



Cross grouping strategy based 2DPCA method for face recognition



Ü.Ç. Turhal^{a,*}, A. Duysak^b

^a Electric and Electronics Engineering Department, Bilecik Şeyh Edebali University, Turkey

^b Computer Engineering Department, Dumlupınar University, Turkey

ARTICLE INFO

Article history:

Received 28 November 2013

Received in revised form 11 August 2014

Accepted 12 January 2015

Available online 19 January 2015

Keywords:

Facial image recognition

Grouping strategy

2DPCA

Cross-covariance

Singular value decomposition

ABSTRACT

Grouping strategy exactly specifies the form of covariance matrix, therefore it is very essential. Most 2DPCA methods use the original 2D image matrices to form the covariance matrix which actually means that the strategy is to group the random variables by row or column of the input image. Because of their grouping strategies these methods have two main drawbacks. Firstly, 2DPCA and some of its variants such as A2DPCA, DiaPCA and MatPCA preserve only the covariance information between the elements of these groups. This directly implies that 2DPCA and these variants eliminate some covariance information while PCA preserves such information that can be useful for recognition. Secondly, all the existing methods suffer from the relatively high intra-group correlation, since the random variables in a row, column, or a block are closely located and highly correlated. To overcome such drawbacks we propose a novel grouping strategy named cross grouping strategy. The algorithm focuses on reducing the redundancy among the row and the column vectors of the image matrix. While doing this the algorithm completely preserves the covariance information of PCA between local geometric structures in the image matrix which is partially maintained in 2DPCA and its variants. And also in the proposed study intra-group correlation is weak according to the 2DPCA and its variants because the random variables spread over the whole face image. These make the proposed algorithm superior to 2DPCA and its variants. In order to achieve this, image cross-covariance matrix is calculated from the summation of the outer products of the column and the row vectors of all images. The singular value decomposition (SVD) is then applied to the image cross-covariance matrix. The right and the left singular vectors of SVD of the image cross-covariance matrix are used as the optimal projective vectors. Further in order to reduce the dimension LDA is applied on the feature space of the proposed method that is proposed method + LDA. The exhaustive experimental results demonstrate that proposed grouping strategy for 2DPCA is superior to 2DPCA, its specified variants and PCA, and proposed method outperforms bi-directional PCA + LDA.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The automated face recognition is an active research area because of the growth in its application in many areas such as face identification and intelligent user interfaces. The face recognition is basically comparing an input face with the face models stored in a database of known faces and matching the input face with a known face in the database. One of the difficult tasks in the face recognition procedure is the high dimensionality of the facial images. Researchers have developed different algorithms, known as subspace methods, for dimension reduction. One of the most popular approaches to reduce image dimension is the Principal Component Analysis (PCA) [1]. This method

produces the most expressive subspace for face representation and recognition. LDA [2] and Fisherfaces [3] methods are the examples of most discriminating subspace methods and other subspace methods [4–6] are modifications or extensions of these methods.

PCA method reduces the redundancy between image pixels. In this method, image matrix has to be transformed into a vector. Transforming an image matrix into a vector increases computational time required to obtain eigenvectors of the image covariance matrix. Recently, to overcome this problem, 2DPCA is proposed by Yang et al. [7]. In 2DPCA, the image matrix is not transformed into a vector but the image covariance matrix is constructed directly using the original image matrix. That is why the computational complexity is reduced in 2DPCA compared to PCA. PCA, on the other hand, requires fewer coefficients than 2DPCA for image representation and therefore is more efficient than 2DPCA in terms of storage requirements. Xu et al. [8] provided that 2DPCA is a

* Corresponding author. Tel.: +90 2282141553.

E-mail address: ucigdem.turhal@bilecik.edu.tr (Ü.Ç. Turhal).

localized PCA. In their work, a lower dimensional representation of the image matrix is obtained by first applying 2DPCA on the row vectors of the image matrix and then applying 2DPCA on the column vectors of the projected image matrix. Wang et al. [9] showed that 2DPCA is a special case of the block image based PCA which has been used for face recognition in a number of systems. Specifically, the blocks are the lines of the raw images. In their experimental studies, they also showed that 2DPCA achieves higher accuracy than typical rectangle block based PCA. Xu et al. [10] used Coupled Subspace Analysis algorithm to overcome the curse of dimensionality dilemma and small sample size problem. Xu et al. [11] also proposed an algorithm, which aims to reduce the redundancy between row vectors and between column vectors by using 2nd or even higher order tensors. Zuo et al. [12] proposed a fast facial-feature extraction technique, BDPCA + LDA, for face recognition. In their work, they first transform an image matrix into BDPCA feature subspace. They do this transformation by using the row projection and the column projection matrices constructed from the application of 2DPCA on the row of the images and the application of 2DPCA on the column of the images, respectively. Then, they apply LDA on this low dimensional BDPCA subspace. In their work, they showed that the BDPCA + LDA framework has a number of significant advantages over the PCA + LDA framework. Nousath et al. [13] applied 2DPCA followed by image matrix based on linear discriminant analysis in order to obtain a lower dimensional representation of the image matrix. Ye [14] proposed Generalized Low Rank Approximation of Matrices (GLRAM) claiming that this method is superior to 2DPCA both in terms of compression and classification performance. However GLRAM lacks a criterion to automatically determine the dimensionality of the projected image matrix. Liu and Chen [15] proposed non-iterative version of GLRAM (NGLRAM) to overcome disadvantages of GLRAM. Yang and Liu [16] presented a horizontal and vertical 2DPCA-based discriminant analysis (HVDA) method for the face verification. In their work, they provide some theoretical justifications for 2DPCA. One of the theoretical justification is the property of 2DPCA that leads to robust transform matrices to image translations and pose variations (mirror images). In their work, they showed that 2DPCA-based methods are less sensitive to imprecise eye detection and face cropping and improved the traditional PCA-based methods for face verification.

Another 2DPCA based recognition algorithm can be found in [17]. In this study, a new face recognition scheme combining 2D and 3D information was proposed. Their method is based on 2DPCA and performs 2DPCA algorithm on the column vectors of the image matrix instead of the row vectors as in 2DPCA. Thus, this method reduces the redundancies between columns of the image matrix. It requires 3D facial data in the training process but can process 2D pictures in the recognition stage. According to the experimental results, this method is very robust for pose variation scenarios.

The above mentioned studies show that subspace methods, used for recognition and representation of images, try to group the elements in some manner and then reduce the redundancy exists in these groups. Examples of most common grouping strategies are analyzed in [18]. Grouping strategy exactly specifies the form of covariance matrix, therefore it is very essential. PCA method, for example, is used to eliminate redundancy among the pixel values of the 1D image vector while 2DPCA method is used to reduce redundancy among row vectors of the image matrix. Most 2DPCA based methods [8–16,18–21] use the original 2D image matrices to form the covariance matrix which actually means that the strategy is to group the random variables by row or column of the input image. In this paper a novel grouping strategy is proposed. The main contributions of this paper are as follows:

1. A definition of image cross-covariance matrix is proposed by developing a new grouping strategy.
2. This new covariance matrix preserves the covariance information of PCA between local geometric structures in the image which is partially maintained in 2DPCA and its variants.
3. In this new grouping strategy intra-group correlation is weak according to the 2DPCA and its variants because the random variables spread over the whole face image. This also makes the algorithm less sensitive to the lighting condition changes because the new grouping strategy reduces the intra-group correlation.
4. Further in order to reduce the dimension, LDA is applied on the feature space of the proposed method. This also increased the recognition performance according to bi-directional PCA + LDA.

