Proceedings of the World Congress on Engineering 2007 Vol I
WCE 2007, July 2 - 4, 2007, London, U.K.

Face Recognition using Discrete Cosine Transform plus Linear Discriminant Analysis

M. Hajiarbabi, J. Askari, S. Sadri, and M. Saraee

***Abstract*—Face recognition is a biometric identification method which among the other methods such as, finger print identification, speech recognition, signature and hand written recognition has assigned a special place to itself. In principle, the biometric identification methods include a wide range of sciences such as machine vision, image processing, pattern recognition neural networks and has various applications in film processing, control access networks and etc. There are several methods for recognition and appearance based methods is one of them. One of the most important algorithms in appearance based methods is linear discriminant analysis (LDA) method. One of the drawbacks for LDA in face recognition is the small sample size (SSS) problem so it is suggested to first reduce the dimension of the space using methods among which, principal component analysis (PCA) is the most popular one. In this paper we show that there exist stronger methods such as discrete cosine transform (DCT).**

***Index Terms*— Discrete cosine transform, Face recognition, Linear discriminant analysis, Principal component analysis, Radial basis function neural network**

I. Introduction

Human identification recognition has attracted the scientists from so many years ago. During these years and due to increasing in terrorism the needs for such systems have increased much more. The most important biometric systems which have been used during these years we can name fingerprint recognition, speech recognition, iris, retina, hand geometry and face recognition. For comparing biometric systems four features have been considered: intrusiveness, accuracy, cost and effort. The investigation has shown that among the other biometric systems, face recognition is the best one [1].

A face recognition system has three parts:

1. Face localization
2. Feature extraction
3. Classification

Manuscript received March 3, 2007.

M. Hajiarbabi, is an M.S. student in the Electrical and Computer Engineering Department, Isfahan University of Technology, Iran (corresponding author to provide phone: 09151107128; e-mail: m_arbabi@ ec.iut.ac.ir).

J. Askari, is with the Electrical and Computer Engineering Department, Isfahan University of Technology, Iran. (e-mail: j-askari@cc.iut.ac.ir).

S. Sadri is with the Electrical and Computer Engineering Department, Isfahan University of Technology, Iran. (e-mail: sadri@cc.iut.ac.ir).

M. Saraee is with the Electrical and Computer Engineering Department, Isfahan University of Technology, Iran. (e-mail: saraee@cc.iut.ac.ir).

In face localization part, the background and other parts of the image that may influence the recognizing process will be removed from the image. For this reason the face will be found in the image and the system will just work on this part of the image. For simplicity we have ignored this part of the system. In the feature extraction part, the unique patterns of the face will be extracted from the image and in classification part these patterns will be placed in the class that it belongs to. Each class shows a person identity. The process of extracting the most discriminating features is very important in every face recognition system. LDA is an efficient method for extracting features but it has some drawbacks. One of the drawbacks that LDA encounters in face recognition is the small sample size problem. In order to avoid singularity of the S_w matrix the number of the training images must be much more than the dimension of the subspace, a situation that rarely occurs in face recognition problem. In order to avoid the singularity problem first we have to reduce the dimension of the problem and then apply LDA. PCA [2] is the most popular method which has been used for dimension reduction. But as we will show in this paper some methods such as DCT could have better results than PCA when applying to images before applying LDA.

The rest of the paper is organized as follows: in section 2 PCA, in section 3 DCT and in section 4 LDA will be reviewed. In section 5 the RBF classifier will be introduced and finally in section 6 the new method results will be compared with PCA on the ORL [3] database.

II. PRINCIPAL COMPONANT ANALYSIS

PCA is a method to efficiently represent a collection of sample points, reducing the dimensionality of the description by projecting the points onto the principal axes, where an orthonormal set of axes points in the direction of maximum covariance in the data. These vectors best account for the distribution of face images within the entire image space. PCA minimizes the mean squared projection error for a given number of dimensions, and provides a measure of importance (in terms of total projection error) for each axis.

