Proceedings of the Fourth International Conference on Machine Learning and Cybernetics, Guangzhou, 18-21 August 2005

WEIGHTED PCA SPACE AND ITS APPLICATION IN FACE RECOGNITION

HUI-YUAN WANG, XIAO-JUAN WU

School of Information Science and Engineering, Shandong University , Jinan 250100, China
E-MAIL: hywang@sdu.edu.cn, xiaojwu@sdu.edu.cn

Abstract:

In this paper, we propose a new PCA based subspace approach for pattern recognition. The conventional PCA feature space is first converted to a WPCA feature space with unit variance by weighting the features and then face recognition is performed in the new space. Detailed theoretical derivation and analysis are presented and simulation results on AR and ORL face databases are given. The simulation results indicate that the proposed approach is superior to conventional PCA approach in recognition accuracy under the same computation complexity.

Keywords:

Principal component analysis (PCA); Weighted principal component analysis (WPCA); Eigenface; Face recognition

1. Introduction

Principle component analysis (PCA), also called Karhunen-Loeve transform, is an important feature extraction method in pattern recognition. It was introduced to the field of face recognition in the early 1990's and became the most popular method since then, i.e., the eigenface method [1,2]. It is this method that brought about a break through to the research in face recognition and make it an extremely active field ever since.

It has been shown from research that under relatively ideal imaging condition, i. e., strictly controlled pose, illumination and expression, the conventional PCA can achieve a recognition ratio of as high as more than 90%. This makes it a benchmark in face recognition. However, when used in dimension reduction, PCA follows the rule of minimizing the mean square error between the original data and the reconstructed data. Therefore, it is more suitable for data representation in the reduced feature space than for classification. For this reason, linear descrimination analysis (LDA)[3,4] method was put forward, which minimizes the classification error. Furthermore, kernelized PCA, kernelized LDA and independent component analysis (ICA) approaches were proposed to make full use of the higher order correlation between the pixels of face images [6,7]. All these approaches perform better than PCA when there are variations in pose, illumination and expression in the face images. But it should be pointed out that this is achieved under the cost increase of algorithm complexity.

According to the statistical characteristics in transformed domain of conventional PCA approach, we propose a weighted PCA (WPCA) approach in this paper. It is shown by theoretical analysis that the proposed approach is more suitable to the requirement of reducing classification errors without increasing computation complexity. Furthermore, experimental results on AR and ORL databases validate the advantages of the proposed approach over the conventional PCA.

2. The conventional PCA approach

PCA was proposed to be used in face image representation first in [1] and was first used in face recognition in [2], where it was called eigenface approach. What follows is a simple introduction to the basic PCA approach.

Given N facial image samples denoted by n-D vectors $\{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N\}$, belonging to C classes $\{\omega_1, \omega_2, \cdots, \omega_c\}$, we are to find a linear transform to map the original n-D image space into an m-D feature space, where it satisfies that m<n and

$$\mathbf{y}_k = \mathbf{W}^t(\mathbf{x}_k - \boldsymbol{\mu}) \qquad k = 1,2,\cdots,N, \tag{1}$$

where $\mathbf{y}_k \in \mathrm{R}^m$ is a centralized feature vector. $\boldsymbol{\mu} \in \mathrm{R}^n$ is the mean vector of all the image samples, and $\mathbf{W} \in \mathrm{R}^{n \times m}$ is a transform matrix with orthonormal column vectors.