All these make the proposed algorithm superior to PCA, 2DPCA and 2DPCA variants in terms of recognition accuracy. Beside, if we evaluate the PCA, 2DPCA, 2DPCA variants and the proposed method for the time required and computational complexity, following results revealed. The length of the object forming the group specifies the covariance matrix dimension, and the covariance matrix dimension specifies the time required and the computational complexity of this algorithm. In 2DPCA method, its variants and the proposed method groups are constructed using the row or/and column vectors so the time requirements and the computational complexity in all these methods are close to each other; however all are much more smaller than PCA.

The developed algorithm constructs image cross-covariance matrix from the summation of the outer products of the column and the row vectors of the image matrix. As a second step, singular value decomposition (SVD) is applied to the image cross-covariance matrix. The resulting right and left singular vectors are used as the optimal projection vectors for the feature extraction.

The remaining of this paper is organized as follows: Section 2 introduces the standard 2DPCA method. Section 3.1 compares the proposed method with 2DPCA and its variants. Cross-covariance matrix of the face images is obtained and diagonalized in Section 3.2. The covariance matrix relationships between the proposed method and PCA, 2DPCA and 2DPCA variants are determined in Section 3.3. Feature extraction and image recognition are demonstrated in Section 3.4. Feature vectors are calculated in bi-directional way by applying LDA after the proposed method (proposed method + LDA) as is given in Section 3.5. Finally, Section 4 provides experimental results in order to show the performance of the proposed method.

2. The 2DPCA method

2DPCA is a linear subspace method, which is used for feature extraction. The idea behind this method is to find the optimal projection directions, which can maximize the scatter of the transformed images. For the optimal projection directions, eigenvectors corresponding to the maximum d eigenvalues of the image covariance matrix are used [7].

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 of 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 image matrices A_i , $i = 1, 2, \dots, M$ with the dimension of $m \times n$

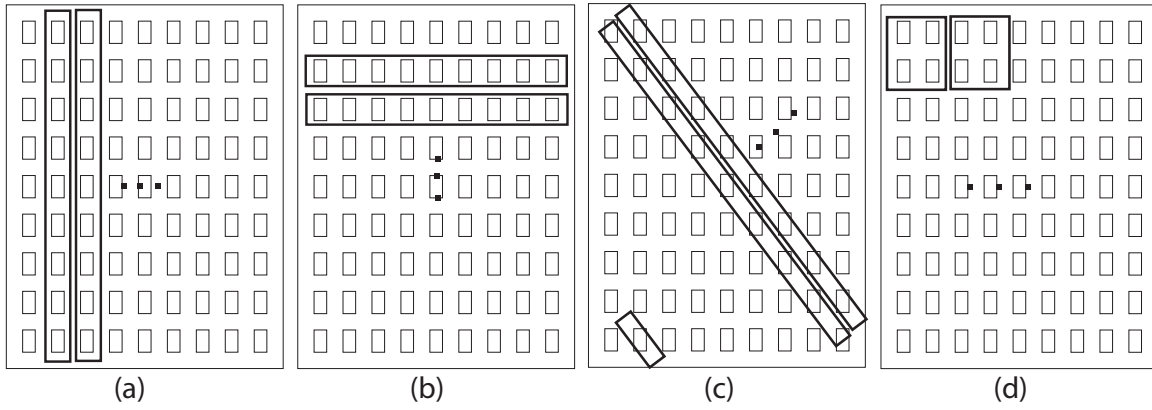


Fig. 1. Different grouping strategies of existing 2DPCA methods. (a) 2DPCA; (b) A2DPCA; (c) DiaPCA; (d) MatPCA.

and the average of the training images is denoted by \bar{A} . Then, the covariance matrix of the images G_t can be computed as:

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

$$\bar{A} = \frac{1}{M} \sum_{i=1}^M A_i.$$

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, \dots, x_d . The optimal projection axes are the orthonormal eigenvectors corresponding to the first d largest eigenvalues of G_t . Then the set of feature vectors, which represent the image matrix, is obtained by the following linear transformation:

$$y_k = A_i x_k \quad (2)$$

where $k = 1, \dots, d$. With this transformation, each image of a person in the training set is represented by d feature vectors. The following section describes the proposed algorithm in detail.

3. An efficient algorithm based on cross variations between row and column vectors

In this study a novel algorithm for face recognition is introduced. In this proposed algorithm variations between the row and column vectors of face images are used for face recognition. The proposed algorithm is also based on the traditional 2DPCA but it differs from the 2DPCA and its variants in grouping strategies.

3.1. Comparison of the proposed method with 2DPCA and its variants

In statistics and probability theory, the concept of “covariance” of a random vector X is used as the matrix of covariances between the scalar components of X . But the term cross-covariance is sometimes used to refer to the covariance between two random vectors (assumed to be centered), $X = [x_1 x_2 \dots x_m]^T$ and $Y = [y_1 y_2 \dots y_n]^T$. So the cross-covariance matrix Σ_{XY} between X and Y [22] is the $m \times n$ dimensional matrix and is defined as follows:

$$\Sigma_{XY} = E[XY^T]. \quad (3)$$

The scope of this study is to use the variations among the column and row vectors of image matrices for face recognition. That is why the two random vectors X and Y in the above definition correspond to the column vector and to the row vector of the

image matrix respectively. The elements of $E[XY^T]$ are the cross-covariances between the components x_i and y_j of the vectors X and Y .

The proposed method and many others given in the introduction section are based on 2DPCA. Most of them try to group the elements of the image matrix in some manner. They aim to reduce the redundancy which exists in these groups. Examples of most common grouping strategies are analyzed in [18]. As they mentioned in their study, grouping strategy exactly specifies the form of covariance matrix, therefore it is very essential. Different grouping strategies of most of the existing methods, based on 2DPCA, are illustrated in Fig. 1. In these figures, each pane corresponds to a pixel (a random variable) in the image and the panes in a round rectangle constitute a group. As it can be seen from Fig. 1, most common groupings are given as column based grouping (Fig. 1a), row based grouping (Fig. 1b), diagonal based grouping (Fig. 1c), and block based grouping (Fig. 1d).

In 2DPCA, each column vector of the image matrix is supposed to be a random vector so image groups consisted of the column vectors. In A2DPCA [19] each row image, in DiaPCA [20] a diagonal of the image matrix and in MatPCA [21] partitioned blocks are supposed to be random vectors, which form the image groups.

In contrast to traditional PCA, in which any two extracted random variables are uncorrelated, there still exist two kinds of correlation on the feature space [18]. These two correlations are defined as inter-group and intra-group correlations. While inter-group correlation refers to the correlation between the random variables belonging to different selected objects, for example the row vectors of different images in 2DPCA, intra-group correlation refers to the correlation between the random variables belonging to the same selected object for example the row vectors of the same image in 2DPCA. As for normalized face images, neighboring pixels are highly correlated, so the 2DPCA method and its variants given in Fig. 1 suffer from the relatively high intra-group correlation because rows, columns, diagonals or blocks in an image matrix are highly correlated.

In this study a new grouping strategy, where the intra-group correlation is relatively weak, is proposed. In the proposed strategy, a group consists of a column and a row vector as given in Fig. 2, where the panes in a round with straight line constitute a group and the panes in a round with dotted line constitute another group. The intra-group correlation is weak according to the most existing grouping strategies because these random variables spread over the whole face image. This makes the proposed algorithm superior to 2DPCA and its variants mentioned above. This can also be seen from the experimental studies (Figs. 6–8). The proposed method achieves higher recognition accuracies compared to 2DPCA.

