

Medical Image Retrieval using Two Dimensional PCA



Srinivasa Reddy. K , Jaya. T

Abstract: Medical image analysis will be used to develop image retrieval system to provide access to image databases using extracted features. Content Based Image Retrieval (CBIR) is used for retrieving similar images from image databases. During the last few years, medical images are grown and used for medical image analysis. Here, we are proposed that medical image retrieval using two dimensional Principal Component Analysis (2DPCA). For extracting medical image features, 2DPCA has advantageous that evaluates accurate covariance matrix easily as much smaller and also requires less time for finding Eigen vectors. Medical image reconstruction is performed with increased values of 2DPCA and observed from results that reconstruction accuracy improves with increase of principal component values. Retrieval is performed for transformed image space by calculating the Euclidean Distance(ED) between 2DPCA values of unknown images with database images. Minimum distance classifier is used for retrieval, which is simple classifier. Simulation results are reported by considering different medical images and showed that simulation results provide increased retrieval accuracy. Further, Segmentation of retrieved medical images is obtained using k-means clustering algorithm.

Keywords: Covariance matrix, Feature Extraction, Medical Image Retrieval, 2DPCA.

I. INTRODUCTION

CBIR method is used in image processing and computer vision to search and browse for similar images from database. Image retrieval systems are developed to provide access to image database based on their visual appearance or information such as color, shape as well as text more efficiently and effectively. The importance of content-based retrieval used in different applications, including with art galleries and museum archives to picture collections [1][4], criminal investigation, medical [2][3] and geographic databases as the fastest growing research fields in information technology. Therefore, many content-based retrieval applications has developed for commercial and research purposes. This finds application in medical imaging [5], satellite imaging [6].

Revised Manuscript Received on February 28, 2020.

* Correspondence Author

Srinivasa Reddy. K.*, Ph.D. Research Scholar, ECE Dept., Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India. Email: ksrm1gnlgap@yahoo.co.in

Jaya.T. Assistant Professor, ECE Dept., Vels Institute of Science, Technology and Advanced Studies (VISTAS), Chennai, India. Email :jaya.se@velsuniv.ac.in

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>

There are two important methods to access an image data base namely a query driven approach and a browsing driven approach. A query driven approach allows the user to specify either a text query or an image query. This methodology is more appropriate for experts or users who do not have any difficulties in formulating a query. A browsing driven methodology is adopted for users who are not familiar with the image characteristics. Before formulating their query, there is need to have more knowledge about characteristics of image in data base.

Weidong Cai et al. [7] developed the efficient 3-D neurological retrieval method for images with texture. Ashish oberoi et al. [8] used different techniques in the designing of CBIR system on different kinds of images and evaluated its performance. Y. Fanid Fathabad et al. [9] proposed a method, how CBIR systems are used to retrieve images of brain using data base. Senthil Kumar et al. [10] proposed the way for improving overall efficiency of CBIR used in medical field. Ms. Anchal et al. [11] proposed various methods used for medical image retrieval system. A.R. Mahajan et al. [12] proposed the different aspects of medical images with the growth of medicine, computers and imaging technology. Ashish et al. [13] suggested basic principles used in CBIR with features to match function measurement. Marjan Mansourvar et al. [14] proposed the technique to solve retrieval problems in data base with various features which also provides application of CBIR for its medical applications. Mohammed Hamzah Abed et al. [15] proposed an experiment using Matlab and tested three different methods on color histogram based search.

During the last few years, there has been a rapid increase in data, images, videos and information are generated, stored and transmitted which is difficult for accessing and needs an efficient method to search, browse and also for retrieving purposes. CBIR is developed for retrieval of images by indexing images with texture and spatial layout. Conventional image retrieval systems are based on textual annotation where key information about the image is stored.

