Journal of Southeast University (English Edition) Vol. 22, No. 1, pp. 26 – 30 Mar. 2006 ISSN 1003—7985

Modified algorithm of principal component analysis for face recognition

Luo Lin[1] Zou Cairong[1] Yang Fengfan[2]

([1]Department of Radio Engineering, Southeast University, Nanjing 210096, China)

([2]Department of Electronics Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China)

Abstract: In principal component analysis (PCA) algorithms for face recognition, to reduce the influence of the eigenvectors which relate to the changes of the illumination on abstract features, a modified PCA (MPCA) algorithm is proposed. The 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 the method. The simulation results show that, for front face and even under the condition of limited variation in the facial poses, the proposed method results in better performance than the conventional PCA and linear discriminant analysis (LDA) approaches, and the computational cost remains the same as that of the PCA, and much less than that of the LDA.

Key words: face recognition; principal component analysis; linear discriminant analysis

Statistical techniques have been widely used for face recognition and facial analysis to extract the abstract features of the face patterns. Principal component analysis (PCA)[1] and linear discriminant analysis (LDA)[2,3] fall into this category. Compared to the PCA method, the computational cost of the LDA is much higher[2,4].

The simulations in Ref. [2] show an improved performance using the LDA method when compared to the PCA approach. However, the work contained in Ref. [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. In this paper, we only consider the case of a training set with a large number of samples. Pentland et al.[6] 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[7] 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 does not increase the cost of computation.

Received 2005-07-26.

Biography: Luo Lin (1969—), female, doctor, associate professor, luolin@seu.edu.cn.

1 Reviews of the PCA Algorithm

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

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 $\boldsymbol{W}_{\mathrm{pca}}$ of the matrix $\boldsymbol{S}_{\mathrm{T}}$ associated with the nonzero eigenvalues $\boldsymbol{\Lambda}$ by solving the problem of the eigenstructure decomposition, $\boldsymbol{S}_{\mathrm{T}}\boldsymbol{W}_{\mathrm{pca}} = \boldsymbol{W}_{\mathrm{pca}}\boldsymbol{\Lambda}$, where the eigenvectors form the feature subspace.

However, a direct computation of $\boldsymbol{S}_{\mathrm{T}}$ is impractical because of the large size ($D\times D$) of the images. In the original image space, the number of training samples is usually less than the dimension of the image ($M<D$), and only $M-1$ meaningful eigenvectors are useful. Thus we can construct the matrix $\boldsymbol{R} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} \in \mathbf{R}^{M\times M}$, and

obtain the eigenvectors $\boldsymbol{V}_{\mathrm{pca}} \in \mathbf{R}^{M \times r}$ by solving the eigenstructure decomposition, $\boldsymbol{R}\boldsymbol{V}_{\mathrm{pca}} = \boldsymbol{V}_{\mathrm{pca}}\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_0, \lambda_1, \dots, \lambda_{r-1}\} \in \mathbf{R}^{r \times r}(\lambda_0 \geqslant \lambda_1 \geqslant \dots \geqslant \lambda_{r-1})$, $\lambda_i(i=0,1,\dots,r-1)$ are the nonzero eigenvalues of $\boldsymbol{R}$. Then the PCA subspace $\boldsymbol{W}_{\mathrm{pca}}$ is formed by multiplying the matrix $\boldsymbol{A}$ with the eigenvectors $\boldsymbol{V}_{\mathrm{pca}}$, i. e., $\boldsymbol{W}_{\mathrm{pca}} = \boldsymbol{A}\boldsymbol{V}_{\mathrm{pca}}\boldsymbol{\Lambda}^{-\frac{1}{2}}$, where $\boldsymbol{W}_{\mathrm{pca}} = [\boldsymbol{w}_0, \boldsymbol{w}_1, \dots, \boldsymbol{w}_{r-1}] \in \mathbf{R}^{D \times r}$, and its column vectors $\boldsymbol{w}_i(i=0,1,\dots,r-1)$ associated with the eigenvalue λ_i are orthonormal. Therefore, the feature vector $\boldsymbol{y}$ of an image $\boldsymbol{x}$ is acquired by projecting $\boldsymbol{x}$ into the coordinate system defined by the PCA subspace, where $\boldsymbol{y} = \boldsymbol{W}_{\mathrm{pca}}^{\mathrm{T}}(\boldsymbol{x} - \boldsymbol{\mu})$.