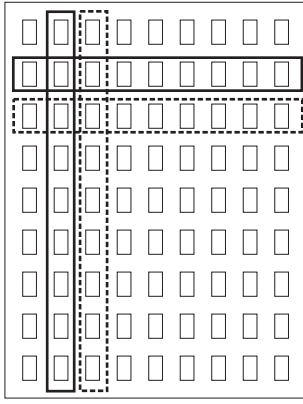


Fig. 2. The proposed cross grouping strategy.

The proposed method is performed in two steps. First, image cross-covariance matrix is defined from the summation of the outer products of the column and the row vectors of the image matrix. This new matrix represents the variations between row and column vectors of the image matrix and it has the same dimension with the original image matrix. In the second step, singular value decomposition is applied to the image cross-covariance matrix. The left and right singular vectors are computed and used as the optimal projection vectors. The feature extraction and image recognition are then achieved.

3.2. Obtaining and diagonalizing the image cross-covariance matrix

Suppose that an $m \times n$ dimensional image matrix is denoted by A and M images are used in the training set. The mean image \bar{A} can be computed as follows:

$$\bar{A} = \frac{1}{M} \sum_{i=1}^M A_i, \quad i = 1, 2, \dots, M. \quad (4)$$

Before the construction of the image cross-covariance matrix, the mean image is subtracted from each image in the training set, resulting in mean subtracted image matrix $\bar{\bar{A}}_i$,

$$\bar{\bar{A}}_i = A_i - \bar{A} = \begin{bmatrix} \bar{\bar{a}}_{11} & \bar{\bar{a}}_{12} & \dots & \bar{\bar{a}}_{1n} \\ \bar{\bar{a}}_{21} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{\bar{a}}_{m1} & \dots & \dots & \bar{\bar{a}}_{mn} \end{bmatrix} = \begin{bmatrix} r_1^i \\ r_2^i \\ \dots \\ r_m^i \end{bmatrix} = [c_1^i \quad c_2^i \quad \dots \quad c_n^i] \quad (5)$$

where $r_j^i, (j = 1, 2, \dots, m)$ denotes the row vectors of $\bar{\bar{A}}_i$, and $c_k^i, k = 1, 2, \dots, n$ denotes the column vectors of $\bar{\bar{A}}_i$.

The image cross-covariance matrix (C_{rc}) with the size of $m \times n$ can be computed from the summation of the outer products of column and row vectors of the image matrices as follows:

$$C_{rc} = \frac{1}{M} \sum_{i=1}^M \frac{1}{mn} \sum_{j=1}^m \sum_{k=1}^n (c_k^i) \otimes (r_j^i) = \frac{1}{M} \sum_{i=1}^M (r_{avg}^i) \otimes (c_{avg}^i)^T \quad (6)$$

where \otimes denotes the outer product operation and $r_{avg}^i = [\bar{r}_1^i \bar{r}_2^i \dots \bar{r}_m^i]^T$ and $c_{avg}^i = [\bar{c}_1^i \bar{c}_2^i \dots \bar{c}_n^i]^T$ are the row and column mean vectors with \bar{r}_j^i and \bar{c}_k^i which are the j th row and k th column means

respectively. The elements of the row and the column mean vectors are given below,

$$\begin{aligned} \bar{r}_j^i &= \frac{1}{n} \sum_{k=1}^n (\bar{a}_{jk}^i); \quad j = 1, \dots, m \\ \bar{c}_k^i &= \frac{1}{m} \sum_{j=1}^m (\bar{a}_{jk}^i); \quad k = 1, \dots, n. \end{aligned} \quad (7)$$

Diagonalization of the cross-covariance matrix given in Eq. (6) can be achieved by using SVD [23]. SVD generalizes the eigendecomposition to rectangular matrices. With this operation a rectangular matrix is decomposed into three simple matrices: Two orthogonal matrices and one diagonal matrix. SVD uses the eigendecomposition of a semi-definite matrix to derive a similar decomposition for a rectangular matrix. Let X and Y be $n \times p$ and $n \times q$ data matrices and C_{xy} be the $p \times q$ matrix, whose elements are the cross-covariances between X and Y . The goal of SVD is to find linear combinations of the data Xa_i and Yb_i [$i = 1, \dots, r; r = \min(p, q)$], with the maximum covariance, subject to the $p \times 1$ vectors a_i and $q \times 1$ vectors b_i , satisfying the orthogonality constraints

$$a_i^T a_j = b_i^T b_j = \begin{cases} 1 & i = j \\ 0 & i \neq j. \end{cases} \quad (8)$$

The solution is found by taking the SVD of C_{xy} , denoted here by $C_{xy} = ADB^T$, where A is a $p \times r$ semiorthogonal matrix, D is an $r \times r$ diagonal matrix, and B is a $q \times r$ semiorthogonal matrix. The i th columns of A and B contain the left and right singular vectors a_i and b_i , and the i th singular value of D is the covariance of Xa_i and Yb_i . The vectors Xa_i and Yb_i are referred to as the i th pair of SVD expansion coefficients. It is desirable to find the directions in p space and q space so that when the variables in the two sets of data are projected onto these axes, they are as similar as possible. The next step is to find a second set of directions, orthogonal to the first, with similar properties, and so on up to r such pairs of directions. Stated another way, a $p \times r$ matrix A and $q \times r$ matrix B are sought, such that the quantity

$$\|XA - YB\|^2 \quad (9)$$

is minimized subject to A and B being semiorthogonal. The notation $\|\cdot\|$ refers to the Frobenius norm of a matrix.

A solution to the problem is to let A and B be the matrices of left and right singular matrices of SVD of C_{xy} .

SVD is applied to the image cross-covariance matrix C_{rc} in Eq. (6) whose elements are the cross-covariances between the image row and column mean vectors. Then the right singular matrix ($C_{rc} C_{rc}^T$) with the size of $m \times m$ and the left singular matrix ($C_{rc}^T C_{rc}$) with the size of $n \times n$ of the SVD of the image cross-covariance matrix are obtained. $n \times 1$ dimensional eigenvectors (left singular vectors), $\{\varphi_s\}$ of $C_{rc}^T C_{rc}$ matrix and $m \times 1$ dimensional eigenvectors (right singular vectors) $\{\psi_s\}$ of $C_{rc} C_{rc}^T$, that is,

$$\begin{aligned} C_{rc}^T C_{rc} \varphi_s &= \lambda_s \varphi_s, \quad s = 1, \dots, R \\ C_{rc} C_{rc}^T \psi_s &= \lambda_s \psi_s, \quad s = 1, \dots, R \end{aligned} \quad (10)$$

are found. Here X in Eq. (9) corresponds to the image row mean vector while Y is corresponding to the image column mean vector which are given in Eq. (6). And also the left and the right singular matrices A and B in Eq. (9) correspond to the left singular matrix

$(C_{rc}^T C_{rc})$ and the right singular matrix $(C_{rc} C_{rc}^T)$. The algorithm then uses the right and the left singular vectors for feature extraction.

3.3. Covariance matrix relationships between PCA and 2DPCA based methods (2DPCA, 2DPCA variants and proposed method)

It is stated in [24] that 2DPCA can work in either row or column direction of the images. The covariance matrix given for 2DPCA in Eq. (1) exactly represents the row covariances that each row is considered as an object in \mathfrak{R}^n .

