Study on Ultrasound Kidney Images Using Principal Component Analysis: A Preliminary Result

C.Karthikeyini,
Lecturer,Dept. of IT,
PSNA College of Engg. and Tech.,
Dindigul – 624 622, INDIA.
ckarthi_raja@yahoo.com

K.Bommanna Raja,
Asst. Prof., Dept. of ECE,
PSNA College of Engg. and Tech.,
Dindigul – 624 622, INDIA.
bommanna_raja@psnacet.org

Dr.M.Madheswaran,
Prof. & Head, Dept. of ECE,
PSNA College of Engg. and Tech.,
Dindigul – 624 622, INDIA.
madhi_eswaran@psnacet.org

Abstract

An effort has been made to test the potential of principal component analysis (PCA) method for quantifying and classifying the ultrasound kidney images. For our analysis two different classes of kidney images namely normal (NR) and medical renal diseases (MRD) are considered. The eigen values and vectors are derived for a set of 40 images. The weight vectors (WV) are estimated from the obtained eigen vectors. The result indicates that the mean value of WV for NR is 0.4205 and for MRD is -0.9983. The sum value of WV for NR and MRD is 8.4103 and -19.9657 respectively. The regression analysis shows that WV's of two classes are negative correlated and are in weak moderate association (-0.4271). The student t-test specifies that the eigen value and WV are much significant (p<0.005) in separating the classes of kidney image under study. These analysis shows that there exists an appreciable measure of relevance for this parameter weight vector in classifying the kidney images.

.

1. Introduction

The image processing techniques are usually applied for medical images to enhance the quality of representation and better understanding of hidden information for proper objective diagnosis. By using such techniques, it is also possible to extract some parameters or features that will be very helpful for the diagnosis of the medical images. The main intention of searching for an efficient algorithm and procedure has certain advantages like, (i). to establish a quantitative reference for the medical images under study, (ii). to retrieve images based on feature values, (iii). to make comparative study on images for better decision making, and (iv). to develop an expert system that automatically recognize the extent of pathology or normality of the biological system being examined. The earlier work done by the researchers in this context shows much promising result [10, 3, 14, and 5].

Hence in this paper, we have pioneered to use Principal Component Analysis (PCA) method to quantify and classify the ultrasound kidney images for the first time. The work done by Turk and Pentland [14], suggests that eigenface method based on PCA is very much useful in detecting and recognizing the human faces. Jorge de la *et. al* [8], suggests that the eigen image method based on PCA is used for image analysis and classification of Galaxies present in the Universe. Also the work on ultrasound kidney images by K.Bommmanna Raja *et .al* [3,4], shows feature extraction from ultrasound kidney images is possible and these features are highly effective in defining the kidney disorders.

To establish efficient feature parameters, other than specified in [3,4], for developing an automated system that aims to meet the intentions given before, this paper concentrate on using the eigen values and vectors obtain by PCA method to quantify and classify the ultrasound kidney image. The main idea in this method is to decompose kidney images into a small set of characteristic feature images called eigen images, which may be thought of as principal components of the original images. These eigen image function also forms the orthogonal basis vectors of a linear subspace called "kidney space". The weight vectors of the eigen images are calculated for the two sets of kidney images considerd for our study. The general study on this weight vector is carried out to sort out the possiblity of using this weight vector for the analysis of ultrasound kidney images. It is believed that searching for more such proposed methods and optimizing the accuracy of quantifiation and classification will definitely help to develop a procedure that may become diagnostically relevant in objective decision making.

2. Image Database and Acquisition

The ultrasound images of 20 normal and 20 medical renal diseases with a mean age of 46 ($\pm 13.58^{*}$), 50.43 ($\pm 14.59^{*}$) respectively [*represent standard deviation]

are acquired from male and female subjects by using the scanning systems (a). ATL HDI 5000 curvilinear probe with transducer frequency 5-240 MHz. (b). WiproGE LOGIQ 400 curvilinear probe with transducer frequency 3-5MHz. Longitudinal cross section of the kidneys are acquired, by fixing transducer frequency at 4MHz. During the image acquisition, sonographer looks for better visualization of the image in screen and freeze to store the image. No initial ultrasound-probe orientation and gain fixation is made. The images are obtained from medical center, M/s. Mediscan systems (P) Ltd., Chennai, India.