2 MPCA Algorithm

In this section, we will analyze 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 $\bar{\boldsymbol{x}}_i = \boldsymbol{x}_i - \boldsymbol{\mu}$ into the PCA eigenface subspace and obtain the feature vector $\boldsymbol{y}_i$ of the training image as

$$\boldsymbol{y}_i = \boldsymbol{W}_{\mathrm{pca}}^{\mathrm{T}}\bar{\boldsymbol{x}}_i = \{y_{i0}, y_{i1}, \dots, y_{i(r-1)}\}^{\mathrm{T}} \tag{1}$$

Thus, $\bar{\boldsymbol{x}}_i$ can be expressed as a linear combination of the basis vectors of the PCA subspace:

$$\bar{\boldsymbol{x}}_i = \sum_{j=0}^{r-1} y_{ij}\boldsymbol{w}_j \qquad i = 1,2,\dots,M \tag{2}$$

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

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

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

$$E[y_{i0}y_{i0}] \geqslant E[y_{i1}y_{i1}] \geqslant \dots \geqslant E[y_{i(r-1)}y_{i(r-1)}] \tag{4}$$

From Eq. (4), 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 Eq. (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 $\bar{\boldsymbol{x}}_i$, the eigenvectors related to the large eigenvalues have more impact on the feature vector elements. However, the eigenvectors associated with the largest eigenvalues are empirically regarded as representing 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 j-th element y_{ij} of the i-th feature vector $\boldsymbol{y}_i$ with respect to its standard deviation, $\sqrt{\lambda_j}$. Hence, the new feature vector $\boldsymbol{y}_i'$ is rewritten as

$$\boldsymbol{y}_i' = \left\{\frac{y_{i0}}{\sqrt{\lambda_0}}, \frac{y_{i1}}{\sqrt{\lambda_1}}, \dots, \frac{y_{i(r-1)}}{\sqrt{\lambda_{r-1}}}\right\}^{\mathrm{T}} \tag{5}$$

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

The steps of the MPCA algorithm which is based on the conventional PCA are as follows:

① Get training set data matrix $\boldsymbol{X}_{\mathrm{Tr}}$ and test set data matrix $\boldsymbol{X}_{\mathrm{Te}}$ from the face database;

② Calculate global mean $\boldsymbol{\mu}$, centered data $\boldsymbol{A} = \boldsymbol{X}_{\mathrm{Tr}} - \boldsymbol{\mu}$, covariance matrix $\boldsymbol{R} = \boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$, the Jacobi decomposition of $\boldsymbol{R}$, and the eigenface space

$$\boldsymbol{W}_{\mathrm{pca}} = \boldsymbol{A}\boldsymbol{V}\boldsymbol{\Lambda}^{-\frac{1}{2}} \tag{6}$$

③ Calculate the projection of training set $\boldsymbol{P}_{\mathrm{Tr}} = \boldsymbol{W}_{\mathrm{pca}}^{\mathrm{T}}(\boldsymbol{X}_{\mathrm{Tr}} - \boldsymbol{\mu})$ and the normalizing feature vector

$$\boldsymbol{P}_{\mathrm{Tr}}' = \boldsymbol{P}_{\mathrm{Tr}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-\frac{1}{2}} \tag{7}$$

④ Calculate the projection of testing set $\boldsymbol{P}_{\mathrm{Te}} = \boldsymbol{W}_{\mathrm{pca}}^{\mathrm{T}}(\boldsymbol{X}_{\mathrm{Te}} - \boldsymbol{\mu})$ and the normalizing feature vector

$$\boldsymbol{P}_{\mathrm{Te}}' = \boldsymbol{P}_{\mathrm{Te}}^{\mathrm{T}}\boldsymbol{\Lambda}^{-\frac{1}{2}} \tag{8}$$

⑤ Calculate the Euclidean distance classifier $d_{ij} = \sqrt{(p'^{i}_{\mathrm{Te}} - p'^{j}_{\mathrm{Tr}})(p'^{i}_{\mathrm{Te}} - p'^{j}_{\mathrm{Tr}})}$, where i is the number of testing images, j is the number of training classes;

⑥ Calculate the receiver operation characteristics (ROC).

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; these are shown as Eq. (7) and Eq. (8). We then calculate the distance between the training and the testing images. In the actual programme, the normalization Eqs. (6) to (8) are combined into one step, $\boldsymbol{W}_{\mathrm{mpca}} = \boldsymbol{A}\boldsymbol{V}_{\mathrm{pca}}\boldsymbol{\Lambda}^{-1}$. The remaining steps are the same as in the case of the conventional PCA. 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 are analyzed.

3 Simulation Results

Two face databases are used in our simulations, namely, the Yale face database and Yale face database B. The 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 size of each image is 243 pixels (width) × 320 pixels (height), and the sample images are shown in Fig. 1.

