Face recognition using eigen-faces, fisher-faces and neural networks

Conference Paper · October 2008

DOI: 10.1109/UKRICIS.2008.4798953 · Source: IEEE Xplore

2 authors:

Hossein Sahoolizadeh
Islamic Azad University Tehran Science and Research Branch

Youness Aliyari
Scotiabank

Face Recognition using Eigen-faces, Fisher-faces and Neural Networks

Hossein Sahoolizadeh[2], Youness Aliyari Ghassabeh[1]

***Abstract*—In this paper, a new face recognition method based on PCA (principal Component Analysis), LDA (Linear Discriminant Analysis) and neural networks is proposed. This method consists of four steps: i) Preprocessing, ii) Dimension reduction using PCA, iii) feature extraction using LDA and iv) classification using neural network. Combination of PCA and LDA is used for improving the capability of LDA when a few samples of images are available and neural network classifier is used to reduce number misclassification caused by not-linearly separable classes. The proposed method was tested on Yale face database. Experimental results on this database demonstrated the effectiveness of the proposed method for face recognition with less misclassification in comparison with previous methods.**

***Index Terms*-Face recognition, Linear discriminant analysis, Neural networks, Principal component analysis.**

I. Introduction

In the last years, Face Recognition [1] has become one of the most challenging tasks in the pattern recognition field. The recognition of faces is very important for many applications such as: video surveillance, retrieval of an identity from a database for criminal investigations and forensic applications. Among various solutions to the problem [2] the most successful seems to be those appearance-based approaches, which generally operate directly on images or appearances of face objects and process the image as two-dimensional patterns. These methods extract features to optimally represent faces belong to a class and separate faces from different classes. Ideally, it is desirable to use only features having high separability power while ignoring the rest. Most effort in the literature have been focused mainly on developing feature extraction methods [5]-[7] and employing powerful classifiers such as probabilistic [8], hidden Markov models (HMMs) [9] neural networks (NNs) [10], [16], [19] and support vector machine (SVM) [11]. The main trend in feature extraction has been representing the data in a lower dimensional space computed through a linear or non-linear transformation satisfying certain properties. Statistical techniques have been widely used for face recognition and in facial analysis to extract the abstract features of the face patterns. Principal component analysis (PCA) [3]-[5] and linear discriminant analysis (LDA) [6] and discrete cosine transform (DCT) [16]-[18] are three main techniques used for data reduction and feature extraction in the appearance-based approaches. DCT [16]-[18], -faces [5] and fisher-faces [6] built based on these three techniques, have been proved to be very successful. DCT remove the some high frequency details and in this way reduce the size of images [16]-[18]. LDA algorithm selects features that are most effective for class separability while PCA selects features important for class representation. A study in [12] demonstrated that PCA might outperform LDA when the number of samples per class is small and in the case of training set with a large number of samples, the LDA still outperform the PCA. Compared to the PCA method, the computational cost of the LDA is much higher [13] and PCA is less sensitive to different training data sets. However, simulations reported in [13] demonstrated an improved performance using the LDA method compared to the PCA approach. When dimensionality of face images is high, LDA is not applicable and therefore we deprive from its advantage to find effective features for class separability. To resolve this problem we combine the PCA and LDA methods. By applying PCA to preprocessed face images, we get low dimensionality images which are ready to extract LDA features. Finally to decrease the error rate in spite of Euclidean distance criteria which was used in [4], we implement a neural network to classify face images based on computed LDA features.

II. Feature Extraction Algorithms

PCA and LDA are two powerful tools used for dimensionality reduction and feature extraction in most of pattern recognition applications. In this section, a brief review of PCA and LDA fundamentals is given.

A. Principal Component Analysis

Let $\{\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_N\}$, $\mathbf{x} \in \Re^n$ be N samples from L classes $\{\omega_1, \omega_2, ..., \omega_L\}$, and $p(\mathbf{x})$ denotes their mixture distribution. In a sequel, it is assumed that a priori probabilities $P(\omega_i)$, $i = 1,2,...,L$, are known. Consider $\mathbf{m}$ and $\mathbf{\Sigma}$ denote mean vector and covariance matrix of samples, respectively. PCA algorithm can be used to find a subspace whose basis vectors correspond to the maximum variance directions in the original n dimensional space. PCA subspace can be used for presentation of data with

