A MODIFIED PCA ALGORITHM FOR FACE RECOGNITION

Lin Luo, M.N.S. Swamy, Fellow, IEEE, Eugene I. Plotkin, Fellow, IEEE
ECE Dept., Concordia University, Montreal, H3G 1M8, Canada
{Lin, Swamy, Eugene}@ece.concordia.ca

Abstract

In principal component analysis (PCA) algorithm for face recognition, the eigenvectors associated with the large eigenvalues are empirically regarded as representing the changes in the illumination; hence, when we extract the feature vector, the influence of the large eigenvectors should be reduced. In this paper, we propose a modified principal component analysis (MPCA) algorithm for face recognition, and this method is based on the idea of reducing the influence of the eigenvectors associated with the large eigenvalues by normalizing the feature vector element by its corresponding standard deviation. The Yale face database and Yale face database B are used to verify our method and compare it with the commonly used algorithms, namely, PCA and linear discriminant analysis (LDA). The simulation results show that the proposed method results in a better performance than the conventional PCA and LDA approaches, and the computation at cost remains the same as that of the PCA, and much less than that of the LDA.

Keywords: *Face recognition; pattern recognition; principle component analysis; linear discriminant analysis.*

1. Introduction

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) [1,2] and linear discriminant analysis (LDA) [3,4] fall into this category. The simulations in [3] show an improved performance using the LDA method compared to the PCA approach. However, the work contained in [5] demonstrates that the PCA might outperform the LDA when the number of samples per class is small, and in the case of a training set with a large number of samples, the LDA still outperforms the PCA. Compared to the PCA method, the computation of the LDA is much higher [3,6]. In this paper, we only consider the case of a training set with a large number of samples. Pentland et al. [7] have empirically shown that the results of superior face recognition can be achieved when the first three eigenvectors, which are associated with the first three largest eigenvalues, are not included (because the first three eigenvectors seem to represent the changes in the illumination). However, it has been demonstrated in a recent study [8] that the elimination of more than three eigenvectors will, in general, worsen the results, due to the loss of some useful information. In this paper, we present a modified principal component analysis (MPCA) method, which is based on the idea of reducing the influence of the eigenvectors associated with the changes in the illumination. The simulation results show that our method leads to an improvement in the recognition performance in contrast to the traditional PCA and the LDA, and do not increase the cost of computation. In Section 2, the PCA algorithm is reviewed. The MPCA algorithm is derived in Section 3. Simulation results and the analysis of these results are presented in Section 4. Section 5 contains the conclusion.

2. Reviews of the PCA algorithm

Consider a training set with the following parameters: the face training set has M images $\mathbf{x}_i \in \Re^D, i = 1,2,\cdots,M$ belonging to N subjects (classes), and D is the number of pixels in the image. The total scatter matrix $\mathbf{S}_T \in \Re^{D\times D}$ is defined as $\mathbf{S}_T = \sum_{i=1}^{M}(\mathbf{x}_i - \mu)(\mathbf{x}_i - \mu)^T = \mathbf{AA}^T$, where μ is the global mean image of the training set, computed by $\left(\sum_{i=1}^{M}\mathbf{x}_i\right)/M$, and $\mathbf{A} = [\ \mathbf{x}_1 - \mu\ \mathbf{x}_2 - \mu\ \cdots\ \mathbf{x}_M - \mu\ ] \in \Re^{D\times M}$.

CCECE 2003-CCGEI 2003, Montréal, May/mai, 2003

The aim of the PCA is to identify the subspace of the image space spanned by the training object image data and to decorrelate the pixel values. This can be achieved by finding the eigenvectors $\mathbf{W}_{pca}$ of the matrix $\mathbf{S}_T$ associated with the nonzero eigenvalues Λ by solving the problem of the eigenstructure decomposition, $\mathbf{S}_T\mathbf{W}_{pca}=\mathbf{W}_{pca}\Lambda$, where the eigenvectors form the feature subspace.

