Neurocomputing 72 (2009) 1111–1120

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Two-dimensional subspace classifiers for face recognition

Hakan Cevikalp [a], Hasan Serhan Yavuz [a,*], Mehmet Atif Cay [a], Atalay Barkana [b]

[a] Department of Electrical and Electronics Engineering, Eskisehir Osmangazi University, 26480 Eskisehir, Turkey
[b] Department of Electrical and Electronics Engineering, Anadolu University, Eskisehir, Turkey

A R T I C L E   I N F O

Article history:
Received 15 June 2007
Received in revised form
22 February 2008
Accepted 28 February 2008
Communicated by T. Heskes
Available online 20 March 2008

Keywords:
Face recognition
Subspace classifiers
2D subspace classifiers
Feature extraction
Image tensor

A B S T R A C T

The subspace classifiers are pattern classification methods where linear subspaces are used to represent classes. In order to use the classical subspace classifiers for face recognition tasks, two-dimensional (2D) image matrices must be transformed into one-dimensional (1D) vectors. In this paper, we propose new methods to apply the conventional subspace classifier methods directly to the image matrices. The proposed methods yield easier evaluation of correlation and covariance matrices, which in turn speeds up the training and testing phases of the classification process. Utilizing 2D image matrices also enables us to apply 2D versions of some subspace classifiers to the face recognition tasks, in which the corresponding classical subspace classifiers cannot be used due to high dimensionality. Moreover, the proposed methods are also generalized such that they can be used with the higher order image tensors. We tested the proposed 2D methods on three different face databases. Experimental results show that the performances of the proposed 2D methods are typically better than the performances of classical subspace classifiers in terms of recognition accuracy and real-time efficiency.

1. Introduction

The subspace classifiers are pattern recognition methods, where the primary model for each class is a linear subspace of the Euclidean sample space [20]. The motivation behind the subspace classifiers is that each class has its own set of representative features differing from those of the other classes. Therefore, the most conspicuous features are extracted from each class by using the corresponding training set samples in the hope that those features also carry the most important discriminatory information. Even though this assumption is seldom valid, good recognition rates can be achieved when the dimensionality of the sample space is sufficiently large [17]. In subspace methods, it is assumed that the vector distribution of a class corresponds to a lower-dimensional subspace of the original sample space. The subspaces representing classes are defined in terms of basis vectors that are linear combinations of the sample vectors of each class. Therefore, basis vectors spanning those subspaces must first be computed. Also, determining the dimension of each subspace is a major issue since subspace dimensions have a strong influence on the performance of the subspace classifier. In particular, large subspace dimensions lead to a low recognition performance due to the overlapping regions among classes, whereas small subspace dimensions increase the error rates because of a poor resulting approximation [13,17]. Once the basis vectors spanning those subspaces are computed, a test sample vector from an unknown class is classified based on the lengths of the projections of that sample onto each of the subspaces or, alternatively, on the distances of the test vector from these subspaces.

Watanabe et al. proposed the first subspace method, the Class-Featuring Information Compression (CLAFIC), for pattern classification [28]. This method employs the Principal Component Analysis (PCA) to compute the basis vectors spanning subspace of each class. A variation of the CLAFIC, which is called as the CLAFIC-μ, was proposed in Ref. [17]. In contrast to the CLAFIC method, the CLAFIC-μ method utilizes the class-specific means during classification. Fukunaga and Koontz [11] proposed a new method, which enables us to select the basis vectors in such a way that the projections onto the so-called rival subspaces are minimized. Gulmezoglu et al. [12] proposed the Common Vector (CV) method for classification tasks, where the number of samples in each class is smaller than or equal to the dimensionality of the sample space. It is not always reasonable to assume that the most representative features carry the most discriminatory information during the construction of subspaces. Especially, if the number of outlying samples in the tails of the probability density functions of classes is large, these representative features will no longer yield good separation among the classes. Therefore, the Learning Subspace (LS) method was proposed in Ref. [14], in which the subspaces are iteratively modified in order to diminish the number of misclassifications. However, it turned out that the final computed basis vectors obtained using the LS method are sensitive

* Corresponding author. Tel. +90 222 2393750; fax: +90 222 2393613.
E-mail address: hsyavuz@ogu.edu.tr (H.S. Yavuz).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.02.015

to the presentation order of training set samples. This problem is resolved in Ref. [16] by the introduction of the Averaged Learning Subspace (ALS) method, in which the correction of the subspaces are carried out in a batch fashion. This process increased the statistical stability since the computed basis vectors representing classes are independent of the presentation order of the training set samples. Recently, the kernel based subspace methods [13,25], the Kernel CLAFIC [2], and the Kernel CV [4], have been proposed to extract nonlinear features of classes.

In order to apply the subspace methods to appearance based face recognition problems, 2D image matrices must be transformed into 1D vectors by concatenating rows or columns. The resulting image feature vectors typically lead to a high-dimensional sample space, which in turn forms a suitable environment for application of subspace classifiers. It is because most of the assumptions, upon which the subspace classifiers are founded, hold in high-dimensional sample spaces. However, some of the subspace classifiers cannot be applied in these high-dimensional sample spaces since they require the use of large correlation matrices or orthogonal projection operators explicitly. Fortunately, Yang et al. introduced a new method, coined the 2D-PCA method, which applies the classical PCA method directly to the image matrices [31]. This procedure leads to easier evaluation of covariance matrices since the size of the image covariance matrices using 2D-PCA is much smaller. Additionally, it has been reported that the recognition performance of 2D-PCA is superior to the classical PCA (Eigenface) method [5,31,33]. Similarly, 2D extension of LDA, called 2D-LDA, was introduced in Refs. [18,32]. Based on a similar idea as in 2D-LDA, Kong et al. [15] proposed 2D-FDA method for dimensionality reduction. Xu et al. [29] proposed matrix based marginal Fisher analysis to handle 2D gray level images. Note that all 2D based approaches use image matrices, which are in fact second-order image tensors. More recently, there has been a growing interest in using higher order image tensors other than the second-order image matrices. Yan et al. [30] proposed multilinear discriminant analysis in which they used filtered Gabor images (third-order image tensor) for face recognition. Tao et al. [24] also used high-order image tensors for human gait recognition. Moreover, same authors extended two famous margin based classifiers, Support Vector Machine (SVM) and Minimax Probability Machine, such that they can be used with the high-order image tensors [22,23]. In this paper, motivated by these techniques, we propose new subspace classifier methods, which will be referred to as the 2D-CLAFIC, the 2D-CLAFIC-μ, and the 2D-ALS in order to apply the classical subspace methods directly to the face image matrices.

The remainder of the paper is organized as follows: In Section 2, we first review the classical subspace methods, CLAFIC, CLAFIC-μ, and ALS. Section 3 introduces our proposed 2D methods. In Section 4, we generalize CLAFIC and CLAFIC-μ methods such that they can be used with higher order image tensors. In Section 5, experimental results are given. Finally, we draw our conclusion based on the experimental results in Section 6.

2. Subspace classifiers

When subspace classifiers are used in the context of face recognition, 2D face image matrices must be transformed into 1D vectors if the gray level values are used as feature vectors. However, this process ignores some spatial information among image pixel values which may be useful for discrimination. On the other hand, classical 1D subspace classifiers can be modified such that they directly operate on 2D image matrices while preserving spatial information as in Refs. [5,31]. As a result, the classification performance of classical subspace classifiers can be improved. Moreover, computational efficiency will also increase since correlation and covariance matrices are evaluated easily in 2D approaches. In this paper, we revise 1D subspace classifiers such that they directly operate on 2D image matrices.

