Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT

PERGAMON

Pattern Recognition 37 (2004) 1873–1885

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY
www.elsevier.com/locate/patcog

Independent component analysis in a local facial residue space for face recognition

Tae-Kyun Kim[a,*], Hyunwoo Kim[a], Wonjun Hwang[a], Josef Kittler[b]

[a] *Human Computer Interaction Laboratory, Samsung Advanced Institute of Technology, San 14-1, Nongseo-ri, Kiheung-eup, Yongin, Kyungki-do 449-712, Republic of Korea*

[b] *Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, Surrey, GU2 7XH, UK*

Received 13 January 2003; received in revised form 18 September 2003; accepted 26 January 2004

Abstract

In this paper, we propose an Independent Component Analysis (ICA) based face recognition algorithm, which is robust to illumination and pose variation. Generally, it is well known that the first few eigenfaces represent illumination variation rather than identity. Most Principal Component Analysis (PCA) based methods have overcome illumination variation by discarding the projection to a few leading eigenfaces. The space spanned after removing a few leading eigenfaces is called the "residual face space". We found that ICA in the residual face space provides more efficient encoding in terms of redundancy reduction and robustness to pose variation as well as illumination variation, owing to its ability to represent non-Gaussian statistics. Moreover, a face image is separated into several facial components, local spaces, and each local space is represented by the ICA bases (independent components) of its corresponding residual space. The statistical models of face images in local spaces are relatively simple and facilitate classification by a linear encoding. Various experimental results show that the accuracy of face recognition is significantly improved by the proposed method under large illumination and pose variations.
© 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Face recognition; Feature extraction; ICA; PCA; Illumination invariance; Pose invariance; Eigenfaces; Residual space; Facial components; Local space

1. Introduction

Recently, methods for face recognition and facial expression analysis have been extensively developed. In video processing and analysis, the human face is a key object of interest for visual discrimination and identification. For face retrieval and person identification in video streams, face images should be described by a compact and discriminative feature set. Challenging problems are listed as follows: (1) features should be insensitive to large variations of light and pose, (2) the matching complexity should be kept low for applications involving huge databases on the Internet, and (3) no prior knowledge about a query person is given and this means that the statistics for feature extraction should be previously learned from training groups which do not have images of the query person. From these points of view, efficient feature extraction and selection schemes are crucial for a successful face recognition system.

Based on the observation that the principal components corresponding to leading eigenvalues represent illumination variation rather than person identity [1–3], as the variations between face images of the same person due to illumination are almost always larger than image variations reflecting the changes in face identity, eigenfaces excluding the first few eigenvectors have been generally used for face recognition. Wang and Tan [4] introduced the second-order PCA method. In their method, the images reconstructed from the leading principal components are subtracted from the input images. The difference is called "residual" images which contain

* Corresponding author. Human Computer Interaction Laboratory, Samsung Advanced Institute of Technology, San 14-1, Nongseo-ri, Kiheung-eup, Yongin, Kyungki-do 449-712, Republic of Korea. Tel.: +82-31-280-1746; fax: +82-31-280-9257.
E-mail address: taekyun@sait.samsung.co.kr (T.-K. Kim).

0031-3203/$30.00 © 2004 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2004.01.019

high-frequency components. They are more insensitive to illumination change. They performed a second PCA on the residual images and referred to it second-order PCA.

Recently the second-order PCA was adopted for face recognition in Ref. [5]. However, the second principal components are the same as a subset of the original principal components, as shown in the appendix. That is, there is no difference between the method of second-order PCA and the conventional eigenface method which discards a few leading principal components. Besides the second-order PCA, there are several conventional methods [6–8] which utilize the residual space. However, they only utilize the magnitude of a residual vector and stop analyzing the residual space further.

Compared to the face feature extraction methods like PCA [4–6,18,26], fisher linear discriminant (FLD) [1,3,23,25], and local feature analysis (LFA) [27], which consider only second-order statistics of face images, ICA provides a better representation of face images for recognition by virtue of exploiting high-order statistics of the input face data [9,10,16,17,19–22,24]. Bartlett [9] insisted that much important information for face recognition was contained in high-order statistics of images. Similarly, ICA was compared to PCA in terms of face recognition performance in Ref. [10]. The study demonstrated that ICA delivers better results than PCA in some experimental conditions. However, it is noted that ICA does not always outperform PCA. This depends on the given training or test database. In the case of Gaussian distributions, ICA, which is based on high-order statistics, loses its merit. There are a number of factors making a face space to be non-Gaussian distributed and consequently ICA encodes the space better than PCA. Pose variation in face database is probably one major factor.

To overcome the challenges arising from geometrical variations in face data, several local feature schemes, which represent a face image as the collection of facial component features, have been developed. In Ref. [11], Nefian and Hayes used an embedded HMM for face modeling. Facial components are modeled by HMMs implicitly and they are matched with the image blocks segmented from a face image using an optimization technique. In Ref. [12], Wiskott et al. represented facial components by labeled graphs based on a Gabor wavelet transform and the phase of the complex Gabor wavelet coefficients which estimates the location of the nodes of the corresponding facial features accurately. In Ref. [13], Heisele et al. worked with facial components to compensate for pose change. A geometrical configuration classifier based on the support vector machine (SVM) approach was then applied. Even if each component is not aligned to reflect geometrical variations, local encoding schemes can benefit from data decomposition. In a divided local space, the data distributions become simpler and can be efficiently captured by a linear encoding.

In this paper, independent components, which form non-orthogonal axes, describe the residual spaces of local facial components. In these spaces, a face image is represented by a collection of independent component features to achieve robust recognition to illumination and pose changes. Whereas the principal components in residual spaces are only a subset of the original principal components, ICA of residual images generates a new set of independent components. These new components are more suitable for robust face representation to illumination changes than the conventional ICA, because the source of the ICA, residual images are more insensitive to light changes. The proposed method deals with pose variation as well as illumination change by utilizing high-order statistics. It was observed that ICA outperforms PCA in the recognition of face images subject to large pose variations and this is because pose variation makes a face space more complex not just Gaussian-distributed. ICA in residual spaces retains the benefits of both robustness to illumination of residual spaces and robustness to pose by capturing high-order statistics. In addition, advantages of local features are also exploited by adopting ICA in the residual spaces of localized facial components. We separate a face image into several facial components such as eyes, nose and mouth, and each local space is represented by the independent components of its corresponding residual space. ICA learns more effective axes, yielding features with simplified statistical structures which are amenable to a linear class separation.

ICA in a residual space is explained in Section 2 and a comparison of PCA and ICA in the residual space is given in Section 3. Section 4 explains the independent component analysis in local spaces. In Section 5, feature selection and similarity matching for face recognition are explained. In Section 6, experimental results supporting the claimed behavior of the proposed method under illumination and pose changes are presented.

2. ICA in face residual space

ICA is a generalization of PCA in the sense that ICA decorrelates the high-order moments of input while PCA encodes the second-order moments only. In the task of face recognition, ICA can be superior to PCA owing to its ability to represent the high-order statistics of face images. While the reconstructed face images with a few leading eigenfaces lose the details and look like low-pass-filtered versions, the corresponding residue images contain high-frequency components and are less sensitive to illumination variation. Since these residue images still contain rich information for the individual identities, face features are extracted from these residue faces by ICA.

