Independent Component Analysis in a Facial Local Residue Space

Tae-Kyun Kim[†‡], Hyunwoo Kim[†], Wonjun Hwang[†], Seok-Cheol Kee[†],
and Josef Kittler[‡]

[†] : Human Computer Interaction Lab., Samsung Advanced Institute of Technology, Korea.
[‡] : Centre for Vision, Speech and Signal Processing, University of Surrey, U.K.
taekyun@sait.samsung.co.kr

Abstract

In this paper, we propose an ICA(Independent Component Analysis) 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 PCA(Principal Component Analysis)-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.

1. Introduction

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. Features should be insensitive to large variations of light and pose, and matching complexity should be kept low for applications involving huge databases on the internet. 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.

Based on the observation that principal components corresponding to leading eigenvalues represent illumination variation rather than person identity, eigenfaces excluding the first few eigenvectors have been generally used for face recognition. Wang and Tan introduced the 2nd-order PCA method [10]. 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 high frequency components. They are more insensitive to illumination change. They performed a second PCA on the residual images and referred to it 2nd-order PCA.

Recently the 2nd-order PCA was adopted for face recognition in [9]. 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 2nd-order PCA and the conventional eigenface method which discards a few leading principal components. Besides the 2nd-order PCA, there are several conventional methods [7,8,11] which utilize the residual space. However, they only utilize the magnitude of a residual vector and stop to analyze the residual space further.

Compared to the face feature extraction methods like PCA, LDA (Linear Discriminant Anslysis), and LFA (Local Feature Analysis), 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 [1]. Bartlett [1] 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 [3]. 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 probable 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 [12], Nefian and Davies model facial components by HMMs implicitly. In [13], Heisele et al. work with facial components to compensate for pose change. A geometrical configuration classifier based on the support vector machine (SVM) approach is 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 in 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. Where principal components in residual spaces are only a subset of the original principal components, ICA in residual images generates a new set of independent components and these new independent components are more suitable for robust face representation to illumination changes than the conventional ICA [1]. 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 amenably to 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 Residue Space

While the reconstructed face images with a few leading eigenfaces lose details of images 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 of ICA Representation

Suppose that we are given a set of M training images $\boldsymbol{\varphi}_i, i = 1,...,M$, each represented by an N-dimensional vector obtained by a raster. We assume that $M < N$. The mean vector of the image set is defined by $\mathbf{m} = (1/M) \cdot \sum_{i=1}^{M} \boldsymbol{\varphi}_i$. After subtracting the mean vector from all images, i.e., $\mathbf{x}_i = \boldsymbol{\varphi}_i - \mathbf{m}$, we can construct an $N \times M$ matrix $\mathbf{X} = [\mathbf{x}_1,...,\mathbf{x}_M]$ with a zero mean and the covariance matrix $\mathbf{X}\mathbf{X}^{\mathbf{T}}$. Generally, ICA aims to find an $N \times N$ invertible matrix $\mathbf{W}^{(0)}$ such that the columns 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 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 the linear combination of the basis images $\mathbf{u}_i \quad (i = 1,...,N)$.

In face recognition, generally 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. The first $M_1 << M$ 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

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

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

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

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

where $\mathbf{W}_{M_1}$ denotes a $M_1 \times M_1$ invertible matrix such that the columns of $\mathbf{U}_{M_1}$ are statistically independent. The weight matrix $\mathbf{W}_{M_1}$ is estimated by Bell and Sejnowski's algorithm [2]. The reconstructed face image $\hat{\mathbf{X}}$ is computed by multiplying both sides of Equation (1) by $\mathbf{P}_{M_1}$ and it is represented by

$$\begin{aligned}
\hat{\mathbf{X}} &= \mathbf{P}_{\mathbf{M}_1}\mathbf{R}_{\mathbf{M}_1} = \mathbf{P}_{\mathbf{M}_1}\mathbf{P}_{\mathbf{M}_1}^{\mathbf{T}}\mathbf{X} \\
&= \mathbf{U}_{\mathbf{M}_1}(\mathbf{W}_{\mathbf{M}_1}^{-1})^{\mathbf{T}}\mathbf{P}_{\mathbf{M}_1}^{\mathbf{T}}\mathbf{X} \\
&= \mathbf{U}_{\mathbf{M}_1}(\mathbf{P}_{\mathbf{M}_1}\mathbf{W}_{\mathbf{M}_1}^{-1})^{\mathbf{T}}\mathbf{X}
\end{aligned} \tag{3}$$

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 $(\mathbf{P}_{\mathbf{M}_1}\mathbf{W}_{\mathbf{M}_1}^{-1})^{\mathbf{T}}\mathbf{X}$ denoted by $\mathbf{B}_{M_1}$. As a result, the ICA transformation matrix is computed by $\mathbf{T}_{M_1} = \mathbf{P}_{\mathbf{M}_1}\mathbf{W}_{\mathbf{M}_1}^{-1}$.