Using the row identification in Eq. (5), the image covariance matrix in Eq. (1) can be rewritten as

$$G_t = \frac{1}{M} \sum_{i=1}^M (\bar{A}^i)^T (\bar{A}^i) = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^m (r_j^i)^T (r_j^i) = G_{11} + G_{22} + \dots + G_{mm} \tag{11}$$

where G_{jk} is the covariance of the j th and k th image row vectors

$$G_{jk} = \frac{1}{M} \sum_{i=1}^M (r_j^i)^T (r_k^i). \tag{12}$$

In the light of this definition the covariance matrix in traditional PCA can be rewritten as

$$S = E[XX^T] = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1m} \\ S_{21} & S_{22} & \dots & S_{2m} \\ \dots & \dots & \dots & \dots \\ S_{m1} & S_{m2} & \dots & S_{mm} \end{bmatrix}_{m \times m} \tag{13}$$

where X is the $mn \times 1$ dimensional image vector which is obtained by concatenating the image rows into a vector form and S_{ij} is an $n \times n$ dimensional cross-covariance matrix between the row vectors r_j and r_k for $i \neq j$ and $G_{ii} = S_{ii}$, $i = 1, \dots, m$. Each row of the covariance matrix in Eq. (13) includes the cross-covariances between the corresponding row and all other rows and also the covariance of that row.

From Eqs. (11) and (13), it can easily be seen that the covariance matrix of 2DPCA preserves the elements only in the main diagonal of the covariance matrix of PCA [25]. It is also stated in [25] that the covariance matrix of 2DPCA is equivalent to the average of the main diagonal of the covariance matrix of PCA. This implies that 2DPCA eliminates the covariance information other than the main diagonal that can be useful for recognition.

The relationships between the covariance matrices of the 2DPCA variants given in [18] with PCA can be defined in the similar manner as in 2DPCA case. While 2DPCA is using a row image matrix as an object, its variants use a column, a diagonal or a block of the image matrix as an object. 2DPCA and its variants differ from each other according to the selected object, but their feature extraction schemes are similar. Thus it is considered that covariance matrices of 2DPCA variants preserve only the covariance information between the elements of the selected objects. For example if we take the column vectors as an object, the image covariance matrix in 2DPCA can be written as;

$$G_t = \frac{1}{M} \sum_{i=1}^M (\bar{A}^i)(\bar{A}^i)^T = \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^n (c_j^i)(c_j^i)^T = G_{11} + G_{22} + \dots + G_{nn} \tag{14}$$

where G_{jk} is the covariance of the j th and k th image column vectors

$$G_{jk} = \frac{1}{M} \sum_{i=1}^M (c_j^i)(c_k^i)^T. \tag{15}$$

The relationship between traditional PCA and 2DPCA which works on column direction can be obtained in a similar manner. Thus the covariance matrix of traditional PCA can be rewritten as

$$S = E[XX^T] = \begin{bmatrix} S_{11} & S_{12} & \dots & S_{1n} \\ S_{21} & S_{22} & \dots & S_{2n} \\ \dots & \dots & \dots & \dots \\ S_{n1} & S_{n2} & \dots & S_{nn} \end{bmatrix}_{m \times m} \tag{16}$$

where X is the $mn \times 1$ dimensional image vector which is obtained by concatenating the image columns into a vector form and S_{ij} is an $m \times m$ dimensional cross-covariance matrix between the column vectors c_j and c_k for $i \neq j$ and $G_{ii} = S_{ii}$, $i = 1, \dots, n$. Each row of the covariance matrix in Eq. (16) includes the cross-covariances between the corresponding column and all other columns and also the covariance of that column.

In this paper, we proposed a framework for investigating the information which may possibly be useful but neglected by 2DPCA and its variants. With this proposed method we use the variations between the image rows and columns. In order to do this we construct the image cross-covariance matrix as in Eq. (6). Then the eigenvectors which diagonalize the image cross-covariance matrix are used as the optimal projection vectors. In our case the image cross-covariance matrix is a rectangular matrix. So SVD is used for this eigendecomposition. With SVD of the image cross-covariance matrix the right and left singular vectors obtained from the right and left singular matrices' eigendecomposition respectively are used as the optimal projection vectors. From Eq. (6), the image cross-covariance matrix is simply formed from the summation of the outer product of the row and column mean vectors with the size of $m \times 1$ and $1 \times n$ respectively.

$$C_{rc} = \frac{1}{M} \sum_{i=1}^M (r_{avg}^i) \otimes (c_{avg}^i)^T = \frac{1}{M} \sum_{i=1}^M \begin{bmatrix} \bar{r}_1^i \\ \bar{r}_2^i \\ \dots \\ \bar{r}_m^i \end{bmatrix} \otimes [\bar{c}_1^i \quad \bar{c}_2^i \quad \dots \quad \bar{c}_n^i] \tag{17}$$

Now we are applying SVD to the image cross-covariance matrix. So $C_{rc} C_{rc}^T$ is the right singular matrix of this decomposition. Recall that the right singular matrix is defined

as,

$$C_{rc}C_{rc}^T = \frac{1}{M} \sum_{i=1}^M \begin{bmatrix} \bar{r}_1^i \bar{c}_1^i & \bar{r}_1^i \bar{c}_2^i & \dots & \bar{r}_1^i \bar{c}_n^i \\ \bar{r}_2^i \bar{c}_1^i & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{r}_m^i \bar{c}_1^i & \dots & \dots & \bar{r}_m^i \bar{c}_n^i \end{bmatrix} \begin{bmatrix} \bar{r}_1^i \bar{c}_1^i & \bar{r}_2^i \bar{c}_1^i & \dots & \bar{r}_m^i \bar{c}_1^i \\ \bar{r}_1^i \bar{c}_2^i & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{r}_1^i \bar{c}_n^i & \dots & \dots & \bar{r}_m^i \bar{c}_n^i \end{bmatrix} \tag{18}$$

$$= \frac{1}{M} \sum_{i=1}^M (c^i) \begin{bmatrix} \bar{r}_1^i \bar{r}_1^i & \bar{r}_1^i \bar{r}_2^i & \dots & \bar{r}_1^i \bar{r}_m^i \\ \bar{r}_1^i \bar{r}_2^i & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{r}_m^i \bar{r}_1^i & \dots & \dots & \bar{r}_m^i \bar{r}_m^i \end{bmatrix} \tag{19}$$

where $(c^i) = (\bar{c}_1^i)^2 + (\bar{c}_2^i)^2 + \dots + (\bar{c}_n^i)^2$ is a coefficient obtained from the summation of squares of the elements of the image column mean vector and $\bar{r}_j^i, j = 1, 2, \dots, m$ are the elements of the row mean vector $r_{avg}^i = [\bar{r}_1^i \bar{r}_2^i \dots \bar{r}_m^i]^T$. As a further step the right singular matrix can be rewritten as

$$C_{rc}C_{rc}^T = \frac{1}{M} \sum_{i=1}^M (c^i) [(r_{avg}^i)(r_{avg}^i)^T] \tag{19}$$

That is why the right singular matrix which is obtained from the SVD of the image cross-covariance matrix is in fact consisting of the covariance of the row mean vector (r_{avg}^i). Substituting the equivalence of (r_{avg}^i) as $[(\bar{r}_1^i) (\bar{r}_2^i) \dots (\bar{r}_m^i)]^T$ and $(\bar{r}_j^i, j = 1, 2, \dots, m)$ as given in Eq. (7) into Eq. (19) the equality of the right singular matrix in Eq. (19) can be rearranged as follows:

$$C_{rc}C_{rc}^T = \frac{1}{M} \sum_{i=1}^M (c^i) (S_{11} + S_{12} + \dots + S_{1n} + S_{21} + \dots + S_{n1} + \dots + S_{nn}) \tag{20}$$

where S_{ij} is the cross-covariance of the i th column vector with j th column vector of the image matrix. The right singular matrix in Eq. (20) includes all the cross-covariances between each column and all other columns and also the covariances of each column as in Eq. (16). Thus the right singular vectors in fact diagonalize the covariance matrix, which is constructed taking each column vector of the image matrix as an object. While doing this it also preserves all the covariance information included in PCA given in Eq. (16) which 2DPCA and its variants do not preserve. As it is proved in [25] the covariance matrix of 2DPCA is equivalent to the average of the main diagonal of the covariance matrix of PCA. Similarly in our case the proposed approach expands the averaging so as to include the covariance information which is given in PCA (Eq. (16)) totally. This can be seen from Eq. (20).