If the covariance matrix from the image vector samples is estimated by

$$\mathbf{C}_x = \sum_{k=1}^{N} (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^t \tag{2}$$

then $\mathbf{W}$ is a matrix formed by m eigenvectors

$\{\mathbf{u}_1, \mathbf{u}_2, \cdots, \mathbf{u}_m\}$ corresponding to the m largest eigenvalues $\{\lambda_1, \lambda_2, \cdots, \lambda_m\}$ of $\mathbf{C}_x$. That is $\mathbf{W} = (\mathbf{u}_1 \mathbf{u}_2 \cdots \mathbf{u}_m)$, and

$$\mathbf{C}_x \mathbf{u}_k = \lambda_k \mathbf{u}_k \qquad k = 1,2,\cdots,m. \tag{3}$$

Or in matrix form

$$\mathbf{C}_x \mathbf{W} = \mathbf{W\Lambda}, \tag{4}$$

where $\mathbf{\Lambda} = diag(\lambda_1, \lambda_2, \cdots, \lambda_m)$. It should be noted that the number of non-zero eigenvalues of $\mathbf{C}_x$ is less than or equal to N-1. The covariance matrix of the feature vectors in the PCA space can be estimated by

$$\begin{aligned}\mathbf{C}_y &= \sum_{k=1}^{N} \mathbf{y}_k \mathbf{y}_k{}^t \\ &= \sum_{k=1}^{N} \mathbf{W}^t (\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^t \mathbf{W}. \\ &= \mathbf{W}^t \mathbf{C}_x \mathbf{W} \\ &= \mathbf{\Lambda}\end{aligned} \tag{5}$$

The centralized $\mathbf{x}_k$ can be reconstructed from all the $\mathbf{y}_k$ s as shown by the following equations.

$$\hat{\mathbf{x}}_k - \boldsymbol{\mu} = \mathbf{W}\mathbf{y}_k = \sum_{i=1}^{m} \mathbf{y}_k(i)\mathbf{u}_i \tag{6}$$

Equation (6) indicates that in PCA space, a centralized estimation of the original image vector can be obtained by the sum of the normalized eigenvectors (eigenfaces) weighted by the corresponding components of the feature vectors in PCA space. It is easy to show that the mean square error produced by the estimation is

$$\varepsilon = \left\| \mathbf{x}_k - \hat{\mathbf{x}}_k \right\|^2 = \sum_{i=m+1}^{n} \lambda_i \tag{7}$$

That is why we choose the m eigenvectors corresponding to the largest eigenvalues to form the transform matrix.

3. Weighted PCA (WPCA) space and its properties

It is known from above discussion that PCA, as a feature compression approach, is optimal in terms of MSE in facial image representation or reconstruction. It is seen from equation (5) that in PCA space, the variance of each component of a feature vector equals to the corresponding eigenvalue and an component in PCA space related to a larger eigenvalue is more important to the representation or reconstruction of facial images according to equation (7). Therefore, when the eigenfaces are ordered in the same sequence of related eigenvalues from the largest to the smallest, the importance of each component of a feature vector as the mapping of a facial image in the PCA space will be ordered in the same sequence.

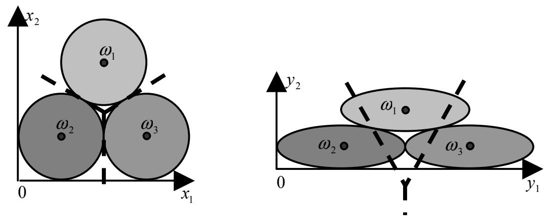
However, in face classification, what are significant are those components that can clearly distinguish faces of different individuals. Unfortunately, there is no clear conclusion so far as to which components are more important to classification in PCA space. As a result, conventional PCA approach equates the importance of each component in classification to that in representation or reconstruction. So that it may not be optimal in theory.

In PCA space, the statistical mean of the squared Euclidian distance between a feature vector and its mean vector can be found by equation (8)

$$\begin{aligned}&E\left\{D_E^2(\mathbf{y}, \mathbf{m}_y)\right\} = E\left\{(\mathbf{y} - \mathbf{m}_y)^t (\mathbf{y} - \mathbf{m}_y)\right\} \\ &= Tr(\mathbf{C}_y) = \sum_i \lambda_i\end{aligned} \tag{8}$$

where $Tr(.)$ denotes the trace of a matrix. It is seen that the contribution of each component of feature vector in PCA space to the squared Euclidian distance is just the corresponding eigenvalue. Therefore, a few components that are related to larger eigenvalues, but are less significant in classification may dictate the similarity metric, i.e., the Euclidian distance and reduce or completely submerge the effect of those with smaller eigenvalues, but more significant in classification.