[1] Youness Aliyari Ghassabeh is with the electrical and computer engineering department, Tehran University, Tehran, Iran.
E-mail: Y.aliyari@ece.ut.ac.ir.
[2] Hossein Sahoolizadeh is with the electrical engineering department, Islamic Azad University, Arak branch, Arak, Iran.
E-mail: Hosein_Sahooli@yahoo.com.

minimum error in reconstruction of original data. Let $\mathbf{\Phi}_{PCA}^{p}$ denote a linear $n \times p$ transformation matrix that maps the original n dimensional space onto a p dimensional feature subspace where $p < n$. The new feature vectors $\mathbf{y}_i \in \Re^p$ are defined by:

$$\mathbf{y}_i = (\mathbf{\Phi}_{PCA}^{p})^t \mathbf{x}_i \quad i = 1,2,...\mathrm{N} \tag{1}$$

It is easily proved that if the columns of $\mathbf{\Phi}_{PCA}^{p}$ are the eigenvectors of the covariance matrix corresponding to its p largest eigen-values in decreasing order, the optimum feature space for the representation of data is achieved. The covariance matrix and mean vector are estimated by:

$$\hat{\mathbf{\Sigma}} = (\sum_{i=1}^{N} (\mathbf{x}_i - \hat{\mathbf{m}})(\mathbf{x}_i - \hat{\mathbf{m}})^t)/(N-1) \tag{2}$$

$$\mathbf{m}_k = \mathbf{m}_{k-1} + \eta_k (\mathbf{x}_k - \mathbf{m}_{k-1}) \tag{3}$$

Where $\mathbf{m}_k$ is the estimation of the mean value at k-th iteration, $\mathbf{x}_k$ is the *k-th* input image and η_k is learning rate. PCA is a technique to extract features effective for representing data such that the average reconstruction error is minimized. In the other word, PCA algorithm can be used to find a subspace whose basis vectors correspond to the maximum variance directions in the original n dimensional space. PCA transfer function is composed of significant eigenvectors of covariance matrix. The following equation can be used for incremental estimation of covariance matrix:

$$\mathbf{\Sigma}_k = \mathbf{\Sigma}_{k-1} + \eta_k (\mathbf{x}_k \mathbf{x}_k^t - \mathbf{\Sigma}_{k-1}) \tag{4}$$

Where $\mathbf{\Sigma}_k$ is the estimation of the covariance matrix at k-th iteration, $\mathbf{x}_k$ is the incoming input vector and η_k is the learning rate.

B. Linear Discriminant Analysis

LDA searches directions for maximum discrimination of classes in addition to dimensionality reduction. To achieve this goal, within-class and between-class matrices are defined. A within-class [14] scatter matrix is the scatter of the samples around their respective class means $\mathbf{m}_i$:

$$\mathbf{\Sigma}_w = \sum_{i=1}^{L} P(\omega_i) E[(\mathbf{x} - \mathbf{m}_i)(\mathbf{x} - \mathbf{m}_i)^t \mid \omega_i] = \sum_{i=1}^{L} P(\omega_i) \mathbf{\Sigma}_i \tag{5}$$

where $\mathbf{\Sigma}_i$ is the covariance matrix of i-th class. The between-class scatter matrix is the scatter of class means $\mathbf{m}_i$ around the mixture mean $\mathbf{m}$, and is given by:

$$\mathbf{\Sigma}_b = \sum_{i=1}^{L} P(\omega_i)(\mathbf{m}_i - \mathbf{m})(\mathbf{m}_i - \mathbf{m})^t \tag{6}$$

Finally, the mixture scatter matrix is the covariance of all samples regardless of class assignments, and is defined by [14]:

$$\mathbf{\Sigma} = E[(\mathbf{x} - \mathbf{m})(\mathbf{x} - \mathbf{m})^t] = \mathbf{\Sigma}_{\mathbf{w}} + \mathbf{\Sigma}_b \tag{7}$$

