(fMRI) takes the technology one step farther.The fMRI is a modified MRI that measures the brain activity with blood flow.For drug study, teaching, and research activities, medical image database systems are emerging as an important component of Picture Archiving and Communication Systems (PACS).The fMRI not only can help in diagnosing diseases of thebrain,it might also allow doctors to get information of our mental processes to determine what we're thinking and feeling. fMRI might even be able to detect whether we're telling the truth.As medical image analysis is very content and composed of completely different minor structures, with medical imaging it is possible for doctors to see the interior portion of the human body, with extreme clarity, detail, thus providing easy detection and diagnosis of various diseases[1]. So there's a requirement for feature extraction and categorization of pictures for simple and simple retrieval [4].In a Content Based Image Retrieval system, for every image, a feature on its pixel values is computed, the image represented by a signature, the components of the signature is called features[2]. The images are compared and retrieved by querying the image data.The reason for using the signature is to improve the connection between image and semantic structure of the image. The Euclidean distance method is used for calculating distances between each pair of signatures. The index is used to locate signatures that are near to the query point and matched images are displayed to the user. Our proposal is an automatic image based retrieval system of which depends on some particular properties such as shape and texture, color, composition, this process is made through using the principle component analysis (PCA) algorithm. Principal Component Analysis(linear subspace projection technique) used to down-sample high-dimensional datasets and to minimize the re-projection error [5,6]. After retrieval the relevant image data are produced as output. Medical image retrieval has several important applications, particularly in medical research, education and diagnosis. Medical image retrieval for diagnostic functions is very important as a result of the historical pictures of various patients in medical centers have valuable information for the coming diagnosing, build additional correct diagnosing and judge on acceptable treatment. The Content Based Image Retrieval for fMRI finds its applications in many domains such as medical diagnostics toDetermine how well the brain functions afterinjuries,Analyzingemotions and Brain mapping of cognitive functions.

This paper is organized as follows: the related works are presented in section 2. The proposed method is presented in section 3. Implementation and Result analysis is described in section 4. Finally, this paper is concluded in section 5.

RELATED WORKS

In PACS(Picture Archiving and Communication System), image information is retrieved by victimisationlimited text keyword in special fields within the image header (e.g. Patient identifier).CBIR (Content-based image retrieval) has received important attentionwithin the literature as a promising technique to facilitate improved image management in PAC system [1,2].AutomaticSearchandchoiceEnginewithRetrievalTools (ASSERT) [3] is a content based retrieval system that focus on the study of textures in high resolution CT scan of the lung.The Image Retrieval for Medical Applications project focus in providing visually rich image management through CBIR techniques which are applied to medical images using texture measures and intensity distribution over the entire image.Most of the Content-based image retrieval systems application are basically depending on extracting some characteristics that can fetch certain visual properties of an image either globally for the entire image or locally for its regions [8] [9]and deciding the features that can effectively discriminate images and help in matching the most similar ones is the most challenging issue in CBIR systems .On the other hand, depending on them as a sole factor for deciding the images similarity will usually result it retrieving images with comparable colour distributions regardless their contents similarity.So instead of extracting texture vectors that are represent the spatial arrangement of pixels in colour, extracting from gray level becomes an essential step for retrieve more precise results[9]. Support vector machines (SVM) are learning models examine data responsible for classification and regression the performances of SVM depend on the tuning of a number of parameters. Each image is valued according the nearest neighbour distances. This approach allow retrieving a higher percentage of images with respect to SVM-based techniques [10].The Spine Pathology and Image Retrieval System (SPIRS) [13,14,15] is a 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 identical.The Image Retrieval for Medical applications (IRMA) aims to provide visually rich image management through CBIR for medical images using texture and intensity distribution which is taken globally over the entire image. This approach allows queries on a various image collection and helps in recognizingimages that are similar with respect to global features. The Image Retrieval for Medical applications system lacks the ability for finding particular pathology that may be localized in particular regions within the image.

PROPOSED SYSTEM

The query images are fetched from the user, pre-processing of the image takes place in this process the RBG scale image is converted to a gray scale image. An image matrix is created with first N pixels as first row, next N in the second row. Mean image vector is calculated, another matrix is created and inserted data of subtraction of all image vectors from the mean image vector. Calculation of Eigenimage take place, if Eigenvalue is greater than 1, then eigenvector is chosen for creating Eigenimage. Comparisons take place by projecting the image in image space and by measuring the Eucliddistance between them. Extracted the features of the images in the database. Calculating and comparing the Euclidian distance of all projected images in the database from the projected query image. Image which having the minimum distance is the recognized image. The related images of the recognized image is displayed as the related images of the query image ,because in PCA the dataset consist of the similar images (the images that belong to same class) .