In medical images it forms an essential component on a patient's record. However, in many occasions this information is very often lost as consequences of image compression or human error. Also, given the number of different standards adopted for medical image annotation, building comprehensive ontology regarding medical terms is not always consensual. Recently, advances in Content Based Image Retrieval prompted researchers towards new approaches in information retrieval for image databases. In medical applications it already met some degree of success in constrained problems. PCA and 2DPCA used as classical techniques for representation of data and also to extract features in Computer Vision,

Medical Image Retrieval using Two Dimensional PCA

Pattern Recognition applications. PCA algorithm used as more accurate image classifier compared to other algorithms by extracting principal components as features of image which are integrated in class of predefined or in module.

During last few years, there are many methods suggested for representing and recognizing the faces. Representation in face recognition such as Eigen faces, holons and local feature analysis depends on PCA. Independent Component Analysis is the generalization method for PCA to separate as additive subcomponents from multivariate signal.

Number of Algorithm's for performing ICA have been proposed for face representation and proved as better than PCA using cosines for similarity measure. ICA is computationally more expensive than PCA. Signal and image processing, uses discrete cosine transform with real numbers with poor results than PCA.

Compensating the above problems, straight forward transformation method for images uses 2DPCA [1] which uses 2D matrices instead of 1D vector. In this, Covariance matrix is obtained directly with matrix of original image instead of previously transforming image matrices into vector form.

Comparing with 1DPCA, 2DPCA has advantageous for evaluating accurate covariance matrix easily as much smaller and also requires less time for finding Eigen vectors. Because of large number of advantages of 2DPCA, a medical image retrieval system using 2DPCA is presented in this paper. For testing and performance evaluation of 2DPCA, simulation results are carried out for medical image reconstruction from computed 2DPCA.

In the first part of the paper, medical image reconstruction from 2DPCA is attempted. It is observed that quality of reconstructed image increases with increase in principal components. Another experiment is performed for medical CBIR.

The retrieval can be performed in the projection space, by computing the Euclidian distance between principle components of unknown images with the retrieved images. Minimum distance classifier is used for retrieval. Euclidian distance corresponds to minimum is selected as the retrieved image.

The organization of medical image retrieval using 2DPCA is presented in different parts. Part 2 presents an overview on PCA along with 2DPCA. Part 3 deals with the derivation of proposed 2DPCA based medical image retrieval algorithm. Experimental or simulations results are shown along with computed and evaluated values in tabular form are indicated in part 4. Last part presents regarding conclusions about the simulation and future scope of work

II. AN OVERVIEW OF PCA AND 2DPCA

A. 1DPCA:

PCA is the linear transformation technique applied for compression and analysis of data [1].

Let r is the vector represented as

$$r = (r_1, r_2, \dots, r_n)^T$$

Then, mean value can be obtained as

$$\mu_x = \text{mean}\{r\}$$

The covariance matrix (M) of the data can be written as

$$M = \text{mean}\{(r - \mu_r)(r - \mu_r)^T\}$$

From the vectors, r_1, \dots, r_M , mean and covariance matrix can be evaluated. From M, calculate eigen values and corresponding eigen vectors.

From r , we obtain transformation of data as

$$y = F(r - \mu_r)$$

From y , reconstruction can be obtained as

$$r = F^T y + \mu_r$$

Some values of first largest Eigen vectors can be used for transformation.

B. 2Dimensional Principal Component Analysis (2DPCA):

In PCA method of image recognition, 2D image is first converted in to 1D image vector [1]. This produces large vector space. Hence, computation of M becomes difficult as the size is more. To overcome this drawback, 2DPCA method is developed in the literature which can be used to compute M directly.

Assuming P which represents N Dimensional (ND) vector. The transformed image F of size $m \times n$ can be obtained

$$Z = FP \quad (1)$$

Using the above equation, one can get ND projected vector Z. The trace can be obtained for getting M of projected feature vector. Then the expression (1) becomes

$$J(P) = \text{Trace}(M) \quad (2)$$

Where M denotes the covariance matrix.