3. Image Pre Processing

The image preprocessing methods are used to get the efficient results from the analysis. The images are subjected to four types of image preprocessing techniques. 1. ***Cropping*** - This is used to remove the unwanted echoes in the scanned image thereby our region of interest, kidney is separated. 2. ***Rotation*** - Usually in the scanned image, the elliptical kidney portion is inclined at an angle with respect to the horizontal axis, because the ultrasound –probe position is not fixed while acquisition. To have certain reference axis, the images for study are rotated, after cropping, so that their major axis is aligned at 0°. 3. ***Edge detection*** - The presence of speckle noise and echoes from the surrounding tissue of the kidney restrict the use available conventional edge detection algorithm. Hence, as an alternate, the cublic-spline interpolation technique is used to connect the coordinate of points that are identified as edges through manual assessment. 4. ***Background subtraction*** - After the above preprocessing steps the image is rectangular with dimension of m x n that includes elliptical kidney region. The unwanted intensity values at the four corners of the images will be incorporated in the calculation of the feature parameters. To avoid this background subtraction is made.

The sample image after preprocessing is shown below. It can be seen that image contains only the kidney region.

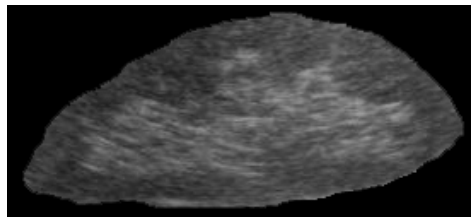

Fig. 1. The preprocessed ultrasound kidney image for the normal (NR) subject #15

4. Principal Component Analysis

The central idea of PCA is to reduce the dimensionality of a data set which consists of a large number of interrelated variables, while retaining as much as possible the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are ordered so that the first PC retains most of the variation present in all of the original variables.

PCA is a kind of Karhunen-Loeve transform which aims to find an orthonormal basis to compress the information. The geometric explanation is that PCA method try to change the coordinate axes so that the one lie along the directions on which the original data set has the largest variations. The two main important features of PCA are dimension reduction and coordinate transformation. Dimension reduction removes the redundant information to achieve a compact representation and coordinate transformation rotates the axis to get a best view angle regarding data variation.

5. Feature Extraction

The algorithm to obtain the WV for the kidney images is:

Step1: Find the average image of NR and MRD
Step2: Compute difference image from average image
Step3: Construct the covariance matrix
Step4: Compute the eigen values and eigen vectors
Step5: Formation of eigen images in eigen space
Step6: Estimate the weight vector

A set of kidney images (20NR and 20MRD) are resized to 103 by 210 pixels. These images could be represented as two dimensional (103 by 210) arrays of pixel intensity data. Let the set of training images be represented by $\Gamma_1, \Gamma_2, \Gamma_3, ..., \Gamma_M$ Where, M is the number of kidney images. In our case the number of kidney images is 40. Each image is represented by a vector of length N, where N is the number of pixels in the image.

5.1 Computation of Average Images

The average image (ψ) is computed for the total (both NR & MRD), NR and MRD images by using the following equations.

For the total kidney images,

$$\psi_T = \frac{1}{M}\sum_{n=1}^{M}\Gamma_n \qquad (1)$$

The average image for the total images is shown here.

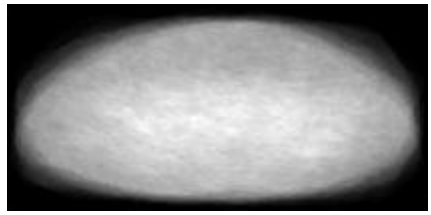

Fig.2. Average image for total kidney images

For the NR images,

$$\psi_{NR} = \frac{2}{M}\sum_{n=1}^{M/2}\Gamma_n \qquad (2)$$

The average image for the normal kidney images is shown below.