Based on above discussion, we believe that without enough a priori information about the importance of each component in classification, every independent component



(a) (b)

Figure 1. Effect of weighting feature space on classification

should be dealt with equally, or increased classification error could be resulted in. Obviously, changing the relative importance of each component can be achieved by weighting it. A simple illustration is given in Fig 1, where, (a) shows the distribution of a supposed three class samples. As is seen, there is no classification error when a minimum distance classifier is used. (b) is the result of weighting (a) with a larger weight to component x_1 and a smaller one to x_2. Now, classification error appears if a minimum distance classifier is still used.

3.1. Unit variance PCA space

From equation (5), it is seen that the component variances of feature vectors in conventional PCA space vary with the related eigenvalues. In order to treat each component equally, let each component have equal variance and this is realized by weighting the conventional PCA space. Specifically, let

$$\begin{aligned}\mathbf{z}_k &= diag(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \cdots, \lambda_m^{-1/2})\mathbf{y}_k \\ &= \mathbf{\Lambda}^{-1/2}\mathbf{y}_k \qquad (k = 1,2,\cdots,N)\end{aligned} \tag{9}$$

Now, the conventional PCA space is transformed to a weighted PCA space, which we call WPCA space in this paper. Then, the covariance matrix of a feature vector in WPCA space becomes

$$\begin{aligned}\mathbf{C}_z &= \sum_{k=1}^{N}\mathbf{z}_k\mathbf{z}_k{}^t \\ &= \sum_{k=1}^{N}\mathbf{\Lambda}^{-1/2}\mathbf{y}_k\mathbf{y}_k{}^t\mathbf{\Lambda}^{-1/2} \\ &= \mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}\mathbf{\Lambda}^{-1/2} \\ &= \mathbf{I}\end{aligned} \tag{10}$$

It is seen from above equation that all the component of $\mathbf{z}_k$ are not only independent each other, but also of unit variance.

In addition, it seems that a linear transform defined by equation (9) is required to convert the conventional PCA space to WPCA space, but in fact, this transform can be incorporated into equation (1). That is

$$\begin{aligned}\mathbf{z}_k &= \mathbf{\Lambda}^{-1/2}\mathbf{y}_k \\ &= \mathbf{\Lambda}^{-1/2}\mathbf{W}^t(\mathbf{x}_k - \boldsymbol{\mu}) \\ &= \mathbf{V}^t(\mathbf{x}_k - \boldsymbol{\mu}) \qquad (k = 1,2,\cdots,N)\end{aligned} \tag{11}$$

where $\mathbf{V} = \left(\lambda_1^{-1/2}\mathbf{u}_1\ \lambda_2^{-1/2}\mathbf{u}_2 \cdots \lambda_m^{-1/2}\mathbf{u}_m\right)$. Therefore, the proposed WPCA transform has the same computational complexity with that of conventional PCA.

3.2. Unchanged reconstruction error

The original image can be estimated in the compressed WPCA space by

$$\hat{\mathbf{x}}_k = \mathbf{W}\mathbf{\Lambda}^{1/2}\mathbf{z}_k + \boldsymbol{\mu} \qquad (k = 1,2,\cdots,N). \tag{12}$$

The error of above estimation is obviously

$$\begin{aligned}\Delta\mathbf{x}_k &= \mathbf{x}_k - \hat{\mathbf{x}}_k(m) \\ &= \mathbf{W}(n)\mathbf{\Lambda}^{1/2}(n)\mathbf{z}_k(n) - \mathbf{W}(m)\mathbf{\Lambda}^{1/2}(m)\mathbf{z}_k(m) \\ &= \sum_{i=m+1}^{n}\mathbf{z}_k(i)\lambda_i^{1/2}\mathbf{u}_i \qquad (k = 1,2,\cdots,N)\end{aligned} \tag{13}$$