Before introduction of our proposed methods, we provide an overview of the CLAFIC, the CLAFIC-μ and the ALS methods in this section.

2.1. The CLAFIC method

Suppose there are C classes denoted by $\omega^{(1)}, \omega^{(2)}, \ldots, \omega^{(C)}$ where the ith class contains N_i samples. Let $x_j^i \in \Re^d$ be a d-dimensional column vector, which denotes the jth sample of the ith class. Let $L_1, L_2, \ldots, L_C$ are the subspaces representing classes. Each subspace is spanned by l_i orthonormal basis vectors $\{w_1^i, \ldots, w_{l_i}^i\}$ in $\Re^d$. The CLAFIC method employs the PCA or the Karhunen–Loeve transform to compute the basis vectors $\{w_1^i, \ldots, w_{l_i}^i\}$ spanning each subspace L_i. The basis vectors are computed through eigen-decomposition of class correlation matrices $R^{(i)} \in \Re^{d\times d}$ defined as

$$R^{(i)} = \frac{1}{N_i}\sum_{j=1}^{N_i} x_j^i (x_j^i)^{\mathrm{T}} = \frac{1}{N_i}\Phi^{(i)}\Phi^{(i)^{\mathrm{T}}}, \quad i = 1, \ldots, C, \tag{1}$$

where $\Phi^{(i)}$ is the matrix whose columns are the sample vectors of the ith class. Note that the mean vectors μ_i of classes are not subtracted. The correlation matrix $R^{(i)}$ is a positive semi-definite matrix, hence all eigenvalues are larger than or equal to 0. The l_i eigenvectors corresponding to the largest eigenvalues of $R^{(i)}$ are chosen as basis vectors for the subspace L_i. The number of basis vectors determines the dimensionality of each subspace. In the CLAFIC method, the number of basis vectors cannot exceed $\min(d, N_i)$ for each class. There are different strategies to choose the subspace dimensions l_i. One way is to set all l_is to be equal to a fixed value l. Then, the optimal value of l can be chosen from the error curves as described in Ref. [17]. The other way employs eigenvalues for choosing the dimensions of subspaces. Let the eigenvalues of $R^{(i)}$ be ordered as

$$\lambda_1^i \geqslant \lambda_2^i \geqslant \cdots \geqslant \lambda_{r_i}^i > 0, \tag{2}$$

where r_i is the rank of the matrix $R^{(i)}$. The dimension of L_i is selected as the value by which the ratio of cumulative sums $\kappa_i = \sum_{j=1}^{l_i}\lambda_j^i / \sum_{j=1}^{r_i}\lambda_j^i, i = 1, \ldots, C$, exceeds a threshold. Typical values of the threshold lie between $0.9 \leqslant \kappa_i \leqslant 1$. After the basis vectors of subspaces are determined, one can proceed with the classification phase. To classify a test sample vector, x_{test}, it is projected onto each subspace separately and it is assigned to the class, in which the projection has the maximum length. This is equivalent to assigning the test sample to the class that achieves the best reconstruction (the smallest reconstruction error) of the test sample through a linear combination of the samples of corresponding class. More formally, for classification of the test vector, we use the following criterion function:

$$g(x_{\text{test}}) = \underset{i=1,\ldots,C}{\arg\max} \left\| P^{(i)} x_{\text{test}} \right\|^2 = \underset{i=1,\ldots,C}{\arg\max} \left\| W^{(i)^{\mathrm{T}}} x_{\text{test}} \right\|^2, \tag{3}$$

where $P^{(i)}$ is the orthogonal projection operator of the ith class, and $W^{(i)}$ represents the transformation matrix whose columns are the basis vectors w_k^i of the ith class. Since the basis vectors are orthonormal, the following relation holds between $P^{(i)}$ and $W^{(i)}$:

$$P^{(i)} = W^{(i)} W^{(i)^{\mathrm{T}}}, \quad i = 1, \ldots, C. \tag{4}$$

2.2. The CLAFIC-μ method

A variation of the CLAFIC, which is called as the CLAFIC-μ method, uses the class-specific means, μ_i, $i = 1, \ldots, C$, in classification [17]. In this approach, each class is modeled as a linear manifold centered at the mean of the corresponding class. Therefore, instead of using class correlation matrices, class covariance matrices are employed to compute the basis vectors. The covariance matrix of each class is given as

$$\Sigma^{(i)} = \frac{1}{N_i}\sum_{j=1}^{N_i}(x_j^i - \mu_i)(x_j^i - \mu_i)^{\mathrm{T}} = \frac{1}{N_i}\bar{\Phi}^{(i)}\bar{\Phi}^{(i)^{\mathrm{T}}}, \quad i = 1,\ldots,C, \tag{5}$$

where $\bar{\Phi}^{(i)}$ is the matrix whose columns are the centered sample vectors of the ith class. Similar to the correlation matrices, the covariance matrices are also positive semi-definite. As in the previous case, the eigenvectors corresponding to the maximum l_i eigenvalues are chosen as the basis vectors. However, the above classification rule given in Eq. (3) can no longer be used in this method. Instead, the minimum distance of the centered test vector from each subspace determines class label. In contrast to the CLAFIC, this classification rule is equivalent to assigning the test sample to the class achieving the best reconstruction of the test sample through an *affine* combination of the corresponding class samples. More formally, the following criterion is used during classification:

$$\begin{aligned} g(x_{\text{test}}) &= \underset{i=1,\ldots,C}{\arg\min} \left\|(I - P^{(i)})(x_{\text{test}} - \mu_i)\right\|^2 \\ &= \underset{i=1,\ldots,C}{\arg\min}\left(||x_{\text{test}} - \mu_i||^2 - \left\|W^{(i)^{\mathrm{T}}}(x_{\text{test}} - \mu_i)\right\|^2\right), \end{aligned} \tag{6}$$

where I represents the identity matrix.

A special case of the CLAFIC-μ method occurs when the dimensionality of the sample space is larger than the number of training set samples in each class. In this case, $(I{-}P^{(i)})$ will be the orthogonal projection operator onto the null space of $\Sigma^{(i)}$ if the basis vectors are chosen as the eigenvectors corresponding to all nonzero eigenvalues. Then, all distances from training samples to their corresponding subspace become zero, and hence 100% classification accuracy with respect to the training set data can be obtained. The method using this approach is called as the CV method, and it can only be used if the dimensionality of the sample space is larger than the number of training set samples of each class [12].

2.3. The Averaged Learning Subspace (ALS) method

The ALS can be considered as an iterative learning version of the CLAFIC method. The basic idea behind this method is to revise the subspaces based on the misclassified samples. The modification of subspaces is accomplished through rotation of subspaces. When a misclassification error occurs, the correct subspace is rotated towards the misclassified sample vector and the wrong subspace is rotated away from it. As a result, this process yields a better selection of subspaces in terms of classification accuracy. In the ALS method, the unnormalized correlation matrix $S^{(i)}$ of each class is modified after errors. The ALS algorithm can be summarized as follows:

- *Step 1*: Compute the unnormalized initial correlation matrices of classes for the epoch $k = 0$ as

$$S_o^{(i)} = \sum_{j=1}^{N_i} x_j^i (x_j^i)^{\mathrm{T}}, \quad i = 1,\ldots,C. \tag{7}$$

  Then compute the basis vectors of each class using these unnormalized correlation matrices as in the CLAFIC.
- *Step 2*: Classify all training set samples using Eq. (3) and compute the misrecognized samples. Then, update the correlation matrices as

$$S_{k+1}^{(i)} = S_k^{(i)} + \alpha \sum_{x_j^i \in A_k^{(i)}} x_j^i (x_j^i)^{\mathrm{T}} - \beta \sum_{x_l^i \in B_k^{(i)}} x_l^i (x_l^i)^{\mathrm{T}}, \quad i = 1,\ldots,C, \tag{8}$$

  where $A_k^{(i)}$ is the set of vectors in the ith class, which are erroneously assigned to some other class during the training epoch k, and $B_k^{(i)}$ is the set of vectors that are classified erroneously as the ith class in that epoch. Here the parameters α and β are some positive numbers.
- *Step 3*: Compute the new basis vectors using the modified correlation matrices and continue with Step 2 unless a predefined number of epochs has been reached, or the classification accuracy starts to decrease.