The covariance matrix (M) can be derived as

$$\begin{aligned} M &= \text{mean}\{(Z - \text{mean}(Z))(Z - \text{mean}(Z))^T\} \\ &= \text{mean}\{[FP - \text{mean}(FP)][FP - \text{mean}(FP)]^T\} \\ &= \text{mean}\{[(F - \text{mean}(F))P][(F - \text{mean}(F))P]^T\} \end{aligned} \quad (3)$$

So,

$$\text{Trace}(M) = P^T \{[\text{mean}(F - \text{mean}(F))]^T (F - \text{mean}(F))\} P \quad (4)$$

Using Trace (M), Covariance Matrix can be rewritten as

$$M = \frac{1}{N} \sum_{j=1}^N (F_j - F')^T (F_j - F') \quad (5)$$

Where F' represents mean of F.

The optimal projection axes can be obtained by maximizing the criteria which corresponds to largest eigen values. It is necessary to select projection axes F_1, \dots, F_d which are known as eigen vectors of M of first 1 to 20 highest eigen values.

C. Feature Extraction

The transformation vector of 2DPCA, P_1, \dots, P_d are used for extracting features of an image F. Then, for image F

$$Z_t = F P_t, \quad t = 1, 2, \dots, d$$

The transformed vectors Z_1, \dots, Z_d , are known as principal component (vectors) for image F.

Then, the Scalar matrix is formed using 2DPCA and obtained features are arranged in the form of matrix

$$C = [Z_1, \dots, Z_d]$$

D. Classification

For each image F, the transformed matrix is obtained using 2DPCA. The transformed values are applied to nearest neighborhood classifier (NN) for classification purpose. The formula for NN classifier for feature matrices,

$C_i = [Z_1^{(i)}, Z_2^{(i)}, \dots, Z_d^{(i)}]$ and $C_j = [Z_1^{(j)}, Z_2^{(j)}, \dots, Z_d^{(j)}]$, are calculated using the formula,

$$d(C_i, C_j) = \sum_{t=1}^d \| C_t^{(i)} - C_t^{(j)} \|_2$$

where $\| C_t^{(i)} - C_t^{(j)} \|_2$ represents ED between $C_t^{(i)}$ and $C_t^{(j)}$.

III. PROPOSED METHOD

2DPCA-BASED IMAGE RETRIEVAL ALGORITHM:

1. Form the training set using images F1, F2,.....,F5.
2. Calculate mean of images F1, F2,.....,F5.

$$mean = \frac{F1 + F2 + F3 + F4 + F5}{5}$$

$$= \frac{(F1 - \bar{F})(F1 - \bar{F})^T + (F2 - \bar{F})(F2 - \bar{F})^T + \dots + (F5 - \bar{F})(F5 - \bar{F})^T}{5}$$

3. Determine co-variance matrix (M)
4. Determine P1.....Pd for Matrix M.
5. Let assuming 20 highest eigen values are used as features.
6. Evaluate 2DPCA values using $Z_t = FP_t \quad t = 1, 2, 3, \dots, 20$
7. C values of unknown images with C values of all images in database.
8. Finally, Arranging Z_t obtain feature matrix C
 $C = [Z1, Z2, \dots, Z20]$
9. Determine C for unknown images as well as all images in the database.
10. Compute the ED between C values of unknown images with C values of all images in database.
11. Finally, retrieves the images matching with minimum distance consecutively.

IV. SIMULATION RESULTS

The MATLAB software is used for the simulations since it is powerful computing language used for Scientific and Engineering calculations and simulations used for image processing applications.

For testing 2DPCA and performance evaluation, simulation results are carried out for medical image reconstruction from computed 2DPCA. In the first part of the paper, medical image reconstruction from 2DPCA attempted.

It is observed that, reconstructed image quality increases with increasing principal components. Another experiment is performed for medical CBIR.

The retrieval can be performed in the projection space, by computing the Euclidian distance between principle components of unknown images with the retrieved images. Minimum distance classifier is used for retrieval. Euclidian distance corresponds to minimum is selected as the retrieved image.

For testing 2DPCA, IRMA medical image database images which are freely available online are used for simulation. This database consists of different medical images. Some sample images presented in Figure 1.