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 i th residual image is represented by $\Delta\mathbf{x}_i = \mathbf{x}_i - \hat{\mathbf{x}}_i$, where $\hat{\mathbf{x}}_i$ denotes the i th column of $\hat{\mathbf{X}}$. The residual matrix corresponding to the residual images is defined by $\mathbf{\Gamma} \equiv \mathbf{X} - \hat{\mathbf{X}} = [\Delta\mathbf{x}_1,...,\Delta\mathbf{x}_M]$. Similarly to the conventional ICA transformation, we perform ICA on $\mathbf{P}_{\mathbf{M}_2}'^T = [\mathbf{p}_1',...,\mathbf{p}_{M_2}']^T$, where $\mathbf{p}_i'$ denotes the eigenvector corresponding to the i th largest eigenvalue of the residual data $\mathbf{\Gamma}$. The ICA reconstruction $\hat{\mathbf{\Gamma}}$ of the residual images is represented by

$$\hat{\mathbf{\Gamma}} = \mathbf{U}_{\mathbf{M}_2}'(\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1})^{\mathbf{T}}\mathbf{\Gamma}, \tag{4}$$

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

$$\begin{aligned}
\hat{\mathbf{\Gamma}} &= \mathbf{U}_{\mathbf{M}_2}'(\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1})^{\mathbf{T}}(\mathbf{X} - \hat{\mathbf{X}}) \\
&= \mathbf{U}_{\mathbf{M}_2}'(\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1})^{\mathbf{T}}[\mathbf{X} - \mathbf{U}_{\mathbf{M}_1}(\mathbf{P}_{\mathbf{M}_1}\mathbf{W}_{\mathbf{M}_1}^{-1})^{\mathbf{T}}\mathbf{X}] \\
&= \mathbf{U}_{\mathbf{M}_2}'\big[(\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1})^{\mathbf{T}} - (\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1})^{\mathbf{T}}\mathbf{U}_{\mathbf{M}_1}(\mathbf{P}_{\mathbf{M}_1}\mathbf{W}_{\mathbf{M}_1}^{-1})^{\mathbf{T}}\big]\mathbf{X}
\end{aligned} \tag{5}$$

Thus, the high frequency components of faces, $\hat{\mathbf{\Gamma}}$, are spanned by the independent basis images $\mathbf{U}_{M_2}'$ and are represented by the ICA coefficients $\big[(\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1}) - \mathbf{P}_{\mathbf{M}_1}\mathbf{W}_{\mathbf{M}_1}^{-1}\mathbf{U}_{\mathbf{M}_1}^{\mathbf{T}}\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1}\big]^{\mathbf{T}}\mathbf{X}$ denoted by $\mathbf{B}_{M_2}'$. As a result, the ICA transformation matrix of the residual space is computed by

(a) Original face images

(b) The reconstructed images from the first 10 principal components

(c) The residual images

Figure 1. Face residue space

$\mathbf{T}_{M_2}' = \mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1} - \mathbf{P}_{\mathbf{M}_1}\mathbf{W}_{\mathbf{M}_1}^{-1}\mathbf{U}_{\mathbf{M}_1}^{\mathbf{T}}\mathbf{P}_{\mathbf{M}_2}'\mathbf{W}_{\mathbf{M}_2}'^{-1}$. Examples of the original face images $\mathbf{X}$, its reconstructed images $\hat{\mathbf{X}}$, and the residual images $\mathbf{\Gamma}$ are shown in Figure 1.

3. PCA vs. ICA in a 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. All residual 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 the residual space of the first principal axis are shown in Figure 2. While the 2nd-order PCA (PCA of the residual) finds the same principal component as the conventional analysis, independent component in the residual space is different from the original independent components. In this example, the independent component in the residual space is identical to the second principal component. However, it is only because the original space has just two-dimensions and the residual space for the first principal
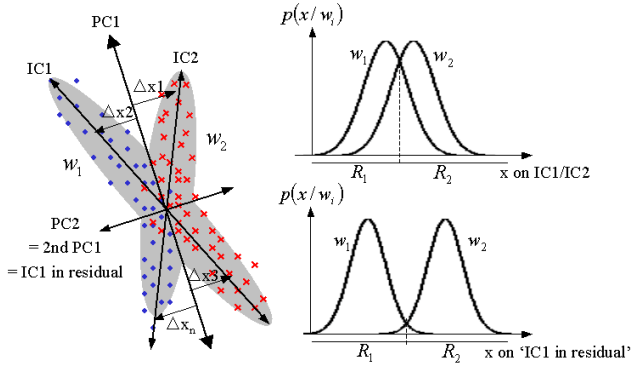
Figure 2. Density estimation in the residual space. For the data distribution which has two classes, w_1 (blue dot), w_2 (red cross), the principal components (PC1/PC2) and independent components (IC1/IC2) are shown. PCA and ICA in the residual space of the first principal axis (PC1) produce the 2nd PC and 'IC1 in the residual' space respectively.

component is one directional space. Imagine a higher dimensional case like the residual space of a 3-dim original space corresponding to the first eigen-component. The residual vectors of all data points are orthgonal to the first eigen-component but have various complex directions. Clearly PCA and ICA will produce different probability densities.