Different objective functions have been used as LDA criteria mainly based on a family of functions of scatter matrices. For example, the maximization of the following objective functions have been previously proposed [14]:

$$J_1 = tr(\mathbf{\Sigma}_w^{-1} \mathbf{\Sigma}_b) \tag{8}$$

$$J_2 = \ln\left|\mathbf{\Sigma}_w^{-1} \mathbf{\Sigma}_b\right| = \ln\left|\mathbf{\Sigma}_b\right| - \ln\left|\mathbf{\Sigma}_w\right| \tag{9}$$

$$J_3 = tr(\mathbf{\Sigma}_b) - \mu[tr(\mathbf{\Sigma}_w) - c] \tag{10}$$

$$J_4 = tr(\mathbf{\Sigma}_b)/tr(\mathbf{\Sigma}_w) \tag{11}$$

In LDA, the optimum linear transform is composed of $p(\leq n)$ eigenvectors of $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}_b$ corresponding to its p largest eigen-values. Alternatively, $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}$ can be used for LDA, simple analysis shows that both $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}_b$ and $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}$ have the same eigenvector matrices. In general, $\mathbf{\Sigma}_b$ is not full rank and therefore not a covariance matrix, hence we shall use $\mathbf{\Sigma}$ in place of $\mathbf{\Sigma}_b$. The computation of the eigenvector matrix $\mathbf{\Phi}_{LDA}$ of $\mathbf{\Sigma}_w^{-1}\mathbf{\Sigma}$ is equivalent to the solution of the generalized eigen-value problem $\mathbf{\Sigma}\mathbf{\Phi}_{LDA} = \mathbf{\Sigma}_w \mathbf{\Phi}_{LDA} \mathbf{\Lambda}$ where $\mathbf{\Lambda}$ is the generalized eigen-value matrix. Under assumption of positive definite matrix $\mathbf{\Sigma}_w$, there exists a symmetric $\mathbf{\Sigma}_w^{-1/2}$ such that the problem can be reduced to a symmetric eigen-value problem:

$$\mathbf{\Sigma}_w^{-1/2} \mathbf{\Sigma} \mathbf{\Sigma}_w^{-1/2} \mathbf{\Psi} = \mathbf{\Psi}\mathbf{\Lambda} \tag{12}$$

where $\mathbf{\Psi} = \mathbf{\Sigma}_w^{1/2} \mathbf{\Phi}_{LDA}$.

III. Proposed Face Recognition method

The proposed face recognition method consists of four separate parts:

i) Pre processing
ii) Dimensionality reduction using PCA
iii) Feature extraction for class separability using LDA
iv) Classification using neural network

A. Preprocessing

In this part, at first we manually cut the input images to 40×40 images in order to remove the background information and have only face details (face images in most databases contain background information that is not useful for recognition, figure (1) shows some face images with background information). After that we histogram equalize all input images in order to spread energy of all pixels inside the image and then normalize them to equalize amount of energy related to each face image.

Fig. 1. Sample Face images that used in our experiment.

Fig. 2. Mean images related to 10 different classes.

As a next step, we subtract mean images from face images to mean center all of them. Figure (2) shows mean images of face images that used in our experiments. Finally all preprocessed face images change to a vector (1600×1 vector) and go to the next step. . Figure (3) shows block diagram of operations made in this section. For complexity and memory size reduction, mean image is computed adaptively using (3). [15]

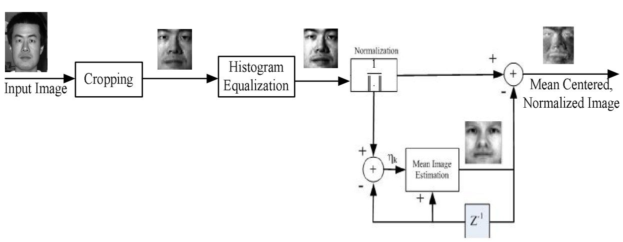

Fig. 3. Preprocessing Block diagram that includes manual cropping, histogram equalization, mean image estimation and mean centering.

B. Dimensionality Reduction

As mentioned in the previous section, we cropped every input image to 40×40 image; as a result the input of this stage is a preprocessed 1600×1 vector. We used these vectors to estimate the covariance matrix [14]. After estimation of the covariance matrix, significant eigenvectors of the covariance matrix are computer. Number of eigenvector depend to our application and accuracy that we need, it is clear that if we compute large number of eigenvectors accuracy of the method will improve but computational complexity will increase in this step and next step . In this stage, we computed 100 most significant eigenvectors and related eigen-faces. By projection of every input image on these eigen-faces, they will convert to a reduced size 100 × 1 vectors which will go to LDA feature extraction part. Figure 4 demonstrated operation done in this stage (covariance estimation using (4) and computation of the eigen-faces). We repeated our experiment with different number of significant eigenvectors equal to 20, 40, and 60 and 80 and compared the performance of the proposed face recognition method.