Fig. 1 Sample images of the Yale face database

The Yale face database B contains 5 760 single light source images of 10 subjects each seen under 576 viewing conditions (9 poses × 64 illumination conditions). For every subject in a particular pose, an image with ambient (background) illumination is also captured. Hence, the total number of images is actually 5 760 + 90 = 5 850. The acquired images are 8 bit (grey scale) PGM raw format. The size of each image is 640 pixel(width) × 480 pixel (height), and Fig. 2 shows the sample images.

Fig. 2 Sample images of the Yale face database B

There are various facial expressions and illumination conditions in the Yale face database. For our simulations, the protocol combined with "leave-one-out" (LOO) strategy and rotation is used. The original images are registered and cropped into size of 50 pixel (weight) ×60 pixel (height) without photometric normalization. The minimum distance classifier based on the Euclidean distance is applied as the matching scheme in this study.

The simulation results are presented in terms of the receiver operating characteristics to show the relationship between the false rejection (FR) and false acceptance (FA) as a function of the decision threshold. Fig. 3 depicts the ROC curve of the face recognition experiments on the Yale face database. Tab. 1 gives a summary of the results for the equal error rate (EER) point on the ROC curves, where the EER point is the decision boundary with the trade-off FA equalling FR. It can be easily seen from Fig. 3 and Tab. 1 that the MPCA algorithm achieves the best performance among the three algorithms, namely, the PCA, LDA and MPCA.

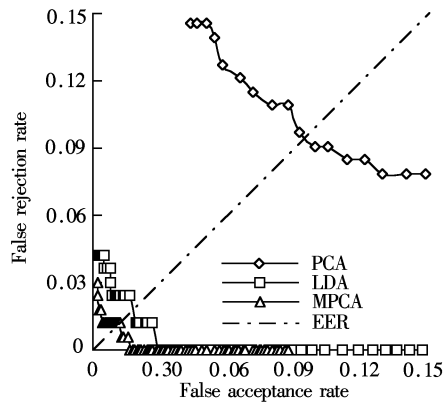

Fig. 3 ROC curves on the Yale face database

Tab. 1 EER of face recognition experiments on the Yale face database

Algorithms	PCA	LDA	MPCA
EER/%	9.52	1.82	1.21

In the Yale face database B, the facial poses and illumination conditions are changed. All the images are registered and cropped into a size of 50 pixel (width) × 60 pixel (height). 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°) 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° from the camera optical axis (i. e., relative to pose 0), while poses 6, 7, and 8 are approximately 24° from the camera optical axis. We choose poses 0, 1, 2, 3, 4, 5 as the training and testing sets in our simulations.

Fig. 4 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 corresponding to Figs. 4 (a) to (e), respectively.

The EER for all the simulations on the Yale face database B are given in Tab. 2. The first row contains the results on the pose 0 set, where the "leave-one-out" procedure is adopted. The remaining rows correspond to the results with respect to poses 1, 2, 3, 4, 5, respectively. Fig. 4 and Tab. 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.

