An algorithm to minimize within-class scatter and to reduce common matrix dimension for image recognition

Ümit Çiğdem TURHAL[1,*]**, Alpaslan DUYSAK**[2]
[1]*Department of Electrical and Electronics Engineering, Bilecik University, Bilecik-TURKEY*
e-mail: ucigdem.turhal@bilecik.edu.tr
[1]*Department of Computer Engineering, Dumlupınar University, Kütahya-TURKEY*

Abstract

In this paper, a new algorithm using 2DPCA and Gram-Schmidt Orthogonalization Procedure for recognition of face images is proposed. The algorithm consists of two parts. In the first part, a common feature matrix is obtained; and in the second part, the dimension of the common feature matrix is reduced. Resulting common feature matrix with reduced dimension is used for face recognition. Column and row covariance matrices are obtained by applying 2DPCA on the column and row vectors of images, respectively. The algorithm then applies eigenvalue-eigenvector decomposition to each of these two covariance matrices. Total scatter maximization is achieved taking the projection of images onto d eigenvectors corresponding to the largest d eigenvalues of column covariance matrix, yielding the feature matrix. The each column of the feature matrix represents a feature vector. Minimization of within class scatter is achieved by reducing the redundancy of the corresponding feature vectors of the different images in the same class. A common feature vector for each dth *eigenvector direction is obtained by applying Gram-Schmidt Orthogonalization Procedure. A common feature matrix is established by gathering d common feature vectors in a matrix form. Then, the dimension of common feature matrix is reduced to d × d taking the projection of common feature matrix onto d eigenvectors which corresponds to the largest d eigenvalues of row covariance matrix. The performance of the proposed algorithm is evaluated experimentally by measuring the recognition rates. The developed algorithm produced better recognition rates compared to Eigenface, Fisherface and 2DPCA methods. Ar-Face and ORL face databases are used in the experimental evaluations.*

Key Words: *2DPCA, Gram-Schmidt orthogonalization, common feature matrix, face recognition, total scatter maximization, within-class scatter minimization*

1. Introduction

Image recognition is a specific hard case for object recognition. It has attracted extensive attention within the past 20 years in computer vision because of the potential applications in many fields, such as identity authentication, surveillance and human-computer interface. Since the most common appearance of faces roughly

---

*Corresponding author: Department of Electrical and Electronics Engineering, Bilecik University, Bilecik-TURKEY

look alike and facial image data is high dimensional, the task of face recognition is extremely difficult. Therefore face image representation plays an important role in image recognition.

Among the existing face recognition techniques, subspace methods are widely used to extract low dimensional features [1]. In the literature, a variety of linear and nonlinear subspace methods can be found for low dimensional feature extraction [2, 3]. Principal Component Analysis (PCA) is one of the most popular linear subspace methods, which uses Karhunen-Loéve Transform (KLT) to produce a most expressive subspace for face representation and recognition [4]. PCA, known as Eigenfaces have played a fundamental role in dimensionality reduction and have demonstrated excellent performance. PCA encodes information in an orthogonal linear space and maximizes the total scatter. This means that the Eigenfaces can not get rid of much within-class variations occur due to the illumination changes or changes in facial expressions.

Linear Discriminant Analysis (LDA) is an another linear subspace method, which can achieve relatively high classification accuracy compared to PCA [5]. This method aims to minimize the within-class scatter while maximizing the between-class scatter. Unlike PCA, it encodes discriminatory information in a linearly separable space, of which bases are not necessarily orthogonal.

Fisherface is a PCA+LDA method with high popularity within the image recognition community [3]. Using PCA, high dimensional data is projected to a low dimensional space and then LDA is performed in this PCA subspace. Generalization capability of LDA is improved when only few samples per person are available. Li et al. [2] provided a comparative study for these subspace analysis methods. PCA and LDA are linear methods, which limits their applicability. Recently, their kernelized versions such as Kernel PCA [6], Kernel (FDA) [7] have appeared in the literature, where the limitation of linearity is overcome by construction of a high dimensional feature space with a nonlinear mapping.