2.1. Review on ICA representation

Suppose that we are given a set of M training images φ_i, $i = 1, \ldots, M$, each represented by an N-dimensional vector obtained by a raster scan. The mean vector of the image set is defined by $\mathbf{m} = (1/M) \cdot \sum_{i=1}^{M} \varphi_i$. After subtracting

the mean vector from all images, i.e., $\mathbf{x}_i = \varphi_i - \mathbf{m}$, we can construct an $N \times M$ matrix $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_M]$ with a zero mean and the covariance matrix, $\mathbf{X}\mathbf{X}^{\mathrm{T}}$. Generally, ICA aims to find an $N \times N$ invertible matrix $\mathbf{W}^{(0)}$ such that the rows of $\mathbf{U}^{(0)} = \mathbf{W}^{(0)}\mathbf{X}$ are statistically independent and the face images $\mathbf{X}$ are represented by independent columns $\mathbf{U}^{(0)}$, used as the basis images, i.e., $\mathbf{X} = \mathbf{W}^{(0)^{-1}}\mathbf{U}^{(0)}$. The ICA reconstruction of a face image $\mathbf{x}$ can be represented by a linear combination of the basis images u_i $(i = 1, \ldots, N)$.

In face recognition, ICA is generally applied to eigen-subspaces to control the number of independent components and to facilitate learning by reducing the dimensionality of the input space without the loss of high-order image statistics. We perform PCA on $\mathbf{X}$ and extract M eigenvalues and eigenvectors. Usually $M < N$. But when a large scale face recognition problem is met, M is larger than N and PCA yields N eigenvalues and eigenvectors. The first $M_1 \ll M$ eigenvectors corresponding to the largest eigenvalues are selected and the projection of the data on the M_1 leading eigenvectors $\mathbf{R}_{M_1}$ is computed as

$$\mathbf{R}_{M_1} = \mathbf{P}_{M_1}^{\mathrm{T}}\mathbf{X}, \tag{1}$$

where $\mathbf{P}_{M_1} = [\mathbf{p}_1, \ldots, \mathbf{p}_{M_1}]$ is the set of the selected eigenvectors and $\mathbf{p}_i$ denotes the eigenvector corresponding to the ith largest eigenvalue.

ICA is performed on $\mathbf{P}_{M_1}^{\mathrm{T}}$ instead of $\mathbf{X}$. It gives M_1 independent basis images $\mathbf{U}_{M_1}$, represented by

$$\mathbf{U}_{M_1}^{\mathrm{T}} = \mathbf{W}_{M_1}\mathbf{P}_{M_1}^{\mathrm{T}}, \tag{2}$$

where $\mathbf{W}_{M_1}$ denotes an $M_1 \times M_1$ invertible matrix such that the rows of $\mathbf{U}_{M_1}$ are statistically independent. The weight matrix $\mathbf{W}_{M_1}$ is estimated by Lee and Sejnowski's algorithm [14]. The reconstructed face image $\hat{\mathbf{X}}$ is computed by multiplying both sides of Eq. (1) by $\mathbf{P}_{M_1}$ and it is represented by

$$\begin{aligned}\hat{\mathbf{X}} &= \mathbf{P}_{M_1}\mathbf{R}_{M_1} = \mathbf{P}_{M_1}\mathbf{P}_{M_1}^{\mathrm{T}}\mathbf{X} \\ &= \mathbf{U}_{M_1}(\mathbf{W}_{M_1}^{-1})^{\mathrm{T}}\mathbf{P}_{M_1}^{\mathrm{T}}\mathbf{X} \\ &= \mathbf{U}_{M_1}(\mathbf{P}_{M_1}\mathbf{W}_{M_1}^{-1})^{\mathrm{T}}\mathbf{X}.\end{aligned} \tag{3}$$

Note the reconstructed images $\hat{\mathbf{X}}$ are spanned by the independent basis images $\mathbf{U}_{M_1}$ and are represented by the ICA coefficients $(\mathbf{P}_{M_1}\mathbf{W}_{M_1}^{-1})^{\mathrm{T}}\mathbf{X}$ denoted by $\mathbf{B}_{M_1}$. As a result, the ICA transformation matrix is computed by $\mathbf{T}_{M_1} = \mathbf{P}_{M_1}\mathbf{W}_{M_1}^{-1}$.

2.2. ICA in residual space

The residual images are computed by subtracting the reconstructed images from the original face images. ICA is then applied to the residual images. The ith residual image is represented by $\Delta\mathbf{x}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$, where $\hat{\mathbf{x}}_i$ denotes the ith column of $\hat{\mathbf{X}}$. The residual matrix corresponding to the residual images is defined by $\boldsymbol{\Gamma} \equiv \mathbf{X} - \hat{\mathbf{X}} = [\Delta\mathbf{x}_1, \ldots, \Delta\mathbf{x}_M]$. Similarly to the conventional ICA transformation, we perform ICA on $\mathbf{P}_{M_2}'^{\mathrm{T}} = [\mathbf{p}'_1, \ldots, \mathbf{p}'_{M_2}]^{\mathrm{T}}$, where $\mathbf{p}'_i$ denotes the eigenvector corresponding to the ith largest eigenvalue of the residual data $\boldsymbol{\Gamma}$. The ICA reconstruction $\hat{\boldsymbol{\Gamma}}$ of the residual images is represented by

$$\hat{\boldsymbol{\Gamma}} = \mathbf{U}'_{M_2}(\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2})^{\mathrm{T}}\boldsymbol{\Gamma}, \tag{4}$$

where $\mathbf{U}'_{M_2}$ denotes M_2 independent basis images and $\mathbf{W}'_{M_2}$ is an $M_2 \times M_2$ invertible weight matrix such that the columns of $\mathbf{U}'_{M_2}$ are statistically independent. Using $\boldsymbol{\Gamma} = \mathbf{X} - \hat{\mathbf{X}}$, Eq. (4) can be rewritten with respect to the original matrix $\mathbf{X}$ as

$$\begin{aligned}\hat{\boldsymbol{\Gamma}} &= \mathbf{U}'_{M_2}(\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2})^{\mathrm{T}}(\mathbf{X} - \hat{\mathbf{X}}) \\ &= \mathbf{U}'_{M_2}(\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2})^{\mathrm{T}}[\mathbf{X} - \mathbf{U}_{M_1}(\mathbf{P}_{M_1}\mathbf{W}_{M_1}^{-1})^{\mathrm{T}}\mathbf{X}] \\ &= \mathbf{U}'_{M_2}[(\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2})^{\mathrm{T}} - (\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2})^{\mathrm{T}}\mathbf{U}_{M_1} \\ &\quad \times (\mathbf{P}_{M_1}\mathbf{W}_{M_1}^{-1})^{\mathrm{T}}]\mathbf{X}.\end{aligned} \tag{5}$$

Thus, the high-frequency components of faces, $\hat{\boldsymbol{\Gamma}}$, are spanned by the independent basis images $\mathbf{U}'_{M_2}$ and are represented by the ICA coefficients $[(\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2}) - \mathbf{P}_{M_1}\mathbf{W}_{M_1}^{-1}\mathbf{U}_{M_1}^{\mathrm{T}}\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2}]^{\mathrm{T}}\mathbf{X}$ denoted by $\mathbf{B}'_{M_2}$. As a result, the ICA transformation matrix of the residual space is computed by $\mathbf{T}'_{M_2} = \mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2} - \mathbf{P}_{M_1}\mathbf{W}_{M_1}^{-1}\mathbf{U}_{M_1}^{\mathrm{T}}\mathbf{P}'_{M_2}\mathbf{W}'^{-1}_{M_2}$. Examples of the original face images $\mathbf{X}$, its reconstructed images $\hat{\mathbf{X}}$, and the residual images $\boldsymbol{\Gamma}$ are shown in Fig. 1.

Fig. 1. (a) Original face images. (b) The reconstructed images from the first 10 principal components. (c) The residue space.