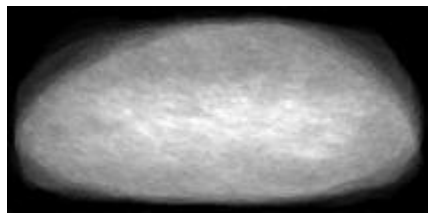

Fig.3. Average image for normal kidney images

For MRD kidney images

$$\psi_{MRD} = \frac{2}{M}\sum_{n=1}^{M/2}\Gamma_n \qquad (3)$$

The average image which is computed for MRD is shown below.

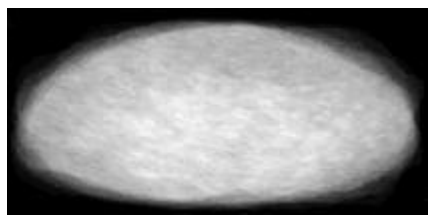

Fig.4. Average image for MRD kidney images

5.2 Computation of Difference Image

Following the computation of the average images for normal, medical renal diseases and total kidney images, the difference image (ϕ) which is the difference of each image from the average image of normal and MRD is computed.

$$\phi_{NR} = \Gamma_n - \psi_{NR} \qquad (4)$$

$$\phi_{MRD} = \Gamma_n - \psi_{MRD} \qquad (5)$$

5.3 Construction of covariance matrix

The covariance matrix (C) is constructed using the following equation.

$$C_{NR} = \frac{2}{M}\sum_{n=1}^{M/2}\phi_{NR}\phi_{NR}^T = AA^T \qquad (6)$$

$$C_{MRD} = \frac{2}{M}\sum_{n=1}^{M/2}\phi_{MRD}\phi_{MRD}^T = BB^T \qquad (7)$$

Where,

$$A = [\phi_{NR1,}\phi_{NR2},....,\phi_{NR\frac{M}{2}}]$$

$$B = [\phi_{MRD1},\phi_{MRD2},....,\phi_{MRD\frac{M}{2}}]$$

5.4 Computation of Eigen Values and Eigen Vectors

The matrix C is N * N (in our case 21630*21630) that can be calculated from AA^T and BB^T.The determination of the N eigen values and eigen vectors is impractical for typical image sizes. However, if M < N, then there will only be M meaningful eigenvectors. This M meaningful vectors can be found by calculating $A^T A$ and $B^T B$. The eigen vectors of $A^T A$ is v_i. To prove that the eigen values and eigen vectors obtained for $A^T A$ and AA^T are same:

$$A^T A v_i = \mu_i v_i \qquad (8)$$

Multiplying both sides by A,

$$AA^T A v_i = \mu_i A v_i \qquad (9)$$

$$CAv_i = \mu_i A v_i \text{, where } C = AA^T \qquad (10)$$

or

$$Cu_i = \mu_i u_i \text{, where } u_i = Av_i \qquad (11)$$

Thus, AA^T and $A^T A$ have the same eigen values and their eigen vectors and they are related as follows:

$$u_i = Av_i \qquad (12)$$

The M best eigen vectors are computed using the above relationship. The eigen vectors obtained for the images are arranged in the ascending order according to the eigen value. The first eigen vector in the series corresponds to the highest eigen value. This procedure is carried out because the eigen vector that corresponds to the highest eigen value is of immense important as it preserves high degree of variation present in all original images. The fig. (5) shows the eigen values in increasing order attained for NR and MRD images.

5.5. Representation of Images in Eigen Space

Each kidney image set (either NR or MRD) can be projected into the eigen space and each set can be represented as a linear combination of the best eigen vectors. The number of best eigen vectors are selected according to our convenience. If the number increases the computation time increases. In our case we considered all 20 eigen vectors (which is the maximum value, ∵ M/2 = 20).