Figure1. Some medical images (100x100 pixels) used for the simulation

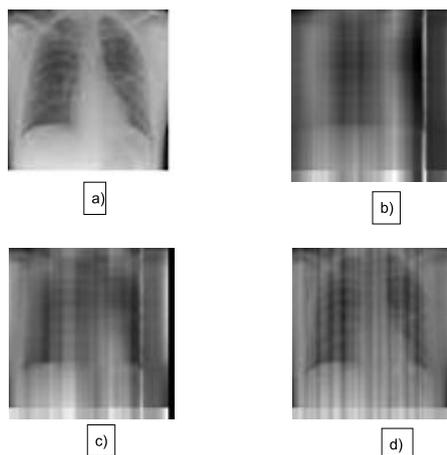


Figure2. Original and reconstructed images

- a) Original image
- b) Reconstructed image 20 PCA
- c) Reconstructed image 30 PCA
- d) Reconstructed image 40 PCA

Table 1: 2DPCA values for figure 2. (a)

Principal Components	2DPCA Values	Principal Components	2DPCA Values
Z_1	3.6	Z_{11}	3
Z_2	24	Z_{12}	4
Z_3	3	Z_{13}	46
Z_4	41	Z_{14}	4
Z_5	3	Z_{15}	2
Z_6	17	Z_{16}	1
Z_7	34	Z_{17}	9
Z_8	11	Z_{18}	25
Z_9	3	Z_{19}	11
Z_{10}	13	Z_{20}	17

Medical Image Retrieval using Two Dimensional PCA

Medical image reconstruction is performed with the increasing retrieved image. The performance of proposed method has been evaluated using retrieval accuracy. Ratio of number of reconstructed image quality increases with increasing principal components. The same experiment is repeated with brain images and the results are shown in Figure 3.

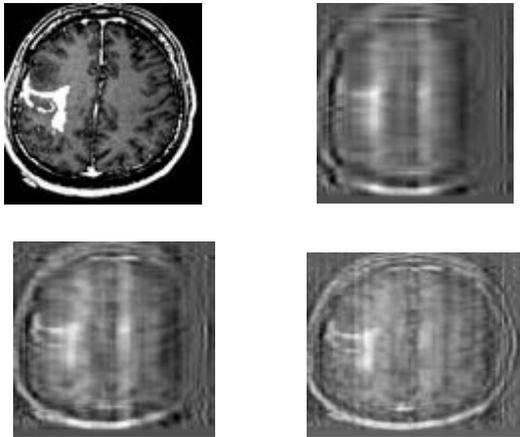


Figure3. Original and reconstructed images

- Original image
- Reconstructed image 20 PCA
- Reconstructed image 30 PCA
- Reconstructed image 40 PCA

Table 2: 2DPCA values for figure 3. (a)

Principal Components	2DPCA Values	Principal Components	2DPCA Values
Z_1	103	Z_{11}	155
Z_2	117	Z_{12}	161
Z_3	132	Z_{13}	130
Z_4	141	Z_{14}	128
Z_5	143	Z_{15}	119
Z_6	148	Z_{16}	108
Z_7	159	Z_{17}	127
Z_8	152	Z_{18}	170
Z_9	147	Z_{19}	207
Z_{10}	148	Z_{20}	255

Another experiment is performed for medical CBIR. The features are computed using 2DPCA. Euclidean distance is used for matching. The retrieval can be performed in the projection space, by computing the Euclidian distance between principal components of unknown images with the retrieved images. Minimum distance classifier is used for retrieval. Euclidian distance corresponds to minimum is selected as the

retrieved image. The performance of proposed method has been evaluated using retrieval accuracy. Ratio of number of correctly retrieved images to the total images used for retrieval is known as retrieval accuracy. Performance of 2DPCA is evaluated and obtained Retrieval accuracy is 90%. After retrieving the medical images the images are segmented for further analysis using K-means clustering shown in Figure 4 and figure 5 respectively.

