A Novel Face Recognition Method Using PCA, LDA and Support Vector Machine

U. Raghavendra[1], P.K. Mahesh[2], and Anjan Gudigar[2]

[1] Manipal Institute of Technology, Manipal, India
raghu_u109@rediffmail.com
[2] Department of Electronics and communication, M.I.T.E., Moodbidri, India
{mahesh24pk,anjangudigar83}@gmail.com

Abstract. Here an efficient and novel approach was considered as a combination of PCA, LDA and support vector machine. This method consists of three steps: I) dimension reduction using PCA, ii) feature extraction using LDA, iii) classification using SVM. Combination of PCA and LDA is used for improving the capability of LDA when new samples of images are available and SVM is used to reduce misclassification caused by not linearly separable classes.

Keywords: Dimension Reduction, Feature Extraction, Classification, Support Vector Machine.

1 Introduction

The face is the primary focus of attention in the society. The ability of human beings to remember and recognize faces is quite robust. Automation of this process finds practical application in various tasks such as criminal identification, security systems and human-computer interactions [6]. Various attempts to implement this process have been made over the years. In the last years face recognition has become one of the most challenging tasks in the pattern recognition field. The recognition of the face is very important for much application such as video surveillance, retrieval of an identity from a data base for criminal investigation and forensic application.

Appearance based approaches gives the most successful solution since it operate directly on appearance or images of face objects and process the image as two dimensional space. These method extract features to optimally represent faces belongs 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 has been focused mainly on developing feature extraction methods and employing power full classifiers. The normal approach in feature extraction is representing the data in a higher dimensional space to lower dimensional space. Principal component analysis (PCA), linear discriminant analysis (LDA) [2] and discrete cosine transform (DCT) are three main techniques used for data reduction and feature extraction in the appearance-based approaches. DCT, Eigen-faces and fisher-faces 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.LDA algorithm selects features that are most effective for class separability

N. Meghanathan et al. (Eds.): CCSIT 2012, Part II, LNICST 85, pp. 241–249, 2012.

while PCA selects features important for class representation. A study has 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 computation of the LDA is much higher and PCA is less sensitive to different training data sets. Oftenly LDA is not applicable for the face images with high dimensional 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 pre-processed face images, we get low dimensionality images which are ready for applying LDA. Oftenly intention is to decrease the rate of change of error. We implement a Support vector machines (SVMs) to classify face images based on its computed LDA features.

2 Related Work

As we know dimensionality reduction is a very essential part in real time processing. The most commonly used techniques are PCA and LDA. The following section gives a brief review of these methods.

A. Principal Component Analysis

Let $\{X1, X2.......XN\}, X \in \Re_n$ be N samples from L classes $\{W1, W2...WL\}$, and $P(X)$ their mixture distribution [1]. Let us assume the probabilities $P(Wi), i = 1, 2, 3..., L$ are known in advance. Now consider m as a mean vector and Σ is the covariance matrix of samples. Now, find out a subspace by calculating its basis vector which corresponds to maximum variance direction in the space. This sub space is very essential and accurate for presenting original data with minimum error. Consider a linear $n \times p$ transformation matrix, it is possible to map original n dimensional space on to the p dimensional feature subspace where $p < n$. Now we can define a new feature vector $y_i = \Re^n$ as

$$y_i = (\Phi^p_{PCA})^t X_i, i = 1, ..., N \quad (1)$$

And it has proved that in Φ^p_{PCA}, if Eigen vector of covariance matrix corresponds to p largest Eigen values and which are in decreasing order, then a new space is available for data representation.

$$\widehat{\Sigma} = (\textstyle\sum_{i=1}^{N} (x_i - \widehat{m})(x_i - \widehat{m})^t)/(N - 1) \quad (2)$$

Where m in (2) can be estimated by:

$$m_k = m_{k-1} - \eta(X_k - m_{k-1}) \quad (3)$$

Where m_k is estimation of mean value at kth iteration and x_k is the *kth* input image. PCA is also a very popular method for extracting features for data representation. So the transfer function of the PCA consists of Eigen vectors of covariance matrix [7]. The incremental estimation of the covariance matrix can be done using following equation.

$$\sum_k = \sum_{k-1} - \eta_k (X_k X_k^t - \sum_{k-1}) \tag{4}$$

Where $\sum_k$ the estimation of the covariance matrix at k-th iteration is X_k is the incoming input vector and η_k is the learning rate.