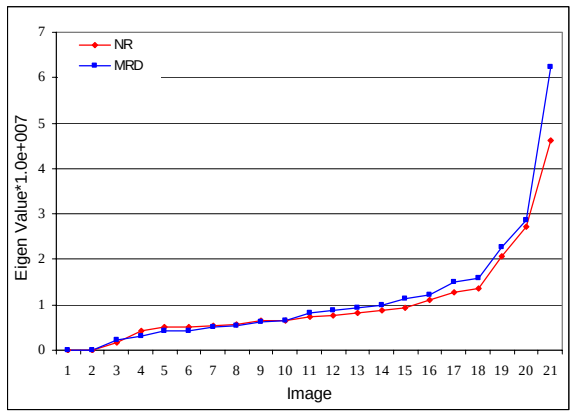

Fig.5. Eigen Values obtained for NR and MRD Images after sorting

These 20 eigen vectors of each image are represented as an eigen image. The resultant eigen image by considering the first set of eigen vectors corresponding to the highest eigen value is depicted in the fig. (6).

Fig.6. Eigen image of the subject 09

In total 40 eigen images (NR=20; MRD=20) can be obtained.

5.6. Estimation of Weight Vector

The end step is the computation of the weight vector and it is calculated by using the following equation.

$$W_j = u_j^T \phi_i \qquad (13)$$

Each image ϕ_i is represented by a vector omega.

$$\Omega_i = \begin{bmatrix} w_1^i \\ w_2^i \\ \dots \\ w_k^i \end{bmatrix}, \text{ i =1, 2… M/2} \quad \text{k=M/2} \qquad (14)$$

Hence the size of the omega matrix is 20* 20. For both NR and MRD images this weight vectors are calculated separately.

6. Results and Discussion

The importance in estimation of weight vector lies in the fact that it provides the useful quantified transformation based on weight values of image set into the image space. As expected the result obtained shows that weight values for NR and MRD varies significantly. The variation of weight values upon considering the first weight value (w_1) for the images is shown in the fig.(7). Some statistical parameters

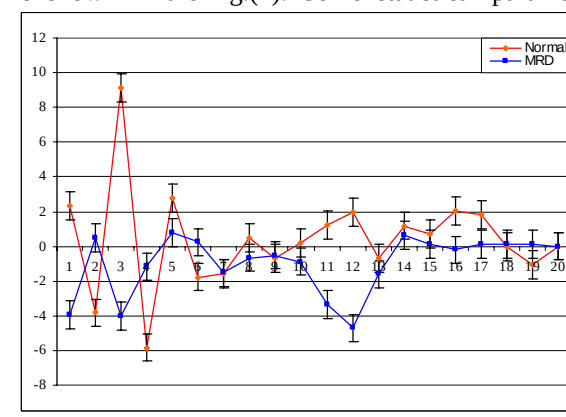

Fig.7. The variation of weight value (w_1) for NR and MRD

obtained with weight value are shown in the Table 1. It can be seen that mean value is positive for NR and is negative for MRD. Though this may not provide the useful interpretation about the weight value, the computation of maximum (9.1431; 0.7631 for NR and MRD respectively) and minimum (-5.8311; -4.7126 for NR and MRD respectively) weight value shows that the weight value of NR dominates MRD. This can also be verified with the sum of weight value. The correlation study reveals, the weight values of NR and MRD are negatively correlated and are in weak moderate association. The calculation of covariance shows that the average of the products of deviations for each data set pair is above -2.0164. In addition to maximum, minimum and sum, this in term ensure that weight values appears to be relevant in classification.