A similar relationship between the covariance matrix of PCA and the left singular matrix is defined as,

$$C_{rc}^T C_{rc} = \frac{1}{M} \sum_{i=1}^M \begin{bmatrix} \bar{r}_1^i \bar{c}_1^i & \bar{r}_2^i \bar{c}_1^i & \dots & \bar{r}_m^i \bar{c}_1^i \\ \bar{r}_1^i \bar{c}_2^i & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{r}_1^i \bar{c}_n^i & \dots & \dots & \bar{r}_m^i \bar{c}_n^i \end{bmatrix} \begin{bmatrix} \bar{r}_1^i \bar{c}_1^i & \bar{r}_1^i \bar{c}_2^i & \dots & \bar{r}_1^i \bar{c}_n^i \\ \bar{r}_2^i \bar{c}_1^i & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{r}_m^i \bar{c}_1^i & \dots & \dots & \bar{r}_m^i \bar{c}_n^i \end{bmatrix} \tag{21}$$

$$= \frac{1}{M} \sum_{i=1}^M (r^i) \begin{bmatrix} \bar{c}_1^i \bar{c}_1^i & \bar{c}_1^i \bar{c}_2^i & \dots & \bar{c}_1^i \bar{c}_n^i \\ \bar{c}_1^i \bar{c}_2^i & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \bar{c}_n^i \bar{c}_1^i & \dots & \dots & \bar{c}_n^i \bar{c}_n^i \end{bmatrix} \tag{21}$$

where $(r^i) = ((\bar{r}_1^i)^2 + (\bar{r}_2^i)^2 + \dots + (\bar{r}_m^i)^2)$ is a coefficient obtained from the summation of squares of the elements of the image row mean vector and $(\bar{c}_j^i), j = 1, 2, \dots, m$ are the elements of the column mean vector $c_{avg}^i = [(\bar{c}_1^i) (\bar{c}_2^i) \dots (\bar{c}_n^i)]^T$. As a further step the left singular matrix can be rewritten as

$$C_{rc}^T C_{rc} = \frac{1}{M} \sum_{i=1}^M (r^i) [(c_{avg}^i)(c_{avg}^i)^T] \tag{22}$$

That is why the left singular matrix which is obtained from the SVD of the image cross-covariance matrix is in fact consisting of the covariance of the column mean vector (c_{avg}^i). Substituting the equivalence of (c_{avg}^i) as $[(\bar{c}_1^i) (\bar{c}_2^i) \dots (\bar{c}_n^i)]^T$ and $(\bar{c}_k^i, k = 1, 2, \dots, n)$ as given in Eq. (7) into Eq. (22) the equality of the left singular matrix in Eq. (22) can be rearranged as follows:

$$C_{rc}^T C_{rc} = \frac{1}{M} \sum_{i=1}^M (c^i) (S_{11} + S_{12} + \dots + S_{1m} + S_{21} + \dots + S_{m1} + \dots + S_{mm}) \tag{23}$$

where S_{ij} is the covariance of the i th row vector with j th row vector of the image matrix. The left singular matrix in Eq. (23) includes all the cross-covariances between each row and all other rows and also the covariances of each row as in Eq. (13). Thus the left singular vectors in fact diagonalize the covariance matrix which is constructed taking each row vector of the image matrix as an object. While doing this, it also preserves all the covariance information included in PCA given in Eq. (13) which again 2DPCA and its variants do not preserve. Here the covariance matrix given in Eq. (23) expands the averaging so as to include the covariance information which is given in PCA (Eq. (13)) totally. This can be seen from Eq. (23).

3.4. Feature extraction and image recognition

The eigenvectors of the right and the left singular matrix obtained from SVD of the image cross-covariance matrix are used as the optimal projection vectors for feature extraction. These eigenvectors give the information of the variation between row and column vectors of the image matrix. d right eigenvectors and d left eigenvectors, corresponding to d largest eigenvalues of the image cross-covariance matrix, are used for the feature extraction ($d < R$). For a given image A , feature vectors are calculated from the following two linear transformation:

$$Y_k^1 = A\varphi_k \tag{24}$$

$$Y_k^2 = A^T \psi_k$$

where Y_k^1 and Y_k^2 , $k = 1, \dots, d$ are the left and the right feature vectors of image matrix A , respectively. In the classification procedure, a nearest neighbor classifier [26] is used. The distance between the arbitrary feature sets of the classes and the test image, D^c , is calculated from the following formula,

$$D^c = \sum_{k=1}^d \|Y_k^{1c} - Y_k^{1test}\| + \sum_{k=1}^d \|Y_k^{2c} - Y_k^{2test}\| \quad (25)$$

where $c = 1, \dots, C$ and C is the class number in the training set. D^c is minimized if the test image belongs to the class C .

An approximation of a face image matrix (\tilde{A}) can be obtained using d^2 eigenimages as discussed in [27]. An eigenimage, E_{ij} , is calculated from the outer product of the right $\{\psi_s\}$ and the left $\{\varphi_s\}$ singular vector which belongs to the right singular matrix, $C_{rc}C_{rc}^T$, and the left singular matrix, $C_{rc}^T C_{rc}$ respectively.

$$\begin{aligned} \tilde{A} &= \sum_{i=1}^d \sum_{j=1}^d D_{ij} E_{ij} \\ E_{ij} &= \Psi_i \Phi_j^T \\ D_{ij} &= \langle A, E_{ij} \rangle \end{aligned} \quad (26)$$

where $1 \leq i \leq d$ and $1 \leq j \leq d$. Here D_{ij} is the coefficient obtained from the inner product of image A with the eigenimage E_{ij} . Summation of the products of these coefficients with corresponding eigenimages gives the approximation of image A , if d numbers of right and left eigenvectors are used. In order to reconstruct the original face image matrix, all of the right and the left eigenvectors $i = 1, \dots, m$ and $j = 1, \dots, n$ need to be used.

3.5. Proposed method plus LDA

If we use the left and the right singular vectors to perform feature extraction in bi-directional way as stated in [12], we can use the transformation given below to extract the total feature matrix Y of the image matrix A :

$$Y = \psi_k A \varphi_k \quad (27)$$

The feature matrix Y is then transformed into feature vector y by concatenating the columns of Y . The LDA projector W_{LDA} is calculated then to obtain the final feature vector z

$$z = W_{LDA}^T Y. \quad (28)$$

4. Experimental studies

In the experimental study, the AR-Face database [28], ORL face database [29] and FERET face database [30] are used.