3. PCA vs. ICA in residual space

PCA finds orthonormal vectors which maximize the variance of a given distribution. PCA in the residual space does not produce new subspaces due to the orthogonality of its basis vectors. The principal components obtained by PCA in the residual space can be simply chosen among the principal components of the original data space. All residual

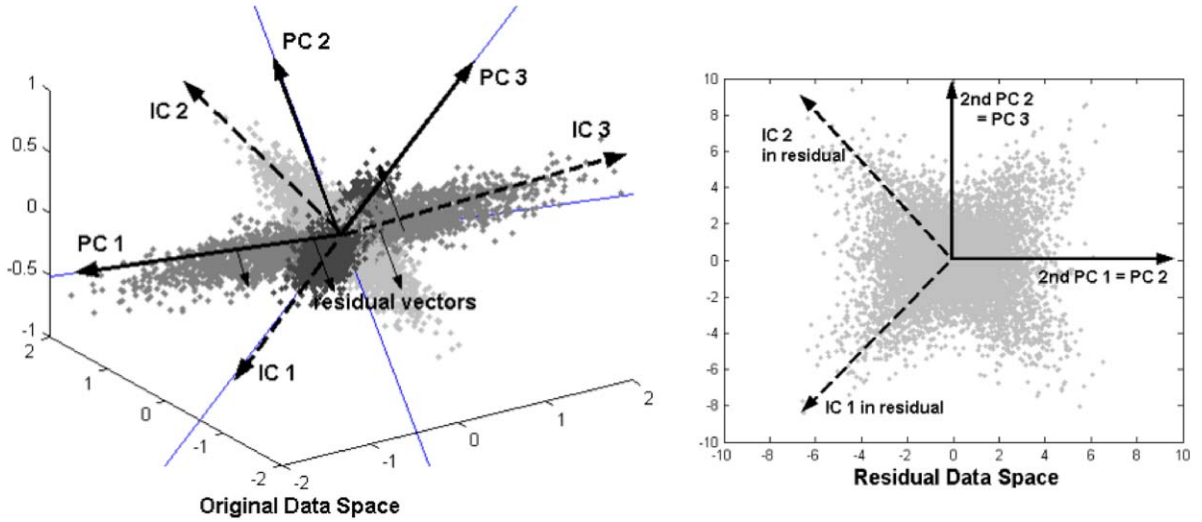

Fig. 2. Density estimation in the residual space. Principal components (PC1/PC2/PC3) and independent components (IC1/IC2/IC3) in the original data space are shown in the left. PCA and ICA in the residual space of the first principal axis (PC1) produce the second PC1, second PC2 and "IC1, IC2 in the residual" space, respectively.

vectors obtained by projecting data points on some leading basis vectors become orthogonal to the leading basis vectors. PCA on these orthogonal residual vectors produces unit basis vectors whose direction is the same as that of the original eigenvectors. The produced set of vectors becomes a subset of the original basis vectors which are orthogonal to the leading basis vectors. On the contrary, ICA, which does not impose any orthogonal constraint for the axes, extracts basis vectors which are independent from the original ICA basis vectors when it is applied in the residual space.

Let us give a simple example. Examples of a principal component and an independent component in both the original and the residual space obtained from the first principal axis are shown in Fig. 2. The left figure shows the original data distribution and the components found by PCA and ICA. Some residual vectors of data points for the first principal component, PC 1, are also given in the figure. The two-dimensional space of these residual vectors and the components in the residual space are shown in the right figure. While the second-order PCA (PCA in the residual space) finds the same principal component as the conventional analysis, independent component in the residual space is different from the original independent components. Moreover, an independent component in the residual space can produce a more efficient basis vectors for class discrimination rather than the original one because they are obtained from a data space freed from contaminating information. We believe that the new features are likely to be better for face recognition, since they are extracted from the residual face images shown in Fig. 1, which do not seem to be affected by illumination changes compared to the original face images. There is ample prior experimental evidence suggesting that PCA without some leading eigenfaces gives better recognition result [1–3]. A proof that a set of eigenvectors of the residual space is a subset of the original eigenvectors is given in the appendix.

4. Linear encoding in local facial space

The face description we propose is based on local facial space analysis. A face image is separated into several facial components corresponding to forehead, eyes, nose and mouth. Compared with holistic image representation, this approach is more robust to illumination and/or pose variation in face encoding, and it is more flexible in similarity matching and in alignment adjustment.

First, image variation due to pose and/or illumination change is smaller in each local region compared with that in the whole image space, so it can be approximately linearized and this simplifies pre-processing. Generally, holistic approaches based on PCA/ICA/LDA encode the gray-scale correlation among every pixel position statistically. Thus, any image variation due to changes of lighting and camera geometry results in a severe change of face representation. However, since our scheme encodes the facial components separately, image variations are limited to each local region. As a local space exhibits less statistical complexity than the whole face space, the linear encoding like PCA/ICA/LDA in a local space will be more robust to illumination changes than the whole face region. In the proposed description, separated facial components have an overlap with the neighboring components as shown in Fig. 3 and they encode their mutual relationships. Thus, important relationships describ-

Fig. 3. Facial component separation.

ing personal characteristics for identification are preserved. The experimental results show that the local space encoding followed by a simple sum of matching scores of the components outperforms holistic encoding methods in person identification.

Second, facial components with large variation are less weighted in the matching stage. Since each facial component can be considered as a separate classifier, the outputs can be weighted according to their discriminability and prior knowledge. Furthermore, when the component positions are well aligned by facial component detection or dense matching methods, the geometrical variation can be compensated and this results in further accuracy improvements. In Ref. [13], the recognition accuracy was improved after component alignment.

As shown in Fig. 3, we separate a face image into 14 facial components. To avoid the dependency on data set, the components defined by Heisele et al. [13] are mainly used: eyebrows, eyes, nose, and mouth. The additional components like forehead, cheeks, and chin are selected similarly to Nefian and Hayes' work [11]. The position and scale of each component is fixed relatively to the eye positions here. In the experimental section, the results of the manually aligned and the fixed case are compared in terms of recognition rate. The results show that if the components are well aligned, the recognition performance is better.

5. Feature selection and similarity matching for face recognition

5.1. Residual ICA

To reduce the bit-rate and improve the performance of the ICA representation, subsets of ICA coefficients of cardinality K_1, K_2 exhibiting the highest class discriminability as defined by the ratio of between-class to within-class variances (9) are selected among those determined by the independent basis images $\mathbf{U}_{M_1}$, $\mathbf{U}'_{M_2}$. The associated bases are denoted by $\mathbf{U}_{K_1}$, $\mathbf{U}'_{K_2}$, respectively. Their corresponding transformation matrices, $\mathbf{T}_{K_1}$ and $\mathbf{T}'_{K_2}$, are different from $\mathbf{T}_{M_1}$ and $\mathbf{T}'_{M_2}$ in permutation and dimension, but are the same in meaning. Fig. 4 shows the PCA basis images and that of ICA in the residual space. The proposed ICA representation consists of basis images $\mathbf{U} = [\mathbf{U}_{K_1} \quad \mathbf{U}'_{K_2}]$ and coefficient matrices represented by

$$\mathbf{B} = \mathbf{TX}, \tag{6}$$

where $\mathbf{T} = [\mathbf{T}_{K_1} \quad \mathbf{T}'_{K_2}]^{\mathrm{T}}$ denotes the transformation matrices. Note that since the basis images $\mathbf{U}$ are fixed, a face image $\mathbf{X}$ is represented by the ICA coefficients $\mathbf{B}$ from Eq. (6), where $\mathbf{T}$ is pre-computed from a training image set.