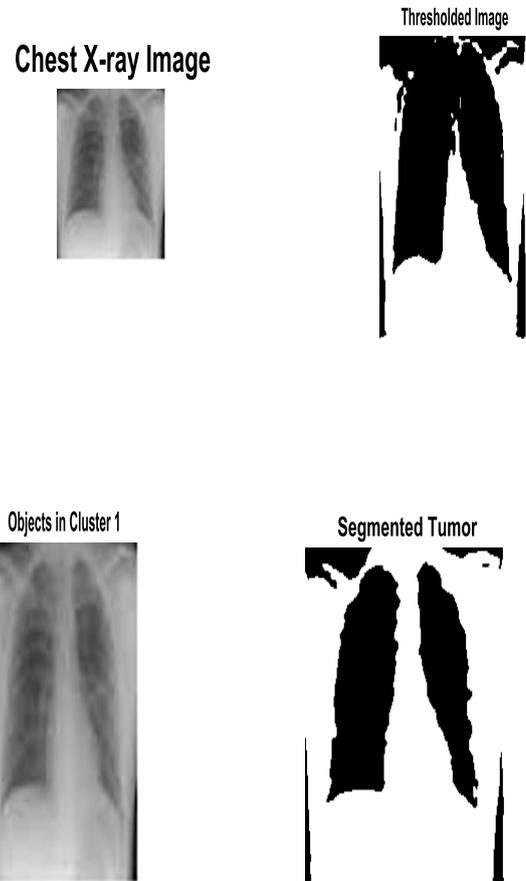
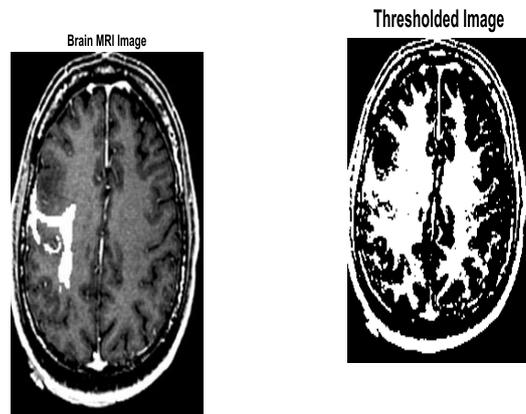


Figure4. Segmented images using K-means clustering



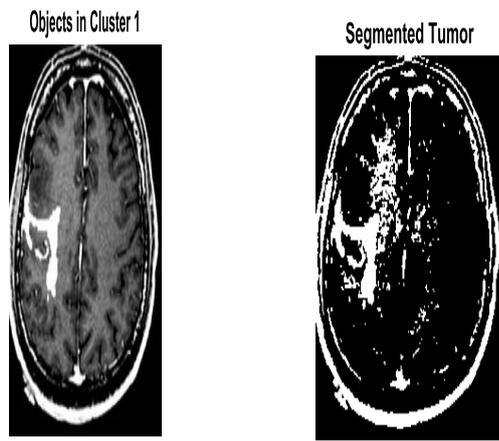


Figure 5. Segmented images using K-means clustering

V. CONCLUSION AND FUTURE SCOPE

Because of the number of advantages of 2DPCA, a medical image retrieval system using 2DPCA is presented in this paper. For testing and performance evaluation, simulation results are carried out for medical image reconstruction from computed 2DPCA. In the first part of the paper, medical image reconstruction from 2DPCA attempted. It is observed that, quality of reconstructed image increases by increasing principal components. Another experiment is performed for medical CBIR. The retrieval can be performed in the projection space, by computing the Euclidian distance between principle components of unknown images with the retrieved images. Minimum distance classifier is used for retrieval. Euclidian distance corresponds to minimum is selected as the retrieved image. Better retrieval accuracy is achieved through simulation results. After retrieving the medical images the images are segmented for further analysis using K-means clustering effectively. To retrieve the images under different image plane transformations is scope of future work.