It follows that the mean square error is

$$\varepsilon = E\left\{\left\|\Delta\mathbf{x}_k\right\|^2\right\} = \sum_{i=m+1}^{n}\lambda_i \tag{14}$$

Comparing above equation with (7), it is clear that when an original image is to be reconstructed in WPCA space, the same rule as in PCA space should be followed, that is, choose the m weighted eigenvectors corresponding to the m largest eigenvalues to form the transform matrix in order to minimize the mean square error of reconstruction.

3.3. Mahalanobis distance

Let $\boldsymbol{\mu}_z \in \mathrm{R}^m$ be the mean vector in WPCA space, then the squared Euclidean distance between any feature vector $\mathbf{z}$ and the mean vector given by

$$\begin{aligned}D_E^2(\mathbf{z}, \boldsymbol{\mu}_z) &= (\mathbf{z} - \boldsymbol{\mu}_z)^t(\mathbf{z} - \boldsymbol{\mu}_z) \\ &= (\mathbf{z} - \boldsymbol{\mu}_z)^t\mathbf{C}_z^{-1}(\mathbf{z} - \boldsymbol{\mu}_z) \\ &= D_M^2(\mathbf{z}, \boldsymbol{\mu}_y)\end{aligned} \tag{15}$$

And it also follows that

$$\begin{aligned}D_E^2(\mathbf{z}, \boldsymbol{\mu}_z) &= (\mathbf{y} - \boldsymbol{\mu}_y)^t\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}^{-1/2}(\mathbf{y} - \boldsymbol{\mu}_y) \\ &= (\mathbf{y} - \boldsymbol{\mu}_y)^t\mathbf{C}_y^{-1}(\mathbf{y} - \boldsymbol{\mu}_y) \\ &= D_M^2(\mathbf{y}, \boldsymbol{\mu}_y)\end{aligned} \tag{16}$$

Equation (15) and (16) show that Euclidean distance and Mahalanobis Distance are the same in WPCA space. Thus, when a minimum Euclidean distance classifier is used, statistical distribution information of the patterns is also taken into consideration. Furthermore, classification according to minimum Euclidean distance criterion in WPCA space is equivalent to classification according to minimum Mahalanobis distance criterion in PCA space. The readers can refer to [8] about the comparison of using different distance metrics in PCA space.

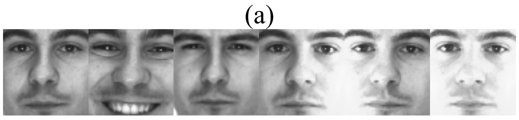
3.4. On the contribution of each feature component to pattern classification

In PCA space, the variances of different feature component are different because of different related eigenvalues. But this difference is not equivalent to that of the importance of different component to pattern classification. Similarly, in WPCA space, all the components are regarded equally and given equal variances, however, it dose not mean that the effect of each component in classification is completely the same. In fact, the contribution of each component is determined by the specific problem itself. For instance, some components may correspond to the common characteristics of all the classes and so that are less significant in classification. While some other components may correspond to the characteristics of individual classes and so that are more significant in classification. In view of the different variances of different components, if a larger variance is mainly caused by the differences between different classes, it will be more significant in classification. However, if it is caused mainly by the disparity within classes, then its effect in classification will be negative. We found in the experiment that although evidently better recognition results are generally obtained in WPCA space, much different results are obtained when different combination of components are chosen. Generally, better results can be obtained using components corresponding to larger eigenvalues.