Let us now describe the PCA algorithm. Consider that Z_i is a two dimensional image with size $m \times m$. First we convert the matrix into a vector of size m^2. The training set of the n face can be written as:

$$Z = (Z_1, Z_2, ..., Z_n) \subset \Re^{m^2 \times n} \quad (1)$$

Each of the face images belongs to one of the c classes. In face recognition the total images that belong to one person is considered as one class. For the training images the covariance matrix can be computed by:

$$\Gamma = \frac{1}{n}\sum_{i=1}^{n}\left(Z_i - \overline{Z}\right)\left(Z_i - \overline{Z}\right)^T = \Phi\Phi^T \tag{2}$$

where $\Phi = \left(\Phi_1, \Phi_2, ..., \Phi_n\right) \subset \Re^{m^2 \times n}$ and $\overline{Z} = \left(1/n\right)\sum_{i=1}^{n} Z_i$ is the average of the training images in the database. As can be easily seen the face images have been centered which means that the mean of the images is subtracted from each image.

After computing covariance matrix, the eigenvectors and eigenvalues of the covariance matrix will be computed. Consider that $U = \left(U_1, U_2, ..., U_r\right) \subset \Re^{m^2 \times r}\left(r \prec n\right)$ be the eigenvectors of the covariance matrix, only small parts of this eigenvectors, for example r, that have the larger eigenvalues will be enough to reconstruct the image. So by having an initial set of face images $Z \subset \Re^{N^2 \times n}$ the feature vector corresponding to its related eigenface $X \subset \Re^{r \times n}$ can be calculated by projecting Z in the eigenface space by:

$$X = U^T Z \tag{3}$$

III. DISCRETE COSINE TRANSFORM

The DCT transforms spatial information to decoupled frequency information in the form of DCT coefficients. Also it exhibits excellent energy compaction. The definition of DCT for an $N \times N$ image is [4]:

$$C(u,v) = \alpha(u)\alpha(v)\sum_{x=0}^{N-1}\sum_{y=0}^{N-1} Z(x,y)\cos\frac{\pi(2x+1)u}{2N}\cos\frac{\pi(2y+1)v}{2N}$$

$$u,v = 0,1,2,...,N-1 \tag{4}$$

$$\alpha(u) = \begin{cases} \sqrt{1/N} & u = 0 \\ \sqrt{2/N} & u \neq 0 \end{cases}$$

For an $N \times N$ image the matrix coefficients cover the whole frequency space of image components. The DCT coefficients with higher values are placed in the upper left of the matrix. In order to convert the DCT matrix to a vector, we need to start with the upper left side of the matrix and in a zigzag manner put the elements in a vector. As a feature extraction method, DCT would change the face images with high dimension to a subspace with low dimension in which the most important features of face such as the lines belonging to hairs and face, the position of eyes, nose and mouth will remain in the DCT coefficients.

IV. LINEAR DISCRMINANT ANALYSIS

LDA is used for projecting a set of training data. In here these training data are the vectors which have been extracted by PCA and DCT in previous sections. In new face space which is a m^2 dimension and m^2 is the vector dimension, consider that $X = \left(X_1, X_2, ..., X_n\right) \subset \Re^{m^2}$ is a matrix containing the vectors in the training set. In LDA two matrixes within class scatter matrix and between class scatter matrixes is defined. This method finds an optimal subspace in which the between class scatter matrix to the within class scatter matrix will be maximized [5]. The between class scatter matrix is computed by:

$$S_B = \sum_{i=1}^{c} n^i\left(\overline{X}^i - \overline{X}\right)\left(\overline{X}^i - \overline{X}\right)^T \tag{5}$$