Table.1. Statistical parameters for the weight value (w_1) for NR and MRD images

Parameters	NR	MRD
Mean	0.4205	-0.9982
Sum	8.4103	-19.9657
Minimum	-5.8311	-4.7126
Maximum	9.1431	0.7631
SD	2.9437	1.6881
Covariance	-2.0164	
Correlation	-0.4271	

SD - Standard deviation

The regression analysis is performed to develop a regression equation to compute the weight value, if an independent parameter is given. The search to find the independent parameters results in mean intensity and mean difference from average. The distribution of weight values with respect to these independent parameters are shown in the fig. (8) and (9).

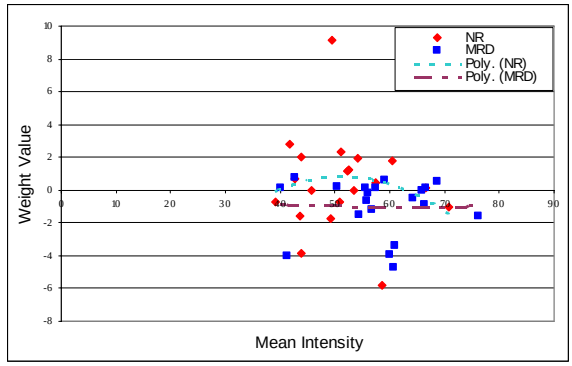

Fig.8.Distribution of weight value with respect to mean intensity of training set images.

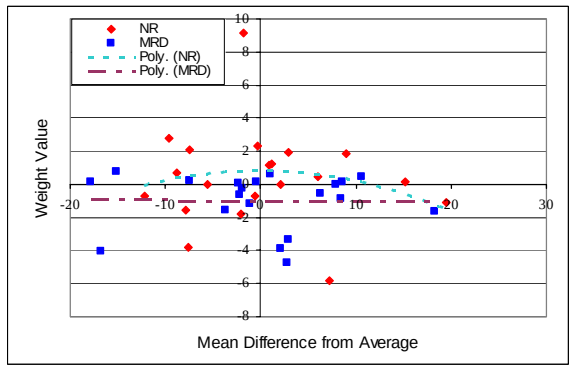

Fig.9. Regression analysis between weight value and mean difference from average of set of images.

The polynomial regression equation of order 2 obtained for NR and MRD is:

Keeping mean intensity as independent variable

For NR images

$$\left.\begin{aligned} Y &= 0.006x^2 + 0.6203x - 15.087 \\ R^2 &= 0.0376 \end{aligned}\right\} \quad (15)$$

For MRD images

$$\left.\begin{aligned} Y &= 0.002x^2 - 0.0273x - 0.1563 \\ R^2 &= 0.0005 \end{aligned}\right\} \quad (16)$$

Keeping mean difference from average as independent variable

For NR image

$$\left.\begin{aligned} Y &= -0.006x^2 - 0.0011x + 0.822 \\ R^2 &= 0.0376 \end{aligned}\right\} \quad (17)$$

For MRD image

$$\left.\begin{aligned} Y &= 0.0002x^2 - 0.0023x - 1.0159 \\ R^2 &= 0.0005 \end{aligned}\right\} \quad (18)$$

From the results of the polynomial regression equation, the coefficient of determination (R^2) is not as much significant, which indicates estimated weight value 'Y' does not corresponds to the actual weight value. This is due to the wide variation in the intensity value of the acquired image. The gain fixation at the time of scanning varies for each subject, as it depends on physiology of the subject being scanned. This change in gain is reflected as intensity variation in the image.

Students *t*-test analysis is performed with the estimated weight values to know measure the significance of this parameter in classification. The result obtained indicate the weight values are much significant ($p<0.005$) and can be used for classification.

7. Conclusion

The study made on ultrasound kidney images using principal component analysis (PCA) results a feature parameter, weight vector. The capability of this parameter in classification of kidney classes taken for study is much appreciable. As the quantified values are significant this parameter may be used to set the reference value for NR and MRD cases. Though the problem of quantification is not thoroughly addressed it is believed from the values obtained that this parameter helps to a greater extent for defining the extent of pathology. The significance of this parameter has to be tested for the huge image database to determine the stability in the weight value for NR and MRD. Also more such parameter can be derived to improve the efficiency of classification. To conclude, PCA provides a reliable parameter that can aid the process of classification of ultrasound kidney images.