4. Experimental results

AR and ORL publicly available databases were used in the experiment. For the AR database [9], we used the images of 117 out of all the 126 individuals, those of the rest being either not complete or not available. For each individual, we used 12 out of the 26 images and avoided the rest, which are either occluded by sunglasses or scuffs, or with closed eyes because of screaming laugh. The images were all cropped and normalized to 64×64 gray ones. Fig. 2 shows the 12 images from one person, where the acquiring time of (a) and (b) is 14 days apart. For the ORL database (http://www.cam-orl.co.uk), we used all the 400 images from 40 individuals and no crop and normalization were performed.

(a)



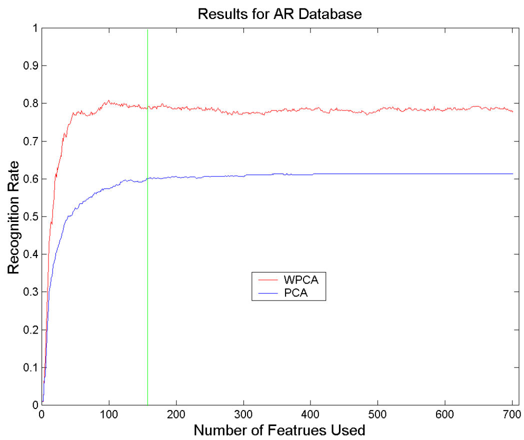
(b)

Figure 2. AR database samples of one person used in the experiment

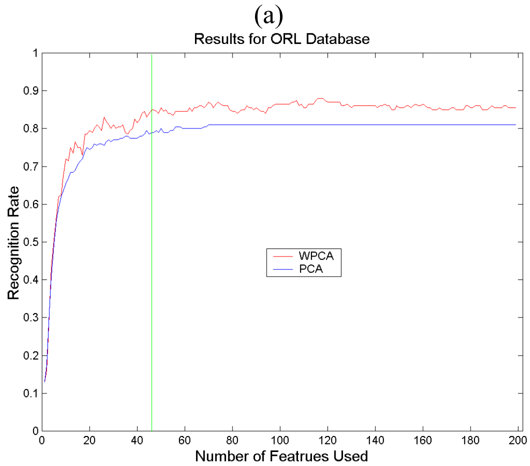
4.1. Experiment on face recognition in WPCA and PCA spaces

Fig. 3 shows the recognition result using transform matrixes formed by different number of eigenvectors corresponding to the largest eigenvalues. Where, for AR database, the 6 images shown in Fig. 2 (a) of each individual were used as the training samples and those in Fig. 2 (b) as the test samples. For ORL database, the first 5 out of 10 images of each person were used as the training samples and the rest as the test samples.

It is seen that WPCA approach achieves higher recognition rate than that of conventional PCA when the same number of features are used. But the feature add ability in WPCA is not as good as that in PCA, because the recognition rate in PCA increases monotonously as the number of features used increases, while in WPCA, vibration appears. The maximum recognition rates of the



(a)



(b)

Figure 3. Comparison of face recognition results in WPCA and PCA spaces

Table 1. Maximum recognition rates

Database	PCA		WPCA	
	Recognition Rate	Number of Features	Recognition Rate	Number of Features
AR	61.4%	471	80.9%	99
ORL	81.0%	70	88.0%	116

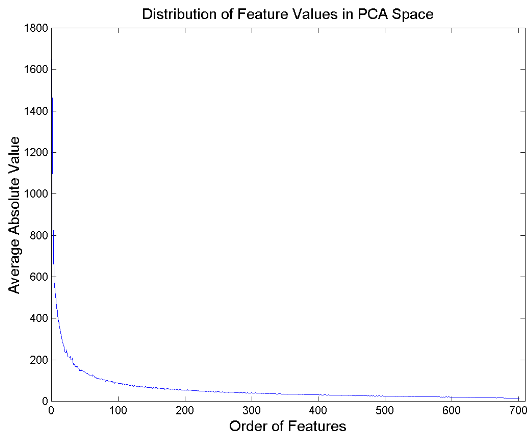
Table 2. Stable recognition rates

Database	PCA		WPCA	
	Recognition Rate	Number of Features	Recognition Rate	Number of Features
AR	60.1%	159	79.2%	159
ORL	79.0%	46	85.0%	46