B. Linear Discriminant Analysis

In addition to dimensionality reduction, LDA searches the directions for maximum discrimination of classes [2, 4]. By defining within-class and between-class matrices this goal achieved, A within-class scatter matrix is the scatter of the samples around their respective class means $\boldsymbol{m_i}$**:**

$$\sum_W = \sum_{i=1}^{L} P(w_i) E[(X - m_i)(X - m_i)^t | w_i] + \sum_{i=1}^{L} P(w_i) \sum_i \tag{5}$$

Where $\sum_{\mathbf{i}}$ denotes the covariance matrix of i-th class. The between-class scatter matrix is the scatter of class means $\boldsymbol{m_i}$ around the mixture mean **m,** and is given by:

$$\sum_b = \sum_{i=1}^{L} P(w_i)(m_i - m)(m_i - m)^t \tag{6}$$

Regardless of class assignments, Finally,the scatter matrix is the covariance of all samples and is defined by:

$$\sum = E[(X - m_i)(X - m_i)^t] = \sum_W + \sum_b \tag{7}$$

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

$$\Sigma_W^{-1/2} \Sigma \Sigma_W^{-1/2} \Psi = \Psi \Lambda \tag{8}$$

Where

$$\Psi = \Sigma_W^{1/2} \Phi_{\boldsymbol{LDA}} \tag{9}$$

3 Proposed Methodology

The proposed face recognition method consists of four separate parts:

i) Dimensionality reduction using PCA
ii) Feature extraction for class separability using LDA
iii) Classification using support vector machine.

In the next sections, we describe role of each part.

A. Dimensionality Reduction

In this section we have used cropped input image of size 40×40; as a result the input of this stage is a pre-processed 1600×1 vector. Here these vectors are used to estimate the covariance matrix. The significant eigenvectors of the covariance matrix are computed.

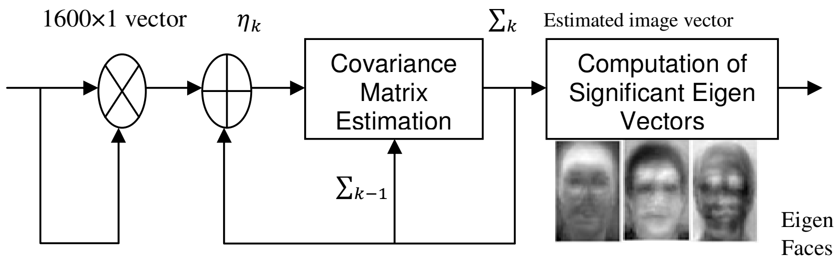

Fig. 1. Dimensionality reduction Block diagram

Computational complexity increased in this step and next step. As number of Eigen-vectors are increased to achieve the accuracy that we need. In this stage, we computed 50 most significant Eigen-vectors and related Eigen-faces. By projection of every input image on these Eigen faces, they will convert to reduce size 50 × 1 vectors which will be go to LDA feature extraction part. Fig. 1 shows the estimation and computation of both the Eigen-faces. We repeated our experiment with different values of significant Eigen-vectors and choose them equal to 20, 30, and 40 and 50 and compared the performance of the proposed face recognition method.