In this algorithm, the positive coefficient parameters α and β control the steepness of the error correction, and hence the convergence speed and the stability of the algorithm heavily depend on the values of these parameters. They are usually made to decrease with the epoch number. The subspace dimensions l_i can be set to a fixed value once and for all before starting the training phase or they can be reselected after each epoch.

2.4. Computational considerations

It is not always practical to store and use the orthogonal projection operators $P^{(i)}$ explicitly for computations in Eqs. (3) and (6), especially if the number of samples in each class is smaller than the dimensionality of the sample space. Instead, the basis vectors are used for computations in such cases. However, we may still have difficulties during computation of these basis vectors that span the subspaces. This problem occurs when the dimensionality of the sample space is large in comparison with the number of samples in each class. In this case, the size of the correlation or covariance matrices will be too large, e.g., in face recognition tasks, images of size 256×256 yield correlation and covariance matrices of size 65536×65536. Computing the eigenvalues and eigenvectors of those huge matrices will be difficult and numerically unstable. To solve this problem, let the number of samples in the ith class be denoted by N_i and the dimension of the subspace is d. Fortunately, the nonzero eigenvalues and the corresponding eigenvectors for each class can be obtained by applying eigen-decomposition to a smaller $N_i \times N_i$ matrix, instead of the $d \times d$ matrix. If A is a $d \times N_i$ matrix composed of all samples in the class, then, the matrix

$$\Psi^{(i)} = A^{(i)}A^{(i)^{\mathrm{T}}}, \quad i = 1,\ldots,C, \tag{9}$$

could be any symmetric positive semi-definite matrix of size $d \times d$, e.g., correlation or covariance matrix of each class. Now, let λ_k^i and v_k^i be the kth nonzero eigenvalue and the corresponding eigenvector of $\Psi^{(i)^{\mathrm{T}}} = A^{(i)^{\mathrm{T}}}A^{(i)}$, where $k \leqslant N_i \ll d$. Then $w_k^i = A^{(i)}v_k^i$ will be the eigenvector corresponding to the kth nonzero eigenvalue of $\Psi^{(i)}$. Although this approach can be easily used in the CLAFIC and the CLAFIC-μ methods, it is not possible to use it in the ALS since the unnormalized correlation matrix $S_k^{(i)}$ at any epoch k cannot be decomposed as in Eq. (9) except at epoch $k = 0$. Therefore, the ALS cannot be used in face recognition tasks where the dimensionality of the sample space is too large. However, as shown in the next section, this drawback can be eliminated if we use face image matrices directly in ALS.

3. Two-dimensional subspace classifiers

The application of classical subspace methods to face recognition problems involves the transformation of original 2D image matrix data into 1D vectors. This transformation is usually accomplished by lexicographic ordering of image matrices into column vectors, which is performed by concatenating rows or columns. As opposed to the conventional subspace methods, our proposed methods utilize image matrices instead of image vectors, and 2D matrix form is preserved in all calculations.

The main purpose of 2D methods is to determine the set of projection vectors which will map an image matrix to a set of feature vectors forming a feature matrix [31]. Classification is performed using these feature matrices. Following this scheme, we will next examine the proposed methods in detail.

3.1. The 2D-CLAFIC method

Let $X_j^i \in \Re^{n\times m}$ denote the jth image matrix of the ith class whose size is $n\times m$ ($n \geqslant m$). In this case, each correlation matrix $R_{2D}^{(i)} \in \Re^{n\times n}$ of the image matrices is defined as

$$R_{2D}^{(i)} = \frac{1}{N_i}\sum_{j=1}^{N_i} X_j^i (X_j^i)^{\mathrm{T}}, \quad i = 1,\dots,C. \tag{10}$$

Note that the size of the image correlation matrix, $n\times n$, is much smaller than the size of the correlation matrix, $d\times d$, obtained using the classical CLAFIC method since $d = mn$. The image correlation matrix $R_{2D}^{(i)}$ is a positive semi-definite matrix, hence all eigenvalues are larger than or equal to 0. The most significant l_i eigenvectors of $R_{2D}^{(i)}$ are chosen as the basis vectors as in the classical CLAFIC method. The number of total basis vectors will be limited by the number of columns of image matrices, which is equal to m for each class. After basis vectors of subspaces are obtained, to classify a test image matrix, X_{test}, we project the image matrix onto the basis vectors of classes to get the image feature matrix

$$Y_{\text{test}}^i = (W^i)^{\mathrm{T}} X_{\text{test}}, \quad i = 1,\dots,C. \tag{11}$$

For the classification of X_{test}, we compute the Frobenius norms of feature matrices by

$$||Y_{\text{test}}^i||_F = \sqrt{\sum_{j=1}^{l_i}\sum_{k=1}^{m} |y_{jk}|^2}, \quad i = 1,\dots,C, \tag{12}$$

for each class. Then, the test sample is assigned to the class, in which the image feature matrix has the maximum Frobenius norm value, i.e.,

$$g(X_{\text{test}}) = \underset{i=1,\dots,C}{\arg\max}\, ||Y_{\text{test}}^i||_F. \tag{13}$$

3.2. The 2D-CLAFIC-μ method

As in the CLAFIC-μ method, we can utilize class-specific means in the 2D-CLAFIC method. The 2D-CLAFIC-μ method uses the class-specific mean images of samples during classification. In this approach, image covariance matrix of each class is used to compute the basis vectors. The image covariance matrix $\Sigma_{2D}^{(i)} \in \Re^{n\times n}$ of each class is defined as

$$\Sigma_{2D}^{(i)} = \frac{1}{N_i}\sum_{j=1}^{N_i} (X_j^i - \bar{X}_i)(X_j^i - \bar{X}_i)^{\mathrm{T}}, \quad i = 1,\dots,C, \tag{14}$$

where $\bar{X}_i$ represents the mean image matrix of the ith class. The most significant l_i eigenvectors corresponding to the largest eigenvalues of each image covariance matrix are used as basis vectors representing classes. Once the basis vectors are obtained, we compute the distance of the centered test image from each subspace by using the Frobenious norm, and we assign the test sample image to the class in which it has the minimum norm value, i.e.,

$$g(X_{\text{test}}) = \underset{i=1,\dots,C}{\arg\min}(||X_{\text{test}} - \bar{X}_i||_F^2 - ||(W^i)^{\mathrm{T}}(X_{\text{test}} - \bar{X}_i)||_F^2). \tag{15}$$

3.3. The two-dimensional Averaged Learning Subspace (2D-ALS) method

The 2D-ALS can be thought as an iterative learning version of the 2D-CLAFIC method. In this method, the basis vectors spanning each subspace are iteratively modified in order to reduce the number of misclassified samples. The modifications of subspaces are accomplished performing rotation of subspaces as in the classical ALS. The 2D-ALS uses the unnormalized correlation matrix $S^{(i)} \in \Re^{n\times n}$ of image samples to compute the basis vectors representing classes. The classical ALS cannot be used if the dimensionality of the sample space is large as mentioned in Section 2.4. However, the 2D-ALS method can be easily used in such cases since the size of the correlation matrices is $n\times n$ as opposed to $d\times d$. The algorithm of this method can be summarized as follows:

- *Step 1*: Compute the unnormalized initial correlation matrices of image samples for the epoch $k = 0$ as

  $$S_o^{(i)} = \sum_{j=1}^{N_i} X_j^i (X_j^i)^{\mathrm{T}}, \quad i = 1,\dots,C. \tag{16}$$

  Then compute the basis vectors of each class using these unnormalized correlation matrices as in the 2D-CLAFIC.
- *Step 2*: Classify all training set samples using Eq. (13) and compute the misrecognized samples. Then, update the correlation matrices as

  $$S_{k+1}^{(i)} = S_k^{(i)} + \alpha \sum_{X_j^i \in A_k^{(i)}} X_j^i (X_j^i)^{\mathrm{T}} - \beta \sum_{X_l^i \in B_k^{(i)}} X_l^i (X_l^i)^{\mathrm{T}}, \quad i = 1,\dots,C, \tag{17}$$

  where $A_k^{(i)}$ is the set of image matrices in the ith class, which are erroneously assigned to some other class during the training epoch k, and $B_k^{(i)}$ is the set of image matrices that are classified erroneously as the ith class in that epoch. Here the positive coefficient parameters α and β control the steepness of the error correction, and they should be properly set as in the conventional ALS.
- *Step 3*: Compute the new basis vectors using the modified correlation matrices and continue with Step 2 unless a predefined number of epochs has been reached, or the classification accuracy starts to decrease.

3.4. Computational considerations

In 2D subspace classifiers, the size of the correlation or covariance matrices is $n\times n$ rather than $d\times d$, where $d = mn$. As a consequence, the eigen-decomposition of these scatter matrices is much easier compared to those of 1D subspace classifiers. Therefore, 2D subspace classifiers are more efficient than 1D subspace classifiers in terms of computational complexity. Especially, testing complexity of our proposed 2D methods is given by $O(nd)$ whereas the testing complexity of the classical subspace classifiers is $O(d^2)$.

4. Subspace classifiers with higher order image tensors

Recent studies showed that higher order image tensor representations of data samples based on Gabor filtering improve the recognition accuracies in several classifications tasks [21–24]. In this section, we give the extensions of CLAFIC and CLAFIC-μ, called multilinear CLAFIC and multilinear CLAFIC-μ methods, which can be used with image tensors of order $N \geqslant 3$. These new methods operate on richer data representations, as a result one may obtain better recognition accuracies. However, it should be kept in mind that high-order image tensor representations of objects such as Gabor function based representations bring additional computational cost over original image matrix based representations.

We first introduce basic tensor definitions which will be used to extend the subspace classifiers to handle image tensors. The definitions were adopted from Refs. [1,6,7,27]. An Nth-order tensor, usually denoted by calligraphic capital letters, is a multidimensional array having N indices: $\mathscr{A} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$. Elements of $\mathscr{A}$ are denoted as $\mathscr{A}_{i_1 i_2 \cdots i_N}$ or $a_{i_1 i_2 \cdots i_N}$. The scalar product of two tensors of the same order $\mathscr{A}, \mathscr{B} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$ is defined as $\langle \mathscr{A}, \mathscr{B} \rangle = \sum_{i_1} \sum_{i_2} \cdots \sum_{i_N} a_{i_1 i_2 \cdots i_N} b_{i_1 i_2 \cdots i_N}$. The Frobenius norm of a tensor $\mathscr{A}$ is defined as $||\mathscr{A}|| = \sqrt{\langle \mathscr{A}, \mathscr{A} \rangle}$. The n-mode product of a tensor $\mathscr{A} \in \Re^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ by a matrix $U \in \Re^{J_n \times I_n}$ is denoted by $\mathscr{A} \times_n U$ and defined as $(\mathscr{A} \times_n U)_{i_1 \cdots i_{n-1} j_n i_{n+1} \cdots i_N} = \sum_{i_n} a_{i_1 \cdots i_N} u_{j_n i_n}$. Mode-$n$ matricizing of an Nth-order tensor $\mathscr{A} \in \Re^{I_1 \times \cdots \times I_n \times \cdots \times I_N}$ is a matrix $A^{(n)} \in \Re^{I_n \times I_1 I_2 \cdots I_{n-1} I_{n+1} \cdots I_N}$ which is the ensemble of vectors obtained by keeping the index i_n fixed and varying other indices.

Let $\mathscr{X}_k^p \in \Re^{I_1 \times I_2 \times \cdots \times I_M}$ denote the pth training tensor sample of the kth class and let the kth class have N_k samples. The tensor including the collection of the class samples are called as the training data tensor of the kth class and denoted by $\mathscr{D}_k \in \Re^{I_1 \times I_2 \times \cdots \times I_M \times I_{M+1}}$ where $I_{M+1} = N_k$. Also it is possible to collect mean-subtracted class samples $(\mathscr{X}_k^p - (1/N_k)\sum_{\forall p} \mathscr{X}_k^p)$ into a mean-subtracted data tensor $\hat{\mathscr{D}}_k \in \Re^{I_1 \times I_2 \times \cdots \times I_M \times I_{M+1}}$. Note that a hat on the tensor letter denotes the data tensor of mean-subtracted samples.

4.1. Multilinear CLAFIC method

In classical CLAFIC, basis vectors of subspaces are determined from the eigenvalue–eigenvector decomposition of the correlation matrix. In multilinear case, they can be determined directly from the higher order singular value decomposition (HOSVD) of the training data tensor. By definition, HOSVD exists for tensors and any tensor $\mathscr{A} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$ can be expressed as the product $\mathscr{A} = \mathscr{S} \times_1 U^{(1)} \times_2 U^{(2)} \times \cdots \times_N U^{(N)}$ in which $\mathscr{S} \in \Re^{I_1 \times I_2 \times \cdots \times I_N}$ is the core tensor and $U^{(n)} \in \Re^{I_n \times I_n}, n = 1, \ldots, N$ are unitary matrices. Here, the core tensor has two properties: (1) If $\mathscr{S}_{i_n = \tau}$ is defined as the subtensor of $\mathscr{S}$ obtained by fixing the nth index to τ, two subtensors $\mathscr{S}_{i_n = \alpha}$ and $\mathscr{S}_{i_n = \beta}$ of $\mathscr{S}$ are orthogonal for all possible values of n, α, and β subject to $\alpha \neq \beta$ ($\langle \mathscr{S}_{i_n = \alpha}, \mathscr{S}_{i_n = \beta} \rangle = 0$ when $\alpha \neq \beta$); (2) Subtensors of $\mathscr{S}$ are ordered as $||\mathscr{S}_{i_n = 1}|| \geqslant ||\mathscr{S}_{i_n = 2}|| \geqslant \cdots \geqslant ||\mathscr{S}_{i_n = I_n}|| \geqslant 0$ for all possible values of n. For more details, one should refer to Ref. [6].

