Neurocomputing 72 (2009) 1152–1159

Contents lists available at ScienceDirect

Neurocomputing

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

Sequential row–column independent component analysis for face recognition

Quanxue Gao [a,b], Lei Zhang [a,*], David Zhang [a]

[a] Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
[b] School of Telecommunication Engineering, Xidian University, 710071 Xi'an, China

ARTICLE INFO

Article history:
Received 10 February 2007
Received in revised form
22 February 2008
Accepted 22 February 2008
Communicated by M.S. Bartlett
Available online 2 March 2008

Keywords:
Independent component analysis (ICA)
Face recognition
Feature extraction

ABSTRACT

This paper presents a novel subspace method called sequential row–column independent component analysis (RC-ICA) for face recognition. Unlike the traditional ICA, in which the face image is transformed into a vector before calculating the independent components (ICs), RC-ICA consists of two sequential stages—an image row-ICA followed by a column-ICA. There is no image-to-vector transformation in both the stages and the ICs are computed directly in the subspace spanned by the row or column vectors. RC-ICA can reduce the face recognition error caused by the dilemma in traditional ICA, i.e. the number of available training samples is greatly less than that of the dimension of training vector. Another advantage of RC-ICA over traditional ICA is that the dimensionality of the recognition subspace is much smaller, which means that the face image can have a more condensed representation. Extensive experiments are performed on the well-known Yale-B, AR and FERET databases to validate the proposed method and the experimental results show that the RC-ICA achieves higher recognition accuracy than ICA and other existing subspace methods while using a subspace of smaller dimensionality.

1. Introduction

Face recognition based on subspace analysis has been widely studied in recent years [3,20,21]. Generally speaking, this class of method seeks for a set of optimal bases under some criteria and represents a face image as a linear combination of them. The task of face recognition is implemented in the space spanned by those bases, which is usually a subspace of the original face space. The most widely used subspace analysis methods include principal component analysis (PCA) [10], which tries to find an optimal subspace in the sense of minimum mean square error (MSE), linear discriminant analysis (LDA) [3], which tries to find a subspace that maximizes the between-class distance and minimizes the within-class distance, and independent component analysis (ICA) [4,6], which tries to find a subspace spanned by a set of independent bases. LDA is a supervised learning technique while PCA and ICA are unsupervised learning techniques.

PCA is the first subspace analysis technology used for face recognition. Kirby and Sirovich [10] first used PCA to represent human faces and found that a face image could be reconstructed approximately as a weighted sum of a small collection of basis facial images plus a mean face image. Based on this research, Turk and Pentland [18] developed the well-known eigenface method. Since then, PCA has been extensively investigated and has become one of the most successful methods in face recognition [17,20]. Most of the PCA-based face recognition methods need to transform the image matrix into a vector before calculating the principle components (PCs). In [21], Yang et al. proposed a two-dimensional PCA (2D-PCA) scheme by projecting the image matrix, but not the transformed image vector, onto a set of basis vectors. This method was further improved by Kong et al. [11]. The drawback, however, of PCA is that it aims to find a set of orthonormal bases maximizing the variance over all samples. In other words, PCA only exploits the second-order statistics of the dataset. The higher-order statistics, which can be very useful to the face representation and recognition, are not exploited in PCA.

ICA [6], as an extension of PCA, was originally developed for blind source separation and it has been widely used in signal processing, medical image analysis and pattern recognition [2,5,6,16,24]. The objective of ICA is to seek for a set of linear bases, which are as independent as possible in the sense of high-order statistics other than the second-order statistics as in PCA. The independent components (ICs) obtained by projecting the face images onto the subspace spanned by these bases can reflect better intrinsic properties and local characteristics of the facial dataset [2,13]. In a word, ICA can remove the high-order statistical dependencies to produce a sparse and independent code for subsequent pattern discrimination.

Bartlett et al. [2] first applied ICA to face recognition and found that high-order statistical information is useful for representing

* Corresponding author.
E-mail address: cslzhang@comp.polyu.edu.hk (L. Zhang).

0925-2312/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.02.007

and identifying faces. They proposed two different ICA architectures for face recognition: architecture I represents the input face images as a linear combination of statistically independent basis images, while architecture II represents the input image by using a set of statistically independent coefficients. Experimental results showed that the recognition accuracy by using ICA is higher than that by using PCA. Since then, ICA has gained more interests in face modeling and recognition [13,14,23]. Moghaddam [15] and Yang et al. [22] compared the performance of PCA with ICA under different similarity metrics in detail and found that ICA is not always better than PCA. To solve this problem, Liu [13] analyzed the performance of ICA and proposed an enhanced ICA (EICA) method, which implements ICA in a reduced PCA space to improve retrieval performance. Pong et al. [23] studied the effect of the number of ICs on recognition accuracy. He argued that not all ICs are useful for recognition and proposed an algorithm to select ICs. Kim et al. [9] used ICA for face recognition and found that ICA is robust to local distortion and partial occlusion. In order to further improve the recognition accuracy, Bach et al. [1] proposed a kernel ICA method. The kernel methods need to map image vectors into an implicit feature space before applying ICA. It is computationally more expensive than ICA.

The above ICA-based methods directly take each pixel as a feature and they need to transform the 2D face image matrix into a 1D vector before computing ICs. The dimensionality of the resulting vector is usually very big (e.g. 16,384 for 128×128 images) and this leads to a dilemma of ICA: the number of available training samples is much less than that of the dimension of the underlying vector. This dilemma makes it very difficult to estimate accurately the statistics of the underlying face vector. The estimation error of the face vector statistics will then deteriorate the accuracy of face recognition.

Inspired by the 2D-PCA scheme [21], where the image matrix but not the transformed image vector is projected onto a set of basis vectors, in this paper we propose a novel subspace feature extraction method called row–column ICA (RC-ICA). The proposed RC-ICA includes two sequential steps: a row-ICA followed by a column-ICA. In contrast to ICA, RC-ICA directly uses the rows and columns of face images, rather than the stretched image vectors, to evaluate the demixing matrix. The computed demixing matrix is then employed to extract the ICs directly from the 2D face images. RC-ICA has three advantages over ICA. First, it greatly dilutes the dimensionality dilemma of ICA because the dimension of the training vector in RC-ICA is much smaller so that the statistics can be more accurately estimated. Secondly, the face image can have a more condensed representation by using RC-ICA. Thirdly, it is computationally more efficient. Our experimental results on the Yale-B, AR and FERET databases show that RC-ICA achieves higher recognition accuracy than ICA.