Principal Component Analysis is a linear subspace projection technique used to down-sample high-dimensional datasets (and minimize the re-projection error).It is also called as Karhunes-Loevetransformation.Steps to find the Principal Component are:

- Create a database and load with medical image.The database consist of N number of image with D*D dimension.
- Convert the medical image in the database to vector which is denoted by Xi.
- Normalise the vector(X_i).
- Calculate the average vector
- Subtract average vector from each image vector in database.

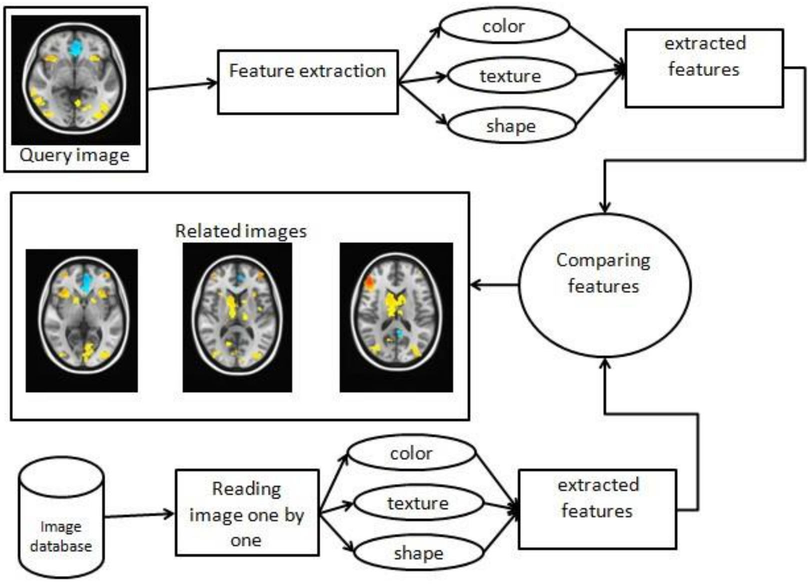

Fig. 1 Data Flow Diagram of Proposed Retrieval System

Normalizing is done to remove the common feature and only unique feature is left behind .common feature are average image vector represented by Ψ after the subtract mean(average)image vector from each vector we get normalized image vector Φ.

Normalized image vector$\Phi i = Xi - \Psi$ Compare the Eigenimage with the use of covariance matrix C.

$$C = AA^T \text{ here } A = \{\Phi_1, \Phi_2, \ldots\ldots \Phi_N\}$$

these are normalized image vectors.

$$\text{Now } A = M^2 * Nr$$

Because dimensions of N images are D*D

$$C = A \qquad A^T$$

$$C = N^2 * M \qquad N^2 * M$$

Here C will become $D^2 * D^2$ dimensions that is very large.

For eg. N=100 and dimensions D*D=50*50 Then D^2= 2500 So covariance matrix become C=2500*2500

That means here 2500 Eigenvectors generated it is large amount. But we need to find K significant Eigenimage because principal of PCA based image recognition is represent each image in training set is linear combination of K selected Eigenimage. Where K ≤ N and N=100 so to find 100 or less than 100 selected features from 2500 it need huge amount of calculation.

- Calculate eigenvectors from a covariance matrix with reduced dimensionality. Here we use formula $C=A^TA$ (this covariance with the reduced dimensionality)

$$C = A^T \qquad A$$
$$C = M * N^2 \qquad M * N^2$$

Here C will be of N*N dimensions means 100*100 Eigenvectors are 100 these are reduced from 2500

- Select K best Eigenimage, such that K<N and can represent the whole training set.
- Convert lower dimensional K eigenvectors into original image dimensionality $U_i = V_i$
  Here U_i= i^{th} vector in higher dimension space and Vi = i^{th} vector in lower space dimension.

By reducing dimensionality we do not only reduce computation but also reduce noise.