In multilinear CLAFIC, for each class, mode-n matrices of the data tensor are calculated by applying HOSVD to the data tensor $\mathscr{D}_k \in \Re^{I_1 \times I_2 \times \cdots \times I_M \times I_{M+1}}$

$$\mathscr{D}_k = \mathscr{S}_k \times_1 U_k^{(1)} \times_2 U_k^{(2)} \times \cdots \times_{M+1} U_k^{(M+1)}, \tag{18}$$

where $\mathscr{S}_k$ is the core tensor and $U_k^{(n)} \in \Re^{I_n \times I_n}$ are unitary mode-n matrices. For the subspace methods, in fact, one does not really need to determine the core tensor. Applying HOSVD to determine mode-n matrices is sufficient because basis vectors of subspaces exist in the columns of $U_k^{(n)}$ matrices. Similar to the classical case, the first $l_{kn'}^{(n)}$ column vectors ($n' \leqslant n$) of $U_k^{(n)}$ are chosen as basis vectors for the subspace L_k, and these basis vectors form the projection matrix $W_k^{(n)} \in \Re^{I_n \times I_{n'}}$ when they are used in the columns of the matrix. Note that $W_k^{(n)} \in \Re^{I_n \times I_{n'}}$ is the reduced form of the matrix $U_k^{(n)} \in \Re^{I_n \times I_n}$ of the second dimension since $n' \leqslant n$. Then, the n-mode projection of the sample tensors for the kth class can be achieved through the projection operator $\mathscr{P}_k$ given with the following relation.

$$\mathscr{P}_k = \times_1 W_k^{(1)^{\mathrm{T}}} \times_2 W_k^{(2)^{\mathrm{T}}} \times \cdots \times_M W_k^{(M)^{\mathrm{T}}}. \tag{19}$$

For the classification of the test sample $\mathscr{X}_{\text{test}} \in \Re^{I_1 \times I_2 \times \cdots \times I_M}$, the following maximization criterion is used:

$$\begin{aligned} g(\mathscr{X}_{\text{test}}) &= \underset{k=1,\ldots,C}{\arg\max} ||\mathscr{X}_{\text{test}} \times \mathscr{P}_k||^2 \\ &= \underset{k=1,\ldots,C}{\arg\max} ||\mathscr{X}_{\text{test}} \times_1 W_k^{(1)^{\mathrm{T}}} \times_2 W_k^{(2)^{\mathrm{T}}} \\ &\quad \times \cdots \times_M W_k^{(M)^{\mathrm{T}}} ||^2. \end{aligned} \tag{20}$$

4.2. Multilinear CLAFIC-μ method

Recall that, the CLAFIC-μ method uses the class-specific means, hence multilinear representation of CLAFIC-μ operates on mean-subtracted data tensor $\hat{\mathscr{D}}_k \in \Re^{I_1 \times I_2 \times \cdots \times I_M \times I_{M+1}}$. First, HOSVD is applied to determine mode-n matrices of the mean-subtracted data tensor:

$$\hat{\mathscr{D}}_k = \hat{\mathscr{S}}_k \times_1 U_k^{(1)} \times_2 U_k^{(2)} \times \cdots \times_{M+1} U_k^{(M+1)}. \tag{21}$$

Then the first $l_{kn'}^{(n)}$ column vectors ($n' \leqslant n$) of $U_k^{(n)}$ are chosen as basis vectors for the subspace L_k to determine mode-n projection matrices $W_k^{(n)} \in \Re^{I_n \times I_{n'}}$ ($n = 1, \ldots, M$). These matrices form the tensor projection operator of the kth class, $\mathscr{P}_k$.

$$\mathscr{P}_k = \times_1 W_k^{(1)^{\mathrm{T}}} \times_2 W_k^{(2)^{\mathrm{T}}} \times \cdots \times_M W_k^{(M)^{\mathrm{T}}}. \tag{22}$$

Since the centered tensors are used in multilinear CLAFIC-μ method, the classification criterion of this method is different from the criterion of the multilinear CLAFIC method. During classification, the class means must be subtracted from the test sample for each class, and hence the following minimization criterion is used:

$$\begin{aligned} g(\mathscr{X}_{\text{test}}) &= \underset{k=1,\ldots,C}{\arg\min} (||\mathscr{X}_{\text{test}} - \bar{\mathscr{X}}_k||^2 - ||(\mathscr{X}_{\text{test}} - \bar{\mathscr{X}}_k) \times \mathscr{P}_k||^2), \\ &= \underset{k=1,\ldots,C}{\arg\min} (||\mathscr{X}_{\text{test}} - \bar{\mathscr{X}}_k||^2 - ||(\mathscr{X}_{\text{test}} - \bar{\mathscr{X}}_k) \times_1 W_k^{(1)^{\mathrm{T}}} \\ &\quad \times_2 W_k^{(2)^{\mathrm{T}}} \times \cdots \times_M W_k^{(M)^{\mathrm{T}}} ||^2) \end{aligned}$$

where $\bar{\mathscr{X}}_k = (1/N_k)\sum_{\forall p} \mathscr{X}_k^p$ is the class mean of data tensors.

4.3. Comparison to related methods

The proposed methods described in this section operate on high-order image tensors and do not suffer from *undersample problem* [24] as in Multilinear Discriminant Analysis (MDA) [30] and General Tensor Discriminant Analysis (GTDA) [24]. However, there is a major difference between these methods (MDA and GTDA) and our proposed methods. This difference is similar to the difference between Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). LDA is built on the assumption that all classes have similar covariance structures. Thus, the class covariance matrix is same for all classes, and it is approximated by the within-class scatter matrix. As a result, LDA produces linear decision boundaries if it is used as a classifier rather than a method for feature extraction. On the other hand, QDA assumes all classes have different covariance structures, and covariance matrix of each class is estimated using class-specific samples. Therefore, QDA yields quadratic decision boundaries as opposed to linear decision boundaries of LDA. Similar to LDA, both MDA and GTDA methods also assume that all classes have identical covariance structures. But, our proposed methods are built on the assumption that each class has different covariance (or correlation) matrix as in QDA. Therefore, MDA and GTDA can be seen as a special case of the proposed multilinear subspace classifiers by using the analogy between LDA and QDA. In the proposed methods, the differences between class covariance (or correlation) matrices of classes are utilized during computation of the distance between the query and subspaces representing classes. The decision boundaries are again quadratic. The same differences apply to the proposed 2D methods and 2D-LDA.

5. Experiments

We performed experiments on three well-known face image databases, namely the AR[1] [19], Yale[2] [3] and ORL (Olivetti-Oracle Research Lab)[3] face databases. The AR and Yale face databases have been employed to evaluate the recognition performances of the proposed methods under conditions where there are variations over time, in facial expressions, and in lighting conditions, whereas the ORL face database has been used to examine the recognition performances of the proposed methods under conditions where the pose is varied. As mentioned before, computational costs of the methods using Gabor filter based representations are higher than those of the methods using original image matrix based representations since image matrices occur naturally without any extra cost as opposed to the Gabor filtered image matrices. Therefore, we performed experiments restricting our attention to 2D based approaches. Beside the proposed 2D subspace classifiers here, we also tested the CLAFIC and CLAFIC-μ subspace classifier methods for comparison. In addition, we also tested the PCA, 2D-PCA and 2D-LDA methods for a better assessment of the recognition performances of our proposed methods [15,26,31]. Note that we could not test the classical ALS method since the dimensionality of the sample space is too large for all face databases. A small set of randomly created training and test sets was employed to compute subspace dimensions [28]. To determine each subspace dimension, we set all subspace dimensions to be equal to a fixed value l for all classes. Then, the optimal value of l was chosen from the recognition rate curves obtained from test sets. Typical recognition rate curves are illustrated in Fig. 1 for the AR face database.