The rest of this paper is organized as follows. Section 2 briefly reviews ICA. Section 3 describes the proposed RC-ICA in detail. Section 4 presents extensive experiments to test the proposed method and Section 5 concludes the paper.

2. Independent component analysis (ICA)

In PCA, the subspace is spanned by the leading eigenvectors of the covariance matrix of the training dataset. PCA exploits only the second-order statistics and it is optimal for datasets which are of Gaussian distribution. In general, however, the distribution of facial images is non-Gaussian. There exist many high-order statistical dependencies among pixels of facial images. These dependencies can be further removed by using ICA, which can be viewed as an extension of PCA to deal with non-Gaussian datasets. What makes ICA superior to PCA is that ICA seeks for a set of linear bases which are statistically independent in an order higher than two to better characterize the intrinsic space of non-Gaussian sources. In this section, we briefly review the concepts and computation procedure of ICA. For more details, please refer to Hyvärinen et al. [6–8].

Denote by $\vec{x} = [x_1, x_2, \ldots, x_n]^{\mathrm{T}}$ a n-dimensional random vector. Suppose $\vec{x}$ can be represented as the linear combination of m ($m \leqslant n$) elements s_1, s_2, …, s_m, which are statistically independent (or as independent as possible), then the noise-free model of ICA can be described as

$$\vec{x} = Q\vec{s} \tag{1}$$

where $\vec{s} = [s_1, \ldots, s_m]^{\mathrm{T}}$ is the vector of ICs, Q is an unknown $n \times m$ mixing matrix.

In general, ICs s_i, $i = 1, 2, \ldots, m$, and the mixing matrix Q are unknown. ICA aims to find a demixing matrix W such that

$$\vec{u} = W\vec{x} = WQ\vec{s} \tag{2}$$

is a good estimation of $\vec{s}$, with possible permutation and rescaling. Fig. 1 illustrates the procedure. If the ICs s_i have unit variances, i.e. $E\{s_i s_i\} = 1$ for all $i = 1, 2, \ldots, m$, the ICs will be uniquely determined except for their signs [6].

To implement ICA, researchers have developed many algorithms based on two criteria: nonlinear decorrelation and maximum non-Gaussianity [6–8]. Among them, the Fast-ICA [7], which is based on maximum non-Gaussianity, has been dominantly used. It is implemented in two steps. In the first step, the mean value is subtracted from the training set and the second-order statistical correlation is removed. This step is called whitened. Denote the covariance matrix of $\vec{x}$ by

$$\Sigma_x = E[(\vec{x} - E[\vec{x}])(\vec{x} - E[\vec{x}])^{\mathrm{T}}] \tag{3}$$

where $E[\bullet]$ is the expectation operator. Let $V = [\vec{\gamma}_1 \vec{\gamma}_2 \cdots \vec{\gamma}_n]$ and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_m)$, where $\vec{\gamma}_i$ and λ_i are the eigenvectors and the corresponding eigenvalues of Σ_x, respectively, and $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_m$. Then the whitened data can be calculated as

$$\vec{y} = (V\Lambda^{-1/2})^{\mathrm{T}}\vec{x} = P^{\mathrm{T}}\vec{x} \tag{4}$$

where $P = V\Lambda^{-1/2}$ is called the whitened matrix. $\vec{y}$ is decorrelated and has an unit variance.

In the second step, the whitened data $\vec{y}$ is used to compute the demixing matrix W via kurtosis such that the m components of $\vec{u} = W\vec{y}$ are independent or almost independent. Denote the column vector of W by $\vec{w}$. The kurtosis of the projection of $\vec{y}$ onto $\vec{w}$ is

$$\mathrm{kurt}(\vec{w}^{\mathrm{T}}\vec{y}) = E[(\vec{w}^{\mathrm{T}}\vec{y})^4] - 3(E[(\vec{w}^{\mathrm{T}}\vec{y})^2])^2 \tag{5}$$

To make the components of $\vec{u}$ as independent as possible, we want to find a $\vec{w}$ to maximize the $\mathrm{kurt}(\vec{w}^{\mathrm{T}}\vec{y})$ under constraint

$$E[(\vec{w}^{\mathrm{T}}\vec{y})^2] = 1 \tag{6}$$

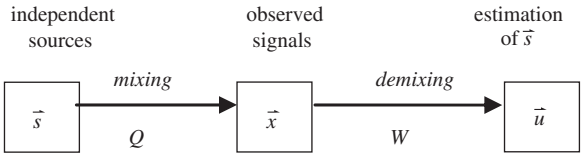

Fig. 1. The model of ICA.

To solve the above optimization problem, we introduce a Lagrangian coefficient λ to (5)

$$\text{kurt}(\vec{w}^{\mathrm{T}}\vec{y}) = E[(\vec{w}^{\mathrm{T}}\vec{y})^4] - 3(E[(\vec{w}^{\mathrm{T}}\vec{y})^2])^2 + \lambda(1 - E[(\vec{w}^{\mathrm{T}}\vec{y})^2]) \quad (7)$$

Differentiate Eq. (7) with respect to $\vec{w}$ and let it be 0, we have

$$\vec{w} = \frac{2}{\lambda}(H^{-1}E[\vec{y}(\vec{w}^{\mathrm{T}}\vec{y})^3] - 3\vec{w}) \quad (8)$$

where $H = E[\vec{y}\vec{y}^{\mathrm{T}}]$. $\vec{w}$ can be calculated iteratively as follows:

$$\vec{w}^*(t) = H^{-1}E[\vec{y}(\vec{w}(t-1)^{\mathrm{T}}\vec{y})^3] - 3\vec{w}(t-1) \quad (9)$$

$$\vec{w}(t) = \frac{\vec{w}^*(t)}{\sqrt{\vec{w}^*(t)^{\mathrm{T}}H\vec{w}^*(t)}} \quad (10)$$

Using Eqs. (9) and (10), we can calculate all the column vectors $\vec{w}_i$, $i = 1,\ldots, m$, of W. Finally, the demixing matrix W is obtained as $W = [\vec{w}_1 \cdots \vec{w}_m]$.

3. Row–column ICA (RC-ICA) for face recognition

3.1. Idea of RC-ICA

As mentioned in introduction, the conventional ICA stretches the face image matrix to a face vector and this matrix-to-vector transformation makes the statistics estimation difficult and inaccurate because the training sample size is relatively very small compared with the high dimensionality of the training vector. Actually, the same problem exists in the conventional PCA-based face recognition. To dilute this small sample size problem, Yang et al. [21] proposed a 2D-PCA scheme. They directly projected the 2D face image onto a set of vectors using $\vec{y} = A\vec{x}$, where A is the face image, $\vec{x}$ is the projection vector and $\vec{y}$ is the projected vector. 2D-PCA computes a set of leading vectors $\vec{x}_i$, $i = 1, 2, \ldots, m$, of A to span the subspace where face recognition is performed.