Yang et al. [8] developed a two-phase KFD framework, i.e. kernel PCA plus Fisher linear discriminant analysis. Their study is based on this framework yielded an algorithm which is called complete kernel Fisher discriminant analysis (CKFD). An other algorithm for kernel PCA is proposed by Kim et al. [9], called the kernel Hebian algorithm (KHA). In this study generalized Hebian algorithm is used and the kernel PCA is iteratively estimated. It is stated that this algorithm is an unsupervised learning technique so the obtained image model can be used for the other tasks, as well.

Eigenface and Fisherface methods aim to find the projection directions based on the second order correlation of image samples. Kernel Eigenface and Kernel Fisherface methods provide generalizations to take higher order correlations into account. Ming-Hsuan Yang et al. [10] compared the performance of kernel methods with the classical methods. In their study, they showed that kernel methods provide better representation and can achieve a better recognition accuracy.

The above mentioned methods are based on transformation of the image matrix into vectors prior to processing them. Yang et al. [11] have proposed a new technique for image feature extraction called as 2DPCA. This method does not require the transformation of image matrices into vectors. In this method, a better recognition performance was obtained over one dimensional PCA. This method has also two important advantages over one dimensional PCA: First, it is easier to accurately evaluate the covariance matrix. Second, less time is required to determine the corresponding eigenvectors, which will be used for optimal projection axis.

Despite its advantages over the traditional PCA, Dong Xu et al. [12] have demonstrated its disadvantages. They have found that, if the training set has typically a small sample number, the traditional PCA gives better results with respect to 2DPCA. The main difference between the traditional PCA and 2DPCA methods is that the projection of the image matrices onto the optimal projection axes is a scalar in traditional PCA while it is

a vector in 2DPCA.

An algorithm for two dimensional image recognition is proposed in reference [13]. In this algorithm, the common properties of images in each individual class of the training set are extracted by eliminating the differences between images. In this algorithm, the common matrix is obtained using two different methods; namely within-class scatter matrix and Gram-Schmidt Orthogonalization [14]. The common matrix obtained is then used for face recognition. In this method, while the within-class scatter minimization is accomplished, the maximization of the total scatter is ignored. In addition, the size of the common matrix is the same as with the original image matrix.

In this proposed algorithm, both the minimization of within-class scatter and the maximization of total scatter, that were ignored in [13], are achieved. In addition, the dimension of the common feature matrix is reduced. The proposed method provides the following distinguishable contributions: 1. A new procedure, using 2DPCA and Gram-Schmidt Orthogonalization to obtain a common feature matrix, is developed. 2. A new method to reduce the dimension of a common feature matrix is provided.

Remainder of this paper is organized as follows. Section 2 briefly describes the well-known 2DPCA method. The proposed method as a two-step algorithm is given in section 3 in detail. A testing procedure for the developed algorithm is also provided in this section. Experimental results are presented in section 4. Finally, various examples and comparisons with the other methods demonstrate the performance of our proposed algorithm.

2. The 2DPCA method

2DPCA is a linear subspace method used for feature extraction processing directly using the image matrices. The idea in this method is to find the optimal projection directions maximizing the scatter of the transformed images. As the optimal projection directions, the eigenvectors corresponding to the maximum d eigenvalues of the image covariance matrix are used. The main difference of 2DPCA from traditional PCA is that 2DPCA does not require transforming image matrices into vectors. Thus, it reduces the computational complexity for the construction of the image covariance matrix and reduces the computation time of the eigenvectors of the covariance matrix.

Suppose that the training set has M images A_i, $i = 1, 2, \ldots, M$ with the dimensions of $m \times n$. If the average of the training images given as $\bar{A} = \frac{1}{M} \sum_{i=1}^{M} A_i$, the covariance matrix of the images G_t can be computed as:

$$G_t = \frac{1}{M} \sum_{i=1}^{M} \{A_i - \bar{A}\}^T \{A_i - \bar{A}\} \tag{1}$$

The optimal projection axis x_{opt} is the eigenvector of G_t corresponding to the largest eigenvalue. In general, it is not enough to have only one optimal projection axis. It usually needs to select a set of projection axes. $x_1, \ldots, x_d$. The optimal projection axes, $x_1, \ldots, x_d$ are the orthonormal eigenvectors corresponding to the first d largest eigenvalues of G_t. The $m \times 1$ dimensional feature vectors representing the image matrices in d projection direction is obtained by the following linear transformation:

$$y_k = A_i x_k, \tag{2}$$

where $k = 1, \ldots$ and $d, i = 1, \ldots, M$.