Where $\overline{X} = \left(1/n\right)\sum_{j=1}^{n} X_j$ is the mean of the vectors in the training set and $\overline{X}^i = \left(1/n^i\right)\sum_{j=1}^{n^i} X_j^i$ is the mean of class i, c is the number of the classes, and n^i is the elements of i th class. The between class scatter matrix defines the average scattering of one class across the average of the total classes. The within class scatter matrix is computed by:

$$S_W = \sum_{i=1}^{c}\sum_{X_i \in n^i}\left(X_i - \overline{X}^i\right)\left(X_i - \overline{X}^i\right)^T \tag{6}$$

The within class scatter matrix defines the data of one class across the average of the class. The optimal subspace is computed by

$$E_{optimal} = \text{argmax}_E \frac{\left\|E^T S_B E\right\|}{\left\|E^T S_W E\right\|} = \left[c_1, c_2, ..., c_{c-1}\right] \tag{7}$$

Where $\left[c_1, c_2, ..., c_{c-1}\right]$ is the set of eigen vectors of S_B and S_W corresponding to $c-1$ greatest generalized eigen value λ_i and $i = 1,2,...,c-1$

$$S_B E_i = \lambda_i S_W E_i \quad i = 1,2,...,c-1 \tag{8}$$

$E_{optimal}$ is an optimal matrix which maximize the proportion of between class scatter matrix to the within class scatter matrix. This means that it maximize the scattering of the data that belongs to different classes and minimize the scattering of the data belonging to the same class.

Thus the most discriminant answer for face images X would be [5]:

$$P = E_{optimal}^T \cdot X \tag{9}$$

V. RADIAL BASIS FUNCTION NEURAL NETWORK

For classification we used a distance measure and also RBF neural network in order to compare their classification power. As a distance measure we selected Euclidean distance. RBF neural network is a powerful classification method for pattern recognition problems. It doesn't have the drawbacks of multi layer perceptron neural networks and trains much faster than it. Fig. 1 shows an RBF neural network.

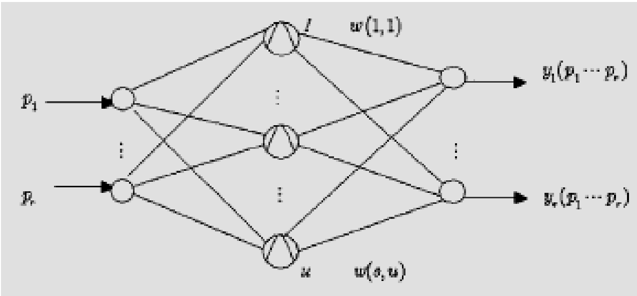

Fig. 1: RBF neural network

Let $P \in \Re^r$ be the input vector and $C_i \in \Re^r$ $(1 \le i \le u)$ be the prototype of the input vectors. The output of each RBF units is as follows:

$$R_i(P) = \exp\frac{-\|P - C_i\|^2}{\sigma_i^2} \tag{10}$$

Where σ_i is the width of the i th RBF unit. The j th output $y_j(P)$ of an RBF neural network is

$$y_i(p) = \sum_{i=1}^{u} R_i(P)*w(j,i) \ + \ w(j,0) \tag{11}$$

Where $R_0 = 1$, $w(j,i)$ is the weight of the i th receptive field to the j th output. The weights of first layer are all equal to one. The number of nodes in the second layer at first equals to the number of classes. Whenever two classes have intersection with each other a node is added to the second layer and a class is split into two subclasses. For further knowledge about RBF neural network the reader can refer to neural network references.

We can summarize the proposed method as follows:

1. Applying DCT on the training set.
2. Computing within class scatter matrix using equation (6).
3. Computing between class scatter matrix using equation (5).
4. Computing optimal subspace using equation (7).
5. Computing most discriminant vectors using equation (9).
6. Applying the new vectors to RBF neural network in order to train the network.
7. Applying DCT to the test data and then compute the most discriminant vector of it.
8. Applying the test vector to neural net in order to classify it.