Given two face images $\mathbf{x}_1$, $\mathbf{x}_2$ represented by ICA coefficients $\mathbf{b}_1$, $\mathbf{b}_2$ ($\mathbf{b}_1 = \mathbf{T}\mathbf{x}_1$, $\mathbf{b}_2 = \mathbf{T}\mathbf{x}_2$) the similarity $d(\mathbf{b}_1, \mathbf{b}_2)$ is measured by calculating cross-correlation between them:

$$d = \frac{\mathbf{b}_1 \cdot \mathbf{b}_2}{\|\mathbf{b}_1\|\|\mathbf{b}_2\|}, \tag{7}$$

where $\mathbf{a} \cdot \mathbf{b}$ denotes the inner product of vectors $\mathbf{a}$ and $\mathbf{b}$; $\|\mathbf{a}\|$ denotes the norm of a vector $\mathbf{a}$.

5.2. Representation of local residual space

Suppose that an image $\mathbf{x}$ is given and it is separated into L local components $\{\mathbf{c}^{(1)}, \ldots, \mathbf{c}^{(L)}\}$. When the residual ICA is performed on the ith local component $\mathbf{c}^{(i)}$ of the image $\mathbf{x}$, the local space is described by a coefficient vector $\mathbf{b}^{(i)}$ obtained by transformation $\mathbf{T}^{(i)}$ with basis image matrix $\mathbf{U}^{(i)}$. Note that $\mathbf{U}^{(i)}$ and $\mathbf{T}^{(i)}$ are pre-computed from a training set of the ith facial component. Finally, a face image $\mathbf{x}$ is represented by a collection of coefficient vectors $\{\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(L)}\}$ with a set of basis images $\{\mathbf{U}^{(1)}, \ldots, \mathbf{U}^{(L)}\}$. Fig. 5 shows the example of local residual ICA basis images.

Given two face images $\mathbf{x}_1$, $\mathbf{x}_2$ represented by ICA coefficients $\mathbf{b}_1$, $\mathbf{b}_2$, the similarity $d(\mathbf{b}_1, \mathbf{b}_2)$ is measured by a weighted sum of cross-correlations between the corresponding components as

$$d = \frac{1}{L}\left\{ w_1 \frac{\mathbf{b}_1^{(1)} \cdot \mathbf{b}_2^{(1)}}{\|\mathbf{b}_1^{(1)}\|\|\mathbf{b}_2^{(1)}\|} + \cdots + w_L \frac{\mathbf{b}_1^{(L)} \cdot \mathbf{b}_2^{(L)}}{\|\mathbf{b}_1^{(L)}\|\|\mathbf{b}_2^{(L)}\|} \right\}, \tag{8}$$

where $\mathbf{b}_1^{(i)}$, $\mathbf{b}_2^{(i)}$ denotes the residual ICA coefficient of the ith local component of the face image $\mathbf{x}_1$, $\mathbf{x}_2$, respectively, and w_i denotes the weighting factor of the ith component. To determine the weighting factor, the class discriminability (9) of each component is computed from the training data set and the factor is then proportional to the discriminability value. Clearly, the proposed method does not consider nonlinear relationships of components. Any merits of nonlinear weighting will be investigated in the future.

6. Experimental results and discussion

6.1. Database

The experimental face database consists of 3175 images of 635 persons (five images of each person), which is the data

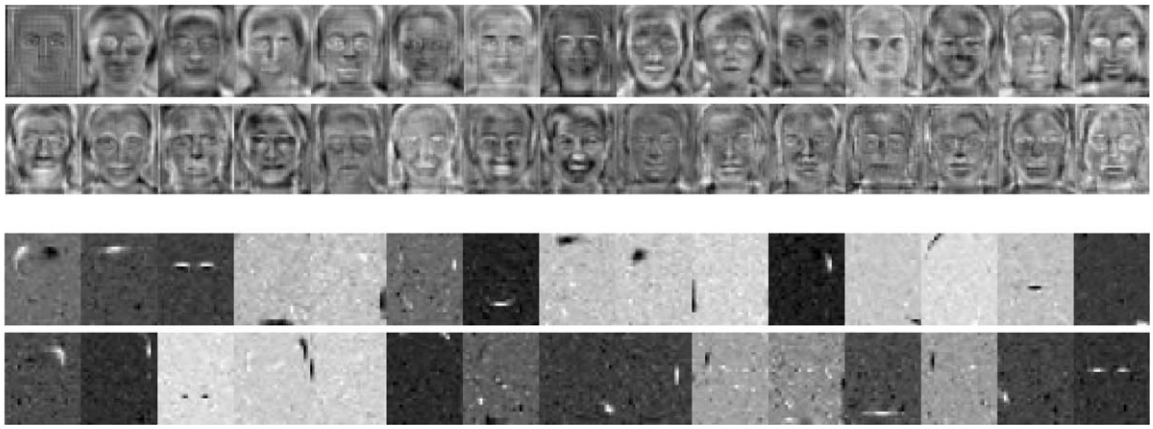

Fig. 4. The 30 most discriminative basis images: (a) the 30 selected PCA basis images, (b) the selected basis images of "ICA in residual".

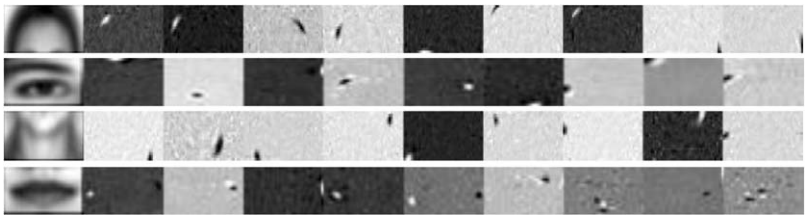

Fig. 5. The local residual ICA basis images.

set adopted for the MPEG-7 VCE-4 (face descriptor) standardization effort [15]. The data set is a collection of various face data sets which include non-public ones. In the experiment comparing PCA and ICA, we utilized a subset (1700 images) of the MPEG-7 data consisting of the well-known public data sets (AR, Yale, ORL, Bern) and FERET because we would like others to duplicate our results easily. The data set has variations of lighting condition and pose as shown in Fig. 6. The original images of the public databases can be obtained from the following URL:

- *AR database*: ⟨http://rvl1.ecn.purdue.edu/v1/ARdatabase/ARdatabase.html⟩
- *Yale database*: ⟨http://cvc.yale.edu/projects/yalefaces/yalefaces.html⟩
- *ORL database*: ⟨http://www.uk.research.att.com/facedatabase.html⟩
- *Bern database*: ⟨http://ftp.iam.unibe.ch/pub/Images/FaceImages⟩

The images in the database are manually normalized to 46×56 pixels^2 giving fixed eye positions. Some images in the database are taken under light variation (light set), and others are taken with the faces at different view angles (pose set). Details of our main database are described in Table 1.

6.2. Protocol

Four different experiments in Table 2 were performed. In Experiment 1, all images for training and test were collected from the light set. Experiment 2 has all images which were from the pose set. Experiments 3 and 4 were done for the mixture of pose and light sets. The database for Experiment 4 is an extended version of the main data set, which consists of 3175 face images of MPEG-7 data set [15]. Five methods—PCA, ICA, the second-order PCA, "ICA in residual space" and "ICA in local residual space"—were applied using the training sets associated with the respective experiments to extract the basis vectors and all images in the test sets were utilized as a query. When a query image is given, the remaining four images of the query person and all other face images in the test set are registered. False identification rate (FIR) is given as a measure of face recognition performance. The recognition is successful when one of the four images of the same person is the top ranked.