A second look on the implementation process of 2D-PCA can reveal that 2D-PCA actually takes each row of the face image as a training vector and then finds a common subspace which applies to all row vectors. Since the dimension of each row vector is significantly less than that of a face image vector, the small training sample size problem is diluted. The experimental results in [21] and the following papers [11,12] validate that 2D-PCA achieves better face recognition performance than PCA, especially when the training sample size is small.

The success of 2D-PCA inspires us to develop a new ICA-based face recognition technique without image-to-vector transformation. We view the 2D face image as a super-class and each row of it as a sub-class. Instead of finding a demixing matrix W for the whole face image, we find a common demixing matrix W_{r} for all the row sub-classes. The row-demixing matrix W_{r} will make the elements within a row vector independent but there are still some dependencies between the row sub-classes. In the second step, the demixed row vectors are reorganized as column vectors and a column-demixing matrix W_{c} is computed to further remove the dependencies. Since the proposed scheme has two steps, i.e. a row-ICA followed by a column-ICA, we call it row–column ICA (RC-ICA).

Fig. 2 illustrates the procedure of the proposed RC-ICA. The face image A is viewed as a set of row sub-classes and is transformed by using the row-demixing matrix W_{r}. The dimensionality of the transformed output is reduced in horizontal direction to get matrix B. Dimensionality reduced matrix B is viewed as a set of column sub-classes and is transformed by using the column-demixing matrix W_{c}. The dimensionality of the output is then reduced in vertical direction to get the final result S, whose dimension is much lower than the original face image A. The ICs in S will be used for face representation and recognition.

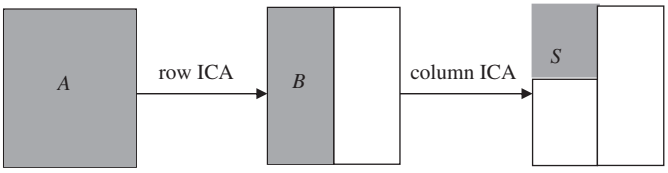

Fig. 2. Procedure of the proposed RC-ICA.

3.2. Implementation of RC-ICA

3.2.1. Row-ICA

Denote by $A \in R^{m\times n}$ a face image and by $T = \{A_i |\ i = 1, 2, \ldots, K\}$ the training set, which has K face image samples. We rewrite A as $A = [\vec{a}_1^{\mathrm{T}}\ \vec{a}_2^{\mathrm{T}}\ \cdots\ \vec{a}_m^{\mathrm{T}}]^{\mathrm{T}}$, where $\vec{a}_j, j = 1, 2, \ldots, m$, is the jth row of A. Unlike in ICA, where the whole face A is directly trained by stretching it to a $m \times n \times 1$ vector, in the first step of RC-ICA, we train the row vectors $\vec{a}_j$. The training set can be rewritten as $T = \left\{\vec{a}_j^i \in R^{1\times n} | i = 1, 2, \ldots, K;\quad j = 1, 2, \ldots, m\right\}$, where $\vec{a}_j^i$ means it is the jth row of the ith face sample. We define a common covariance matrix $G \in R^{n\times n}$ of vector $\vec{a}_j$ as

$$G = \frac{1}{m}\sum_{j=1}^{m} E\left[\left(\vec{a}_j - \bar{\vec{a}}_j\right)^{\mathrm{T}}\left(\vec{a}_j - \bar{\vec{a}}_j\right)\right] = \frac{1}{mK}\sum_{j=1}^{m}\sum_{i=1}^{K}\left(\vec{a}_j^i - \bar{\vec{a}}_j\right)^{\mathrm{T}}\left(\vec{a}_j^i - \bar{\vec{a}}_j\right) \quad (11)$$

where $\bar{\vec{a}}_j = (1/K)\sum_{i=1}^{K}\vec{a}_j^i$. In another form, G can be written as

$$G = \frac{1}{mK}\sum_{i=1}^{K}(A_i - \bar{A})^{\mathrm{T}}(A_i - \bar{A}) \quad (12)$$

where $\bar{A} = \left[\bar{\vec{a}}_1^{\mathrm{T}}\ \bar{\vec{a}}_2^{\mathrm{T}}\ \cdots\ \bar{\vec{a}}_m^{\mathrm{T}}\right]^{\mathrm{T}}$.

Let $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ be the eigenvalues of G and $\vec{v}_1, \vec{v}_2, \ldots, \vec{v}_n$ the corresponding eigenvectors. The whitened matrix P on face image A is

$$P = V\Lambda^{-1/2} \quad (13)$$

where $V = \left[\vec{v}_1\ \vec{v}_2\ \cdots\ \vec{v}_l\right]$ and $\Lambda = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_l)$. We denote by $\vec{z}_j^i$ the whitened vector of each training sample $\vec{a}_j^i$

$$\vec{z}_j^i = \left(\vec{a}_j^i - \bar{\vec{a}}_j\right)P \quad (14)$$

The demixing matrix $W \in R^{n\times l}$ of the whitened data is computed by taking $\vec{z}_j^i$ as inputs. $l \leqslant n$ is the number of ICs we set for $\vec{a}_j^i$. The Fast-ICA algorithm introduced in Section 2 is employed to compute W from $\vec{z}_j^i$. For expression convenience and to be consistent with ICA, we let $W_{\mathrm{r}} = P \cdot W$ and call it the demixing matrix of the original data $\vec{a}_j$.

Projecting the original image A onto W_{r}, we get the row-ICA output as

$$B = \left[\vec{b}_1^{\mathrm{T}}\ \vec{b}_2^{\mathrm{T}}\ \cdots\ \vec{b}_m^{\mathrm{T}}\right]^{\mathrm{T}} = (A - \bar{A}) \cdot W_{\mathrm{r}} \quad (15)$$

where $\vec{b}_j = \left(\vec{a}_j - \bar{\vec{a}}_j\right) W_r$ is the ICs of the jth row of image A. Compared with A, whose dimension is $m \times n$, the row-ICA output B preserves the most important ICs of A but with a smaller dimension $m \times l$. (The feature section and dimensionality reduction for matrix W_r will be discussed in Section 3.2.3.) However, this dimension is still relatively large to represent a face image because of the dependencies between the rows of B. Next we use a subsequent column-ICA to further reduce the dependencies existed in B so that a more condensed representation of A can be obtained.