However, a direct computation of $\mathbf{S}_T$ is impractical because of the huge size ($D \times D$) of the images. Usually, we can construct the matrix $\mathbf{R} = \mathbf{A}^T\mathbf{A} \in \Re^{M\times M}$, $M << D$, and obtain the eigenvectors $\mathbf{V}_{pca} \in \Re^{M\times r}$ by solving the eigenstructure decomposition, $\mathbf{R}\mathbf{V}_{pca}=\mathbf{V}_{pca}\Lambda$, where $\Lambda = diag\ [\lambda_0 \lambda_1 \cdots \lambda_{r-1}] \in \Re^{r\times r}$ ($\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{r-1}$), $\lambda_i (i=0,1,\cdots,r-1)$ are the nonzero eigenvalues of $\mathbf{R}$. Then the PCA subspace $\mathbf{W}_{pca}$ is formed by multiplying the matrix $\mathbf{A}$ with the eigenvectors $\mathbf{V}_{pca}$, that is, $\mathbf{W}_{pca}=\mathbf{A}\mathbf{V}_{pca}\Lambda^{-1/2}$, where $\mathbf{W}_{pca}=[\mathbf{w}_0\ \ \mathbf{w}_1\ \cdots\ \mathbf{w}_{r-1}] \in \Re^{D\times r}$, and its column vectors $\mathbf{w}_i$ ($i=0,1,\cdots,r-1$), associated with the eigenvalue λ_i, are orthonormal. Therefore, the feature vector $\mathbf{y}$ of an image $\mathbf{x}$ is acquired by projecting $\mathbf{x}$ into the coordinate system defined by the PCA subspace, as $\mathbf{y}=\mathbf{W}_{pca}^T(\mathbf{x}-\boldsymbol{\mu})$.

3. MPCA Algorithm

In this section, we will analyse the probability characteristics of the feature vectors of the training set in the PCA eigenface subspace. We project the centralized training face images denoted by the vector $\overline{\mathbf{x}}_i = \mathbf{x}_i - \boldsymbol{\mu}$ into the PCA eigenface subspace and obtain the feature vector $\mathbf{y}_i$ of the training image as,

$$\mathbf{y}_i = \mathbf{W}_{pca}^T \overline{\mathbf{x}}_i = [y_{i0}\quad y_{i1}\quad \cdots\quad y_{i(r-1)}]^T \tag{1}$$

Thus, $\overline{\mathbf{x}}_i$ can be expressed by a linear combination of the basis vectors of the PCA subspace as

$$\overline{\mathbf{x}}_i = \sum_{j=0}^{r-1} y_{ij}\mathbf{w}_j \quad (i=1,2,\cdots,M) \tag{2}$$

where M is the number of samples in the training set, and the coefficients $y_{i0}, y_{i1}, \cdots, y_{i(r-1)}$ are uncorrelated. The second order moment of the feature vector $\mathbf{y}_i$ is evaluated as follows

$$\mathbf{E}[\mathbf{y}_i\mathbf{y}_i^T] = \mathbf{W}_{pca}^T\mathbf{S}_T\mathbf{W}_{pca} = \Lambda \tag{3}$$

It can be easily observed that the variance of the jth element of the feature vector $\mathbf{y}_i$ is the jth eigenvalue, i.e., $\mathbf{E}[y_{ij}y_{ij}] = \lambda_j\ (j=0,1,\cdots,r-1)$. Since $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{r-1}$, we have

$$\mathbf{E}[y_{i0}y_{i0}] \geq \mathbf{E}[y_{i1}y_{i1}] \geq \cdots \geq \mathbf{E}[y_{i(r-1)}y_{i(r-1)}] \tag{4}$$

From the above equation, we can see that after the projection of the training images into the PCA eigenface subspace, the variances of the different elements in the feature vector are ordered as in Equation 4. The variance of the elements associated with the eigenvector, which corresponds to a large eigenvalue, is also large. Thus, when we use the eigenvectors to express $\overline{\mathbf{x}}_i$, the eigenvectors related to the large eigenvalues have more impact on the feature vector element. However, the eigenvectors associated with the largest eigenvalues are empirically regarded to represent the changes in the illumination. Therefore, the influence of the illumination should be reduced before we use these eigenvectors to calculate the feature space.

In our MPCA approach, we reduce the influence of the eigenvectors corresponding to the large eigenvalues by normalizing the jth element y_{ij} of the ith feature vector $\mathbf{y}_i$ with respect to its standard deviation, $\sqrt{\lambda_j}$. Hence, the new feature vector $\mathbf{y}_i'$ is rewritten as,

$$\mathbf{y}_i' = [\frac{y_{i0}}{\sqrt{\lambda_0}}\quad \frac{y_{i1}}{\sqrt{\lambda_1}}\quad \cdots\quad \frac{y_{i(r-1)}}{\sqrt{\lambda_{r-1}}}]^T \tag{5}$$

These normalized feature vectors are used to construct a new feature subspace as explained below.