With this transformation, each image of a person in the training set is represented by a d feature vectors and only the total scatter of the training images are maximized. Here, there is no attempt to minimize within-class scatter, which is the subject of the proposed method given in the following section.

3. An algorithm to minimize within class scatter and to reduce the dimension of common matrix

The developed algorithm can be summarized as follows. It is based on 2DPCA method and common vector approach, and can be divided into two parts. In the first part of the algorithm, 2DPCA is performed both on the column and the row vectors of the image matrices. Thus, two optimal projection vector sets are obtained. The first optimal projection vector set, obtained from the application of 2DPCA on the column vectors of image matrices, is used for the maximization of the total scatter. This part constructs the common matrix for each class. In the second part of the algorithm, the second optimal projection vector set, obtained from the application of 2DPCA on the row vectors of the image matrices, is used for reduction of the common matrix dimension. The mathematics of the algorithm is now studied in detail.

3.1. Obtaining common matrix for each class

Let us assume the training set has C classes and l images in each class. "Class" refers to the ensemble of a certain number of images per person. The main idea in this approach is to spread out the different classes as far apart from each other as possible, and then to reduce the variation within classes. Maximization between class scatters is achieved using the variations between the column vectors of each image in all classes. Then, a common matrix is computed for each class to minimize within-class scatter.

The training set consists of $M = lC$ images in total. The mean image matrix of the training set is denoted as $\bar{A}$ and computed as, $\bar{A} = \frac{1}{M} \sum_{c=1}^{C} \sum_{i=1}^{l} A_i^c$. Here, A_i^c denotes the ith image in class c with the size of $m \times n$. Performing 2DPCA on the column vectors of the image matrices column covariance matrix G_{tc} is then obtained as

$$G_{tc} = \frac{1}{M} \sum_{c=1}^{C} \sum_{i=1}^{l} (A_i^c - \bar{A})(A_i^c - \bar{A})^T. \tag{3}$$

The eigenvalue-eigenvector decomposition is applied to the column covariance matrix. The first d eigenvectors, which corresponds to the largest d eigenvalues, are used as the first optimal projection axis set. By projecting the images onto this set, the maximization in the total scatter is achieved. Thus, a feature vector set is obtained for each image in a class with this projection. As d optimal projection axes are used, the feature vector set of an image consists of d feature vectors. Suppose that y_{ik}^c denotes the feature vector for the ith image in the cth class in the kth optimal projection direction. The feature vector is obtained as

$$y_{ik}^c = (A_i^c)^T x_k, \tag{4}$$

where $i = 1, \ldots, l$, $k = 1, \ldots, d$ and $c = 1, \ldots, C$. x_k is the kth eigenvector which corresponds to the kth largest eigenvalue of the column covariance matrix.

By gathering d feature vector in a matrix form a low dimensional representation of an image is obtained and it is called feature matrix Y_i^c,

$$Y_i^c = [y_{i1}^c, \ldots, y_{id}^c]. \tag{5}$$

Now, the important point is to reduce the redundancy between the low dimensional representation of image matrices in the same class. Firstly, d feature vector set F_k^c is constructed for each class.

$$F_k^c = \{y_{1k}^c, y_{2k}^c, \ldots, y_{lk}^c\}. \tag{6}$$

where F_k^c includes the feature vectors of different images in a class in the kth eigenvector direction of the column covariance matrix. Then, using Gram-Schmidt Orthogonalization procedure a common feature vector $(y_{\text{com},k}^c)$ is obtained for each F_k^c feature vector set. As a result, the within class scatter minimization is achieved along each optimal projection direction separately.

The algorithm, in order to find the common feature vector, can be summarized as follows: Firstly, d feature vector space $F_k^c = \{y_{1k}^c, y_{2k}^c, \ldots, y_{lk}^c\}$ where $k = 1, \ldots, d$ are constructed. In order to reduce the redundancy in F_k^c, difference feature vector subspace is constructed and then Gram-Schmidt Orthogonalization procedure is applied in this subspace. It is achieved by taking any feature vector (y_{ik}^c) from the kth feature vector space as a reference and then subtracting it from the other feature vectors in that space:

$$\begin{aligned} y_{1k}^c &= \text{ref} \\ s_{(j-1)k}^c &= y_{jk}^c - \text{ref} \end{aligned} \tag{7}$$

where $j = 2, \ldots, l$.