REFERENCES

1. Jian Yang, David Zhang, Senior Member, IEEE, Alejandro Frangi, Jing-Yu Yang, "2-D PCA: New Approach to Appearance-Based Face Representation and Recognition," *IEEE Transactions on PA and MI*, vol. 26, Jan, 2004.
2. M. Kirby, L. Sirovich, "Application of KL for Characterization of Human Faces," *IEEE Transactions on PA and MI*, vol. 12, Jan, 1990.
3. Peter Belhumeur, Joao Hespanha, David Kriegman, "Eigenfaces Vs fisher faces: Recognition using Class Specific Linear Projection," *IEEE Transactions on PA and MI*, vol.19, 711-720, July, 1997.
4. Chun.Y, Kim.N, Jang.I, "CBIR Using Multiresolution Color and Texture Features," *IEEE Transactions On Multimedia*, Vol. 10, October, 2008.
5. Prachibhende, Cheran, "CBIR in Medical Imaging". *IJCER* vol.3, issue 8, August, 2013.
6. NourEIDinLabin, MotazEISaban, Ayman Nasr, HodaOnsi, " System refinement for Satellite CBIR". *The Egyptian Journal of RS and SS*, volume 15, issue 1, June, 2012.
7. Weidong Cai, Sidong Liu, Lingfengwen, Stefan Eberyl, Machael Fulham, Dagan Feng, "3D Neurological IR with localized pathology-centric CMRGLC patterns". *IEEE 17th IC on Image Processing* September 26-29, 2010, Hong Kong.
8. Ashish Oberoi, Manpreet Singh, "CBIR-MD Endoscopy, Dental and

9. skull images". *IJCSI*, Vol. 9, Issue 3, May, 2012.
9. Y. Fanid Fathabad, Balafar, "Application of CBIR in diagnosis brain disease". *IJTPE* Volume 4, Issue 13, December, 2012.
10. R.Senthil kumar, Senthilmurugan, "CBIR system in medical applications". *IJERT*, Volume 2, Issue 3, March 2013.
11. Ms. Anchal Solio, Dr.Siddharth Ladhake, " Review of query image in CBIR". *IJARCET*, Vol.2, Issue 4, April, 2013.
12. A.R. Mahajan, S.D. Zade, Pawan Raut "CBIR in medical images: current-status and future directions" *IJAIEM Special issue*, 2013.
13. Ashish Mahan Yadav, BPS Sengar, " Survey on CBIR systems". *IJETAE*, Volume 4, Issue 6, June, 2014.
14. Marjan Mansourvar, Maizatul Akmar Ismail, "CBIR". *IJIT*, vol.20, 2014.
15. Mohammad Hamzah Abed, Dawood Salman Jasim Al-Farhoosi Iraq, "CBIR based on Histogram " *IJCA*, volume 110, January, 2015.

AUTHORS PROFILE



K. Srinivasa Reddy received his B.E. Degree in Electronics Engineering from Dr. Baba Saheb Ambedkar Marathwada University, Aurangabad and M.Tech in Electronics & Communication Engineering with a specialization of Digital Systems & Computer Electronics from Jawaharlal Nehru Technological University, Hyderabad. He has more than 20 years of teaching experience. His research areas of interest are in digital systems, medical image processing, data communication and computer networks. At present, he is pursuing Ph.D. in Dept of ECE, Vels Institute of Science Technology & Advanced Studies (VISTAS), Chennai. He has published 10 research papers in peer reviewed journals and has attended more than 10 workshops including conferences.



T. Jaya received the B.E. degree in Electronics and Communication Engineering from the Anna University, Chennai, India, in 2008 , M.E. in Electronics and Communication Engineering (Applied Electronics) in Sathyabama University and Ph.D in Electronics and Communication Engineering in Institute of Science Technology & Advanced Studies (VISTAS). , Chennai. Currently working as Asst. Professor in Dept of ECE, Vels Institute of Science Technology & Advanced Studies (VISTAS), Chennai. Her research interest includes Wireless Communication, Mobile Ad hoc networks, and Sensor Networks Communication networks. She has published 30 research papers in peer reviewed journals and she has presented more than 25 research papers in international conferences and 5 papers in national conferences.