3.2.2. Column-ICA

After row-ICA in the previous section, the $m \times n$ face image A is transformed into a $m \times l$ image B. For each row of B, i.e. $\vec{b}_j$, $j = 1, 2, \ldots, m$, its l elements are independent. However, for the elements in different rows, the row-ICA does not remove their dependencies. Suppose $\vec{b}_{j_1}(q_1)$ and $\vec{b}_{j_2}(q_2)$ are two elements in different rows, where $j_1 \neq j_2$, they are not independent in general. We explain this conclusion as follows.

If $\vec{b}_{j_1}(q_1)$ and $\vec{b}_{j_2}(q_2)$ are independent, we have

$$E\left[\vec{b}_{j_1}(q_1) \cdot \vec{b}_{j_2}(q_2)\right] = E\left[\vec{b}_{j_1}(q_1)\right] E\left[\vec{b}_{j_2}(q_2)\right] \tag{16}$$

Because $E\left[\vec{b}_{j_1}(q_1)\right] = E\left[\left(\vec{a}_{j_1} - \bar{\vec{a}}_{j_1}\right)\right] \cdot \vec{w}_{q_1} = 0$ and $E\left[\vec{b}_{j_2}(q_2)\right] = E\left[\left(\vec{a}_{j_2} - \bar{\vec{a}}_{j_2}\right)\right] \cdot \vec{w}_{q_2} = 0$, where $\vec{w}_{q_1}$ and $\vec{w}_{q_2}$ are the q_1th and q_2th columns in W_r, we have $E\left[\vec{b}_{j_1}(q_1) \cdot \vec{b}_{j_2}(q_2)\right] = 0$. Then it can be seen that

$$E\left[\vec{b}_{j_1}(q_1) \cdot \vec{b}_{j_2}(q_2)\right] = \vec{w}_{q_1}^{\mathrm{T}} \cdot E\left[\left(\vec{a}_{j_1} - \bar{\vec{a}}_{j_1}\right)^{\mathrm{T}} \left(\vec{a}_{j_2} - \bar{\vec{a}}_{j_2}\right)\right] \times \vec{w}_{q_2} = 0 \tag{17}$$

In general, $\vec{w}_{q_1}$ and $\vec{w}_{q_2}$ are not zero vectors and the rows $\vec{a}_{j_1}$ and $\vec{a}_{j_2}$ in the original image A are correlated, i.e. $E\left[\left(\vec{a}_{j_1} - \bar{\vec{a}}_{j_1}\right)^{\mathrm{T}} \left(\vec{a}_{j_2} - \bar{\vec{a}}_{j_2}\right)\right] \neq 0$. Only in very rarely cases, (17) will hold. Therefore, (16) will not hold in general, which means that the elements $\vec{b}_{j_1}(q_1)$ and $\vec{b}_{j_2}(q_2)$ are not independent.

The above conclusion implies that the feature matrix B still contains redundancies between different rows. B can be further compressed to get a more condensed representation of the face image. We rewrite B as

$$B = \left[\begin{array}{cccc} \vec{b}_1^{\mathrm{T}} & \vec{b}_2^{\mathrm{T}} & \cdots & \vec{b}_m^{\mathrm{T}} \end{array}\right]^{\mathrm{T}} = \left[\begin{array}{cccc} \vec{c}_1 & \vec{c}_2 & \cdots & \vec{c}_l \end{array}\right] \tag{18}$$

where $\vec{c}_q$, $q = 1, 2, \ldots, l$, is a $m \times 1$ column vector containing the qth elements of $b_1, b_2, \ldots, b_m$. Similar to Section 3.2.1, where we found a left multiplication matrix W_r to demix row vectors $\vec{a}_1, \vec{a}_2, \ldots, \vec{a}_m$, here we can find a right multiplication matrix W_c to demix column vectors $\vec{c}_1, \vec{c}_2, \ldots, \vec{c}_l$:

$$S = W_c \cdot B \tag{19}$$

W_c is called the column-ICA demixing matrix. By setting $W_c \in R^{p \times m}$, where $p < m$ is the number of ICs we want to preserve, we can further reduce the dimension of B from to $m \times l$ to $p \times l$.

3.2.3. Subspace selection for feature extraction

With ICA, the discriminability of each column of W_r, or each row of W_c, is not known in prior [2,19]. In order to improve the classification accuracy and further reduce the dimensionality of features, we select the subspaces of W_r and W_c based on the discriminability of their columns and rows. Here we take W_c for example and the similar procedures are applied to W_r. The discriminability of each row of W_c is defined as follows [2]:

$$r = \frac{\sigma_b}{\sigma_w} \tag{20}$$

where $\sigma_b = \sum_j (\bar{s}_j - \bar{s})^{\mathrm{T}} (\bar{s}_j - \bar{s})$ and $\sigma_w = \sum_j \sum_{s \in C_j} (s - \bar{s}_j)^{\mathrm{T}} (s - \bar{s}_j)$ represent the between-class and the within-class scatter, respectively, of the projected vectors s of the training images. $\bar{s}$ denotes the global mean of the projected vectors and $\bar{s}_j$ denotes the mean for the jth class. Based on the magnitude of r, a sub-matrix of W_c, denoted by $U \in R^{k \times m}$, can be determined, where $k (k \leqslant m)$ is the number of selected columns.

Similarly, a sub-matrix of W_r, denoted by $V \in R^{n \times l}$ $(l \leqslant n)$, can be determined in the stage of row-ICA. Given an arbitrary original face image matrix A, by using RC-ICA with demixing matrices U and V, we can extract the ICs of A as

$$S = U(A - \bar{A})V \tag{21}$$

S is much condensed than A while preserving the most important ICs of A. The classification and recognition will then be based on the RC-ICA representation S.

In [19], Vicente et al. analyzed the equivalence of some common linear feature extraction techniques for appearance-based object recognition. From the above development, we can see that if we denote by $\Gamma = P_c(A - \bar{A})P_r$ the PCA-based whitening output of the original data, then the RC-ICA transformation can be denoted as $S = T_c \Gamma T_r$, where T_c and T_r denote the demixing operation by ICA. Since T_c and T_r are orthogonal matrices, if we preserve all the components in S, by using a linear classifier the classification result with S will be equivalent to that with Γ [19]. However, with the feature selection processing (refer to Eqs. (20) and (21)) in the RC-ICA transformation domain, only the selected highly discriminative features in S are used for classification so that better recognition results can be expected even using a linear classifier.