Gathering $(l-1)$ difference vectors, the difference subspace S_k^c is obtained for the kth feature vector space as

$$S_k^c = \{s_{1k}^c, s_{2k}^c, \ldots, s_{(l-1)k}^c\}. \tag{8}$$

Applying Gram-Schmidt Orthogonalization Procedure in the difference subspace obtained in (8), orthonormal bases, $\beta_{1k}^c, \ldots, \beta_{(l-1)k}^c$, which span the difference subspace of the kth feature vector space, are obtained for each class. Here, the notation b_{ik}^c $(i = 2, \ldots, l-1)$ is defined as the difference between s_{ik}^c and projection $\langle s_{2k}^c, \beta_{1k}^c \rangle \beta_{1k}^c$. For $i = 1$, the definition is given as $b_{1k}^c = s_{1k}^c \rightarrow \beta_{1k}^c$. The orthonormal bases are then found as

$$\begin{aligned} b_{1k}^c &= s_{1k}^c \rightarrow \beta_{1k}^c = \frac{b_{1k}^c}{\parallel b_{1k}^c \parallel}, \\ b_{2k}^c &= s_{2k}^c - \langle s_{2k}^c, \beta_{1k}^c \rangle \beta_{1k}^c \rightarrow \beta_{2k}^c = \frac{b_{2k}^c}{\parallel b_{2k}^c \parallel}, \\ &\vdots \\ b_{(l-1)k}^c &= s_{(l-1)k}^c - \langle s_{(l-1)k}^c, \beta_{(l-2)k}^c \rangle \beta_{(l-2)k}^c \rightarrow \beta_{(l-1)k}^c, \\ &= \frac{b_{(l-1)k}^c}{\parallel b_{(l-1)k}^c \parallel}. \end{aligned} \tag{9}$$

Projecting any of the feature vector (y_{ik}^c) from the kth feature vector space onto the orthonormal bases obtained for the kth feature vector space, and subtracting the result from that feature vector, common feature vector $y_{\text{com},k}^c$ of the kth feature vector space is computed. Varying $k = 1, 2, \ldots, d$ common feature vector for

each kth direction in a class is obtained:

$$y_{\text{com},k}^c \quad = \quad y_{1k}^c - \sum_{i=1}^{l-1} \langle y_{1k}^c, \beta_{ik}^c \rangle \beta_{ik}^c. \tag{10}$$

Then, by gathering d common feature vector into a matrix form, a common feature matrix B_{com}^c is constructed:

$$B_{\text{com}}^c \quad = \quad [y_{\text{com},1}^c, \ldots, y_{\text{com},d}^c]. \tag{11}$$

The common feature vector in each feature subspace in each class is independent from the reference feature vector y_{1k}^c in (7), and the subtrahend feature vector y_{1k}^c in (10) [14] Table 1.

Table 1. Summary of the first part of the proposed algorithm.

> · Compute the column covariance matrix G_{tc} equation (3).
>
> · Apply eigenvalue eigenvector decomposition.
>
> · Compute the feature vectors y_{ik}^c equation (4).
>
> · Obtain Y_i^c equation (5).
>
> · Construct F_k^c equation (6).
>
> · Apply Gram-Schmidt Orthogonalization procedure.
>
> · Find the orthonormal bases $\beta_{1k}^c, \ldots, \beta_{(l-1)k}^c$ equation (9).
>
> · Compute the common feature vector $(y_{\text{com},k}^c)$ equation (10).
>
> · Obtain Common Feature Matrix (B_{com}^c) equation (11).