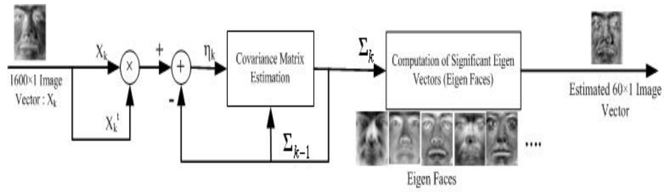

Fig. 4. Dimension reduction Block diagram that includes covariance estimation and computation of eigen-faces.

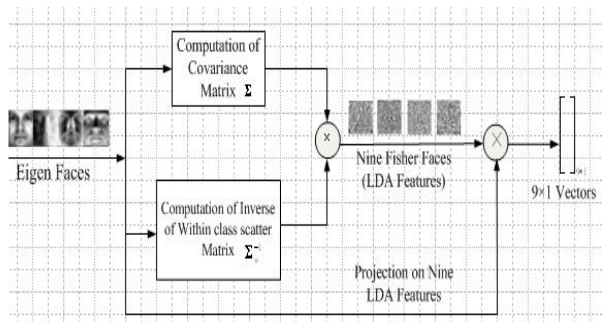

Fig. 5. LDA feature extraction (Fisher faces) Block diagram that includes estimation of covariance matrix and inverse of within class scatter matrix and then computation of the significant fisher-faces.

C. LDA Feature Extraction

Outputs of dimension reduction part are 100×1 vectors which are used to construct within class scatter matrix and covariance matrix. As mentioned in section II, significant eigenvectors of $\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}$ can used for separability of classes in addition to dimension reduction. Using 100×1 vectors, $\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}$ computed and then eigenvectors related to the greater eigen-values are selected. In our experiment we considered 10 classes, therefore there are 9 major eigenvectors (Fisher faces) associated with non-zero eigen-values which have separability capability. It is clear that extracting all of 9 LDA features will increase the discriminatory power of the method. Therefore, this section produce 9×1 vectors which are used as input of MLP three layer neural network. Figure 5 demonstrate operation of this part, at first covariance and within class scatter matrices are estimated and then significant eigenvector of $\boldsymbol{\Sigma}_w^{-1}\boldsymbol{\Sigma}$ are computed (Fisher Faces).

D. Neural Network Classifier

We used a three layer perceptron neural network with 40 neurons in the input layer, 20 neurons in the hidden layer and 10 neurons in the output layer, for classification of the input data. A simple back propagation algorithm is used to update weights according to desired values. Three layers MLP neural network is trained using LDA features and at the output layer, a 10×1 vector will be produced that each elements of that vector is a number between zero and one representing similarity of input face images to each of ten classes. Training LDA features enter the neural network and according to their class, a back propagation error, spread on the network and correct the weights toward the right values. The input LDA features (face images) will classify to the class which has the greatest similarity to it. For example if for a test input face image, row 3 of network's output be greater than other rows, that test face images will classified to class 3. Figure 6 demonstrate process classification using three layer neural network.

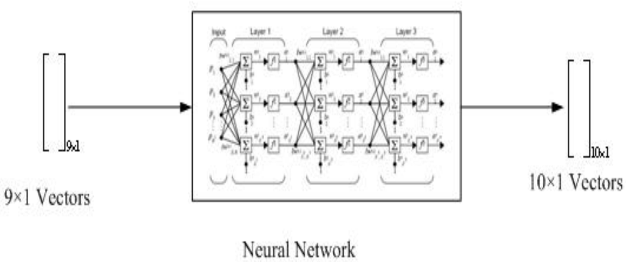

Fig. 6. Classification using three layer MLP neural Network.

IV. Experiment Results

We applied the proposed new face recognition method on YALE face datasets for separation of ten classes. For all experiments, we used Matlab code running on a PC with Intel Pentium 4, 2.8-GHZ CPU and 2048-Mb RAM. Before doing any experiment, we cropped the input images to reduce their size to 40×40. Our selected database contains grayscale images of 10 subjects in GIF format. In these experiments, we considered 60 images per each subject (total 600 images) containing different illumination and different poses, which 40 images of each class are used for training and remaining 20 images used for testing the method. Figure 7 shows some of selected preprocessed subjects in different position and illumination.Then 100 significant eigen-faces are computed in stage 2, where Figure 8 shows first 50of them. As mentioned we repeat the same experiment by extracting 80, 60, 40 and 20 eigen-faces and compared the identification rate of the proposed method, in that cases.