VI. IMPLEMENTATION AND RESULTS

In order to test the algorithms mentioned above we used ORL database. ORL database contains 400 images that belong to 40 people with variety in scale and small variety in pose head. 5 image from every person was used as training set and the rest used as test set. Fig. 2 shows a sample of this database.

Fig. 2: ORL database

First we applied PCA and DCT to the training images and then we applied LDA to the vectors. We did this task for several features and then by using RBF neural network and also Euclidean distance we classified the data. The input size of the neural network is equal to the size of the vector and the output size of it is 40 equal to number of classes. Fig. 3 and 4 show the results and we can observe that when the numbers of extracted features are 20 to 70 features DCT outperforms PCA. DCT also spends about less than half time to extract features than PCA, on the same platform.

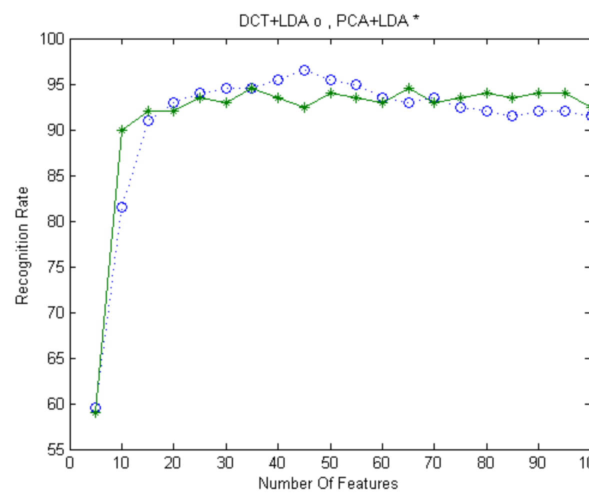

Fig. 3: results for ORL database using Euclidean distance

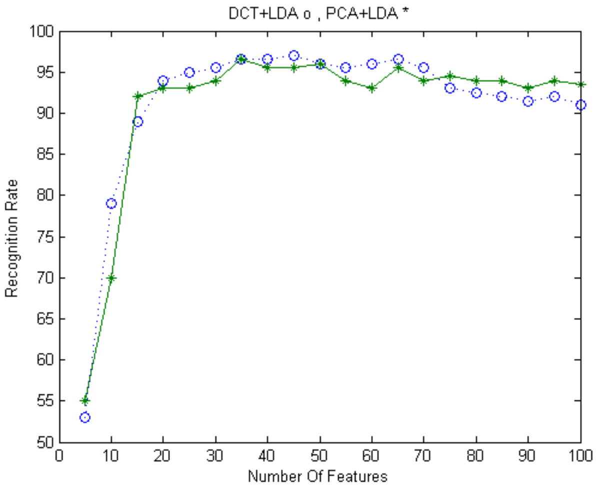

Fig. 4: results for ORL database using RBF

VII. CONCLUSION

In this paper we used DCT as a prior feature extraction for LDA and showed the result on ORL database using Euclidean distance and RBF neural networks as classifiers. The results show that DCT extracts more discriminant features than PCA and also show that RBF classifier is much stronger than distance based classifiers such as Euclidean distance. This work can be applied to other databases to see the result of proposed method on other databases.

REFERENCES

[1] R. Hietmeyer, "Biometric identification promises fast and secure processing of airline passengers," *The Int'l Civil Aviation Organization Journal*, vol. 55, no. 9, 2000, pp. 10–11.

[2] M. Turk, and A. Pentland,, "Eigenfaces for recognition," *J. Cognitiv Neurosci.*, vol. 3, 1991, pp. 71–86.

[3] ORL database: http://www.camorl.co.uk.

[4] P. M. Embree, and B. Kimble, *C language Algorithm for Digital Signal Processing*. Printice Hall, New Jercy , 1991

[5] K. Fukunaga,, *Introduction to statistical pattern recognition*, 2nd ed. San Diego, CA: Academic Press, 1990, pp. 445-450.