6.3. Feature selection scheme

The class discriminability of basis vectors, defined as

$$r = \frac{\sigma_B}{\sigma_W}, \tag{9}$$

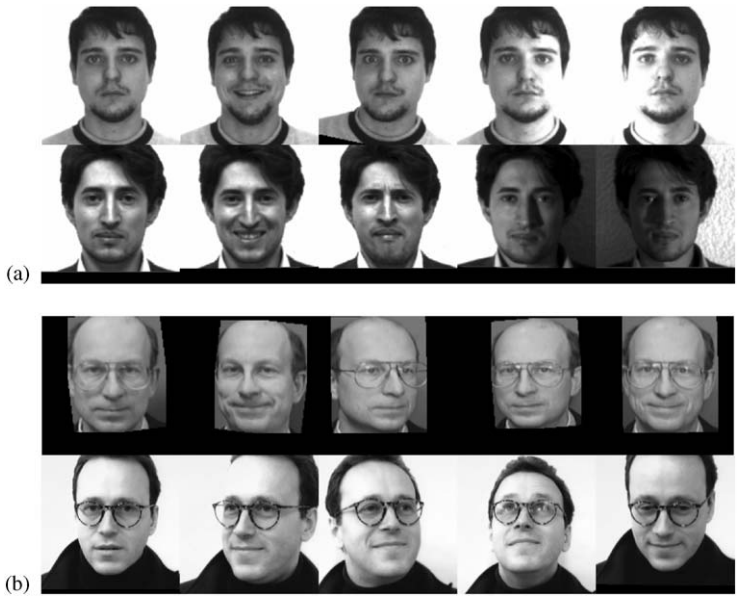

Fig. 6. Examples of the face database: (a) examples of the light set, (b) examples of the pose set.

Table 1
The face database

Filename	No. of persons	Light set	Pose set	Original database	Comment
face_0001–face_0065	65	√		AR DB	Test set
face_0066–face_0115	50	√		AR DB	Training set
face_0116–face_0133	18	√		AR DB	Test set
face_0134–face_0148	15	√		Yale DB	Test set
face_0170–face_0180	11		√	ORL DB	Test set
face_0181–face_0200	20		√	ORL DB	Training set
face_0201–face_0209	9		√	ORL DB	Test set
face_0210–face_0239	30		√	Bern DB	Test set
face_0407–face_0436	30		√	FERET DB	Training set
face_0437–face_0528	92		√	FERET DB	Test set
Total	340	148	192		

where

$$\sigma_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)^2, \qquad \sigma_W = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \mu_i)^2$$

was calculated for a training set and the best combination of the k most discriminative basis vectors was chosen. X_i is the ith class training set, μ_i is the ith class mean and μ is the global mean. Fig. 7 shows the class discriminability of the basis vectors of PCA, ICA, second PCA and residual ICA for the combined training set. The basis vectors were sorted by the magnitude of r. Note that the conventional ICA basis vectors consistently had a greater class discriminability than PCA and this is consistent with the result of Bartlett [9]. Although independent components in residual spaces do not provide better class discrimination than the conventional ICA individually, the combination of independent components of residual spaces performed better in face recognition.

6.4. Results

The residual space was obtained by removing 10 leading eigenfaces. The number, 10, was arbitrarily chosen as an inflection point of the eigenvalue plot. However, changing this number did not affect the recognition result significantly. The results are shown in Table 3.

Table 4 shows the best FIR of each method. In the best case for the light set, ICA was inferior to PCA. As mentioned

Table 2
The summary on training and test data set for the experiments

	Training set	Test set	
		Illumination set	Pose set
Experiment 1 (illumination set)	250 images: • face_0066_01–face_0115_05	490 images: •face_0001_01–face_0065_05 •face_0116_01–face_0148_05	
Experiment 2 (pose set)	250 images: •face_0181_01–face_0200_05 •face_0407_01–face_0436_05		710 images: •face_0170_01–face_0180_05 •face_0201_01–face_0239_05 •face_0437_01–face_0528_05
Experiment 3 (combined set)	500 images: •face_0066_01–face_0115_05 •face_0181_01–face_0200_05 •face_0407_01–face_0436_05	1200 images: •face_0001_01–face_0065_05 •face_0116_01–face_0148_05 •face_0170_01–face_0180_05 •face_0201_01–face_0239_05 •face_0437_01–face_0528_05	
Experiment 4 (extended combined set)	1685 images	1490 images	

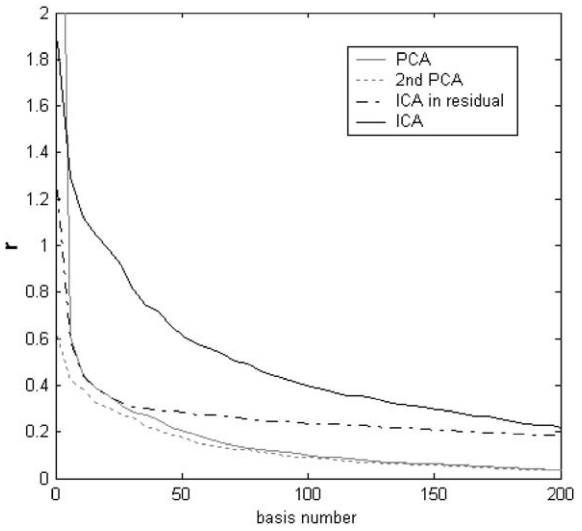

Fig. 7. Class discriminability of basis vectors of PCA, ICA, second-order PCA and "ICA in residual".

earlier, it is difficult to argue that ICA is always better than PCA in face recognition. Both second-order PCA and "ICA in residual space" were much better than PCA and ICA by removing the illumination effect. The statistical characteristics of the face residual space were such that ICA was more suitable to encode them rather than PCA. The ICA in the local residual space also enhanced the result of the residual ICA of a holistic face.

Table 3
Performance comparison of ICA features according to the number of leading eigenfaces which are eliminated for residual space

No. of leading eigenfaces	ICA feature set	FIR
6	ICA in leading eigenfaces ($\mathbf{U}_{K_1}$)	0.750
	ICA in residual space ($\mathbf{U}'_{K_2}$)	0.265
	Combination	0.337
10	ICA in leading eigenfaces ($\mathbf{U}_{K_1}$)	0.597
	ICA in residual space ($\mathbf{U}'_{K_2}$)	0.275
	Combination	0.397
14	ICA in leading eigenfaces ($\mathbf{U}_{K_1}$)	0.558
	ICA in residual space ($\mathbf{U}'_{K_2}$)	0.269
	Combination	0.405

Table 4 also shows the result on the pose set. ICA was much better than PCA due to non-Gaussian distribution of rotated face images. While second-order PCA lost its merits compared with PCA, ICA in residual was still better than ICA. The ICA in the local residual space provided the best recognition result. The alignment of the components improved the result of the ICA in the local residual space. This is because the pose set includes much geometrical variation of faces and the component alignment could compensate it.