B. LDA Feature Extraction

Outputs of dimension reduction part are 50×1 vectors which are used to construct within class scatter matrix and covariance matrix. As mentioned in section 2, significant eigenvectors of $\Sigma^{-1}_{w}\Sigma$ can use for separability of classes in addition to dimension reduction. Using 50×1 vectors, $\Sigma^{-1}_{w}\Sigma$ computed and then Eigen vectors 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 increase the discriminatory power of the method. Therefore, this section produces 9×1 vectors. Fig. 2 demonstrates operation of this part, at first covariance and within class scatter matrices are estimated and then significant eigenvector of $\Sigma^{-1}_{w}\Sigma$ are computed (Fisher Faces).

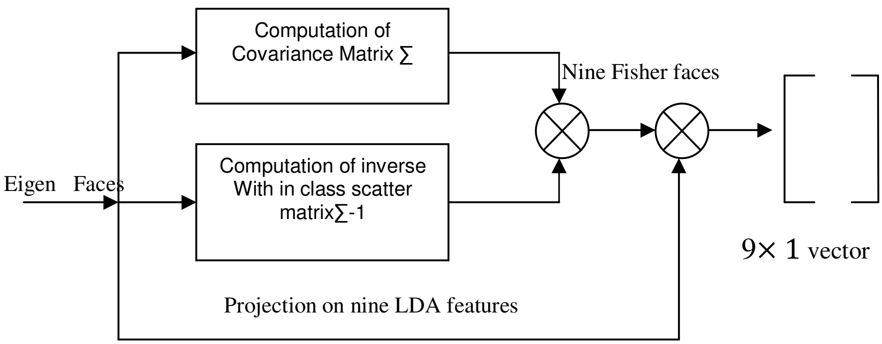

Fig. 2. Generation of Fisher faces

C. Support vector machines (SVMs)

Output of the LDA, which is nothing but nine fisher face is given as the input to the support vector machine as in Fig 3. Then SVM generates support vector for the given nine fisher faces by projecting it to a higher dimension space using Gaussian kernel.

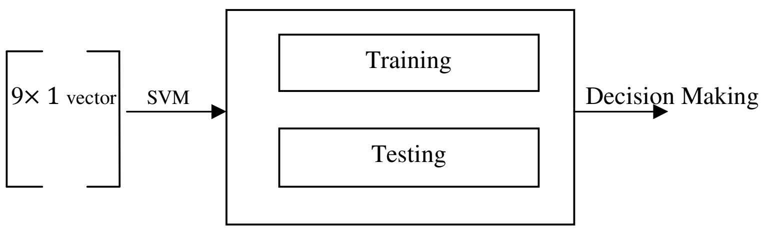

Fig. 3. Classification using Support Vector Machine

For the training process of SVMs, we used the library LIBSVMS [8]. However, in many cases, the data cannot be separated by linear function. A solution is to map the input data into a higher dimensional space as shown in Fig.4. Due to the fact that the training data are used through a dot product, if there was a "kernel function" so that we satisfy K(x_i , x_j) =(ϕ (x_i) , ϕ (x_j)), we can avoid computing ϕ_x explicitly and use the kernel function K(x_i , x_j) .we have used a Gaussian kernel as follows:

$$K\,(X_{i,}X_j\,) = e^{\frac{-||X_{i,}-X_j||^2}{2\sigma^2}} \tag{10}$$

$$f(x) = sgn\,(\textstyle\sum_{i=1}^{N_s} \alpha_i\, y_i K(s_i, x) + b) \tag{11}$$

Where $N_s$the number of support is vectors , and $s_i$are the support vectors.

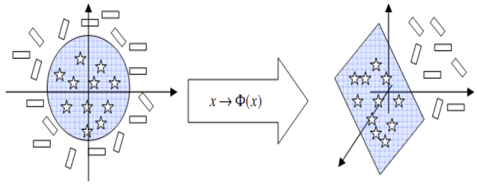

Fig. 4. Mapping of data to a higher dimensional space

Here proper tuning is required for value of σ,either by using Manual tuning or brute force search. An brute force technique could involve stepping through a range of values for σ, perhaps in a gradient ascent optimization, seeking optimal performance of a model with training data. Although this approach is feasible with supervised learning, it is much more difficult to tune σ for unsupervised learning methods. During the testing time the same process has been done and SVM predictor does the classification of the given input data with trained data.

