Modern Applied Science
www.ccsenet.org/journal.html
Vol. 3, No. 11
November 2009

Modular PCA Face Recognition Based on Weighted Average

Chengmao Han (Corresponding author)
Department of Mathematics, Linyi Normal University
Linyi 276005, China
E-mail: hanchengmao@163.com

Abstract

This paper presents an improved modular PCA approach, that is, modular PCA algorithm based on weighted average. This algorithm extracts weighted average for every sub-block of every training sample in each type of training sample, and normally operates the corresponding sub-block in training sample using weighted average, then all standardized sub-blocks constitute the overall scatter matrix, and thus the optimal projective matrix is obtained; From the middle value of sub-blocks in training set, and normally projecting sub-blocks of training samples and test samples to the projective matrix, then we can get identified characteristics; At last, use the recent distance classifier to class. The test results in the ORL face database show that the proposed method in identifying performance is superior to ordinary modular PCA approach.

Keywords: Face recognition, Principal component analysis, Weighted average

Face recognition is an active subject in the current pattern recognition field, which has broad application prospects[Valentin D, 1994-Zhang Cuiping, 1995] and has a lot of algorithms[J.Lu, 2003-Gan Junying, 2007]. In the human face image recognition, principal component analysis (PCA)[Kriby M., 1990] also known as KL transformation, is considered one of the most successful linear discriminant analysis methods, which is still widely used in image recognition field such as human face, etc. PCA method can not only effectively reduce the dimension of human face images, but also retain its key identifying information. However, this method requires the matrix of human face image pre-converted into one-dimensional vector, and then takes vector as the original characteristic to feature extraction. For the dimension of the converted one-dimensional vector is generally higher, extracting the subsequent feature causes difficulty, which makes the following algorithm has higher computational complexity. In addition, in face recognition when the facial expression and illumination conditions change largely, for the ordinary PCA method extracts the global features of image, its identification results are unsatisfactory. In fact, when facial expression and illumination conditions change, only some face significants variate, but little changes in other parts, or even no change. Discriminant analysis for the blocked sub-images can capture local information characteristics of human face, thus help identify.

Chen Fubing[Chen Fubing, 2007] proposed blocked PCA algorithm based on traditional PCA method. This method, first blocks an image, then discriminant analyses the blocked sub-images using PCA. Its characteristics are able to effectively extract the local characteristics of images, especially for the images whose facial expressions and illumination conditions change largely. Compared with the PCA method, blocking the original digital image, can not only easily reduce the dimension of image vector by two powers, but also increase the sub-images number of training samples by two parts, which converts the small sample problem into large sample problem to deal with and can reduce complexity of the problem

average face method proposed by He Guohu[He Guohu,2006] effectively increases the distance between the samples of different categories, while narrowing the distance between the samples, that is, making the distance between classes larger and distance smaller, which is conducive to identification, and improves the correct recognition rate of human face.

However, in small sample cases, average cannot guarantee that average of various types of training samples is the center of this sample distribution. And the projective matrix taken from the average of training sample as the center of this class samples can not guarantee to be optimal.

In order to further improve the recognition performance of PCA method and reduce the influence of taking the optimal projective matrix by average derivation center of training samples. This paper presents an improved modular PCA approach based on the above method and by the adaptive weighted average idea in paper [Yin Hongtao,2006], that is modular PCA approach based on weighted average. This algorithm extracts weighted average for every sub-block of every training sample in each type of training sample, and normally operates the corresponding sub-block in training sample using weighted average. The test results in the ORL face database show that the proposed method in identifying performance is superior to ordinary modular PCA approach.

1. PCA Algorithm

PCA method is a statistical analysis method based on Karhunen-Loeve (KL) transformation, whose principle is that high-dimensional vector projects to low-dimensional vector space by a specific feature vector matrix. Through the vector of low-dimensional representation and the feature vector matrix, we can reconstruct the corresponding original high-dimensional vector. In the face recognition process, after the KL transformation, we can get a set of feature vectors to form a lower dimensional subspace. Any human face image can project it and get a set of coordinates factors. This group of coefficients shows that the image location in the sub-space can be used as a basis for face recognition.