3.3. Classification

After projecting each training images A_i $(i = 1, 2, \ldots, K)$ onto the subspace determined by demixing matrices W_r and W_c, we obtain the independent feature matrices S_i for image A_i. Given an input face image I to be recognized, we can compute its independent feature matrix S^* by using Eq. (21). The nearest neighbor classifier is used for classification. The Euclidian distance between S and S^* is defined by

$$d(S^*, S_i) = \sqrt{\sum_{j=1}^{k} \sum_{k=1}^{q} \left(S^*(j,k) - S_i(j,k)\right)^2} \tag{22}$$

If $d(S^*, S_t) = \min_i d(S^*, S_i)$, then the input image I is judged belonging to the class of face image A_t. It should be noted other distance metrics other than Euclidian distance can also be used, such as cosine, L1 and Mahalanobis distances.

4. Experimental results

The proposed RC-ICA algorithm is tested using three well-known face databases: Yale-B, FERET and AR. The Yale-B database is used to evaluate the performance of the proposed method under the variations of illumination. The FERET database is employed to test the performance when there are variations over time and facial expressions. The AR database is used to examine

the proposed method when there are facial variations over time, different facial expressions, illumination and occlusion, etc.

The popular methods such as ICA, EICA [13], PCA (eigenface) [18], 2D-PCA [21], BDPCA [12] and the whitened BDPCA (W-BDPCA) are used for comparison. In the following experiments, the dimensionality of each low-dimensional subspace varies from 1 to 100 in ICA and PCA, from 1 to 15 in 2D-PCA, BDPCA, W-BDPCA, row-ICA and RC-ICA.

4.1. Experiments on variations of illumination using Yale-B database

We select the Yale-B database (http://cvc.yale.edu/projects/yalefacesB.html) to test the performance of RC-ICA on variations of illumination. Yale-B database contains 5760 single light source images of 10 individuals, each providing 576 viewing conditions (9 different poses and 64 different lighting conditions from negative azimuth to positive azimuth). The objective of this section is to evaluate the performance of the proposed RC-ICA with other well-known unsupervised subspace methods under the variations of illumination. We select 64 frontal images under different lighting conditions for each person in our experiments. Thus, there are 640 images for 10 subjects with each image being manually cropped and resized to 171×128. Fig. 3 shows some images of one subject.

We first compare the proposed row-ICA and RC-ICA with traditional ICA and EICA. The nearest neighbor classifier with Euclidian distance is used for classification. In the first experiment, we select the images of 35 negative azimuths for training and use the remaining images of 29 positive azimuths for testing. Thus the total number of training images is 350 and the number of testing images is 290. Fig. 4 plots the correct recognition rate (CRR) curves of ICA, EICA, row-ICA and RC-ICA versus the number of used features. In the second experiment the training and testing datasets are exchanged. The CRR curves of four methods are also plotted in Fig. 4. Table 1 lists the top CRR values, the corresponding dimensions of feature vector, and the sizes of demixing matrices for these methods in the two experiments.

Fig. 4 shows that the proposed row-ICA and RC-ICA methods achieve better performance than the standard ICA and EICA methods. From Table 1, we know the top recognition rates of row-ICA, RC-ICA, ICA and EICA are 90.69% (92.57%), 91.38% (92.57%), 82.41% (84.57%) and 86.90% (85.14%), respectively. (The values in parentheses denote the results of the second experiment.) Compared with ICA, RC-ICA improves approximately 8% in top recognition accuracy with less dimensionality of IC features. Row-ICA also achieves better recognition accuracy than ICA but it needs more features. RC-ICA has similar recognition accuracy to row-ICA but the required feature dimensionality is significantly reduced. What's more, the full demixing matrices in ICA and EICA

Fig. 3. Some images of one subject in Yale-B database.

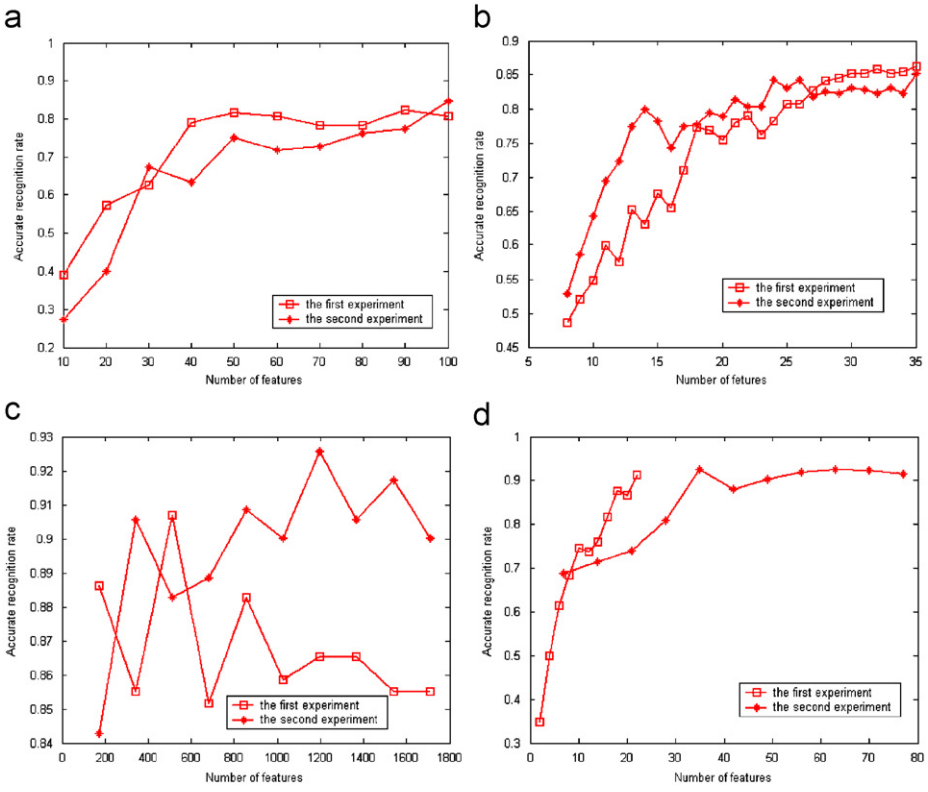

Fig. 4. Recognition accuracies of (a) ICA; (b) EICA; (c) row-ICA and (d) RC-ICA.

Table 1
The top recognition accuracies, the corresponding dimensions of feature vectors and the size of full demixing matrices for ICA, EICA, row-ICA, and RC-ICA

Algorithm	Top recognition accuracy (%)	Dimension of features	Size of full demixing matrix
ICA	82.41 (84.57)	90 (100)	$21{,}888 \times 21{,}888$
EICA	86.90 (85.14)	35 (34)	$21{,}888 \times 21{,}888$
Row-ICA	90.69 (92.57)	513 (1368)	128×128
RC-ICA	91.38 (92.57)	22 (35)	128×128 and 171×171

The values in parentheses are the result for the second experiment.