In the traditional PCA algorithm, after projecting the training and testing sets into the eigenface space, the feature vectors are used to compute the corresponding Euclidean distance. In our approach, we first normalize the feature vectors by the square root of the corresponding eigenvalues, and then calculate the distance between the training and the testing images. In the actual programme, the step of calculating PCA space, $\mathbf{W}_{pca}=\mathbf{A}\mathbf{V}_{pca}\Lambda^{-1/2}$, and the step of the normalization can be combined to be one step, i.e., $\mathbf{W}_{mpca}=\mathbf{A}\mathbf{V}_{pca}\Lambda^{-1}$. The remaining steps are the same as in the case of the conventional PCA. It is very clear that the cost of computation of MPCA is the same as that of the classical PCA algorithm.

In the next section, simulations are carried out based on different face databases using the conventional PCA, LDA and the proposed MPCA, and the results analyzed.

4. Simulation Results

Two face databases are used in our simulations, namely, the Yale Face Database and Yale Face database B [9]. Yale face database contains 165 grey scale images of 15 individuals in the GIF format. There are 11 images per subject, one for each different facial expression or configuration. The sample images are shown in Figure 1 (a).

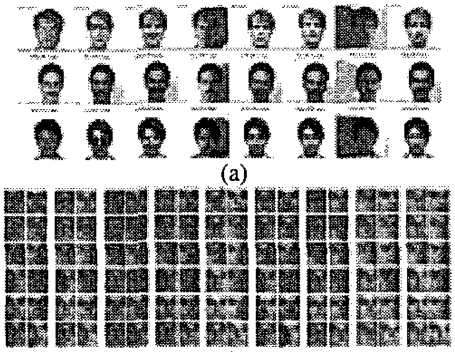

Fig. 1. Sample images: (a) Yale face database
(b) Yale face database B

Yale Face Database B contains 5760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses × 64 illumination conditions). Figure 1(b) describes the registered images of a single individual.

There are various facial expressions and illumination conditions in the Yale face database. For our simulations, the protocol combined with "Leave-One-Out" [10] strategy and rotation is used. The original images are registered and cropped into size 50(w)×60(h), without photometric normalization. The minimum distance classifier based on the Eucludean distance is applied as the matching scheme in this study.

The simulation results are presented in terms of the receiver operating characteristics (ROC) to show the relationship between the false rejection (FR) and false acceptance (FA) as a function of the decision threshold. Figure 2 depicts the ROC curve of the face recognition experiments on the Yale face database. Table 1 gives a summary of the results for the equal error rate (EER) point on the ROC curves. It can be easily seen from Figure 2 and Table 1 that the MPCA algorithm achieves the best performance among the three algorithms, namely, the PCA, LDA and MPCA.