In this method, generating matrix is the total scatter matrix of training samples, i.e:

$$\boldsymbol{S}=\frac{1}{\boldsymbol{M}}\sum_{\boldsymbol{i}=1}^{\boldsymbol{M}}\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i-\overline{\boldsymbol{X}}\right)^{\boldsymbol{T}}=\frac{1}{\boldsymbol{M}}\boldsymbol{X}\boldsymbol{X}^{\boldsymbol{T}} \quad , \tag{1}$$

Where $\boldsymbol{X}_i$ is the image vector of the $\boldsymbol{i}$ -th training sample; Vector dimension is $\boldsymbol{n}$; $\overline{\boldsymbol{X}}$ is the average figure vector for training sample; $\boldsymbol{M}$ is the total number of training samples.

According to the general scatter matrix, we can derive a set of orthogonal eigenvectors $\boldsymbol{u}_1,\boldsymbol{u}_2,\cdots,\boldsymbol{u}_n$, and its corresponding characteristic values are $\lambda_1,\lambda_2,\cdots,\lambda_n$. Through choosing the corresponding eigenvectors of the previous $\boldsymbol{m}\,(\boldsymbol{m}<\boldsymbol{n})$ non-zero eigenvalues as the orthogonal basis, in the new orthogonal sub-space $\boldsymbol{U}$, the face sample $\boldsymbol{X}_i$ can be expressed as:

$$\boldsymbol{Y}_i=\boldsymbol{U}^{\boldsymbol{T}}\boldsymbol{X}_i \tag{2}$$

2. Adaptive weighted average [Yin Hongtao, 2006]

When using PCA algorithm, we first spread the image matrix by row (column) as a vector. Suppose the vector spread by all the image matrix be:

$$\boldsymbol{X}_j^{(i)}=\left(\boldsymbol{X}_{(j,1)}^{(i)},\boldsymbol{X}_{(j,2)}^{(i)},\cdots\boldsymbol{X}_{(j,m)}^{(i)}\right)^{\boldsymbol{T}} \quad , \tag{3}$$

Where $\boldsymbol{i}=1,2,\cdots,\boldsymbol{c}$, $\boldsymbol{c}$ is the type number of training samples; $\boldsymbol{j}=1,2,\cdots\boldsymbol{N}_i$, is the number of the $\boldsymbol{i}$ -th training sample; m is the vector dimension. Because the mean vector of several vectors is taken averagely from the scalar vector of the corresponding dimension, we explain the determined method of weighted value by taking the first dimension as example.

First of all, calculate the distance sum $\boldsymbol{d}_{j,1}^{(i)}(\boldsymbol{j}=1,2,\cdots,\boldsymbol{N}_i)\boldsymbol{d}_{j,1}^{(i)}(\boldsymbol{j}=1,2,\cdots,\boldsymbol{N}_i)$ of every sample in the $\boldsymbol{i}$ -th sample and other sample, then find the minimum of them $\boldsymbol{d}_1^{(i)}=\arg\min_{j}\left(\boldsymbol{d}_{(j,1)}^{(i)}\right)\boldsymbol{d}_1^{(i)}$.