Table 2
Comparison of RC-ICA with other methods on Yale-B database

Algorithms	Top recognition accuracy (%)	Dimension of features
PCA	79.31 (80.29)	42 (45)
2DPCA	82.76 (86.88)	2223 (2394)
BDPCA	82.76 (88.57)	105 (190)
W-BDPCA	87.59 (91.14)	49 (78)
RC-ICA	91.38 (92.57)	22 (35)

The values in parentheses are the result for the second experiment.

are of size $21{,}888 \times 21{,}888$ ($171 \times 128 = 21{,}888$), while the full demixing matrix is of size 128×128 for the row-ICA, and the full demixing matrices are of sizes 128×128 and 171×171 for RC-ICA (the former is for row direction and the later is for column direction). The proposed methods are fast for feature extraction than ICA and EICA.

We then compare the performance of RC-ICA with PCA, 2D-PCA, BDPCA and W-BDPCA. Similarly, two experiments by exchanging the training and testing datasets are performed. The top recognition accuracy results using different methods and the corresponding dimension of features are listed in Table 2. RC-ICA is the best one among them.

4.2. Experiments on partial FERET database

This partial FERET face database comprises 400 gray-level frontal view face images from 200 persons. Each person has two images (fa and fb) which are obtained at different times and with different facial expressions. All the images are cropped manually to the size of 40×40. In the experiment, the fa images are used as gallery for training while the fb images as probes for testing. PCA, 2D-PCA, BDPCA, W-BDPCA, ICA, EICA and the proposed row-ICA and RC-ICA methods are used for feature extraction, and then a nearest neighbor classifier is employed for classification. Fig. 5 plots the CCR curves of these methods. Table 3 lists the top CRR values, as well as the corresponding dimensions of feature vector used in these methods. It can be seen from Fig. 5 and Table 3 that RC-ICA achieves the highest recognition accuracy with a relatively small set of features.

4.3. Experiments on AR database

We use the AR database to evaluate the performance of RC-ICA under other cases including variation of illumination and facial expression. AR database contains over 4000 color face images from 126 people (70 men and 56 women), including frontal views of faces with different facial expression, lighting conditions and occlusions. The pictures of most persons were taken in two sessions, separated by two weeks. Each session contains 13 color images per person and 120 individuals

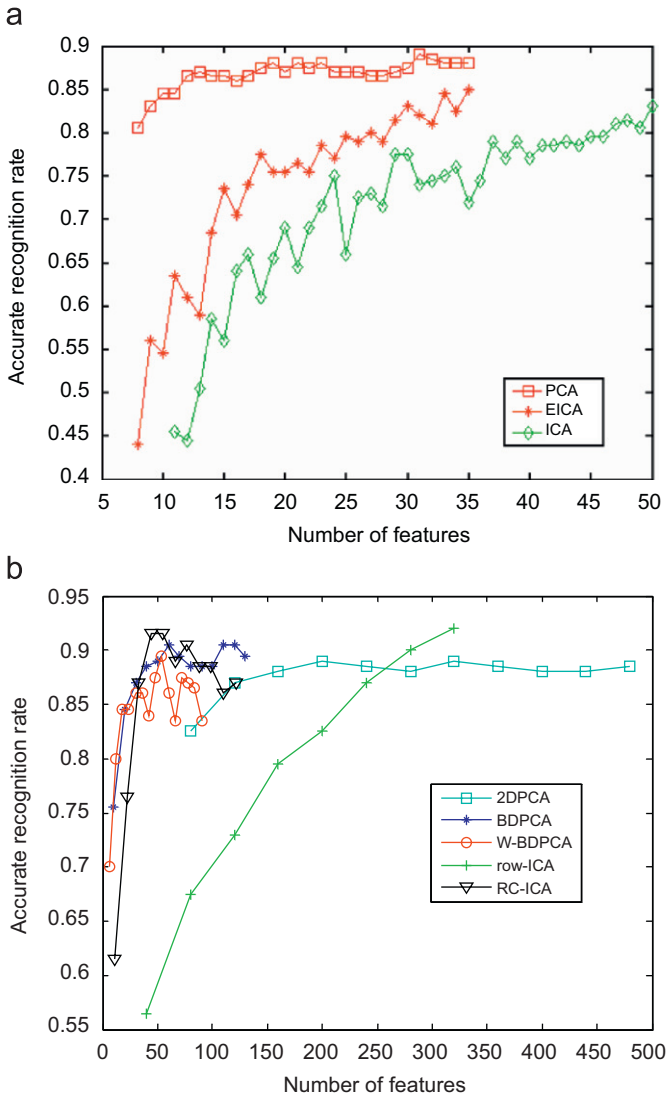

Fig. 5. Recognition accuracies of ICA, EICA, PCA, 2D-PCA, BDPCA, W-BDPCA, row-ICA and RC-ICA.

Table 3
Comparison of RC-ICA with other methods using fafb database

Method	PCA	2DPCA	BDPCA	W-BDPCA	ICA	EICA	RC-ICA
Recognition (%)	89.00	89.00	90.50	89.50	83.00	85.00	91.50
Dimension	31	400	60	48	50	35	44

(65 men and 55 women) participated in both sessions. In our experiment, the images of these 120 individuals are selected and used. The face portions of those images were manually cropped to 50×40 pixels. Fig. 6 shows the cropped images of one person.

To evaluate the performance of RC-ICA comprehensively, we perform experiments by varying the number of training samples per person. In the kth test, we randomly select k images for each person for training and use the remaining samples for testing. The ICA, EICA, PCA, 2D-PCA, BDPCA and W-BDPCA methods are used for comparison. The top CRR values of these methods at different number of training samples and corresponding number of features are listed in Table 4. It is shown that RC-ICA is always much better than ICA, and it works the best

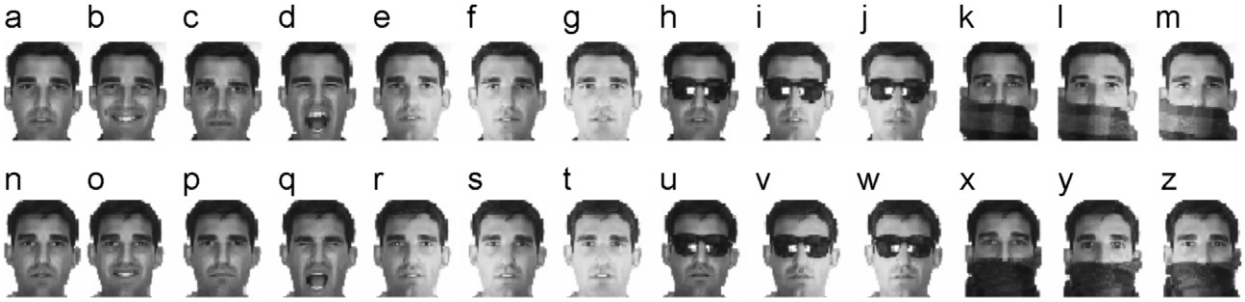

Fig. 6. Images for one subject of the AR database.