Acknowledgement

The authors wish to thank the M/s.Mediscan System (P) Ltd., Chennai, for providing the images used for the study. Also they wish to thank Dr. M.Ramasubba Reddy, Associate professor, Bio-medical Engineering Department, IIT Chennai for his valuable suggestion to improve the quality of the analysis.

References

[1]. Y.Adini, Y.Moses and S.Ullman: Face recognition -the problem of Compensating for changes in Illumination Direction, *IEEE Trans. On pattern Analysis and Machine intelligence, Vol.19, no.7,pp .721-732,1997.*

[2]. P.Belhumeur, J.Hespanha, D.Kriegman: Eigenfaces Vs.Fisherfaces.Recognition using class specific linear projection. *IEEE Transactions on pattern Analysis and Machine intelligence,pp.711-720,1997.*

[3]. Bommanna Raja.K, Ramasubba Reddy.M, Swaranamani.S, and Suresh.S "Analysis of Kidney Disorders using Ultrasound Images by Geometric Moments",*Biomedical Engineering:Recen Developments*,J Vossoughi (Editor), ISBN: 1-930636-01-6, 2002 Medical and Engineering Publishers, Inc., Washington D.C., USA, pp 153 -154, Sep 2002.

[4]. Bommanna Raja.K, Ramasubba Reddy.M, Swaranamani.S, Suresh.S, Madheswaran.M, and Thyagarajah.K, "Study on Ultrasound Kidney Images for Implementing Content Based Image retrieval System using Regional Gray-Level Distribution", *Proc. of International Conference on advances in infrastructures for electronic business, education, science, medicine, and mobile technologies on the internet*, ISBN 88-85280-75-7, L-aquila, Italy, Paper No. 93, January 6 - 12, 2003.

[5]. R.Brunelli, T. Poggie, Face Recognition: Features vs. Templates. *IEEE Trans. On PAMI, Vol.12, No.1, Jan. 1990*

[6]. R.cutler: Face recognition using infrared images andeigenfaces.*Citeseer.nj.nec.com/456378.html, 1996*

[7]. Craw, ı., Costen, N., Kato, T., and Akamatsu, S., How should we represent faces for automatic recognition. *Transactions on Pattern Analysis and Machine Intelligence, vol 21, no.8 , 1999.*

[8]. Jorge de la Calleja and Olac Fuen Morphological galaxy classification. *Computer science Department, National Institute of Astrophysics, Optics and Electronics, Tonantzintla 72840, Puebla,Mexico*

[9]. A.Pentland, B.Moghassom, T.Starner: View – Based and Modular Eigenfaces for Face Recognition. *Proc.of IEEE Conf. on Computer Vision and Pattern recognition,1994.*

[10]. A.Pentland and T.Choudry: Face recognition for smart environments. *Computer, Vol.33,Iss2, Feb. 2000*

[11]. A.Samal and P.Ayengar: Automatic recognition and analysis of human faces and facial expressions:*A survey. Pattern Rrecognition,25(I): 65-77,1992.*

[12]. Steve Lawrence, Peter Yianilos and ingemar cox: *NEC Research,4 independence way,Princeton, NJ 08540.*

[13]. Thomas Heseltine, Nick Pears and Jim Austin:Evaluation of image pre_processing techniques for eigenface based face recognition. *Proc.International conference on Image and Graphics, SPIE vol.4875,pp.677-685,2002.*

[14]. M.Turk and A.Pentland: Face Recognition Using eigenfaces. *Proc, IEEE conf. on Computer Vision and Pattern Recognition, pp.586-591,1991.*

[15]. M.Turk and A.Pentland: Eigenfaces for Recognition. *Journal of Cognitive Neuroscience, vol.3,pp.72-86,1991.*