4.1. Experiments on AR-Face database

AR-Face database contains over 4000 face images of 126 people (70 men and 56 women), including frontal views of face with different facial expressions, lighting conditions and occlusions. The pictures of people were taken in two different sessions separated by two weeks time interval. Each session contains 13 images for each person. In this study, only 37 individuals (20 males and 17 females) were selected and used. Only the non-occluded images were used for each individual. The face portion of the images were cropped and then normalized to 50×40 pixels. The normalized images for one person are shown in Fig. 3.

Four different databases are constructed from AR-Face in order to use in experimental study and four different experiments are performed to evaluate the performance of the proposed method. In the first experimental study, the first image from the first row of

Table 1

The comparison of top-one recognition accuracy of proposed method with PCA and 2DPCA for the first and second experiments on AR-Face database.

	1. Results (%)	2. Results (%)
PCA	84.68 (24)	77.47 (26)
2DPCA	87.39 (4)	79.72 (5)
A2DPCA	73.87 (17)	75.68 (19)
DiaPCA	72.07 (25)	73.42 (25)
MatPCA	74.77 (19)	76.13 (20)
GridPCA	81.98 (16)	74.77 (20)
Proposed method	91.89 (2)	81.53 (2)

Table 2

The comparison of top-one recognition accuracy of proposed method with PCA and 2DPCA for the third and the fourth experiments on AR-Face database.

	3. Results (%)	4. Results (%)
PCA	79.28 (61)	57.91 (32)
2DPCA	58.55 (39)	52.90 (16)
A2DPCA	65.77 (42)	54.05 (25)
DiaPCA	61.26 (60)	49.42 (30)
MatPCA	67.12 (45)	54.05 (23)
GridPCA	77.48 (40)	53.28 (20)
Proposed method	81.98 (31)	61 (13)

Fig. 3 from AR-Face database is used for each person as the training set. The remaining images from the first row of Fig. 3 are used for the test set. In the second experimental study, the performance of the proposed method is evaluated when there are facial expression variations between the images. In this experiment, the first images from each row of Fig. 3 (normal expressions from both session) are used for the training set for each person and the second, the third and the fourth images from each row of Fig. 3 are used for the test set. In the third experimental study, the performance of the proposed method is evaluated when there are variations in lighting conditions. In this experiment, again, the first images from each row of Fig. 3 are used for the training set for each person and the last three images of each row of Fig. 3 are used for the test set. In the fourth experimental study, the performance of the proposed method is evaluated when there are time differences between the training set images and the test set images. That is why images from the first row of Fig. 3 are used as the training set and the images from the second row of Fig. 3 are used as the test set for each person. The same experiments are performed using both PCA and 2DPCA methods in order to compare results with that of the new method. Results are summarized in Tables 1 and 2.

Table 1 presents the top-one recognition accuracy [8], which belongs to first two experiments, and Table 2 presents third and the fourth experimental results. Numbers in brackets are the eigenvectors' number, for which top-one recognition accuracy is obtained. According to the experimental results, using cross variations between the row and column vectors of the image matrix can remove more redundancy than using the variations between the row vectors of the image matrix for face recognition. Also better recognition accuracy can be achieved by using less number of eigenvectors for the proposed method than 2DPCA and PCA. First column of Table 1, for example, reveals that the recognition accuracy is increased 4.5 points in respect of 2DPCA and 7.21 points in respect of PCA.

In each of these four experimental studies, the proposed method outperforms PCA and 2DPCA in context of both in recognition accuracy and the number of eigenvectors used. The maximum recognition accuracy difference between 2DPCA and the proposed method is achieved when there are lighting condition changes between the images. So, the proposed method gives better recognition accuracy results compared to 2DPCA when there are lighting condition changes between images.



Fig. 3. Sample images of a person from AR-Face database.



Fig. 4. Five sample images of a person from ORL database.

Table 3
The comparison of top-one recognition accuracy of proposed method with 2DPCA for the experiments on ORL database.

	1. Results (%)	2. Results (%)	3. Results (%)
2DPCA	77.77 (2)	88.5 (5)	95 (10)
A2DPCA	71.39 (25)	80.63 (27)	94.25 (28)
DiaPCA	64.44 (33)	75.63 (32)	82.5 (33)
MatPCA	71.94 (13)	81.25 (13)	94.5 (13)
GridPCA	64.17 (22)	80.94 (28)	94.25 (24)
Proposed method	78.33 (4)	92 (21)	95.5 (8)

4.2. Experiments on ORL database

This database contains 40 individuals each containing 10 different images. They were taken in two years' time. The images were taken at different lighting conditions, facial expressions and facial details. The background of the images is dark homogeneous. The size of each image is 112×92 pixels, with 256 gray levels per pixel. The images were taken with a tolerance for some tilting and rotation of the face of up to 20° and with some variation in the scale of up to about 10%. Five image samples from ORL database are shown in Fig. 4.

Three different experiments are performed using this database. In the first experiment, only the one image per person is used as the training set and all the remaining images per person are used as the test set. In the second experiment, first two images per person are used as the training set and all the remaining images are used as the test set. Finally, in the third experiment, leave-one-out strategy is applied to the database. The performance of the proposed method in each of these three experiments is compared with 2DPCA method. Top-one recognition accuracy is given for the experiments, Table 3, and numbers in brackets are the eigenvector numbers, for which the top-one recognition accuracy is obtained.

According to the experimental results, using more image in the training set increases the recognition accuracy. In the second experiment, for example, when first two images are used as the training set, the recognition accuracy is increased 3.5 points compared to 2DPCA.

Table 4
The comparison of top-one recognition accuracy of proposed method with 2DPCA and its variants for the experiments on FERET database.

	Results (%)
2DPCA	79.79
A2DPCA	78.87
DiaPCA	79.27
MatPCA	74.54
GridPCA	79.27
Proposed method	92.26

4.3. Experiments on FERET database

The FERET face database is one of the standard databases which is used in testing and evaluating face recognition algorithms. The FERET database contains more than 10,000 images of over 1000 individuals. Within FERET database, the standardized FA contains 994 images, and FB contains 992 images, image sets are used in this work. Example normalized images' face images in the FERET database are given in Fig. 5.

The proposed method was tested on a subset of the FERET database. This subset includes 762 images of 127 individuals (six images per person). This subset involves frontal views with variations in facial expression, illuminations, facial details, and time differences. In our experiments, the facial portion of each original image was cropped based on the localization of eyes to a size of 60×60 . The cosine similarity measure is used to calculate the similarity score between any pair of query and target images. Two experimental studies are performed using this database. In the first experiment leave-one-out strategy is used. As the purpose of this experiment is to compare the recognition accuracy of the proposed method with 2DPCA and its variants, the experimental study is performed for also 2DPCA and its variants which are A2DPCA, DiaPCA, MatPCA and GridPCA. The maximum recognition rates for all of these methods are given in Table 4. Fig. 6 shows the cumulative eigenvalue curves for proposed method, 2DPCA, A2DPCA, DiaPCA, MatPCA and GridPCA against the preserved feature dimension. It is clear that with the same dimension proposed method keeps far more energy than all 2DPCA methods. Our proposed method is remarkably better than the others because its grouping strategy implies lower intra-group correlations. The performance statistics is cumulative match score and it is given in Fig. 7 for all these methods. In the second experiment, first three images are used as the training set and remaining images as the test set. The performance statistics of the second experiment is cumulative match score and it is given in Fig. 8 for all these methods.

We also use the left and the right singular vectors to perform feature extraction in bi-directional way as stated in [12]. Applying LDA after the proposed method for the leave-one-out strategy, the recognition rate, which is (%)92.26 as given in Table 4, is increased



Fig. 5. Six sample images from FERET database.