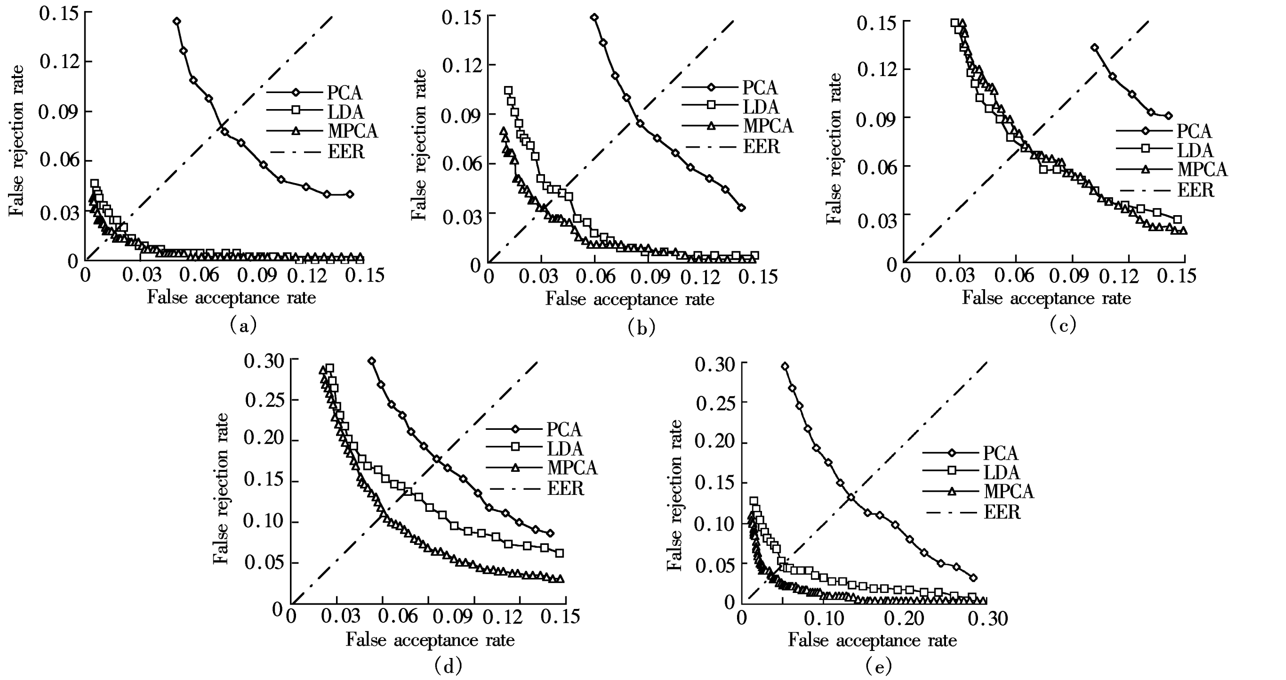

Fig. 4 ROC curves on the Yale face database B for different poses. (a) Pose 1; (b) Pose 2; (c) Pose 3; (d) Pose 4; (e) Pose 5

Tab. 2 EER of face recognition experiments on the Yale face database B %

Pose	PCA	LDA	MPCA
0	8. 98	0. 00	0. 00
1	7. 70	2. 00	1. 58
2	8. 51	4. 20	3. 21
3	11. 34	6. 84	6. 84
4	16. 88	13. 42	10. 48
5	13. 39	5. 01	3. 59

Since in the actual program, we combine Eqs. (6) to (8) into one equation, it is very clear that the computational cost of the method proposed is the same as that of the classical PCA algorithm, and is much lower than that of the LDA algorithm.

4 Conclusion

In this paper, an 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 cost. It also outperforms the LDA, with a much lower computational cost than the LDA.

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

References

[1] Kirby M, Sirovich L. Application of the Karhunen-Loeve procedure for the characterization of human faces [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1990, **12**(1): 103 − 108.

[2] Swets D L, Weng J J. Using discriminant eigenfeature for image retrieval [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1996, **18**(8): 831 − 836.

[3] Belhumeour P N, Hespanha J P, Kriegman D J. Eigenface vs. fisherfaces: recognition using class specific linear projection [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 1997, **19**(7): 711 − 720.

[4] Zhao W, Chellappa R, Krishnaswamy A. Discriminant analysis of principal components for face recognition [C]//*Proc of IEEE International Conference on Automatic Face and Gesture Recognition*. Nara, Japan, 1998: 336 − 341.

[5] Martinez A M, Kak A C. PCA versus LDA [J]. *IEEE Trans on Pattern Analysis and Machine Intelligence*, 2001, **23**(2): 228 − 233.

[6] Pentland A, Starner T, Etcoff N, et al. Experiments with eigenfaces [C]//*Proc of Looking at People Workshop, International Joint Conference on Artificial Intelligence*. Chamberry, France, 1993: 1 − 6.

[7] Moon H, Phillips P J. Analysis of PCA-based face recognition algorithms[C]//Bowyer K J, Phillips P J. *Empirical Evaluation Techniques in Computer Vision*. California: IEEE Computer Science Press, 1998: 57 − 71.

改进的人脸识别主分量分析算法

罗　琳[1]　邹采荣[1]　仰枫帆[2]

([1] 东南大学无线电工程系,南京 210096)
([2] 南京航空航天大学电子工程系,南京 210016)

摘要:在应用于人脸识别领域的主分量分析(PCA)算法中,为了降低与外界光照变化相关的特征向量对提取特征的影响,提出了一种改进的主分量分析(MPCA)算法,利用相对应的标准方差对提取的特征矢量元素进行归一化处理. 采用耶鲁大学的 2 个人脸数据库(Yale face database 和 Yale face database B)进行了验证,实验结果表明,对于正面人脸和具有小角度姿态变化情况下的人脸,提出方法的性能优于传统的 PCA 和 LDA(线性判别分析)算法,而运算量和 PCA 算法相同,大大低于 LDA 算法.

关键词:人脸识别;主分量分析;线性判别分析
中图分类号:TN391