Fig. 7. Preprocessed Sample face Images. 10 face images related to five different classes are shown in the above figure.

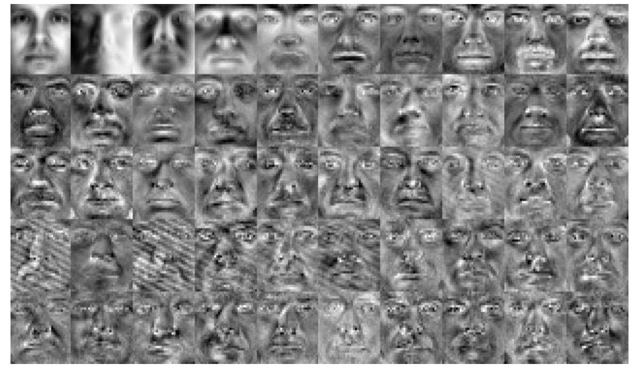

Fig. 8. 50 most significant eigen-faces correspond to significant eigen-values.

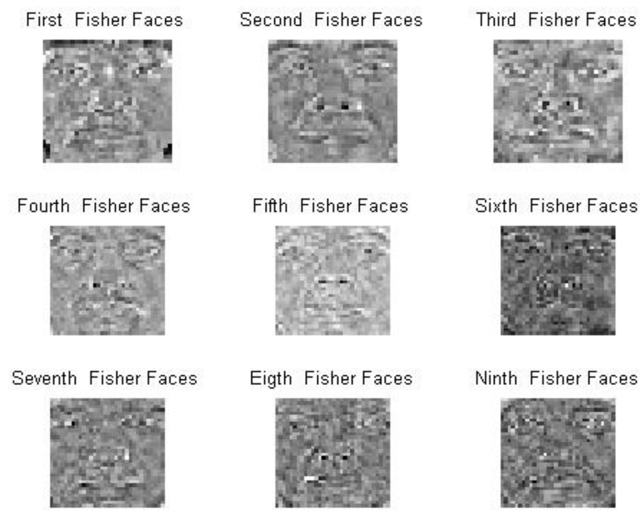

Fig. 9. Computed nine Fisher faces. The first one has the most discriminatory ability and the last one has the least discriminatory power.

In our simulations, we considered 10 subjects. For each selection of eigen-faces we also change the number of selected LDA significant features. We changed number of LDA features from 3 to 9 for each selection of eigen-faces. It means that for example if number of the eigenfaces is equal to 60, we repeated the experiment by LDA features equal to 3 to 9 and compare the error rate. Figure 9 shows 9 estimated fisher faces in the situation that all nine available LDA features are selected. In this case face image projected to nine dimensional space. Figure 10 demonstrate distribution of the face images in the 3 dimensional LDA subspace, some classes are linearly separable and there are overlapping between some other classes. It is clear that overlapped classes may be are separable in the remaining (not shown) 6 dimensional subspaces (it is not possible to demonstrate face images in nine dimensional space resulted by extracting LDA features and only to get a sense about performance of the method we show them in three dimensional space).

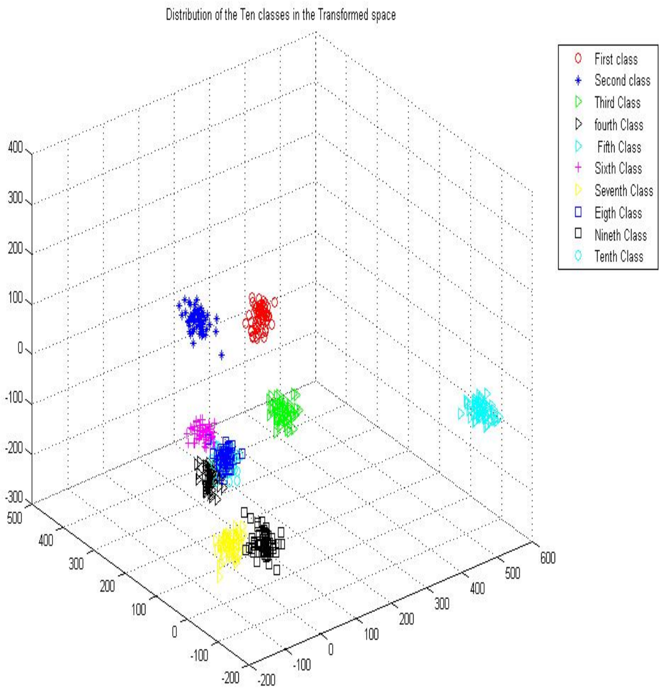

Fig. 10. Distribution of classes in three dimensional LDA feature space. Axes are three most significant LDA features.