- Represent the each image a linear combination of all K Eigen vectors and each image can be represented as a weighted sum of K Eigenimage + mean or average of image.

$$\Omega i = \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

Weighted vector Ω_i is the Eigenimage representation of the i^{th} face. Weight vector for each is calculated.

- Now reorganization using PCA

$$SB = \sum_{i=1}^{c} (xi - \mu)\,(xi - \mu)$$

$$SW = \sum_{i=1}^{c} \sum_{x_k \in x_i} (x_k - \mu_i)\,(x_i - \mu_i)^{\;T}$$

here N_i is the number of training samples in class i, c is the number of distinct classes, μ_i is the mean vector of samples belonging to class i and X_i represents the set of samples belonging to class i with X_k being the k^{th} image of that class. SW represents the scatter of features around the mean of each image class and SB represents the scatter of features around the overall mean for all image classes. Compute the eigenvectors (e_1, e_2, ...,e_d) and corresponding Eigenvalues (λ_1, λ_2, ..., λ_d) for the scatter matrices. The goal is to maximize value of SB while minimizing the value of SW.Sort the eigenvectors by decreasing Eigenvalues and choose k eigenvectors with the largest Eigenvalues to form a d×kdimensional matrix W (where every column represents an eigenvector). Use this d×k eigenvector matrix to transform the samples onto the new subspace. This can be summarized by the mathematical equation: $y = W^T \times X$ (where x is a d×1-dimensional vector representing one sample, and y is the transformed k×1-dimensional sample in the new subspace).

EXPRIMENTAL RESULT AND DISCUSSION

The practical code of proposed system is implemented using MATLAB R2008a on an Intel Core i5 , 4 GHz window based laptop. The system is tested on fMRI dataset having 500 fMRI images.

Performance Parameters

The images are retrieved and measured againstPrecision and Recall[12]

$$\text{Precision} = \frac{\text{number of relevant image retrieved}}{\text{Total number of image retrieved}}$$

$$\text{Recall} = \frac{\text{number of relevant image retrieved}}{\text{Total number of relevant images in the database}}$$

where Precision is the ratio to measure accuracy and Ratio is used to measure robustness.

Test on axial brain fMRI images Dataset

The image retrieval is observed for a database of 900, 192x229 pixel images consisting of 100 different individuals, each individual having 9 images with different characteristic blood flow. Fig 2 shows the query image which is an axial view of brain and 9 relevant images at query image are in the Fig 3,4 and 5.