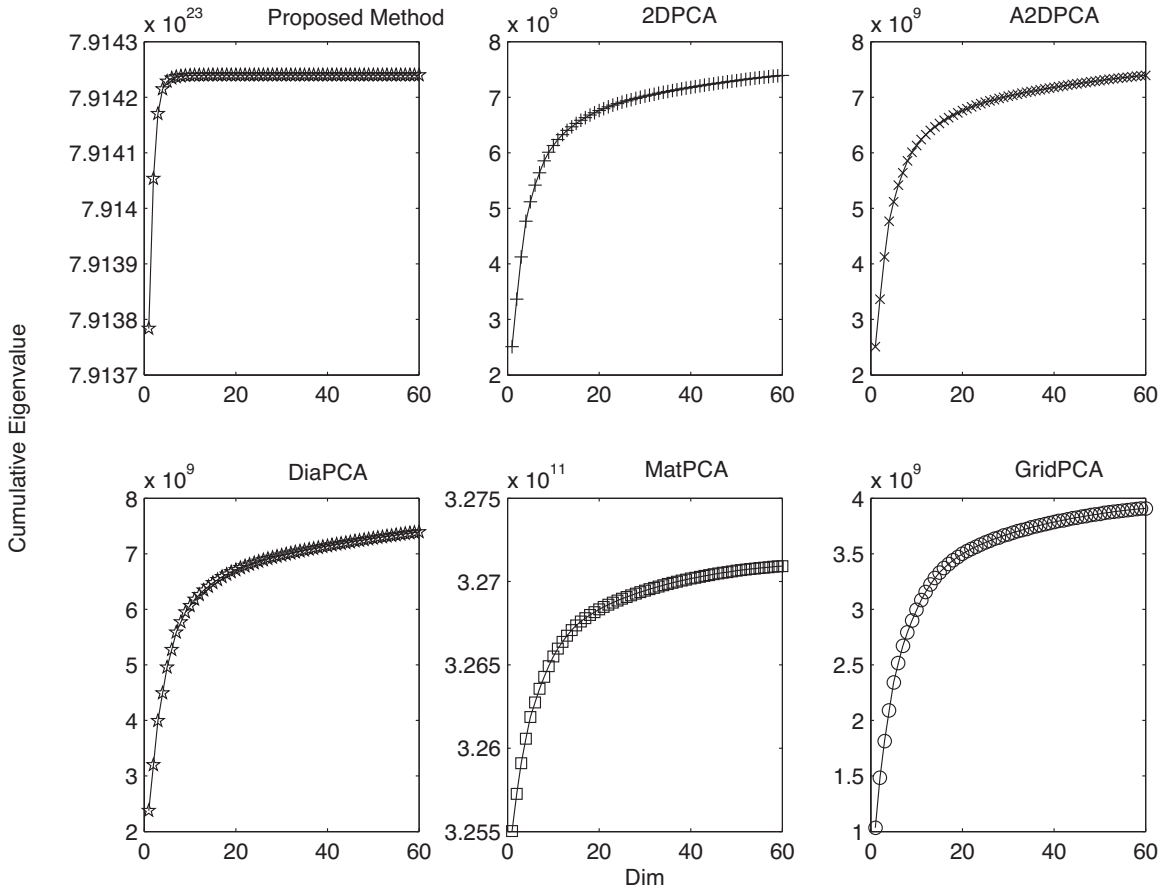


Fig. 6. The cumulative eigenvalue curves for proposed method, 2DPCA, A2DPCA, DiaPCA, MatPCA and GridPCA against the preserved feature dimension.

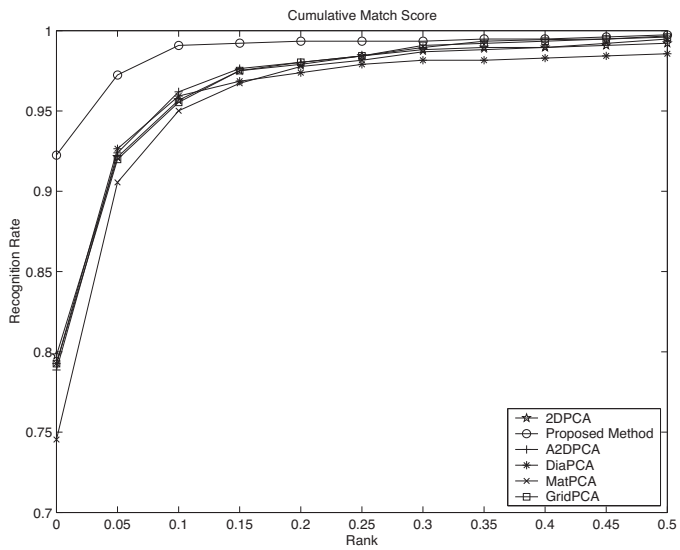


Fig. 7. The cumulative match scores results for proposed method, 2DPCA, A2DPCA, DiaPCA, MatPCA and GridPCA.

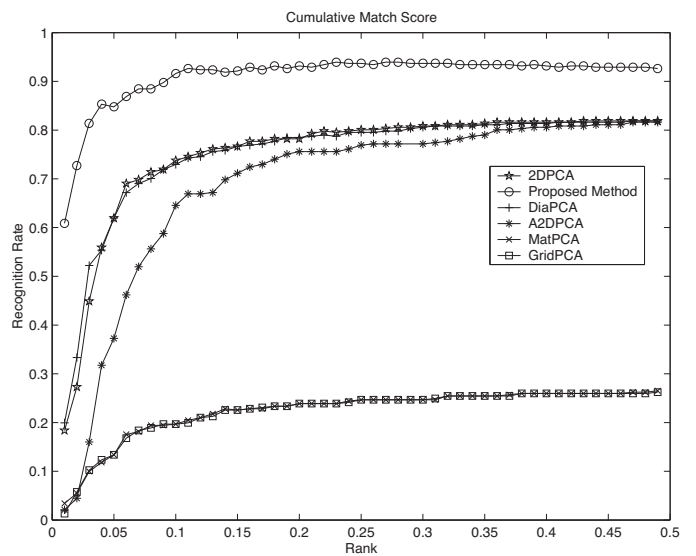


Fig. 8. The cumulative match scores results using proposed method, 2DPCA, A2DPCA, DiaPCA, MatPCA and GridPCA for the second experiment.

to (%)96.59. Another experiment for proposed method + LDA is performed on FERET database. In this experiment randomly selected three images are used as the training set and remaining three images are used as the test set. To reduce the randomness of the experimental results, the average results of 10 trials are reported. For the comparison purposes the experiment is also performed for BDPCA. For the proposed method+LDA the recognition rate is achieved as (%)82.62, for BDPCA+LDA the recognition rate is achieved as (%)67.14 and for PCA+LDA the recognition rate is achieved as (%)64.18.

5. Conclusions

In this work, a new face recognition algorithm, which reduces the redundancy among row and column vectors of the face image matrix, is presented. In order to do this, a novel grouping strategy is proposed, which reduces the intra-group correlations and results in an increase on the recognition performances. Another important point, which contributes to the improvement of the recognition performances, is that the proposed method preserves all of the correlation information of pixel pairs for estimating the image covariance matrix as in PCA. In the proposed algorithm, image cross-covariance matrix is constructed from the summation of the outer products of the column and the row vectors of the image matrix. Singular value decomposition (SVD) is, then, applied to the image cross-covariance matrix. Right and left singular vectors obtained from SVD of the image cross-covariance matrix are used as the optimal projection vectors.