We believe that the sample whose distance sum with the same type sample is larger. It deviates greater from the class center. In calculating class average, it should be given a smaller weight and the weighted value in the first dimension of the $\boldsymbol{j}$ -th sample in the $\boldsymbol{i}$ -th class is

$$\mu_{(j,1)}^{(i)}=1+\beta\frac{\boldsymbol{d}_1^{(i)}}{\boldsymbol{d}_{(j,1)}^{(i)}} \quad , \tag{4}$$

where β is a constant greater than or equal to zero. For regulating the weight extent, when $\beta=0$, the algorithm becomes the traditional averaging method. The average of the first dimension in the $\boldsymbol{i}$ -th sample can be modified to

$$\tilde{\boldsymbol{X}}_1^{(i)}=\frac{\sum_{\boldsymbol{j}=1}^{\boldsymbol{N}_i}\mu_{(j,1)}^{(i)}\boldsymbol{X}_{(j,1)}^{(i)}}{\sum_{\boldsymbol{j}=1}^{\boldsymbol{N}_i}\mu_{(j,1)}^{(i)}} \tag{5}$$

Similarly, we can find the means of other dimensions of training samples.

3. Modular PCA algorithm based on the weighted average

The basic idea of modular PCA algorithm based on the weighted average is as follows: block $m \times n$ image matrix I into $p \times q$ blocked image matrix, namely,

$$I = \begin{pmatrix} I_{11} & I_{12} & \cdots & I_{1q} \\ I_{21} & I_{22} & \cdots & I_{2q} \\ \vdots & \vdots & \ddots & \vdots \\ I_{p1} & I_{p2} & \cdots & I_{pq} \end{pmatrix} \tag{6}$$

Where each sub-image matrix I_{kl} is a $m_1 \times n_1$ matrix, $pm_1 = m$, $qn_1 = n$, then taking the sub-image matrix of all training samples as the image vector of training sample to purpose PCA method. The difference from the traditional PCA algorithm is that we derive all scatter matrix not using the sub-block average of all training samples, but using weighted average of sub-blocks. This can reduce the impact of deriving the optimal projective matrix from the mean deviation in training samples, thus improving the recognition rate.

Algorithm steps are as follows:

For convenience, we first introduce the concept of quantization matrix.

Definition: Suppose $A = (A_1, A_2, \cdots, A_n) \in R^{m \times n}$, $mn \times 1$ vector is defined as

$$Vec(A) = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_n \end{pmatrix} , \tag{7}$$

Where the vector is arranged in turn by column vector of the matrix A, which is called the quantification of matrix A.

1) Suppose the model category is C, the image matrix in the i-th class training sample is $n(i)$; A_{i1}, A_{in_i}, $N = \sum_{i=1}^{C} n(i) N$ is the total number of training samples, and each sample image is $m \times n$ matrix. The $p \times q$ blocked matrix of training sample image A_{ij} is expressed as:

$$A_{ij} = \begin{pmatrix} (A_{ij})_{11} & (A_{ij})_{12} & \cdots & (A_{ij})_{1q} \\ (A_{ij})_{21} & (A_{ij})_{22} & \cdots & (A_{ij})_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ (A_{ij})_{p1} & (A_{ij})_{p2} & \cdots & (A_{ij})_{pq} \end{pmatrix} \tag{8}$$

2) Require overall scatter matrix of sub-image matrix in all the training sample images.

Let $(\eta_{ij})_{kl} = Vec\left((A_{ij})_{kl}\right)$, $k = 1, 2, \cdots, p$, $l = 1, 2, \cdots, q$, then $(\eta_{ij})_{kl} \in R^{m_1 \times n_1}$. So the overall scatter matrix of all training sample sub-blocks is:

$$S = \frac{1}{M} \sum_{i=1}^{C} \sum_{j=1}^{n(i)} \sum_{k=1}^{p} \sum_{l=1}^{q} \left((\eta_{ij})_{kl} - (\eta_{kl})_i\right)\left((\eta_{ij})_{kl} - (\eta_{kl})_i\right)^T . \tag{9}$$

Where $i = 1, 2, \cdots, C$; $j = 1, 2, \cdots n(i)$; $k = 1, 2, \cdots p$; $l = 1, 2, \cdots, q$, $n(i)$ is the number of each class of training samples;

$$M = \left(\sum_{i=1}^{C} n(i)\right) pq = Npq \tag{10}$$

is the total number of training sample sub-images matrix. .