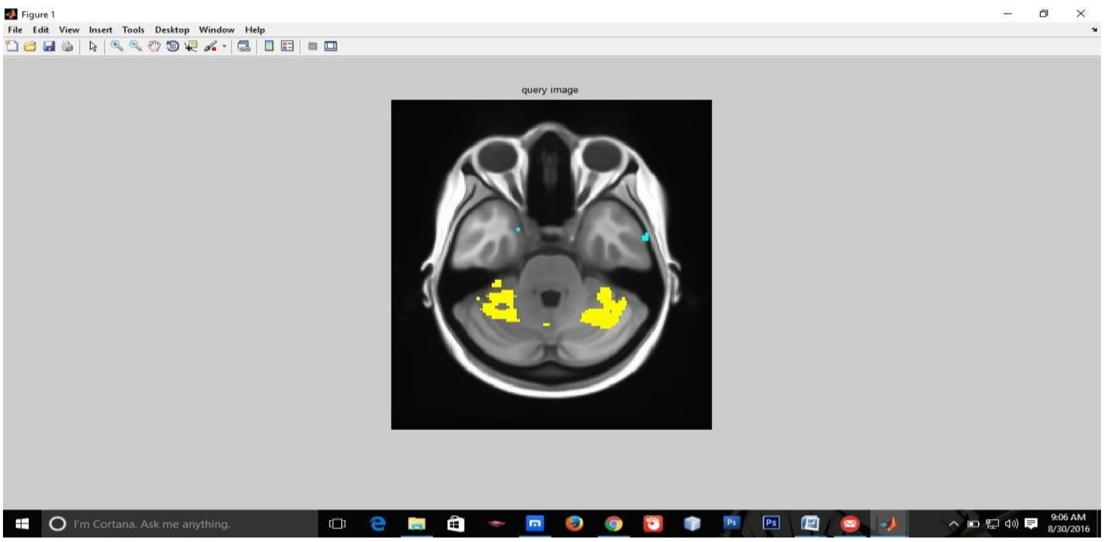

Fig.2Query Image for Axial Brain fMRI Dataset

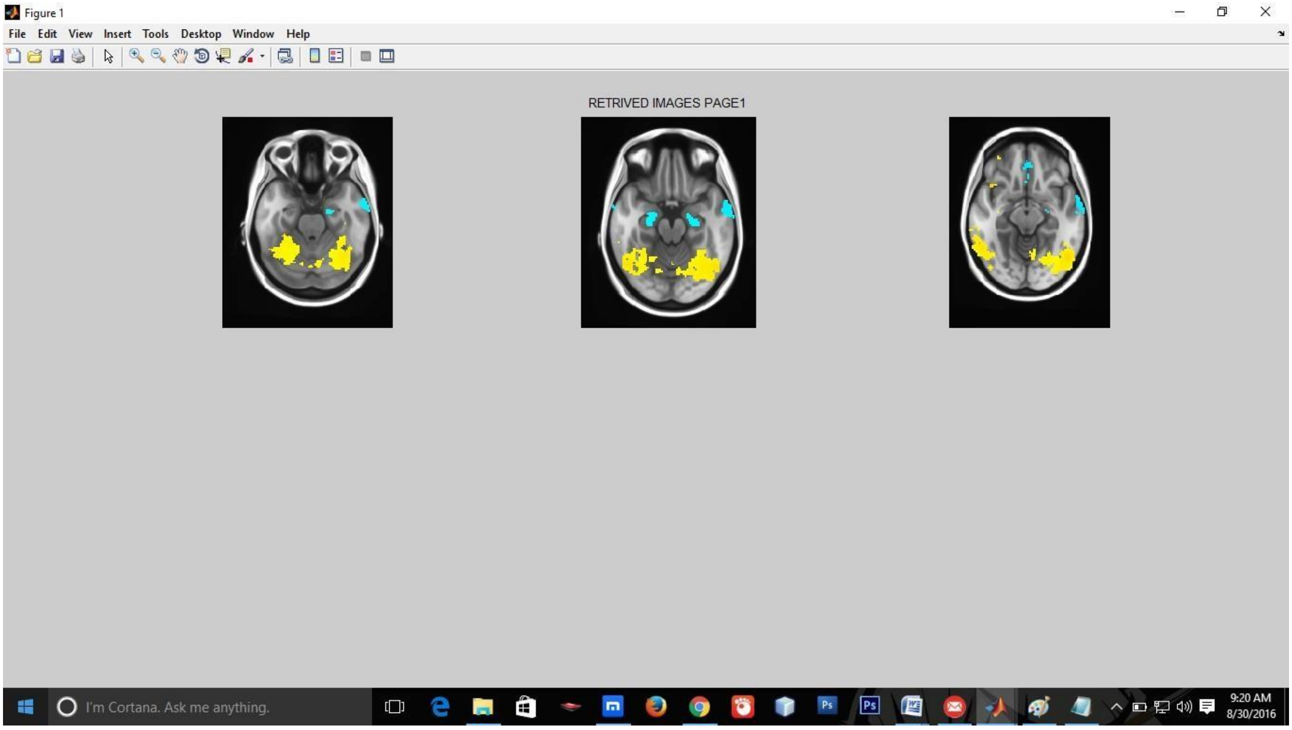

Fig.3 Retrieved Relevant Images for the Query Image

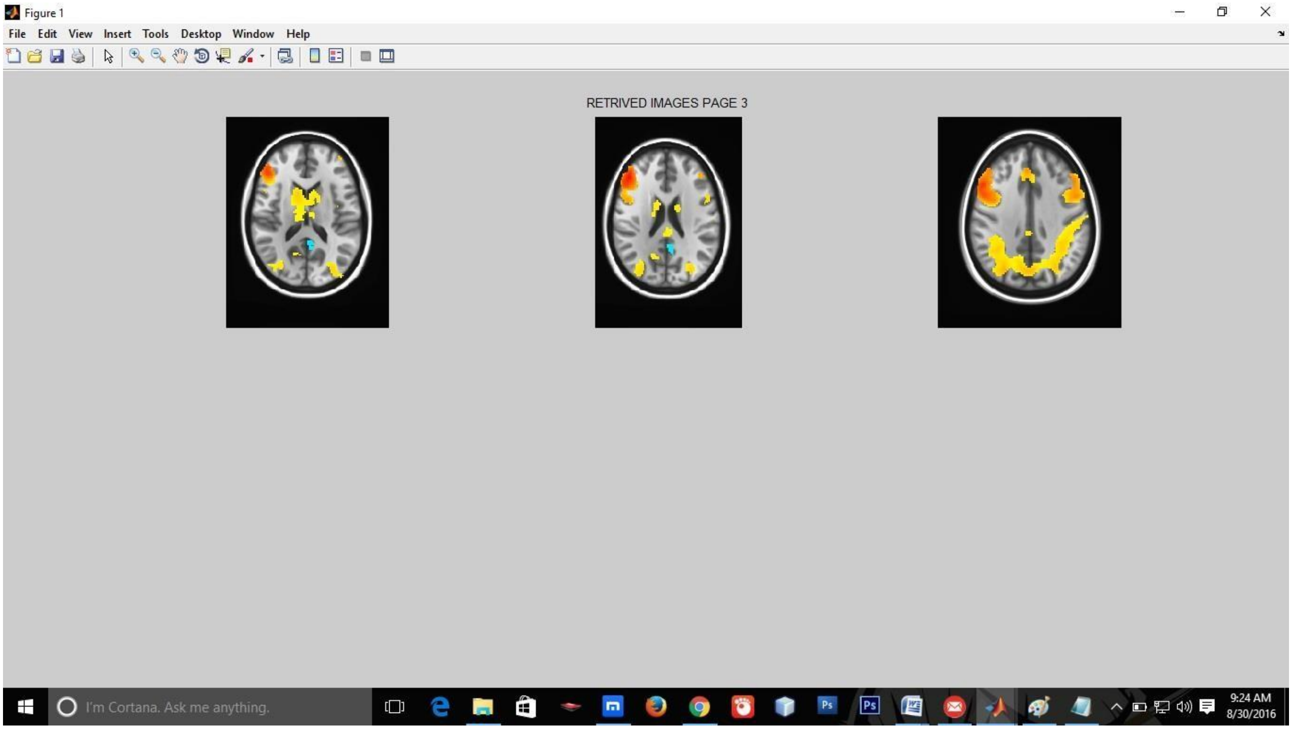

Fig.4 Retrieved Relevant Images for the Query Image

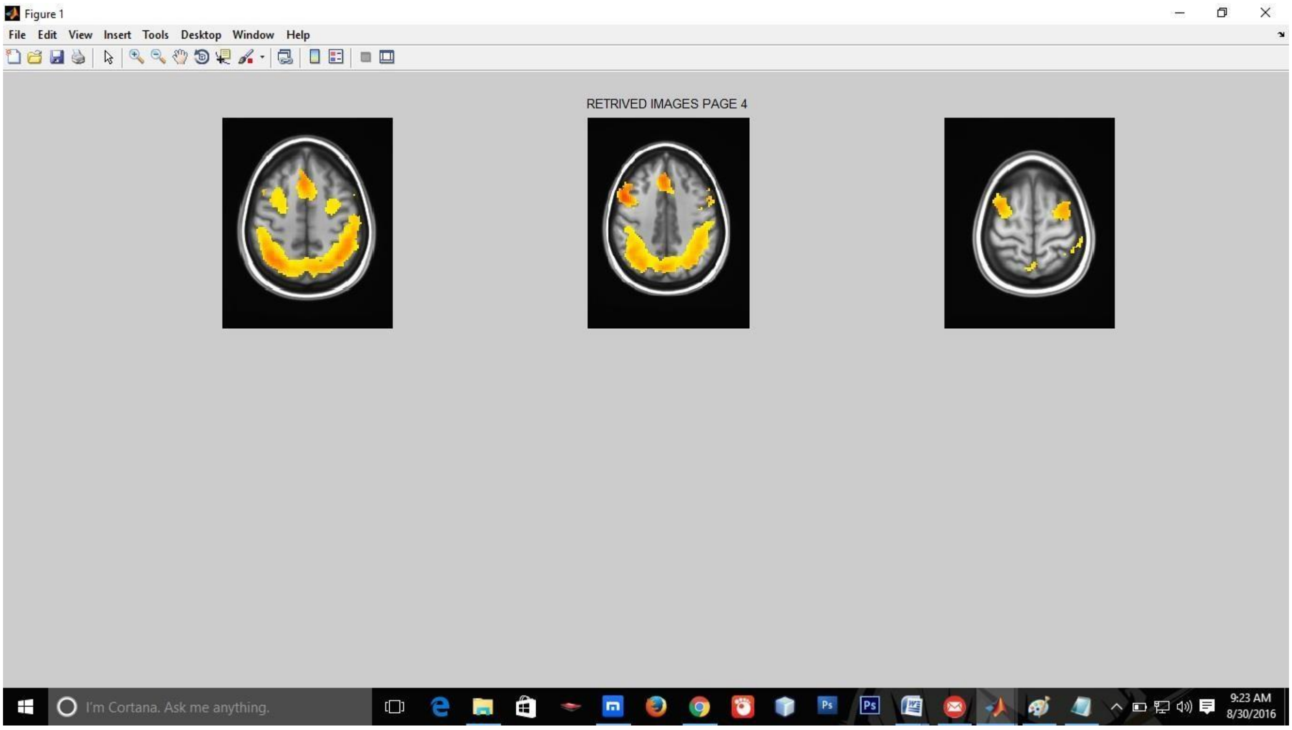

Fig.5 Retrieved Relevant Images for the Query Image

CONCLUSION& FUTURE SCOPE

A system that effectively uses most of the information from an imageis an efficient contentbased image retrieval system for medical databases.In this proposal ,we developed a content based image retrievalfor fMRI images based upon the feature extraction using Principal Component Analysis.The testing is performed on fMRI dataset in order to find the accuracy of the system.The PCA possess a faster result in terms of good accuracy result. Future work may be carried out in the field of image enhancement and there is a need for GUI based CBIR for creating better interface to interact and work efficiently with the system.

REFFERENCES