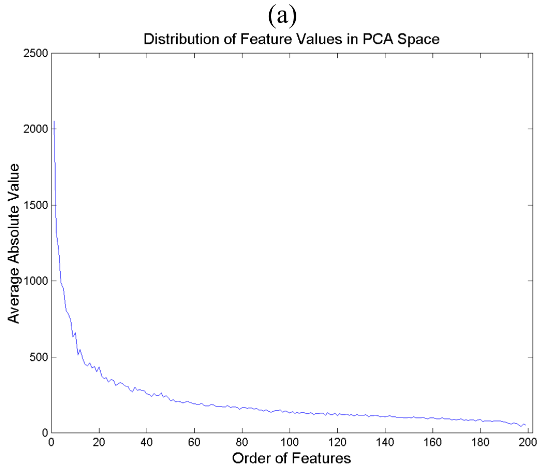
two approaches are given in Tab. 1. And the stable recognition rates marked by the vertical lines in Fig. 3 are given in Tab. 2.

4.2. Comparison of WPCA and PCA in feature distribution

To have some idea about the distribution of the features and to study the contribution of each feature in both PCA and WPCA spaces in more detail, we present some result on above databases in Fig. 4 and Fig. 5. Fig. 4 shows the average distribution of the absolute value of features in PCA space and Fig. 5 shows that in WPCA space. It is seen from Fig. 4 that as is pointed out earlier, in PCA space, a few features corresponding to the first few largest eigenvalues have much larger statistical means of

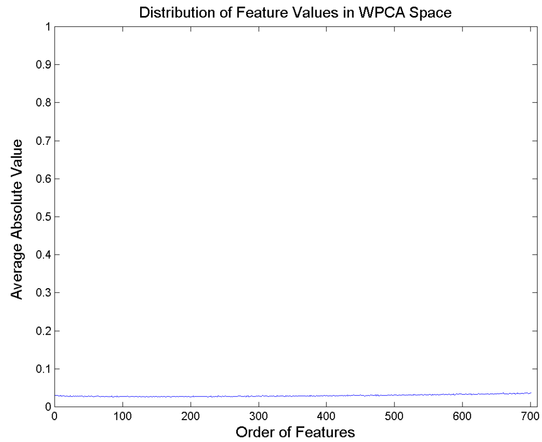


(a)

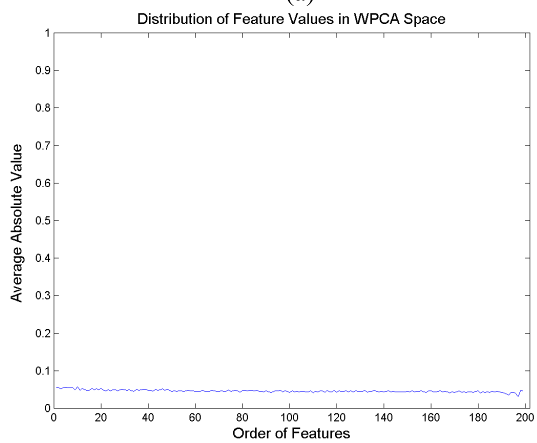


(b)

Figure 4. Distribution of the average absolute values of features in PCA space. (a) AR database (b) ORL database

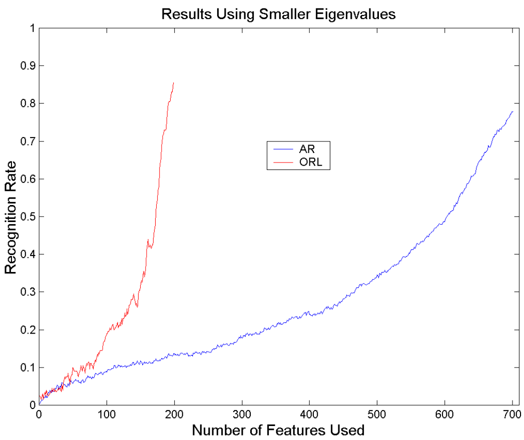


(a)