4 Experimental Results

We applied the proposed new face recognition method on ORL face datasets for separation of ten classes. For all experiments, we used Mat lab code running on a PC with Intel dual core CPU and 4096-Mb RAM. Before doing any experiment, we have used cropped images of size to 40×40.

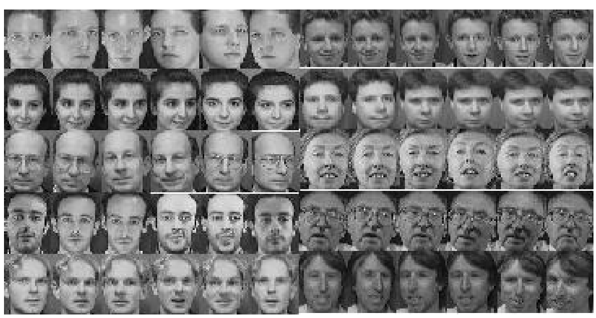

Fig. 5. Pre-Processed Sample Face Images

Our selected database contains gray scale images of 10 subjects in PNG format. In these experiments, we considered 6 images per each subject (total 60 images) containing different illumination and different poses, which 60 images of each used for training and remaining 40 images used for testing the method. Fig. 5 shows some of selected pre-processed subjects in different position and illumination. Then 50 significant Eigen faces are computed in stage II, where Fig.6 shows those 50 significant Eigen faces.

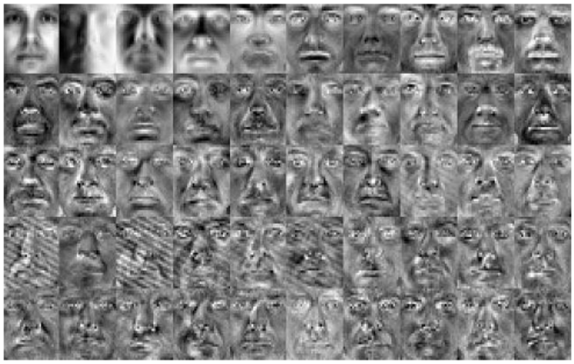

Fig. 6. .50 Most Significant Eigen Faces

As mentioned we repeat the same experiment by extracting 50, 40, 30 and 20 Eigen faces and compared the identification rate of the proposed method, in that cases. 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 for the case, Eigen faces equal to 50, we repeated the experiment by LDA features equal to 3 to 9 and compared the error rate. Fig. 7 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 spaces.

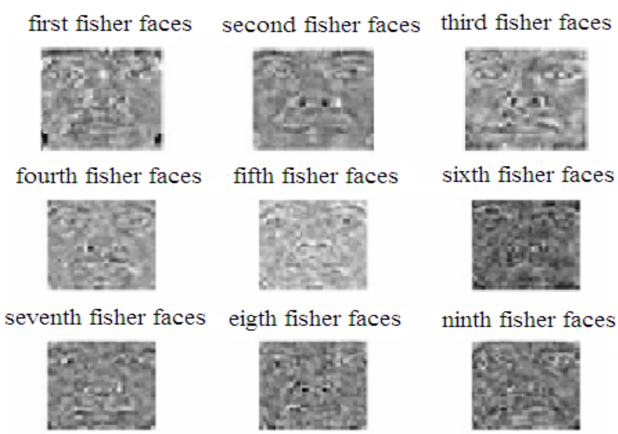

Fig. 7. Computed nine fisher faces

Fig.8 compares average recognition rates by changing the number of PCA and LDA features, it is clear that in the case of PCA features equal to 50 and LDA feature equal to 8 or 9, we get the recognition rate equal to 96.67% (in average two