Fig. 8 shows the comparative results with the methods based on holistic representation using the same number (bit-rate) of basis vectors (60). For the illumination test, PCA method has a local minimum of FIR at 40 bases and

Table 4
Best results of the light set and the pose set

Method	Experiment 1, the light set		Experiment 2, the pose set	
	No. of basis vectors	FIR	No. of basis vectors	FIR
PCA	40	0.153	130	0.454
ICA	80	0.179	60	0.360
Second PCA	220	0.061	130	0.476
ICA in residual	220	0.053	60	0.311
ICA in local residual (1/4 subsampling)	8×100	0.044	10×40	0.184
ICA in local residual (1/4 subsampling) (+manual alignment)	8×80	0.048	10×60	0.094

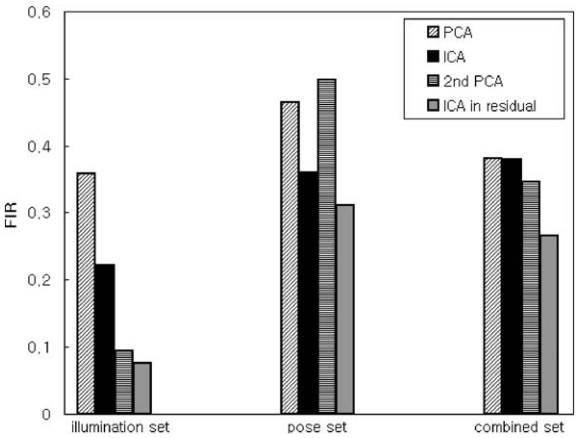

Fig. 8. FIR for the same number (bit-rate) of components.

the accuracy gets worse with the increasing number of components with the exception of this case. The results of the methods in all three conditions of illumination, pose and combined one, are similar to the best results of each experiment.

To check that the proposed ICA in the residual space conserves the benefits of residual space and the high-order statistics of rotated faces, a combined set of light and pose variant images were trained and tested. The results are shown in Table 5. Fig. 9 shows the recognition result as a function of the number of basis vectors. The result of superiority and inferiority of holistic representations for the combined set is consistent with the number of components. The curves of cumulative FIR vs. rank are also presented in Fig. 10. FIR is measured when the case that one of the four images of the query identity is within the rank is successful. The proposed two methods, ICA in residual and ICA in local residual, also keep their superiority in the cumulative measure of face recognition performance.

Finally, Table 6 gives the computational complexity and size of descriptors of the proposed ICA in the residual space and the ICA in the local residual space. Sixty independent components in the holistic face residual space were utilized and each independent component coefficient was coded with four bits. For the ICA in the local residual space, 50 independent components were used for each facial component. The proposed description based on ICA provided an efficient encoding and achieved a redundancy reduction of the feature space and robust pattern classification. The matching complexity of the linear classifier (cross-correlation) used here is favorably low compared to other nonlinear classifiers.

7. Conclusion

In this paper, "ICA in the residual space" and "ICA in the local residual space" were proposed for the representation of face images for compact face recognition which is robust to illumination and pose changes. While PCA in the residual space gives the same principal components as those of the conventional PCA, "ICA of residual images" provides a new set of independent components and this new feature set exhibits a more robust face recognition performance under illumination variation. This is achieved by utilizing the high-order statistics of the data in the face residual space. It was observed that the method based on ICA outperformed PCA in terms of recognition of face images including large pose variations, which make the face space non-Gaussian. The experimental results show that the proposed description based on ICA in facial residual space achieves better recognition as compared with the conventional ICA/PCA and second-order PCA methods. Moreover, the recognition result was further enhanced by splitting the face space into several local spaces and combining them. The lower statistical complexity of local region makes the linear encoding with ICA more effective.

Appendix

Here we prove that a set of eigenvectors of a residual space is a subset of the original eigenvectors. Let $\mathbf{X}$ be the

Table 5
Best results on the combined set

The type of experiment	Method	No. of basis vectors	FIR
Experiment 3	PCA	80	0.37
	ICA	40	0.362
	Second PCA	160	0.320
	ICA in residual	60	0.266
	ICA in local residual (1/4 subsampling)	14×50	0.143
Experiment 4	Second PCA	50	0.306
	ICA in residual	50	0.205
	ICA in local residual (1/4 subsampling)	14×50	0.112
	ICA in local residual (1/16 subsampling)	14×50	0.145

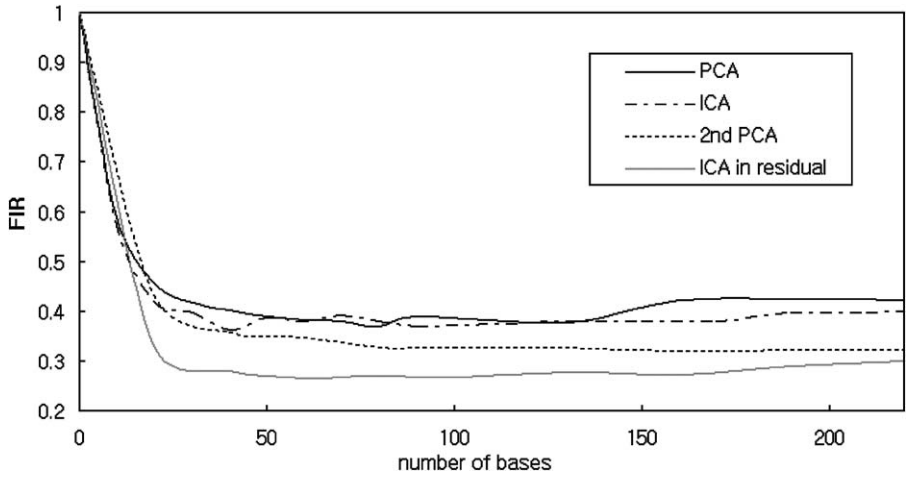

Fig. 9. Top FIR for the number of selected basis vectors.

Table 6
Computational complexity and descriptor size

	ICA in residual	ICA in local residual
Feature extraction complexity		
Additions	$N_0(N-1) = 154,500$	$14N_1(N_{\text{avg}}-1) = 188,300$
Multiplications	$N_0 N = 154,560$	$14N_1 N_{\text{avg}} = 189,000$
Matching complexity		
Additions	$3(N_0-1) = 177$	$14 \times 3(N_1-1)+4 = 2,062$
Multiplications	$3N_0 = 180$	$14(3N_1+1) = 2,114$
Size of descriptor in bits	60×4	$14 \times 50 \times 4$

N_0: the number of elements of the feature vector (=60), N: input image size(=46×56), N_1: the number of elements of one component feature vector(=50), N_{avg}: average size of a component input image (=270).