In the 2D-ALS method, we used the same values for both α and β parameters as recommended in Ref. [17]. To determine the best value for α and β, we covered the values between 0 and 5 with an increment 0.5 during the selection phase in all experiments. After computing the best value for α and β parameters, this value was forced to decrease with the epoch number. The maximum number of epochs in the algorithm was chosen as 10 in order to avoid over-fitting problem. The subspace dimensions were set to a fixed value l, and they were kept the same during all epochs in the 2D-ALS method. In contrast to the other subspace classifiers, the data samples were projected onto a unique subspace in PCA, 2D-PCA, and 2D-LDA methods. The dimensionality of this space was found using recognition curves as described above. Then, the nearest-neighbor algorithm has been employed during classification.

5.1. Experiments on the AR face database

The AR face database includes 26 images with different facial expressions, illumination conditions, and occlusions for 126 subjects. All individuals are in an upright, frontal position. Images were recorded in two different sessions 14 days apart. Thirteen images were recorded under controlled circumstances in each session. The size of the images in the database is 768×576 pixels, and each pixel is represented by 24 bits of RGB color values.

We randomly selected 50 individuals (30 males and 20 females) for the experiment. Only nonoccluded images ((a)–(g) and (n)–(t) as in Fig. 2) were chosen for every subject. Thus, our face database size was 700 with 14 images per subject. First, these images were converted to grayscale images. Second, we preprocessed these images by aligning and scaling them so that the distances between the eyes were the same for all images, and also ensuring that the eyes were located in the same coordinates of the image. The resulting image was then cropped. The final size of the images was 299×222. Finally, based on empirical observations, we decreased the dimensionality of the sample space to 134×99 by down-sampling.

The training set consisted of seven images randomly selected from each subject, and the rest of the images were used for the test set. Note that there is no overlap between the training and test sets. This process was repeated 25 times, and 25 different training and test sets were created. The first five data sets were used for subspace dimensionality and parameter selections, and the remaining 20 sets were used to evaluate performance. Thus, the final recognition rates for the experiment were found by averaging these 20 rates obtained in each trial. We set initial α and β parameter values to 1. The computed recognition rates, corresponding standard deviations, empirically selected subspace dimensions are shown in Table 1. In the same table, we also provided the testing times for classifying a single test sample.

As can be seen in the table, the best recognition accuracy is achieved by 2D-LDA. Our proposed method, 2D-CLAFIC-μ, achieved the best recognition rate among all subspace classifiers tested here. Although the 2D-CLAFIC-μ method outperformed the CLAFIC method, the 2D-CLAFIC method could not outperform its classical counterpart. Using class-specific means in 2D approach significantly improved the recognition rates since the 2D-CLAFIC-μ method yielded better recognition rates than the 2D-CLAFIC method. There were only a few misrecognized samples in the training sets, thus subspaces could not be modified substantially. As a result, the 2D-ALS method did not show much improvement over 2D-CLAFIC. It should be noted that all subspace classifiers outperformed the classical PCA method.

Testing time is the consumed time that is required to classify a new test image. To classify a test image, we have to evaluate the criterion functions given in Eqs. (3), (6), (13) and (15) depending on the chosen method. This process involves the projection of test samples onto the basis vectors spanning each subspace. The subspace dimensions of classical subspace classifiers and their corresponding 2D classifiers are similar. Therefore our proposed methods are also more practical than classical subspace classifiers for real-time image recognition applications since the testing complexity of our proposed methods is given by $O(nd)$ whereas the testing complexity of the classical subspace classifiers is $O(d^2)$, where image size is $n \times m$ and $d = mn$. Note that we omitted the subspace dimensionality and the number of the classes from the complexity term for the sake of clear comparison between the classical and 2D subspace approaches. PCA, 2D-PCA and 2D-LDA

[1] Available at: http://cobweb.ecn.purdue.edu/~aleix/alexi_face_DB.html.
[2] Available at: http://cvc.yale.edu/projects/yalefaces/yalefaces.html.
[3] Available at: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

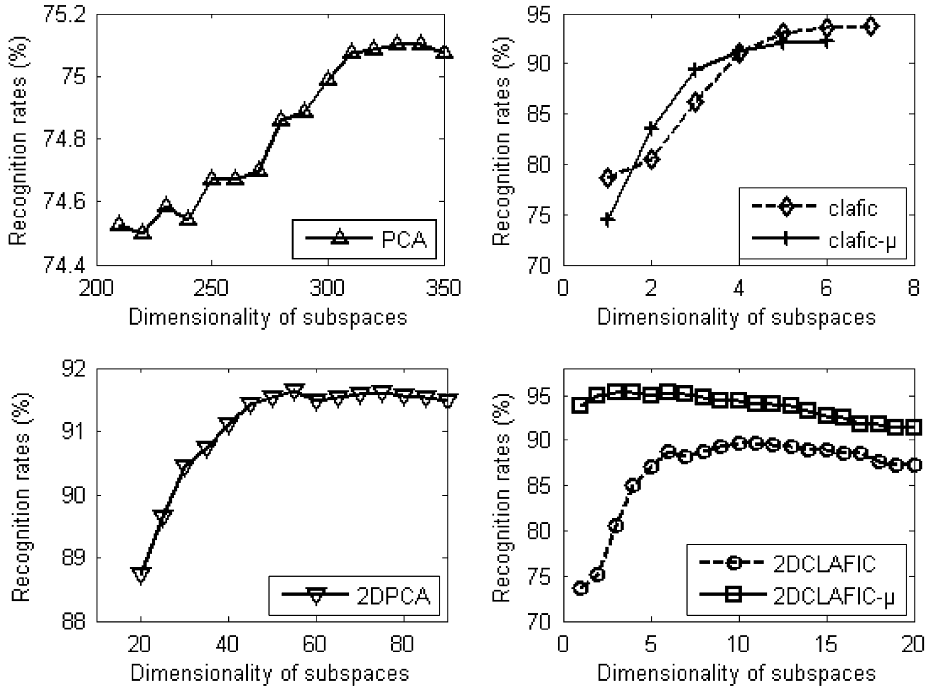

Fig. 1. The recognition rates as a function of subspace dimensionalities on the AR face database: in each method the dimension of the subspace is chosen as the value yielding the highest recognition rate on the validation set.

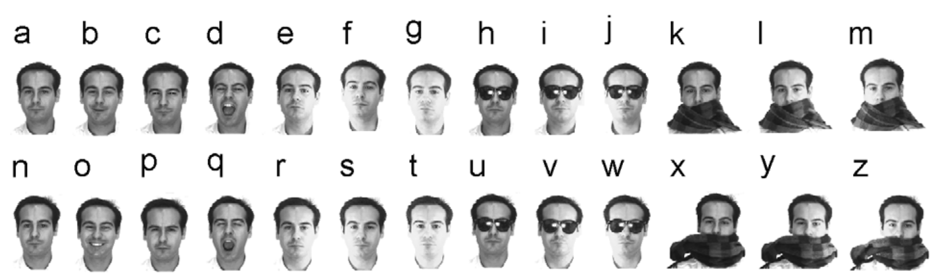

Fig. 2. Images of one subject in the AR face database. First 13 images (a)–(m) were taken in one session and the others (n)–(z) in another session. Only nonoccluded images (a)–(g) and (n)–(t) were used in our experiments.

Table 1
Recognition rates on the AR face database

Methods	Subspace dimensions	Recognition rates (%)	Standard deviation	Testing time (ms)
PCA	$l = 330$	75.1	1.56	9.57
2D-PCA	$l = 55$	91.64	1.46	8.96
CLAFIC	$l = 7$	93.73	1.45	7.43
CLAFIC-μ	$l = 6$	92.27	1.55	9.97
2D-CLAFIC	$l = 10$	89.79	1.68	4.43
2D-CLAFIC-μ	$l = 4$	95.39	0.98	5.92
2D-ALS	$l = 10$	90.39	1.45	4.45
2D-LDA	$l = 18$	97.86	0.81	6.68

methods are used as feature extraction methods in the experiments. Therefore, following the projection of data samples onto the chosen axes, one has to search for the nearest neighbor in the reduced space to label the test sample. As a result, the testing times of these methods are typically higher than the proposed 2D subspace classifiers as given in Table 1.