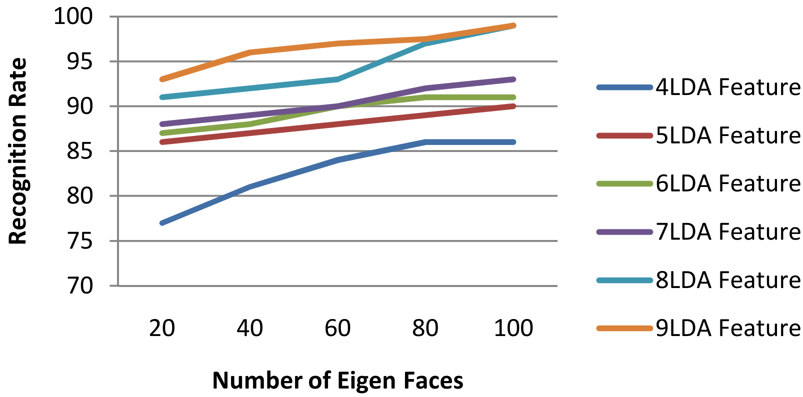

Fig. 8. Comparison of Recognition Rate for Choosing Different Values of Pca and Lda Features

misclassification for 60 test face images). The top line in the Fig. 8 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) increases (for a fixed number of LDA features) again accuracy of recognition rate improves.

Although choosing 50 PCA features and 8 or 9 LDA features improve the recognition rate in comparison to previous method, but it also increase the computational cost.

5 Conclusion

Here an efficient and novel approach was considered as a combination of PCA, LDA and support vector machine. We can use these algorithms to construct efficient face recognition method with a high recognition rate. Proposed method consists of three parts: i) dimension reduction using PCA that main features that are important for representing face images are extracted ii) feature extraction using LDA that significant features for class separability are selected and iii) svm classifier that classify input face images into one of available classes. Simulation results using ORL face datasets demonstrated the ability of the proposed method for optimal feature extraction and efficient face classification. In our simulations, we chose 10 persons and considered 6 training image and 4 test image for each person (totally 60 training and 40 test face images). Experimental results show a high recognition rate equal to 96.67% (in average two misclassification for 60 face images) which demonstrated an improvement in comparison with previous methods. The new face recognition algorithm can be used in many applications such as security methods.

References

[1] Turk, M., Pentland, A.: Eigen faces for recognition. Journal of Cognitive Neuroscience 3, 72–86 (1991)
[2] Etemad, K., Chellappa, R.: Discriminant Analysis for Recognition of Human Face Images. Journal of Optical Society of America A, 1724–1733 (1997)
[3] Belhumeur, P., Hespanha, J., Kriegman, D.: Eigen faces vs. Fisher-faces: Recognition Using Class Specific Linear Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 19(7), 711–720, 0162-8828 (1997)
[4] Yan, S., Xu, D., Yang, Q., Zhang, L.: Multi linear Discriminant Analysis for Face Recognition. IEEE Transactions on Image Processing 16(1) (January 2007)
[5] Aizerman, M., Braverman, E., Rozonoer, L.: Theoretical foundations of the potential function method in pattern recognition learning. Automation and Remote Control 25, 821–837 (1964)
[6] Jain, A., Bolle, R., Pankanti, S. (eds.): BIOMETRIC – Personal Identification in Networked Society. Kluwer Academic Publishers, Boston (1999)
[7] Solar, J.R., Navarreto, P.: Eigen space-based face recognition: a comparative study of different approaches. IEEE Tran., Systems Man and Cybernetics- Part C: Applications 35(3) (2005)
[8] Chang, C., Lin, C.: LIBSVM: A Library for Support Vector Machines (2001), http://www.csie.ntu.edu.tw/~cjlin/libsvm