The neural network consists of three layers, 40 neurons in the first layer, 20 neurons in the hidden layer and 10 neurons in the output layer. Tangent sigmoid function is used as activation function in the first and second layers and pure linear function is used at the output layer and simple back-propagation learning rule is used for training the weights. Output of the neural network is a 10×1 vector that each element of that vector represents similarity of the input image to one of 10 available classes. For example the first element of the output vector represents similarity to the first class and the other element represents the similarity to other classes. Similarity is between zero and one and the input image is classified to the class that has the greatest similarity to it. 40 images from each class (totally 400 images) are used for network training and remaining 20 images from each class (totally 200 images) are used for testing the proposed method. As mentioned before, the above experiment repeated for different values of PCA eigenvectors and LDA eigenvectors. As the number of selected PCA features increase the reconstruction error of a face images decrease and we get a more precise estimation of it and as the number of LDA features increase separability power raise and error rate reduce. Figure 11 compares average recognition rates by changing the number of PCA and LDA features, x axis represent the number of selected eigen-faces and y axis shows the recognition rate. It is clear that in the case of PCA features equal to 100 and LDA feature equal to 8 or 9, we get the recognition rate equal to 99.5% (in average one misclassification for 200 test face images). The top line in the figure 11 related to the case that all nine LDA feature are computed and as we expected in this case we get the highest recognition rates in comparison with cases that less LDA features are selected. It is also obvious that as number of selected PCA features (eigen-faces) increase (for a fixed number of LDA features) again accuracy of recognition rate improve. Table I compares recognition rate for the selection of different values of LDA features in the case of PCA features equal to 100, 80, 60, 40 and 20. Although choosing 100 PCA features and 8 or 9 LDA features improve the recognition rate in comparison in previous method, but it also increase the computational cost.

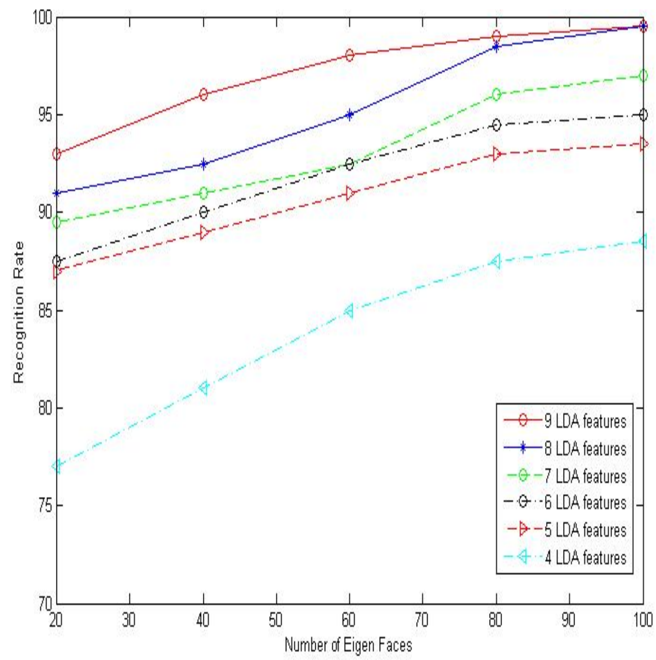

Fig. 11. Comparison of recognition rate for choosing different values of PCA and LDA features. X-axis shows number of selected eigen-faces and y-axis shows recognition rate.

V. Concluding Remarks

In this paper, a new Face recognition method is presented. The new method was considered as a combination of PCA, LDA and neural networks. We used these algorithms to construct efficient face recognition method with a high recognition rate. Proposed method consists of four parts: i) image preprocessing, ii) dimension reduction using PCA iii) feature extraction using LDA and iv) neural network classifier. Simulation results using YALE face datasets demonstrated the ability of the proposed method for optimal feature extraction and efficient face classification. Getting a high recognition rate equal to 99.5% (one misclassification for each 200 face images) demonstrated an improvement in comparison with previous methods. The new face recognition algorithm can be used in many applications such as security applications.