$(\eta_{kl})_i$ is the weighted average image between the i-th sample image and the kl-th block. The specific calculation method is to spread all sub-blocks by row to column vector, then calculates its weighted mean vector by equations (4) and (5), and reverts the middle measures to a matrix.

Easily, we can prove S is a $m_1 n_1 \times m_1 n_1$ non-negative definite matrix.

3) Seek optimal projective matrix

Take corresponding orthonormal eigenvectors (discriminant vectors) $Z_1, Z_2, \cdots, Z_r$ of the r largest eigenvalue of S to constitute

$$Q = [Z_1, Z_2, \cdots, Z_r],$$

4) Require weighted average vector of all training sample sub-blocks matrix

In order that test samples and training samples are comparable, standardize them by the same weighted average matrix. So we must calculate weighted average matrix η of all training sub-block samples.

5) Feature extraction of training samples.

Each block of training samples $A_{ij} = \begin{pmatrix} (A_{ij})_{11} & (A_{ij})_{12} & \cdots & (A_{ij})_{1q} \\ (A_{ij})_{21} & (A_{ij})_{22} & \cdots & (A_{ij})_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ (A_{ij})_{p1} & (A_{ij})_{p2} & \cdots & (A_{ij})_{pq} \end{pmatrix}$ is quantified by equation (7) and normalized, then obtain the characteristics matrix of A_{ij} after projecting to $Q = [Z_1, Z_2, \cdots, Z_r]$:

$$B_{ij} = \begin{pmatrix} Q^T\left((\eta_{ij})_{11} - \eta\right) & Q^T\left((\eta_{ij})_{12} - \eta\right) & \cdots & Q^T\left((\eta_{ij})_{1q} - \eta\right) \\ Q^T\left((\eta_{ij})_{21} - \eta\right) & Q^T\left((\eta_{ij})_{22} - \eta\right) & \cdots & Q^T\left((\eta_{ij})_{2q} - \eta\right) \\ \vdots & \vdots & \vdots & \vdots \\ Q^T\left((\eta_{ij})_{p1} - \eta\right) & Q^T\left((\eta_{ij})_{p2} - \eta\right) & \cdots & Q^T\left((\eta_{ij})_{pq} - \eta\right) \end{pmatrix}_{pr \times q} . \tag{11}$$

6) Feature extraction of test samples

Each block of test sample image $I_x = \begin{pmatrix} I_{11} & I_{12} & \cdots & I_{1q} \\ I_{21} & I_{22} & \cdots & I_{2q} \\ \vdots & \vdots & \vdots & \vdots \\ I_{p1} & I_{p2} & \cdots & I_{pq} \end{pmatrix}$ is quantified by equation (7) and normalized, then obtain the characteristics matrix of test samples after projecting to $Q = [Z_1, Z_2, \cdots, Z_r]$:

$$B_x = \begin{pmatrix} Q^T(\eta_{11} - \eta) & Q^T(\eta_{12} - \eta) & \cdots & Q^T(\eta_{1q} - \eta) \\ Q^T(\eta_{21} - \eta) & Q^T(\eta_{22} - \eta) & \cdots & Q^T(\eta_{2q} - \eta) \\ \vdots & \vdots & \vdots & \vdots \\ Q^T(\eta_{p1} - \eta) & Q^T(\eta_{p2} - \eta) & \cdots & Q^T(\eta_{pq} - \eta) \end{pmatrix}_{pr \times q} , \tag{12}$$

where $\eta_{kl} = Vec(I_{kl})$, $l = 1, 2, \cdots, q$.

7) Sort

Suppose $B_{ij} = \left[Y_1^{(ij)}, Y_2^{(ij)}, \cdots, Y_q^{(ij)}\right]$, $B_x = \left[Y_1^{(x)}, Y_2^{(x)}, \cdots, Y_q^{(x)}\right]$, carrying the most recent method to sort:

$$d(B_{ij}, B_x) = \sum_{m=1}^{q} \left\| Y_m^{(ij)} - Y_m^{(x)} \right\|_2 \tag{13}$$

$i = 1, 2, \cdots, C$; $j = 1, 2, \cdots n(i)$; x is identified the x -th sample under test.