3.2. Reducing common matrix dimension

In order to reduce the dimension of the common matrix obtained in the previous section, the row covariance matrix G_{tr} is calculated taking each row of the image matrix as an object. It is the same covariance matrix as in 2DPCA. So the row covariance matrix can be calculated as

$$G_{tr} = \frac{1}{M} \sum_{i=1}^{M} \{A_i - \bar{A}\}^T \{A_i - \bar{A}\}. \tag{12}$$

Then eigenvalue-eigenvector decomposition is applied to the row covariance matrix. d eigenvectors corresponding to d largest eigenvalues are used for the optimal projection direction. The common matrix obtained for each class is then projected onto these d eigenvectors. The resulting common matrix with reduced dimension, R_{com}^c is calculated as

$$\begin{aligned} R_{\text{com}}^c \quad &= \quad (B_{\text{com}}^c)^T Z \\ Z \quad &= \quad [z_1, z_2, \ldots, z_d], \end{aligned} \tag{13}$$

where $z_k, k = 1, \ldots, d$, are the eigenvectors obtained for the row covariance matrix.

3.3. Testing procedure

In the testing procedure, feature vectors for the test image y_k^{test} are calculated as in (4). Taking the projections of each feature vector obtained for the test image onto the orthonormal bases obtained for the kth feature vector space for each class and then subtracting the result from that test feature vector. Then, the remaining feature vector for each class is obtained for the test image:

$$
\begin{aligned}
y_k^{\text{test}} &= (A_{\text{test}})^T x_k, \\
y_{\text{rem},k}^{\text{test},c} &= y_k^{\text{test}} - \sum_{i=1}^{l-1} \langle y_k^{\text{test}}, \beta_{ik}^c \rangle \beta_{ik}^c.
\end{aligned}
\tag{14}
$$

Then remaining feature matrix for class c $B_{\text{rem}}^{\text{test},c}$ is obtained by gathering d remaining feature vectors in (14) as

$$
B_{\text{rem}}^{\text{test},c} = [y_{\text{rem},1}^{\text{test},c}, \ldots, y_{\text{rem},d}^{\text{test},c}].
\tag{15}
$$

Then, a second projection is performed on the remaining matrix of the test image using the eigenvectors obtained from the row covariance matrix. Thus, the resulting remaining matrix with reduced dimension for the test image $R_{\text{rem}}^{\text{test},c}$ for each class are calculated as

$$
R_{\text{rem}}^{\text{test},c} = (B_{\text{rem}}^{\text{test},c})^T Z
\tag{16}
$$

$$
Z = [z_1, z_2, \ldots, z_d]
\tag{17}
$$

For classification, a nearest neighbor classifier is used [16]:

$$
argmin_{1 \leq c \leq C} \{ \sum_{k=1}^{d} \parallel r_{\text{com},k}^c - r_{\text{rem},k}^{\text{test},c} \parallel_2 \}
\tag{18}
$$

where $r_{\text{com},k}^c$ and $r_{\text{rem},k}^{\text{test},c}$ are the column vectors of R_{com}^c and $R_{\text{rem}}^{\text{test},c}$ respectively.

If the test image belongs to the class c, the distance R_{com}^c, $R_{\text{rem}}^{\text{test},c}$ should be minimum.

4. Experimental studies

In the experimental study, the Ar-Face [17] and the ORL databases are used [18]. The Ar-Face database contains over 4000 color face images taken from 126 people (70 men and 56 women), including front view of faces with different facial expressions, lighting conditions and occlusions. The pictures of the people were taken in two different sessions separated by two weeks interval. Each session contains 13 color images for each person. In this study only 37 people (20 male and 17 females) were selected and used. Only the non-occluded images were used for each person. The face portion of the first images were cropped and then normalized to 50×40 pixels. The normalized images for one person are shown in Figure 1.

The developed algorithm was also tested with ORL database that contains 40 people having 10 different images each. They were taken a period of two years. The images were taken at different lighting conditions,

Figure 1. Sample images of a person from Ar-Face database.

facial expressions and details. The background of the images was dark homogeneous. The size of the each image was 122×92 pixels, with 256 grey levels per pixel. The images were taken with a tolerance for some tilting and rotation of the face of up to 20 degrees and with some variation in the scale of up to about 10 percent. Five sample images of a person from ORL database are shown in Figure 2.

Figure 2. Five sample images of a person from ORL database.

4.1.   Experiments on Ar-Face database

Three different experimental studies are made by using three different databases, which constructed from Ar-Face database. Each experiment is performed using developed method, Eigenface, Fisherface and 2DPCA algorithms. In the experiments we used the largest d eigenvectors, which contains the 95% of the total energy. In each experiment the eigenvector numbers used are approximately same.