References

[1] A. Jain, R. Bolle, S. Pankanti Eds, "BIOMETRIC – Personal Identification in Networked Society", Kluwer Academic Publishers, Boston/ Dordrecht/ London, 1999.O.

[2] J. R. Solar, P. Navarreto, " Eigen space-based face recognition: a comparative study of different approaches, IEEE Tran. , Systems man And Cybernetics- part c: Applications, Vol. 35, No. 3, 2005 .

[3] D.L. Swets and J.J. Weng, “Using Discriminant Eigen features for image retrieval “, IEEE Trans. Pattern Anal. Machine Intel. , vol. 18, PP. 831-836, Aug. 1996.

[4] P.N. Belhumeur, J.P. Hespanha, and D. J. Kriegman, “Eigen faces vs. Fisher faces: Recognition using class specific linear projection”, IEEE Trans. Pattern Anal. Machine Intel. vol. 19, PP. 711-720, may 1997.

[5] M. Turk, A. Pentland, "Eigen faces for face recognition", Journal cognitive neuroscience, Vol. 3, No.1, 1991.

[6] W. Zhao, R. Chellappa, A, Krishnaswamy, “Discriminant analysis of principal component for face recognition”, .IEEE Trans. Pattern Anal. Machine Intel. , Vol. 8, 1997.

[7] O.Deniz, M. Castrill_on, M. Herrnandez, “Face recognition using independent component analysis and support vector machines” , Pattern Recognition letters, Vol. 24, PP. 2153-2157, 2003.

[8] B. Moghaddam, “Principal manifolds and probabilistic subspaces for visual recognition”, IEEE Trans. pattern Anal. Machine Intel. , Vol. 24, No. 6, PP. 780-788, 2002.

[9] H. Othman, T. Aboulnasr, “A separable low complexity 2D HMM with application to face recognition” IEEE Trans. Pattern. Anal. Machine Intel. , Vol. 25, No. 10, PP. 1229-1238, 2003.

[10] M. Er, S. Wu, J. Lu, L.H.Toh, “face recognition with radial basis function(RBF) neural networks”, IEEE Trans. Neural Networks, Vol. 13, No. 3, PP. 697-710, 1999.

[11] K. Lee, Y. Chung, H. Byun, “SVM based face verification with feature set of small size”, electronic letters, Vol. 38, No. 15, PP. 787-789, 2002.

[12] A. M. Martinez, A. C. Kak, “PCA versus LDA”, IEEE Trans. Pattern Anal. Machine Intel. , Vol. 23, pp. 228-233. 2004.

[13] J. J. Weng, “using discriminant eigen-features for image retrieval”, IEEE Trans. Pattern Anal. Machine Intel. , Vol. 18, No. 8, pp. 831-836, 1996.

[14] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd Edition, Academic Press, New York, 1990.

[15] S. Pang, S. Ozawa, N. Kasabov, ”Incremental linear discriminant analysis for classification of data streams”, IEEE Trans. on Systems, Man and Cybernetics, vol. 35, no. 5, pp. 905-914, 2005.

[16] M. J. Er, W. Chen, S. Wu, “High speed face recognition based on discrete cosine transform and RBF neural network”, IEEE Trans. On Neural Network, Vol. 16, No. 3, PP. 679,691, 2005.

[17] D. Ramasubramanian, Y. Venkatesh, “Encoding and recognition of Faces based on human visual model and DCT”, Pattern recognition, Vol. 34, PP. 2447-2458, 2001.

[18] X. Y. Jing, D. Zhang, “A face and palm print recognition approach based on discriminant DCT feature extraction”, IEEE Trans. On Sys. Man & Cyber. , Vol. 34, No. 6, PP. 2405-2415, 2004.

[19] Z. Pan, A. G. Rust, H. Bolouri,, “ Image redundancy reduction for neural network classification using discrete cosine transform”, Proc. Int. Conf. on Neural Network, Vol. 3, PP. 149,154, Italy, 2000.

TABLE I.
Comparison of recognition rate by choosing different number of PCA and LDA features

	4 LDA features	5 LDA features	6 LDA features	7 LDA features	8 LDA features	9 LDA features
20 PCA features	77%	84%	87.5	87.5%	92%	94.5
40 PCA features	81%	86%	89.5%	90%	93.5%	96.5%
60 PCA features	85%	90.5%	93%	93%	95%	98%
80 PCA features	87.5	92%	94.5	96	98.5%	99%
100 PCA features	88.5%	92.5%	95%	97%	99.5%	99.5%