If $d(B_{nj}, B_x) = \min_i d(B_{ij}, B_x)$, the sample I_x belongs to the i -th category.

4. Experiment and result analysis

Test the method of this paper in ORL (olivetti research laboratory) face database. This face database contains 40 individuals, and each person has 10 images. The image is a positive image of single dark background that contains a certain amount of illumination changes, facial changes (open eyes and closed eyes , laughing or not laughing), facial details changes(wearing glasses or not wearing glasses), and the depth rotation within a certain range. The sizes of these images are 112 × 92 pixels. Other part faces are showed in Figure 1. For each person, randomly selecte five images as training samples and the rest five images are used to test the identification method performance.

The experimental results are shown in Figure 2 and Figure 3. Figure 2 shows experimental result of traditional PCA method, 2×2 modular PCA method and 2 × 2 blocked PCA method based on weighted average. From the figure, we can see that recognition rate of traditional PCA method is lower, that is up to 77 %. Module PCA method improves the recognition rate, while the blocked PCA method based on weighted average is superior to ordinary blocked PCA method. Figure 3 respectively shows the test results of 4 × 2 sub-blocks and 4 × 4 sub-blocks conditions. From the figures we can see that in the 4 × 2 block case, modular PCA method based on weighted average has a higher recognition rate and a more robust than ordinary blocked PCA method; In addition, test result also shows that 4 × 2 blocked approach is superior to 2 × 2 blocked approach. In the blocked mode, the correct recognition rate is greatly decreased. The cause is that the more blocks number of each image is, the more reduced the distinguished information contained in each sub-block. So there will be more similar sub-blocks and it is not conducive to classification, thus correct identification rate has dropped. In this case, modular PCA method based on weighted average is still better than ordinary blocked PCA method. At the same time, we find in experiments the recognition performance of 4×2 sub-blocks is far better than that of 2 × 4 sub-blocks, which is shown in Table 1. The cause is that the difference between different people faces focus on eyes, nose, mouth, chin and other parts, so the vertical multi-block is not conducive to identification.

5. Conclusion

The prominent advantage of face recognition method based on modular PCA is the ability to extract the local features of image, which better reflects the difference between images. We can easily use discriminant analysis method in the smaller image for the process is simple. To further improve the recognition rate, this paper improves face recognition method based on modular PCA and proposes modular PCA algorithm based on weighted average. The experiment on ORL face database shows that this method is superior to the traditional PCA method and ordinary PCA method. For the same database, if the original image has different sub-blocks, the obtained highest recognition rate is generally different. How to find the best sub-blocks acquired highest recognition rate and how to simplify the sub-blocks PCA algorithm have yet to be further studied.

References

Chen Fubing, Yang Jingyu. (2007). Modular PCA and its application in human face recognition. *Computer Enineering and Design*,28(8):1889-1913.

Gan Junying, Li Chunzhi. (2007). 2DICA based on wavelet transformation and application. *Journal of System Simulatio*n,19(3):612-619.

He Guohui, Gan Junying. (2006). Study for class average face method based on PCA in face recognition. *Application Reseach of Computer*, 3:165-169.

J. Lu, K. Plataniotis, A.Venetsanopoulos.(2003). Face Recognition using LDA-based Algorithms[J]. *IEEE Trans*. Neural Networks, 14 (1):195-200.

Jian Yang, David Zhang. (2004). Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition [J]. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(1):131-137.

Keun-Chang Kwak, Witold Pedrycz. (2007). Face Recognition using an Enhanced Independent Component Analysis Approach[J]. *IEEE Trans.* Neural Networks, 18(2):530-541.

Kirby M, Sirovich L.(1990). Application of the KL Procedure for the Characterization of Human Faces[J].*IEEE Trans.* Pattern Analysis and Machine Intelligence, 12(1):103-108.