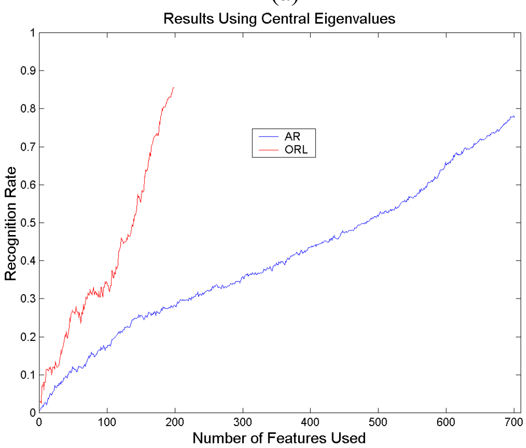


(b)

Figure 5. Distribution of the average absolute values of features in WPCA space. (a) AR database (b) ORL database



(a)



(b)

Figure 6. Results of different feature combinations in WPCA space

computation of the Euclidean distances. While in WPCA space, as is shown in Fig. 5, the statistical means of the absolute values of all the features are very close to each other in a wide range and so the combined effect of the features is ensured.

4.3. Recognition results of different feature combination in WPCA space

Recognition results of two different feature combinations in WPCA space are shown in Fig. 6, where (a) shows the results using features corresponding to smaller eigenvalues. And (b) shows the results using the central part of eigenvalues. Comparing Fig. 6 (a) and (b) with Fig. 3, it is seen that even in WPCA space, features corresponding to [10]

absolute values and then take a dictating role in the

larger eigenvalues are still more important in classification.

5. Conclusions

By weighting the features in PCA space, features in WPCA space with equal variances are obtained to equalize the relative variation amplitudes of different features. Thus, recognition performance is increased by a considerable extent without increasing the computation complexity. Through analysis and experiment, it is pointed out that as in PCA space, eigenvectors corresponding to larger eigenvalues should be chosen to form the transform matrix for the purpose of either representation of recognition of face images. Furthermore, it is observed that a few features that are not very important in classification take much larger values in both amplitude and variance is a possible key factor that results in the decreased performance of classifiers with Euclidean distance criteria such as PCA approach.

References

[1] Kirby M and Sirovich L. “Application of the Karhunen-Loeve Procedure for the Characterization of human Faces”. IEEE Trans., 1990, PAMI-12(1): 103-108.

[2] Turk M and Pentland A. “Eigenfaces for Recognition”. J. Cognitive Neuroscience, 1991,13(1):71-86.

[3] Belhumeur P N, Hespanha J P, and Kriegman D J. “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection”. IEEE Trans., 1997, PAMI-19(5): 711–720.

[4] Martnez A M and Kak A C. “PCA versus LDA”. IEEE Trans. , 2001, PAMI-23(2): 228–233.

[5] Yang M H. “Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods”. Proc. Fifth IEEE Int’l Conf. Automatic Face and Gesture Recognition (RGR’02)[C], 2002. 215-220.

[6] Bartlett M S, Movellan J R, and Sejnowski T J. “Face Recognition by Independent Component Analysis”. IEEE Trans., 2002, NN-13(6):1450-1464.

[7] Yuen P.C and Lai J H. “Face Representation Using Independent Component Analysis”. Pattern Recognition, 2002, 35(6):1247-1257.

[8] Perlibakas V. “Distance measures for PCA-based face recognition”. Pattern Recognition Letters, 2004, 25(6):711–724.

[9] Martinez A M and Benavente R. The AR Face Database[DB/OL]. http:// rvl1.ecn.purdue.edu/ ~aleix/ aleix_face_DB. html, 2003.