5.2. Experiments on the Yale face database

The Yale face database consists of images from 15 different people, using 11 images from each person, for a total of 165 images. The images contain variations with the following facial expressions or configurations: center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised and wink. For subjects numbered 2, 3, 6, 7, 8, 9, 12 and 14, the normal facial expression and the without glasses (or with glasses if subject normally wears glasses) images were copies of each other. Thus, we removed the image without glasses (or with glasses if subject normally wears glasses) from every subject in order to make all classes have an equal number of samples and have all sample images distinct. Thus, we had 10 samples per subject yielding a face database size of 150. We pre-processed these images by aligning and scaling them so that the distances between the eyes were the same for all images, and also ensuring that the eyes occurred in the same coordinates of the image. The resulting image was then cropped. The final image size was 152×126.

We randomly selected five samples from each class for training and the remaining samples were used for testing. This process was repeated 25 times and 25 different training and test sets were created. As in the previous case, the first 5 data sets were used for determining the subspace dimensions and the best parameter values whereas the remaining 20 data sets were used for performance evaluation. The initial α and β parameter values were chosen as 0.5. The computed recognition rates, standard deviations, subspace dimensions and testing times of methods are shown in Table 2.

In terms of classification accuracy, the best recognition rate was attained by the proposed 2D-CLAFIC-μ method. Subspace classifiers typically yielded better recognition rates than the PCA method. As in the previous case, although the 2D-CLAFIC-μ method outperformed its classical counterpart, the 2D-CLAFIC method did not offer any improvement over the CLAFIC method. It should be noted that the 2D-ALS method significantly improved the recognition rates of the 2D-CLAFIC method in this case.

5.3. Experiments on the ORL face database

The ORL face database contains 40 individuals, with 10 images per person. The images are taken at different time instances with different lighting conditions (slightly), facial expressions, and facial details. Some individuals from the ORL face database are shown in Fig. 3. The size of each image is 112×92.

We selected randomly five samples from each class for training and the remaining samples were used for testing. We did not apply any pre-processing to the images. This process was repeated 25 times. As in the other experiments, the first 5 data sets were used for finding subspace dimensionalities and optimal parameter values, and the remaining 20 data sets were used to evaluate the performance. The initial α and β parameters were set to a value of 1. The computed recognition rates, standard deviations, subspace dimensions and testing times are shown in Table 3.

For the ORL face database, the 2D-LDA method achieved the best recognition rate among all methods tested here. The classical subspace methods outperformed our proposed 2D-based subspace classifiers. Furthermore, the 2D-PCA method showed only a little improvement over the PCA method. Therefore, we conclude that the 2D-based subspace classifiers do not offer a significant

Table 2
Recognition rates on the Yale face database

Methods	Subspace dimensions	Recognition rates (%)	Standard deviation	Testing time (ms)
PCA	$l = 75$	75.27	4.34	2.81
2D-PCA	$l = 13$	78.73	3.95	1.93
CLAFIC	$l = 5$	80.06	4.96	2.37
CLAFIC-μ	$l = 4$	77.67	4.19	3.37
2D-CLAFIC	$l = 9$	76.27	4.07	1.57
2D-CLAFIC-μ	$l = 2$	82.67	5.01	2.03
2D-ALS	$l = 3$	80.67	3.25	0.77
2D-LDA	$l = 12$	80.40	3.87	1.93

Fig. 3. Images of some individuals from the ORL face database.

Table 3
Recognition rates on the ORL face database

Methods	Subspace dimensions	Recognition rates (%)	Standard deviation	Testing time (ms)
PCA	$l = 190$	94.20	1.50	4.47
2D-PCA	$l = 16$	94.90	1.47	4.01
CLAFIC	$l = 5$	94.22	1.40	3.51
CLAFIC-μ	$l = 4$	95.05	1.45	5.06
2D-CLAFIC	$l = 6$	91.93	1.62	2.24
2D-CLAFIC-μ	$l = 2$	93.30	1.75	3.10
2D-ALS	$l = 6$	92.40	1.50	2.25
2D-LDA	$l = 10$	95.60	1.26	2.69

improvement over their 1D-based counterparts for image recognition problems where there is a variation in pose. However, it should be noted that as in the previous cases, our proposed methods are better suited for real-time image recognition applications than the conventional 1D subspace classifiers because of their low computation cost.

6. Discussion and conclusion

In this paper we have proposed new 2D subspace classifiers for face recognition problems. The major advantage of two-dimensional subspace classifiers over the conventional subspace classifiers is that they can be applied directly to the image matrices. This process gives rise to correlation and covariance matrices with smaller size. As a result, eigenvalue–eigenvector decomposition of scatter matrices can be performed fast, which in turn speeds up the training and testing phases. Since employing only a few number of basis vectors leads to high recognition rates in 2D subspace classifiers, our proposed methods are more practical than 1D based subspace classifiers for real–time face recognition tasks. In addition, experimental results of different databases demonstrated that the proposed methods are more robust to varying illumination conditions, which is a serious problem that PCA based feature extraction techniques cannot handle well. However, one drawback of the proposed methods is their sensitivity to rotations in poses as observed in experiments on the ORL face database. Nevertheless, when the subspace dimension and the computational efficiency of the proposed methods are considered, the recognition accuracies can be accepted as satisfactory.

Among the proposed three methods, especially 2D-CLAFIC-μ, is superior to other proposed two-dimensional methods. It implies that the mean subtraction scheme which uses the covariance measure is more successful in face recognition problems rather than the methods using correlation matrix. The 2D-ALS method generally achieves better results than 2D-CLAFIC method, but none of these methods show an improvement over their respective 1D subspace classifier methods. In general, 2D-CLAFIC-μ method gives rise to better recognition results with less amount of subspace dimension. These results demonstrate the superiority of the proposed 2D-CLAFIC-μ method in face recognition applications.

The proposed 2D subspace classifiers are typically outperformed by 2D-LDA method followed by the nearest-neighbor classification rule since the face classes have similar covariance structures. But, this does not make the proposed methods uninteresting or useless. Because, there are a lot of applications, in which image classes have different covariance structures. For instance, the objects in Coil100 database have quite different covariance structures and LDA based methods yield poor results compared to subspace classifiers [4]. Moreover, it has recently become more and more popular to use multiple images for face recognition [8–10]. As demonstrated in these studies, although conventional methods using single images, such as LDA, have been successful in controlled environments, these methods typically fail in real world applications. On the other hand, the methods using multiple images are more successful than conventional methods in real world applications and they are all built on subspace classifiers. Therefore, we believe that our proposed methods will find wide application areas in the near future.

Although we introduced multilinear subspace classifiers, which allow us to use subspace classifiers with high-order image tensors, we have not experimented with these methods. In near future, we are planning to focus on this direction. Since multilinear subspace classifiers operate on richer data representations, we expect to improve the recognition accuracies.