In the first experiment, the first images of the first and second row in Figure 1 are used as the training set for each class. The second, third and fourth images taken from each row are used as the test set. As a result, the training set has 74 images. In the second experiment, the images in the first row of Figure 1 are used as the training set for each class and the images in the second row of Figure 1 are used as the test set. Thus the training set has 259 images. In the third experiment, "leave one out" strategy is used. Therefore the training set has 481 images. The results are given in the Table 2.

According to the experimental results shown in Table 2, the best recognition rates are achieved for the proposed method as it is compared with the other well-known methods such as Eigenfaces, Fisherfaces, 2DPCA and the method given in [15]. In the first experimental study, only the faces with normal expressions are used for the training set and the faces with expression differences are used for the test set. The results have shown that the difference between the recognition rates of 2DPCA and the proposed method is small. But the difference

Table 2. Recognition rates obtained using different methods on the three different databases constructed from Ar-Face database.

Methods	1. Results (%)	2. Results (%)	3. Results (%)
Eigenfaces	77.47	57.91	88.22
Fisherfaces	77.92	64.09	95.17
2DPCA	79.72	52.12	72.58
Work in [15]	80.63	66.40	95.94
New Method	82.43	68.72	97.29

between the recognition rates of 2DPCA and the proposed method has increased in the second experiment. In this experiment, the faces that are taken under different illumination conditions are used for the training and the test sets. Therefore the results of the proposed method gives better recognition rates in contrast to the other well-known methods when the illumination conditions are changed for different images. In the third experiment, "leave one out" method has used. The difference between the success of the proposed method and the other methods has increased to a higher value. The reason for this consequence is that the number of the images used for the training set has increased. As a result of the recognition performance evaluation performed, the proposed method gives the highest recognition rates with respect to the other well-known methods and to the work in [15].

In reference [13], a common matrix for face recognition was computed in a different manner. But, it has the same dimension as the original image matrix. In our proposed method the dimension of a common matrix is reduced to $d \times d$ where d is the number of eigenvectors containing the 95% of the total energy. This value is relatively smaller than the original image matrix dimension.

4.2. Experiments on ORL database

Two different experiments are performed using ORL database. In the first experiment, first five images are used as the training set and the remaining images are used as the test set. In the second experiment, "leave-one-out" strategy is applied to the database. The performance of the proposed method in each experiment was compared with 2DPCA method. In the experiments, we used the largest d eigenvectors, which contains 95% of the total energy. In each experiment, the number of eigenvectors used are approximately same. Table 3 summarizes the recognition accuracy obtained in two experiments.

Table 3. Recognition rates obtained using the proposed and 2DPCA methods on the three different databases constructed from ORL database.

Methods	1. Results (%)	2. Results (%)
2DPCA	84.5	94.33
Proposed Method	89.5	98

The effect of the number of images used for the training set is examined in the experimental study which has performed on ORL database. The results obtained using the proposed method and 2DPCA are given in Table 3. According to the experimental results, the proposed method gives better recognition rates in respect of 2DPCA method. It has been shown in the experiments that the dimension of the common matrix was reduced to $d \times d$.

5. Conclusions

In this paper we have presented a new algorithm for the recognition of facial images. Proposed algorithm is based on computation of a common feature matrix and reduction its dimension. The algorithm is performed in two phases. In the first phase of the algorithm, feature vectors (representation of the facial images) are obtained using 2DPCA performed on the column vectors of the image matrices. Then, Gram-Schmidt Orthogonalization Procedure is applied to the feature vectors resulting the common feature matrices. As a result, the maximization of between-class scatter was achieved while minimizing within-class scatter between the feature vectors. In the second phase of the algorithm, as a further step, the dimension of the common feature matrix was reduced by projecting the common feature matrix onto the eigenvectors obtained from the row covariance matrix. The resulting common feature matrix with reduced dimension is used for face recognition.

The algorithm provides following important contributions. Performance improvements: The maximization between-class scatter was achieved while minimizing within-class scatter between the feature vectors. Thus redundancies between the column and row vectors of the image matrices and also between the feature vectors of different images in the same class are reduced. Dimension Reduction: The dimension of the common feature matrix was considerably decreased. Thus both the testing time and the storage requirements of the common feature matrices are also decreased.