Experimental studies show that the proposed method outperforms PCA, 2DPCA and 2DPCA variants in terms of recognition accuracy. The computational cost of the proposed method is superior to 1DPCA, since less time is required in order to find eigenvalues and corresponding eigenvectors. Experiments on the AR-Face, the ORL face and the FERET face database show that the proposed method improves the results and is valid for more general cases. The recognition performances are also evaluated in the cases, where proposed method+LDA is used. This also improves the recognition performances. It should be noted that as there are two feature vector sets obtained in the proposed method for image representation; the storage requirement in the proposed method is more than what is required for the PCA and 2DPCA methods. In the future work, it is planned to address two points. The first one is related about the recognition performance improvements. In order to increase the recognition performance of the proposed method combination of this method with improved LDA based methods given in [31,32] will be investigated. Because as it is stated in [31,32] proposed improved LDA methods are said to be more robust to non-ideal conditions such as variations of expression, illumination, pose and so on. And the second one is related about the computational cost and time requirements. In this sense QR decomposition implementation is considered instead of SVD [33–35] to reduce the computational complexity and the time requirements.

References

- [1] M. Turk, A. Petland, Eigenfaces for recognition, *J. Cogn. Neurosci.* 3 (1991) 72–86.
- [2] P. Belhumeur, J. Hespanha, D.J. Kriegman, Discriminant eigenfeatures for image retrieval, *PAMI* 19 (7) (1997) 711–720.
- [3] W. Zhao, R. Chellappa, A. Krishnaswamy, Discriminant analysis of principle components for face recognition, in: *FG, 3rd International Conference on Face & Gesture Recognition*, 1998, p. 336.
- [4] A. Hyvarinen, Survey on independent component analysis, *Neural Comput. Surv.* 2 (1999) 94–128.
- [5] B. Schelkopf, A. Smola, K. Müller, Nonlinear component analysis as a kernel eigenvalue problem, *Neural Comput.* 10 (1998) 1299–1319.
- [6] S. Mika, G. Rotsch, J. Wetson, B. Schölkopf, K.-R. Müller, Fisher discriminant analysis with kernels, in: Y.-H. Hu, J. Larsen, E. Wilson, S. Douglas (Eds.), *Neural Networks for Signal Processing IX*, IEEE, 1999, pp. 41–48.
- [7] J. Yang, Z. David, A.F. Frangi, J.-Y. Yang, Two dimensional PCA: a new approach to appearance-based face representation and recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (1) (2004).
- [8] D. Xu, S. Yan, L. Zhang, M. Li, W. Ma, Z. Liu, H. Zhang, Parallel image matrix compression for face recognition, in: *11th International Multimedia Modelling Conference (MMM'05)*, 2005, pp. 232–238.
- [9] L. Wang, X. Wang, X. Zhang, J. Feng, The equivalence of two-dimensional PCA to line-based PCA, *Pattern Recognit. Lett.* 6 (1) (2005) 57–60.
- [10] D. Xu, S. Yan, L. Zhang, M. Li, W. Ma, Z. Liu, H. Zhang, Coupled Subspace Analysis, *MSR-TR 106*, 2004.
- [11] D. Xu, S. Yan, L. Zhang, H.-J. Zhang, Z. Liu, H.-Y. Shum, Concurrent subspaces analysis, in: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005.
- [12] W. Zuo, D. Zhang, J. Yang, K. Wang, BDPCA plus LDA: a novel fast feature extraction technique for face recognition, *IEEE Trans. Syst. Man Cybern.* 36 (4) (2006) 946–953.
- [13] S. Noushath, G.H. Kumar, P. Shivakumara, A new hybrid approach for efficient face representation and recognition based on 2DPCA and IMLDA techniques, in: *Proceedings of the International Conference on Cognition and Recognition*, 2005.
- [14] J. Ye, Generalized low rank approximation of matrices, in: *The Twenty-First International Conference on Machine Learning (ICML)*, 2004, pp. 887–894.
- [15] J. Liu, S. Chen, Non-iterative low rank approximation of matrices, *Pattern Recognit. Lett.* (2006) 1002–1008.
- [16] J. Yang, C. Liu, Horizontal and vertical DPCA-based discriminant analysis for face verification on a large-scale database, *IEEE Trans. Inf. Forensics Secur.* 2 (4) (2007) 781–792.
- [17] R. Antonia, T. Frances, P2CA: a new recognition scheme combining 2D and 3D information, in: *IEEE International Conference on Image Processing*, Geneva, Italy, September 11–14, 2005.
- [18] S. Shan, B. Cao, Y. Su, L. Qing, X. Chen, W. Gao, Unified principal component analysis with generalized covariance matrix for face recognition, in: *IEEE Conference on Computer Vision on Pattern Recognition. CVPR 2008*, 2008.
- [19] D. Zhang, Z.-H. Zhou, Two-directional two-dimensional PCA for efficient face recognition and recognition, *Neurocomputing* 69 (2005) 224–231.
- [20] D. Zhang, Z.-H. Zhou, S. Chen, Diagonal principal component analysis for face recognition, *Pattern Recognit.* 39 (2006) 140–142.
- [21] S. Chen, Y. Zhu, D. Zhang, J.-Y. Yang, Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA, *Pattern Recognit. Lett.* 26 (2005) 1157–1167.
- [22] <http://en.wikipedia.org/wiki/Cross-covariance>
- [23] S. Cherry, Some comments on singular value decomposition analysis, *J. Clim.* 10 (1997) 1759–1761.
- [24] L. Wang, X. Wang, J. Feng, On image matrix based feature extraction algorithms, *IEEE Trans. Syst. Man Cybern. B: Cybern.* 36 (2006) 194.
- [25] M. Safayani, M.T. Manzuri-Shalmani, M. Khademi, Extended two-dimensional PCA for efficient face representation and recognition, in: *Proceedings of CoRR*, 2010.
- [26] R.O. Duda, P.E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
- [27] Z. Daoqiang, C. Songcan, L. Jun, Representing image matrices: eigenimages versus eigenvectors, in: *Proceedings of the 2nd International Symposium on Neural Networks, ISNN'05*, 2005.
- [28] A.M. Martinez, R. Benavante, The AR Face Database, *CVC Tech. Report # 24*, 1998.
- [29] ORL Face Recognition Database. <http://www.cam-orl.co.uk>
- [30] P.J. Philips, H. Moon, S.A. Rizvi, P.J. Rauss, The FERET evaluation methodology for face-recognition algorithms, *IEEE Trans. Pattern Anal. Mach. Intell.* 22 (10) (2000) 1090–1104.
- [31] J. Gao, L.Y. Fan, L. Xu, Median null (Sw)-based method for face feature recognition, *Appl. Math. Comput.* 219 (12) (2013) 6410–6419.
- [32] X. Li, S. Fei, T. Zhang, Median MSD-based method for face recognition, *Neurocomputing* 72 (16) (2009) 3930–3934.
- [33] M. Koç, A. Barkana, A new solution to one sample problem in face recognition using FLDA, *Appl. Math. Comput.* 217 (24) (2011) 10368–10376.
- [34] J. Gao, L. Fan, L. Xu, Solving the face recognition problem using QR factorization, *WSEAS Trans. Math.* 11 (8) (2012).
- [35] J.Q. Gao, L.Y. Fan, L. Li, L.Z. Xu, A practical application of kernel-based fuzzy discriminant analysis, *Int. J. Appl. Math. Comput. Sci.* 23 (4) (2013) 887–903.