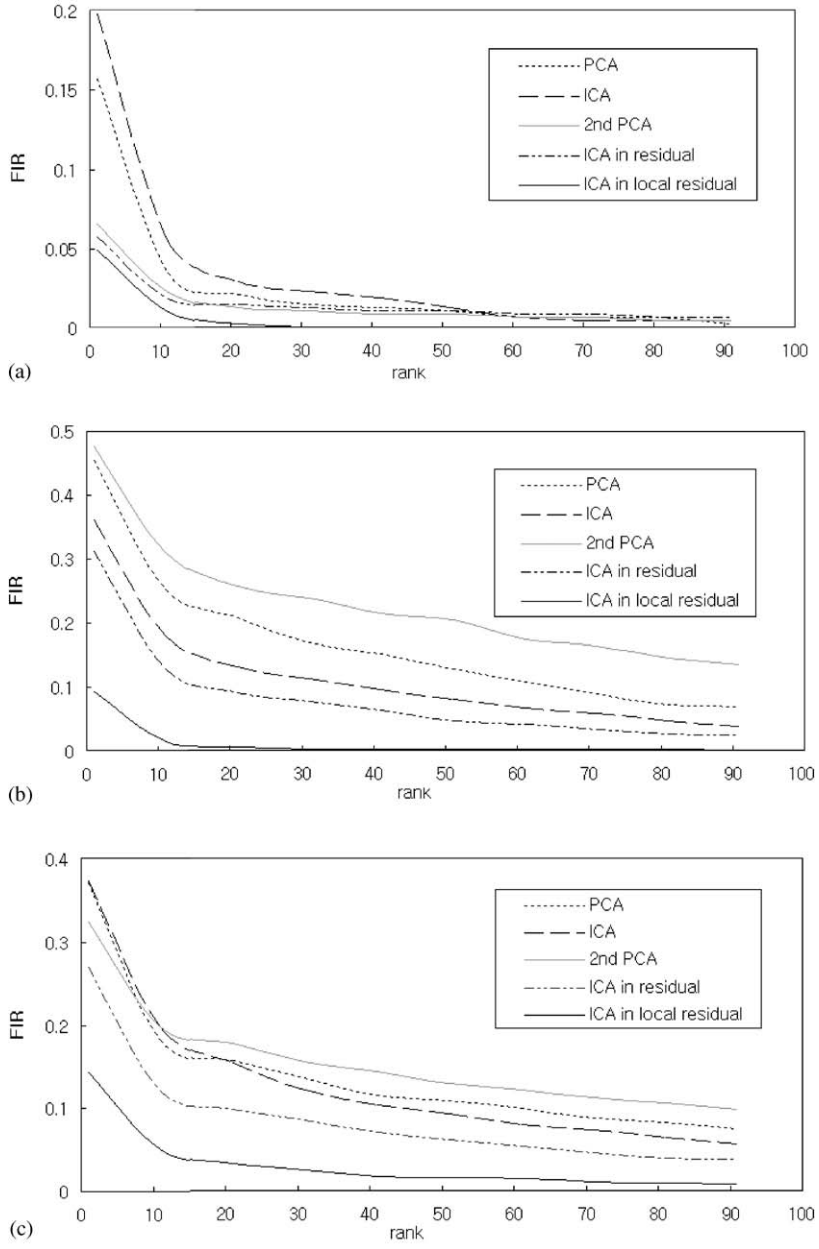

Fig. 10. Cumulative FIR graphs: (a) illumination set, (b) pose set, (c) combined set.

zero mean input matrix such that $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_m]$ where $\mathbf{x}_i$ is a column vector and $\mathbf{x}_i \in \mathbf{R}^N$ with $m < N$. Eigenvectors $\mathbf{v}_i \forall^i$ of the covariance matrix $\mathbf{X}\mathbf{X}^{\mathrm{T}}$ are defined as

$$\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{v}_i = \lambda_i \mathbf{v}_i \quad \text{s.t. } \mathbf{v}_i \perp \mathbf{v}_j \ (i \neq j), \tag{A.1}$$

where λ_i is an eigenvalue. Let Φ be the diagonal matrix whose elements are eigenvalues and let the eigenmatrix $\mathbf{V}$ be defined by $\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{V} = \mathbf{\Phi}\mathbf{V}$. The projection and reconstruction using $\mathbf{m}_1 \leqslant \mathbf{m}$ eigenvectors are obtained by $\mathbf{R}_{m_1} = \mathbf{V}_{m_1}^{\mathrm{T}}\mathbf{X}$ and $\hat{\mathbf{X}} = \mathbf{V}_{m_1}\mathbf{R}_{m_1}$, respectively, where $\mathbf{V}_{m_1} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m_1}]$. By subtracting the reconstruction from the original, we have an $N \times m$ matrix $\overline{\mathbf{X}}$ whose columns are the residual vectors defined by

$$\overline{\mathbf{x}}_i = \mathbf{x}_i - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}}\mathbf{x}_i = (\mathbf{I} - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}})\mathbf{x}_i. \tag{A.2}$$

The residual vector $\overline{\mathbf{x}}_i$ is orthogonal to the m_1 eigenvectors, $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m_1}$ as can be seen by multiplying the above equation by $\mathbf{v}_j^{\mathrm{T}}$:

$$\begin{aligned}\mathbf{v}_j^{\mathrm{T}}\overline{\mathbf{x}}_i &= \mathbf{v}_j^{\mathrm{T}}\mathbf{x}_i - \mathbf{v}_j^{\mathrm{T}}\mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}}\mathbf{x}_i \\ &= \mathbf{v}_j^{\mathrm{T}}\mathbf{x}_i - [0,\ldots,\mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j,\ldots,0][\mathbf{v}_1,\ldots,\mathbf{v}_j,\ldots,\mathbf{v}_{m_1}]^{\mathrm{T}}\mathbf{x}_i \\ &= \mathbf{v}_j^{\mathrm{T}}\mathbf{x}_i - \mathbf{v}_j^{\mathrm{T}}\mathbf{v}_j\mathbf{v}_j^{\mathrm{T}}\mathbf{x}_i = 0. \end{aligned} \tag{A.3}$$

Let $\mathbf{u}_i$ denote an eigenvector of the residual matrix $\overline{\mathbf{X}}$ associated with a nonzero eigenvalue $\mathbf{u}_i$ is described by $\overline{\mathbf{X}}\,\overline{\mathbf{X}}^{\mathrm{T}}\mathbf{u}_i = \lambda_i\mathbf{u}_i$. It is noted that $\mathbf{u}_i$ is also orthogonal to $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{m_1}$ as

$$\mathbf{v}_j^{\mathrm{T}}\overline{\mathbf{X}}\,\overline{\mathbf{X}}^{\mathrm{T}}\mathbf{u}_i = \mathbf{v}_j^{\mathrm{T}}\lambda_i\mathbf{u}_i,$$

$$[\mathbf{v}_j^{\mathrm{T}}\overline{\mathbf{x}}_1, \mathbf{v}_j^{\mathrm{T}}\overline{\mathbf{x}}_2, \ldots, \mathbf{v}_j^{\mathrm{T}}\overline{\mathbf{x}}_m]\overline{\mathbf{X}}^{\mathrm{T}}\mathbf{u}_i = \lambda_i\mathbf{v}_j^{\mathrm{T}}\mathbf{u}_i,$$

$$\mathbf{0} = \lambda_i\mathbf{v}_j^{\mathrm{T}}\mathbf{u}_i, \qquad \therefore\ \mathbf{v}_j \perp \mathbf{u}_i. \tag{A.4}$$

By substituting $(\mathbf{I} - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}})\mathbf{X}$ for $\overline{\mathbf{X}}$ in the equation $\overline{\mathbf{X}}\,\overline{\mathbf{X}}^{\mathrm{T}}\mathbf{u}_i = \lambda_i\mathbf{u}_i$, we have the following equation:

$$(\mathbf{I} - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}})\mathbf{X}\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}})^{\mathrm{T}}\mathbf{u}_i = \lambda_i\mathbf{u}_i. \tag{A.5}$$

The left side of the above equation changes by virtue the orthogonal condition $\mathbf{v}_j \perp \mathbf{u}_i$:

$$\begin{aligned}&(\mathbf{I} - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}})\mathbf{X}\mathbf{X}^{\mathrm{T}}(\mathbf{I} - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}})^{\mathrm{T}}\mathbf{u}_i \\ &= (\mathbf{X}\mathbf{X}^{\mathrm{T}} - \mathbf{V}_{m_1}\mathbf{V}_{m_1}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}})\left(\mathbf{u}_i - \mathbf{V}_{m_1}\begin{bmatrix}\mathbf{v}_1^{\mathrm{T}}\mathbf{u}_i \\ \vdots \\ \mathbf{v}_{m_1}^{\mathrm{T}}\mathbf{u}_i\end{bmatrix}\right) \\ &= \mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{u}_i - \mathbf{V}_{m_1}(\mathbf{V}_{m_1}^{\mathrm{T}}\mathbf{X}\mathbf{X}^{\mathrm{T}})\mathbf{u}_i \\ &= \mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{u}_i - \mathbf{V}_{m_1}(\mathbf{V}_{m_1}^{\mathrm{T}}\boldsymbol{\Phi}_{m_1}^{\mathrm{T}})\mathbf{u}_i \\ &= \mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{u}_i - \boldsymbol{\Phi}_{m_1}^{\mathrm{T}}\mathbf{V}_{m_1}(\mathbf{V}_{m_1}^{\mathrm{T}}\mathbf{u}_i) \\ &= \mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{u}_i. \end{aligned} \tag{A.6}$$

Then, $\mathbf{X}\mathbf{X}^{\mathrm{T}}\mathbf{u}_i = \lambda_i\mathbf{u}_i$ and $\mathbf{u}_i$ becomes an eigenvector of the original covariance matrix. By the fact that $\mathbf{u}_i$ is orthogonal to $\mathbf{v}_j$, $j = 1, \ldots, m_1$, we know that $\mathbf{u}_i$ is one of the other eigenvectors, $\mathbf{v}_j$, $j \neq 1, \ldots, m_1$.