Dimension of the common feature matrix has reduced to the selected eigenvector numbers $d \times d$ which contains the 95% of the total energy. In the experiments performed on Ar-Face database selected eigenvector numbers d has been 16 while original image size is 50×40 pixels and on ORL face database selected eigenvector numbers d has been 26 while original image size is 112×92 pixels. Thus for Ar-Face a common feature matrix with dimension of 16×16 and for ORL a common feature matrix with dimension of 26×26 are used for recognition of facial images.

The experiments have shown that the performance of the proposed method (over the 2DPCA) increases with the increasing number of images used for a class in the training set. Therefore, if databases include more images for a class the recognition rate would be increased. In a future study, the performance of the proposed method will be evaluated using different databases.

References

[1] Zhao W., Chellappa R., Rosenfeld A. and Phillips P., Face Recognition,A Literature Survey, Technical Report (2000)

[2] Li J., Zhou S., and Shekhar C., A Comparison of Subspace Analysis for Face Recognition, ICASSP (2003)

[3] Zhao W., Chellappa R., Krishnaswamy A., Discriminant Analysis of Principle Components for Face Recognition, International Conference on Face and Gesture Recognition, fg, p.336, 3 rd. (1998)

[4] Turk M. and Petland A., Eigenfaces for Recognition, Journal of Cognitive Neutoscience, vol. 3, pp. 72–86 (1991)

[5] Belhumeur P., Hespanha J. and Kreigman D. J., Discriminant Eigenfeatures for Image Retreival, PAMI, 19 (7): 711–720 (1997)

[6] Schölkopf B., Smola A. and Muller K., Nonlinear Component Analysis as a Kernel Eigenvalue Problem, Neural Computation, vol. 10, pp. 1299–1319 (1998)

[7] Mika S., Rotsch G., Wetson J., Schölkopf B., and Mller K.-R., Fisher Discriminant Analysis with Kernels. In Neural Networks for Signal Processing IX, IEEE, Y.-H. Hu, J. Larsen, E. Wilson and S. Douglas, Eds., pp. 41–48 (1999)

[8] Yang J., Frangi A. F., Yang J., Zhang D., KPCA Plus LDA: A Complete Kernel Fisher Discriminant Framework for Feature Extraction and Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligenc, vol. 27, No. 2 (2005)

[9] Kim K. I., Franz M. O. and Schölkopf B. , Iterative Kernel Principal Component Analysis for Image Modelling, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, No. 9 (2005)

[10] Yang M., Kernel Eigenfaces v.s. Kernel Fisherfaces: Face Recognition Using Kernel Methods, Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition (2002)

[11] Yang J., David Z., Frangi A. F. and Yang J-Y., Two Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition, Pattern Analysis and Machine Intelligence, IEEE Transactions on vol. 26, No. 1 (2004)

[12] Xu D., Yan S., Zhang L., Li M., Ma W., Liu Z., Zhang H., Parallel Image Matrix Compression for Face Recognition, mmm, 11.th International Multimedia Modelling Conference (MMM'05), pp. 232–238 (2005)

[13] Turhal Ü. Ç., Gülmezoglu M. B., Barkana A., Face Recognition Using Common Matrix Approach, EUSIPCO, Antalya, on Sept: 4–8 (2005)

[14] Gülmezoglu M. B., Dzhafarov V., Keskin M. and Barkana A., A Novel Approach to Isolated Word Recognition, Speech and Audio Processing, IEEE Transactions on vol. 7, pp. 620–628 (1999)

[15] Turhal Ü. Ç., Duysak A., Gülmezoglu M. B., A Two Stage Algorithm for Face Recognition: 2DPCA and Within-Class Scatter Minimization, (554) Signal Processing, Pattern Recognition, and Applications (2007)

[16] Cover T.M. and Hart P.E., Nearest Neighbor Pattern Classification, IEEE Trans. on Information Theory, pp. 21–27 (1967)

[17] Martinez A. M. and Benavante R., The Ar Face Database, CVC Tech. Report ♯ 24 (1998)

[18] ORL Face Recognition Database, http:www.cam-orl.co.uk