References

[1] B.W. Bader, T.G. Kolda, Algorithm 862: MATLAB tensor classes for fast algorithm prototyping, ACM Trans. Math. Software 32 (4) (2006) 635–653.
[2] T. Balachander, R. Kothari, Kernel based subspace pattern classification, in: International Joint Conference on Neural Networks, vol. 5, 1999, pp. 3119–3122.
[3] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, Eigenfaces vs. Fisherfaces: recognition using class specific linear projection, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 711–720.
[4] H. Cevikalp, M. Neamtu, A. Barkana, The kernel common vector method: a novel nonlinear subspace classifier for pattern recognition, IEEE Trans. Syst. Man Cybern. B: Cybern. 37 (2007) 937–951.
[5] S.C. Chen, Y.L. Zhu, D.Q. Zhang, J.Y. Yang, Feature extraction approaches based on matrix pattern: MatPCA and MatFLDA, Pattern Recognit. Lett. 26 (8) (2005) 1157–1167.
[6] L. De Lathauwer, B. De Moor, J. Vandewalle, A multilinear singular value decomposition, SIAM J. Matrix Anal. Appl. 21 (4) (2000) 1253–1278.
[7] L. De Lathauwer, B. De Moor, J. Vandewalle, Computation of the canonical decomposition by means of a simultaneous generalized Schur decomposition, SIAM J. Matrix Anal. Appl. 26 (2) (2004) 295–327.
[8] W. Fan, D.Y. Yeung, Locally linear models on face appearance manifolds with application to dual-subspace based classification, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2006.
[9] A.W. Fitzgibbon, A. Zisserman, Joint manifold distance: a new approach to appearance based clustering, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003.
[10] K. Fukui, O. Yamaguchi, Face recognition using multi-view point patterns for robot vision, in: 11th International Symposium of Robotics Research, 2003, pp. 192–201.
[11] K. Fukunaga, W.L. Koontz, Application of the Karhunen–Loeve expansion to feature selection and ordering, IEEE Trans. Comput. 19 (4) (1970) 311–318.
[12] M.B. Gulmezoglu, V. Dzhafarov, A. Barkana, The common vector approach and its relation to principal component analysis, IEEE Trans. Speech Audio Process. 9 (6) (2001) 655–662.
[13] S.-W. Kim, B.J. Oommen, On utilizing search methods to select subspace dimensions for kernel-based nonlinear subspace classifiers, IEEE Trans. Pattern Anal. Mach. Intell. 27 (2005) 136–141.
[14] T. Kohonen, G. Nemeth, K.J. Bry, M. Jalanko, H. Riittinen, Spectral classification of phonemes by learning subspaces, in: Proceedings of the 5th International Conference on Acoustics, Speech and Signal Processing, Washington, DC, 1979, pp. 97–100.
[15] H. Kong, E. Teoh, J. Wang, R. Venkateswarlu, Two-dimensional Fisher discriminant analysis: forget about small sample size problem, in: Proceedings of the IEEE Conference on Acoustics, Speech, Signal Processing, 2005, pp. 761–764.
[16] M. Kuusela, E. Oja, The averaged learning subspace method for spectral pattern recognition, in: Proceedings of the 6th International Conference on Pattern Recognition, Munchen, 1982, pp. 134–137.
[17] J. Laaksonen, Subspace classifiers in recognition of handwritten digits, Ph.D. Thesis, Helsinki University of Technology, 1997.
[18] K. Liu, Y. Cheng, J. Yang, Algebraic feature extraction for image recognition based on an optimal discriminant criterion, Pattern Recognit. 26 (6) (1993) 903–911.
[19] A.M. Martinez, R. Benavente, The AR face database, CVC Technical Report #24, 1998.
[20] E. Oja, Subspace Methods of Pattern Recognition, Research Studies Press, 1983.
[21] J. Sun, D. Tao, C. Faloutsos, Beyond streams and graphs: dynamic tensor analysis, in: ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006, pp. 374–383.
[22] D. Tao, X. Li, W. Hu, S. Maybank, X. Wu, Supervised tensor learning, in: Proceedings of the 5th International Conference on Data Mining, 2005.
[23] D. Tao, X. Li, X. Wu, W. Hu, S.J. Maybank, Supervised tensor learning, Knowl. Inf. Syst. 13 (2007) 1–42.
[24] D. Tao, X. Li, X. Wu, S.J. Maybank, General tensor discriminant analysis and gabor features for gait recognition, IEEE Trans. Pattern Anal. Mach. Intell. 29 (2007) 1700–1715.
[25] K. Tsuda, Subspace classifier in reproducing kernel Hilbert space, in: International Joint Conference on Neural Networks, vol. 5, 1999, pp. 3454–3457.
[26] M. Turk, A.P. Pentland, Eigenfaces for recognition, J. Cogn. Neurosci. 3 (1991) 71–86.
[27] M.A.O. Vasilescu, D. Terzopoulos, Multilinear subspace analysis for image ensembles, in: Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR '03), vol. 2, 2003, pp. 93–99.
[28] S. Watanabe, P.F. Lambert, C.A. Kulikowski, J.L. Buxton, R. Walker, Evaluation and selection of variables in pattern recognition, Comput. Inf. Sci. II (1967) 91–122.

[29] D. Xu, S. Yan, D. Tao, S. Lin, H.J. Zhang, Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval, IEEE Trans. Image Process. 16 (11) (2007) 2811–2821.
[30] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H.J. Zhang, Multilinear discriminant analysis for face recognition, IEEE Trans. Image Process. 16 (1) (2007) 212–220.
[31] J. Yang, D. Zhang, 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) 131–137.
[32] J. Ye, R. Janardan, Q. Li, Two-dimensional linear discriminant analysis, Adv. Neural Inf. Process. Syst. 17 (2004) 1569–1576.
[33] D.Q. Zhang, Z.H. Zhou, S.C. Chen, Diagonal principal component analysis for face recognition, Pattern Recognit. 39 (1) (2006) 140–142.

Hakan Cevikalp received the M.S. degree from the Department of Electrical and Electronics Engineering, Eskisehir Osmangazi University, Eskisehir, Turkey, in 2001 and the Ph.D. degree from the Department of Electrical Engineering and Computer Science, Vanderbilt University, Nashville, TN, USA in 2005. He is in the academic staff of the Electrical and Electronics Engineering Department of Eskisehir Osmangazi University. His research interests include pattern recognition, neural networks, image and signal processing, optimization, and computer vision. He is a member of the IEEE.

Hasan Serhan Yavuz received the B.S. and the M.S. degrees from the Electrical and Electronics Engineering Department of Eskisehir Osmangazi University, Eskisehir, Turkey, in 1999 and 2002, respectively. He is currently pursuing a Ph.D. degree in the Electrical and Electronics Engineering Department of Eskisehir Osmangazi University. He is in the academic staff of the Electrical and Electronics Engineering Department of Eskisehir Osmangazi University as a research assistant. His research interests include pattern recognition, image and signal processing, fuzzy logic and computer vision.

Mehmet Atif Cay received the B.S. degree in Electrical Engineering from Istanbul Technical University, Istanbul, Turkey in 1982, M.S. degree in Electrical and Electronics Engineering from Anadolu University, Eskisehir, Turkey in 1986 and Ph.D. degree in Electrical Engineering from Lehigh University, Pennsylvania, USA in 1994. He is an assistant professor in Electrical and Electronics Engineering Department of Eskisehir Osmangazi University, Turkey. His current research interests include digital signal processing, speech and/or image recognition.

Atalay Barkana received the B.S. degree in Electrical Engineering from Robert College, Istanbul, Turkey, in 1969, and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Virginia, Charlottesville, in 1971 and 1974, respectively. From 1974 to 1986, he worked on linear and nonlinear theory. His current interests include speech recognition, pattern analysis, neural networks, and statistical signal processing. Since 2005, he has been with the Electrical and Electronics Engineering Department, Anadolu University, Eskisehir, Turkey, where he is presently a Professor. Dr. Barkana is a member of the IEEE Signal Processing Society.