[1] G.RoslineNesakumari,V.Vijayakumar,B.V.Ramana Reddy,"Generation of An Efficient Digital Watermark Key Based on Honey Comb Polynomial Interpolation Approach", IJCNIS, vol.5, no.3, pp.47-55,2013.DOI: 10.5815/ijcnis.2013.03.06

[2] T.M. Lehmann, M.O. Guld, C Thies,B Fischer, K. Spitzer, and D. Keysers,“ Contentbasedimage retrieval in medical applications”,Methods of Info in Med, IOS Press , Vol. 43,No. 4, 2004, pp. 354–361.

[3] Chi-RenShyu, Carla E. Brodley, AvinashC.Kak, and Akio Kosaka,“ ASSERT:APhysician-in-the-LoopContent-BasedRetrieval System for HRCT Image Databases”,Computer Vision and Image Understanding,Vol. 75, No. 1, 1999, pp. 111–132.

[4] B. Ramamurthy, and K.R. Chandran, “CBMIR:Shapebased Image Retrieval using Canny Edge Detection and K-meansClustering Algorithms for Medical Images”, International Journal of Engineering Science and Technology, Vol. 3, No. 3, 2011, pp. 209-212.

[5] Haykin S. Neural networks: a comprehensive foundation. New York: Prentice Hall; 1999.

[6] Jolliffe IT. Principal component analysis. New York: Springer-Verlag; 1986.

[7] Roy.K and Mukherjee.J,(2013)Image Similarity MeasureUsing Color Histogram, Color Coherence Vector, and SobelMethod.InternationalJournal of Science and Research (IJSR),2,538-543.

[8] Selvarajah.S,andKodituwakku. S.R,(2011)Analysis and Comparison ofTextureFeatures for Content Based ImageRetrieval.International Journal ofLatest Trends in Computing.2,108113.

[9] Kodituwakku.S.R,and Selvarajah.S,(2010)Comparison ofColor Features forImage Retrieval,Indian Journal of Computer Science andEngineering , 1,207-211.

[10] Juan C. Caicedo, Fabio A. Gonzalez and Eduardo Romero., Content-Based Medical Image Retrieval Using Low-Level Visual Features and Modality Identification.

[11] Changjun Zhou⁎, Lan Wang, Qiang Zhang, XiaopengWei "Face recognition based on PCA image reconstruction and LDA" Optik 124 (2013) 5599– 5603.

[12] M. Henning, R. Antoine , and Jean-Paul, "Comparing Feature Sets for content-based Image Retrieval in a Medical Case Database", Proceeding of SPIE Conference on Medical Imaging, 2004, pp. 99-109.

[13] 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.

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

[15] 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; LHNCBCTR-2006-002.