Rama Chellappa, et al.(1995). *Human and Machine Recognition of Faces: A Survey[J]*. Proceedings of the IEEE, 83(5): 705-740.

Valentin D, Abdi H, OToole A J. (1994). *Connectionist Model of Face Processing: A Survey [J]*. Pattern Recognition, 27(9): 1209-1230.

Yin Hongtao, Fu Ping, Meng Shengwei. (2006). Face recognition based on adaptively weghted Fisherface. *Journal of Optoelectronics*. *Laser*,17(11):1405-1408.

Zhang Cuiping, Sun Guangda. (2005). A survey on face recognition. *Journal of Image and Graphics,* 5(11):885-894.

Table 1. recognition rate of 2×4 blocks and 4×2 blocks of method in this paper (%)

recognition rate number	5	10	15	20	25	30	35	40
2×4 blocks	84.5	85	85.5	85.5	85	86	85.5	85
4×2 blocks	91.5	92	91.5	92	91.5	91.5	91.5	91.5

Figure 1. image in ORL face database

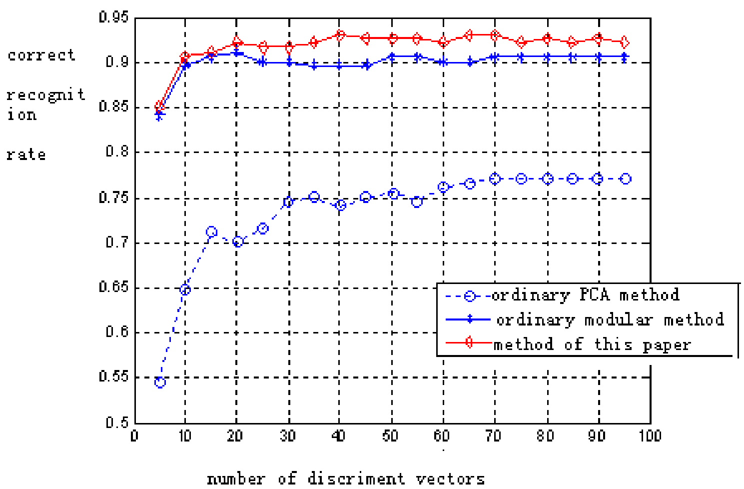

Figure 2. Experimental result of 2 ×2 sub-blocks

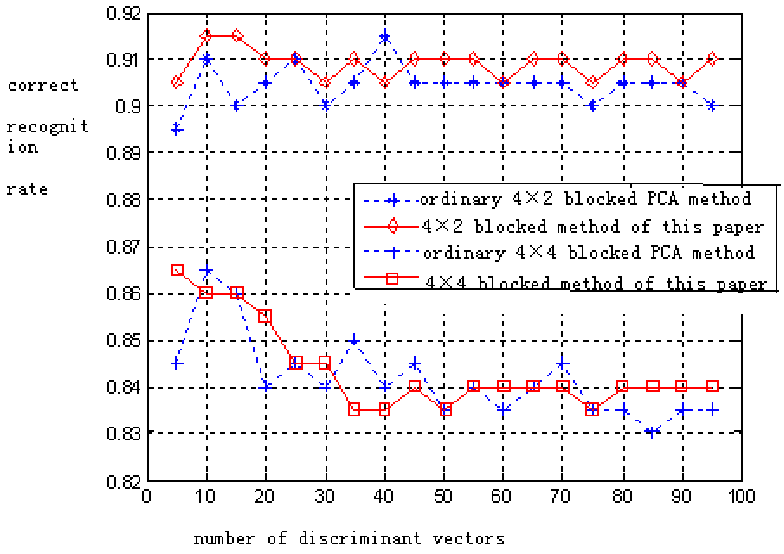

Figure 3. Experiment result of 4×2 blocks and 4×4 blocks