References

[1] P.N. Belhmeur, 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.

[2] Y. Adini, Y. Moses, S. Ullman, Face recognition: the problem of compensating for changes in illumination direction, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 721–732.

[3] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, Wiley-Interscience, New York, 2001.

[4] L. Wang, T.K. Tan, Experimental results of face description based on the 2nd-order eigenface method, ISO/IEC JTC1/SC21/WG11/M6001, Geneva, May 2000.

[5] Hyun-Chul Kim, Daijin Kim, Sung Yang Bang, Face retrieval using 1st- and 2nd-order PCA mixture model, International Conference on Image Processing, Rochester, New York, September 2002.

[6] B. Moghaddam, W. Wahid, A. Pentland, Beyond eigenfaces: probabilistic matching for face recognition, IEEE International Conference on Automatic Face and Gesture Recognition, Nara, Japan, 1998, pp. 30–35.

[7] B. Moghaddam, A. Pentland, Probabilistic visual learning for object representation, IEEE Trans. PAMI 19 (7) (1997) 696–710.

[8] Kah-Kay Sung, Tomaso Poggio, Example-based learning for view-based-human face detection, IEEE Trans. PAMI 20 (1) (1998) 39–51.

[9] M.S. Bartlett, J.R. Movellan, T.J. Sejnowski, Face recognition by independent component analysis, IEEE Trans. Neural Networks 13 (6) (2002) 1450–1464.

[10] A.X. Guan, H.H. Szu, Local face statistics recognition methodology beyond ICA and/or PCA, International Joint Conference on Neural Networks, Vol. 2, Washington, DC, 1999, pp. 1016–1021.

[11] A. Nefian, M. Hayes, An embedded hmm-based approach for face detection and recognition, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol. 6, 1999, pp. 3553–3556.

[12] L. Wiskott, J.-M. Fellous, N. Krüger, C. von der Malsburg, Face recognition by elastic bunch graph matching, IEEE Trans. Pattern Anal. Mach. Intell. 19 (7) (1997) 775–779.

[13] B. Heisele, P. Ho, T. Poggio, Face recognition with support vector machines: global versus component-based approach, in: Proceedings of the IEEE International Conference on Computer Vision, 2001.

[14] T.W. Lee, M. Girolami, T.J. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources, Neural Comput. 11 (2) (1999) 417–441.

[15] T. Kamei, The MPEG-7 face data set, TM-TK-MPEG001, January 18, 2002.

[16] N. Kwak, C. Choi, N. Ahuja, Face recognition using feature extraction based on independent component analysis, International Conference on Image Processing, Rochester, New York, September 2002.

[17] B. Fasel, J. Luettin, Recognition of asymmetric facial action unit activities and intensities, International Conference on Pattern Recognition, Vol. 1, Barcelona, Spain, 2000, pp. 1100–1103.

[18] B. Moghaddam, Principal manifolds and Bayesian subspaces for visual recognition, IEEE International Conference on Computer Vision, Vol. 2, Kerkyra, Greece, September 1999, pp. 1131–1136.

[19] S.Z. Li, X. Lv, H. Zhang, View-subspace analysis of multi-view face patterns, IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Vancouver, BC, Canada, 2001, pp. 125–132.

[20] M.S. Bartlett, Face Image Analysis by Unsupervised Learning, Kluwer Academic Publishers, Dordrecht, 2001.

[21] C. Liu, H. Wechsler, Comparative assessment of independent component analysis (ICA) for face recognition, International

Conference on Audio Video Based Biometric Person Authentication, 1999.
[22] P.C. Yuen, J.H. Lai, Independent component analysis of face images, IEEE Workshop Biologically Motivated Computer Vision, Seoul, Korea, 2000.
[23] Tae-Kyun Kim, Hyunwoo Kim, Wonjun Hwang, Seok Cheol Kee, Jong Ha Lee, Component-based LDA face descriptor for image retrieval, British Machine Vision Conference, Cardiff, UK, September 2–5, 2002.
[24] Tae-Kyun Kim, Hyunwoo Kim, Wonjun Hwang, Seok-Cheol Kee, Josef Kittler, Independent component analysis in a facial local residue space, IEEE International Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, 2003.
[25] Tae-Kyun Kim, Josef Kittler, Hyun-Chul Kim, Seok-Cheol Kee, Discriminant analysis by multiple locally linear transformations, British Machine Vision Conference, East Anglia, Norwich, UK, 2003.
[26] M. Turk, A. Pentland, Eigenfaces for Recognition, Journal of Cognitive Neuroscience 3 (1) (1991) 71–86.
[27] P.S. Penev, J.J. Atick, Local feature analysis: A general statistical theory for object respresentaion, Network: Comput. Neural Syst. 7 (3) (1996) 477–500.

About the Author—TAE-KYUN KIM received a B.Sc. degree and an M.Sc. degree both in electrical engineering from Korea Advanced Institute of Science and Technology (KAIST) in 1994 and 1999, respectively. He is currently a senior researcher in Human Computer Interaction (HCI) laboratory of Samsung Advanced Institute of Technology (SAIT), Korea. His main activities are the development of face detection and description algorithms for video security and MPEG7 standardization respectively.

About the Author—HYUNWOO KIM received a B.S. degree in Electronic Communication Engineering from Hangyang University, Seoul, Korea, in 1994, and M.S. and Ph.D. degrees in Electrical and Computer Engineering from POSTECH, Pohang, Korea, in 1996 and 2001, respectively. In 2000, he worked in the Institute for Robotics and Intelligent Systems at University of Southern California, Los Angeles, CA, as a visiting scientist. In 2001, he joined Samsung Advanced Institute of Technology, Korea, where he is currently a Research Scientist. His current research interests include computer vision, virtual reality, augmented reality and computer graphics.

About the Author—WONJUN HWANG received both B.S. and M.S. degrees from the Department of Electronics Engineering, Korea University, Seoul, Korea, in 1999 and 2001, respectively. In 2002, he worked as an engineer for Samsung Electronics. He is now working as a researcher on face recognition for SAIT. His research interests are in image processing, object detection, object recognition, and robot vision.

About the Author—JOSEF KITTLER is Professor of Machine Intelligence, and Director of the Centre for Vision, Speech and Signal Processing at the University of Surrey. He has worked on various theoretical aspects of Pattern Recognition and Image Analysis, and on many applications including personal identity authentication, automatic inspection, target detection, detection of microcalcifications in digital mammograms, video coding and retrieval, remote sensing, robot vision, speech recognition, and document processing. He has co-authored a book with the title "Pattern Recognition: A statistical approach" published by Prentice-Hall and published more than 400 papers. He is a member of the Editorial Boards of Image and Vision Computing, Pattern Recognition Letters, Pattern Recognition and Artificial Intelligence, Pattern Analysis and Applications, and Machine Vision and Applications.