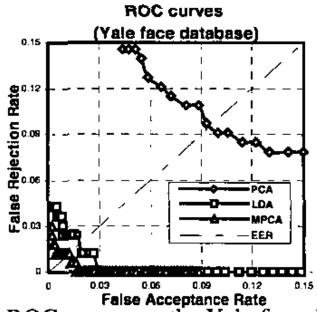

Fig. 2. ROC curves on the Yale face database

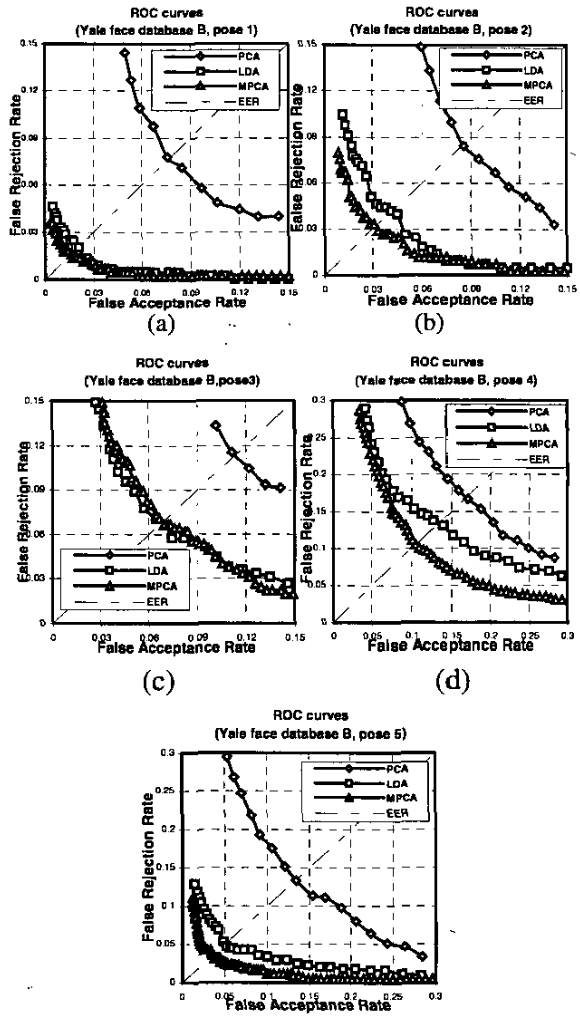

Fig. 3. ROC curves on the Yale face database B

In the Yale face database B, the facial poses and illumination conditions are changed. All the images are registered and cropped into size of 50(w) × 60(h). Due to the large-scale variation in the illumination conditions, the histogram equalization procedure is used to reduce the influence of the illumination prior to the application of the recognition algorithms. For every training and testing set, 45 images (for which the illumination azimuth and elevation angles are less than 70 degrees) are selected from 65 original images per pose set. Pose 0 is the frontal pose, and poses 1, 2, 3, 4, and 5 are about 12 degrees from the camera optical axis (i.e., relative to pose 0). We choose poses 0, 1, 2, 3, 4, 5 to be the training and testing sets in our simulations.

Table 1 EER on Yale face database

	PCA	LDA	MPCA
EER (%)	9.52	1.82	**1.21**

Table 2 EER on Yale face database B

EER (%)	PCA	LDA	MPCA
Pose 0 (Leave-one-out)	8.98	**0.00**	**0.00**
Pose 1	7.70	2.00	**1.58**
Pose 2	8.51	4.20	**3.21**
Pose 3	11.34	**6.84**	**6.84**
Pose 4	16.88	13.42	**10.48**
Pose 5	13.39	5.01	**3.59**

Figure 3 shows the ROC curves of the face recognition simulations on the Yale face database B, where pose 0 set is used to be the training set in all the simulations, and poses 1, 2, 3, 4 and 5 are the testing sets for Figures 3 (a), (b), (c), (d) and (e) respectively.

The EER for all the simulations on the Yale face database B are given in Table 2. The first row contains the results on pose 0 set, where "Leave-One-Out" procedure is adopted. The remaining rows correspond to the results with respect to poses 1,2,3,4,5 respectively. Figure 3 and Table 2 clearly show that the MPCA approach performs substantially better than the conventional PCA and LDA methods, even under the condition of limited variation in the facial poses.

5. Conclusion

In this paper, a MPCA algorithm is proposed for the face recognition. In the proposed algorithm, we reduce the influence of the changes in the illumination by reducing the influence of the eigenvectors associated with large eigenvalues. In our study, we have compared the MPCA with two conventional algorithms, namely, the PCA and LDA. The results obtained show that the MPCA algorithm leads to a better performance compared to the conventional PCA algorithm, without increasing the computational complexity. It also outperforms the LDA, with a much less computational complexity than the LDA.

Acknowledgement

We would like to thank Yale University for making Yale face database and Yale face database B available free in the public domain for purposes of research.

References

[1] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces", *IEEE Trans. on PAMI*, 12 (1), pp. 103-108, 1990

[2] M. Turk and A. Pentland, " Eigenfaces for Recognition", *Journal Cognitive Nero-science*, 3(1), 1991

[3] D.L. Swets and J.J. Weng, "Using Discriminant Eigenfeature for Image Retrieval", *IEEE Trans. PAMI*, 18(8), pp. 831-836, 1996

[4] P. N. Belhumeour, J. P. Hespanha, and D. J. Kriegman, "Eigenface vs. Fisherfaces: Recognition Using Class Specific Linear Projection", *IEEE Trans. on PAMI*, 19(7), pp. 711-720, 1997

[5] Aleix M. Martinez and Avinash C. Kak, "PCA versus LDA", *IEEE Trans. on PAMI*, 23(2), pp. 228-233, 2001

[6] W. Zhao, R. Chellappa, and A. Krishnaswamy, "Discriminant Analysis of Principle Components for Face Recognition", In *IEEE Int. Conf. on Automatic Face and Gesture Recognition*. Nara, Japan, pp. 336-341, 1998

[7] A. Pentland, T. Starner, N. Etcoff, O.Oliyide, and M. Turk, "Experiments with Eigenfaces", Looking at people, *Workshop of IJCAI*, 1993

[8] H. Moon and P.J. Phillips, "Analysis of PCA-Based Face Recognition Algorithms. Empirical Evaluation Techniques in Computer Vision (K.J. Bowyer and P.J.Phillips, Eds.)", *IEEE Computer Science*, 1998

[9] Yale Face Database and Yale Face Database B: `http://cvc.yale.edu/projects`

[10] P. A. Devijver and J. Kittler, "*Pattern Recognition: A statistical Approach*", Prentice-Hall, Englewood Cliffs, N.J., 1982