Table 4
Comparison of the recognition accuracies (%) of RC-ICA, PCA, ICA and 2DPCA on AR database

Training samples	PCA	2DPCA	BDPCA	W-BDPCA	ICA	EICA	RC-ICA
1	53.97 (49)	63.93 (550)	68.93 (240)	69.77 (108)	52.03 (100)	59.03 (50)	68.93 (108)
2	54.90 (50)	63.81 (500)	66. 60 (228)	70.07 (99)	58.37 (90)	62.74 (49)	69. 90 (121)
3	65.62 (50)	71.56 (500)	73.77 (330)	74.86 (132)	69.24 (90)	70.11 (47)	75.54 (130)
4	63.75 (49)	70.91 (500)	73.11 (330)	74.32 (144)	69.70 (90)	70.98 (49)	74.39 (91)
5	62.30 (49)	69.44 (450)	72.22 (270)	73.69 (144)	71.11 (89)	70.95 (44)	73.81 (100)
6	73.92 (50)	85.46 (700)	86.42 (315)	89.17 (120)	86.42 (100)	88.79 (88)	89.46 (120)
7	73.90 (49)	86.05 (700)	86.32 (294)	89.52 (120)	89.52 (99)	90.00 (87)	90.44 (98)
8	72.50 (49)	85.32 (650)	86.44 (294)	90.05 (120)	90.60 (97)	90.65 (88)	
9	70.94 (50)	85.25 (700)	86.37 (315)	90.64 (156)	89.90 (97)	90.00 (90)	92.01 (77)

The values in parentheses are the corresponding number of features.

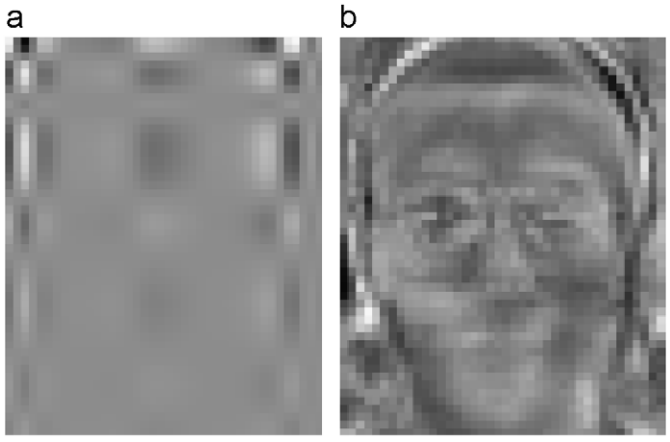

Fig. 7. The basis images of (a) the proposed RC-ICA and (b) the conventional ICA.

among all the methods except when only 1 and 2 training samples are used.

4.4. Face representation

In this section, we present face image using the extracted features to illustrate what kind of features are captured by the proposed method. We used the training images and test images in the 9th test on the AR database in Section 4.3. As it can be seen from Table 4, the proposed method has the top accuracy recognition 92.01% with 77 features, while U and V are of the size 7×50 and 11×40, respectively. Then the basis image can be calculated as

$$B\_I = \sum_j \sum_i U(j,:)^{\mathrm{T}} \times V(i,:)$$

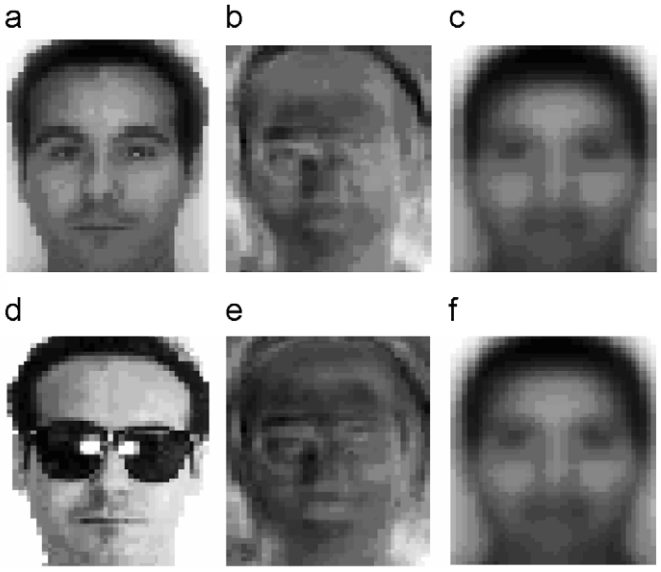

Fig. 8. Original images and reconstructed images. (a) and (d) are two original images selected from the training dataset and test set respectively; (b) and (e) are the reconstructed images of (a) and (d) by using ICA; (c) and (f) are the reconstructed images of (a) and (c) by proposed RC-ICA.

Similarly, the basis image corresponding to the ICA algorithm can be easily calculated.

Fig. 7 shows the basis images corresponding to the features extracted by the propose RC-ICA and conventional ICA. As can be seen, the basis image of ICA looks more like a face than the basis image of RC-ICA because in RC-ICA, the rows and columns, instead of the original face image, are used for training and feature extraction. The basis image for RC-ICA can capture more local features. Fig. 8 shows some example-reconstructed images by the conventional ICA and the proposed RC-ICA. We see that the

reconstructed images by ICA look more likes the original images, which may, however, lead to incorrect classification when some occlusion (e.g. the glass) occurs in the face. The RC-ICA is more suitable and stable for feature extraction and classification.

5. Conclusion and discussions

A new face recognition method called sequential row-column independent component analysis (RC-ICA) was presented in this paper. RC-ICA works directly on the rows and columns of images without image-to-vector stretching as in traditional ICA. RC-ICA has two sequential steps: a row-ICA followed by a column-ICA. In each step, the dimensionality of the training vector is much smaller and the number of training samples is much greater compared with the traditional ICA. Therefore, RC-ICA significantly dilutes the dilemma in ICA: the dimensionality of training vector is very high but the size of training samples is relatively very small. Extensive experiments on Yale-B, AR and FERET databases were conducted to validate the performance of the proposed RC-ICA scheme. The results show that RC-ICA has three advantages over traditional ICA: higher recognition accuracy; lower storage space; and lower computation load.

In each of the two sequential steps, RC-ICA takes image rows and columns as training samples to compute the demixing matrices separately. The computed two demixing matrices (right and left multiplication matrices respectively) only reflect the variations within rows or columns of the face images. Theoretically, demixing by the two matrices will not make the output face features completely independent. Can we find a better transformation than RC-ICA, which can preserve the advantages of RC-ICA while make the face features be more independent? This will be our future research work.

References

[1] F.R. Bach, M.I. Jordan, Kernel independent component analysis, J. Mach. Learn. Res. 3 (2002) 1–48.
[2] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, IEEE Trans. Neural Networks 13 (6) (2002) 1450–1464.
[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 (1997) 711–720.
[4] P. Comon, Independent component analysis, a new concept?, Signal Process. 36 (1994) 287–314.
[5] J.M. Górriz, C.G. Puntonet, F. Rojas, R. Martin, S. Hornillo, E.W. Lang, Optimizing blind source separation with guided genetic algorithms, Neurocomputing 69 (2006) 1442–1457.
[6] A. Hyvärinen, Survey on independent component analysis, Neural Comput. Surveys 2 (1999) 94–128.
[7] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Comput. 9 (7) (1997) 1483–1492.
[8] A. Hyvärinen, E. Oja, Independent component analysis: algorithms and applications, Neural Networks 13 (2000) 411–430.
[9] J. Kim, J. Choi, J. Yi, M. Turk, Effective representation using ICA for face recognition robust to local distortion and partial occlusion, IEEE Trans. Pattern Anal. Mach. Intell. 27 (12) (2005) 1977–1981.
[10] M. Kirby, L. Sirovich, Application of the KL procedure for the characterization of human faces, IEEE Trans. Pattern Anal. Mach. Intell. 12 (1) (1990) 103–108.
[11] H. Kong, L. Wang, E.K. Teoh, X. Li, J.G. Wang, R. Venkateswarlu, Generalized 2D principal component analysis for face image representation and recognition, Neural Networks 18 (Special Issue) (2005) 585–594.
[12] H. Kong, X.C. Li, L. Wang, E.K. Teoh, J.G. Wang, Generalized 2D Principal component analysis, in: Proceedings of International Joint Conference on Neural Networks, Montreal, Canada, July 2005.
[13] C. Liu, Enhanced independent component analysis and its application to content based face image retrieval, IEEE Trans. System, Man Cybern. Part B 34 (2) (2004) 1117–1127.
[14] C. Liu, H. Wechsler, Independent component analysis of gabor features for face recognition, IEEE Trans. Neural Networks 14 (4) (2003) 919–928.
[15] B. Moghaddam, Principal manifolds and probabilistic subspaces for visual recognition, IEEE Trans. Pattern Anal. Mach. Intell. 24 (2002) 780–788.
[16] Z. Shi, C. Zhang, Semi-blind source extraction for fetal electrocardiogram extraction by combining non-Gaussianity and time-correlation, Neurocomputing 70 (7–9) (2007) 1574–1581.
[17] K.-K. Sung, T. Poggio, Example-based learning for View-based Human face detection, IEEE Trans. Pattern Anal. Mach. Intell. 20 (1) (1998) 39–51.
[18] M. Turk, A. Pentland, Eigenfaces for recognition, J. Cognitive Neurosci. 3 (1) (1991) 72–86.
[19] M.A. Vicente, P.O. Hoyer, A. Hyvarinen, Equivalence of some common linear feature extraction techniques for appearance-based object recognition tasks, IEEE Trans. Pattern Anal. Mach. Intell. 29 (5) (2007) 896–900.
[20] X. Wang, X. Tang, A unified framework for subspace face recognition, IEEE Trans. Pattern Anal. Mach. Intell. 26 (9) (2004) 1222–1228.
[21] 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.
[22] J. Yang, D. Zhang, J.Y. Yang, Is ICA significantly better than PCA for face recognition? in: The 10th IEEE International Conference on Computer Vision, vol. 11, 2005, pp. 198–203.
[23] P.C. Yuen, J.H. Lai, Face representation using independent component analysis, Pattern Recogn. 135 (2002) 1247–1257
T.K. Kim, H. Kim, W. Hwang, independent component analysis in a facial local residue space, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03), 2003.
[24] C. Zheng, D. Huang, L. Shang, Feature selection in independent component subspace for microarray data classification, Neurocomputing 69 (2006) 2407–2410.

Quanxue Gao received his Ph.D. degrees from the College of Automation, Northwestern Polytechnical University, Xi'an, China, in 2005. He is a lecture at School of Telecommunication Engineering, Xidian University, China. He is now working as a research associate in the Department of Computing, The Hong Kong Polytechnic University. His current research interests include face recognition and image processing.

Lei Zhang received the B.S. degree in 1995 from Shenyang Institute of Aeronautical Engineering, Shenyang, PR China, the M.S. and Ph.D. degrees in Electrical and Engineering from Northwestern Polytechnical University, Xi'an, P.R. China, respectively in 1998 and 2001. From 2001 to 2002, he was a research associate in the Dept. of Computing, The Hong Kong Polytechnic University. From January 2003 to January 2006 he worked as a Postdoctoral Fellow in the Department of Electrical and Computer Engineering, McMaster University, Canada. Since January 2006, he has been an Assistant Professor in the Department of Computing, The Hong Kong Polytechnic University. His research interests include Image and Video Processing, Biometrics, Pattern Recognition, Multisensor Data Fusion and Optimal Estimation Theory, etc.

David Zhang graduated in computer science from Peking University in 1974 and received his M.Sc. and Ph.D. degrees in Computer Science and Engineering from the Harbin Institute of Technology (HIT), Harbin, PR China, in 1983 and 1985, respectively. He received the second Ph.D. degree in Electrical and Computer Engineering at the University of Waterloo, Waterloo, Canada, in 1994. From 1986 to 1988, he was a Postdoctoral Fellow at Tsinghua University, Beijing, China, and became an Associate Professor at Academia Sinica, Beijing, China. Currently, he is a Professor with the Hong Kong Polytechnic University, Hong Kong. He is Founder and Director of Biometrics Research Centers supported by the Government of the Hong Kong SAR (UGC/CRC). He is also Founder and Editor-in-Chief of the International Journal of Image and Graphics (IJIG), Book Editor, The Kluwer International Series on Biometrics, and an Associate Editor of several international journals. His research interests include automated biometrics-based authentication, pattern recognition, biometric technology and systems. As a principal investigator, he has finished many biometrics projects since 1980. So far, he has published over 200 papers and 10 books.