Moreover, an independent component in the residual space can produce a more efficient basis vectors for class discrimination rather than the original one. The above example generates a component which has a smaller overlap between the two class distributions. 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 Figure 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. 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 proposed description 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 obtained by component analysis on original image dimension, it is more robust to illumination and/or pose variation in face encoding, and it has flexibility in similarity matching and in alignment adjustment.
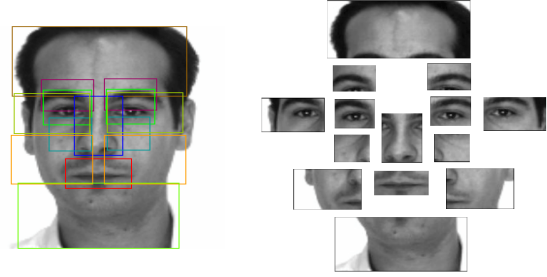

Figure 3. Facial component separation.

First, image variation due to pose and/or illumination change is smaller in each local region compared with that in a whole image space, so it can be approximately linearlized and 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 Figure 3 and they encode their mutual relationships. Thus important relationships describing 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. In the matching stage, since each facial component can be considered as a separate classifier, the outputs can be weighted according to its 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 [13], the recognition accuracy was improved after component alignment.

Figure 3 shows our facial component separation. As shown in the figure, we separate a face image into 14 facial components. To avoid the dependency on dataset, 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 et al.'s work [12]. The position and scale of each component is fixed relatively to the eye positions here. In the experimental section, the results of manually aligned and fixed case are compared in terms of
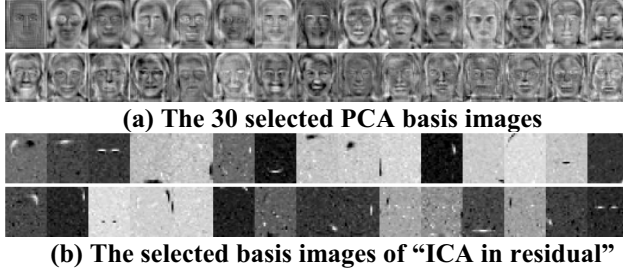
(a) The 30 selected PCA basis images

(b) The selected basis images of "ICA in residual"

Figure 4. The 30 most discriminative basis images

recognition rate. The results show that if the components are well aligned, the recognition performance will be better.

5. Feature Selection and Similarity Matching for Face Recognition

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 (8) are selected among those determined by the independent basis images $\mathbf{U}_{M_1}$, $\mathbf{U}'_{M_2}$ respectively. 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. Figure 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

$$\mathbf{B} = \mathbf{TX}, \qquad (6)$$

where $\mathbf{T} = [\mathbf{T}_{K_1} \quad \mathbf{T}'_{K_2}]^{\mathbf{T}}$ 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 Equation (6), where $\mathbf{T}$ is precomputed from a training image set.

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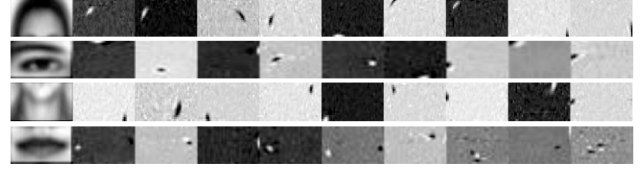

Figure 5. The local residual ICA basis images

basis images $\{\mathbf{U}^{(1)},...,\mathbf{U}^{(L)}\}$. Figure 5 shows the example of local residual ICA basis images.

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

$$d = \frac{1}{L}\left\{ w_1 \frac{\mathbf{b}_1^{(1)} \cdot \mathbf{b}_2^{(1)}}{\left\|\mathbf{b}_1^{(1)}\right\| \left\|\mathbf{b}_2^{(1)}\right\|} + ... + w_L \frac{\mathbf{b}_1^{(L)} \cdot \mathbf{b}_2^{(L)}}{\left\|\mathbf{b}_1^{(L)}\right\| \left\|\mathbf{b}_2^{(L)}\right\|} \right\}, \qquad (7)$$

where $\mathbf{b}_1^{(i)}$, $\mathbf{b}_2^{(i)}$ denotes the residual ICA coefficient of the i th local component of the face image $\mathbf{X}_1$, $\mathbf{X}_2$, respectively, and w_i denotes the weighting factor of the i th component. To determine the weighting factor, the class discriminability (8) 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 factor will be investigated in future.

6. Experimental Results and Discussion

6.1. Database & Protocol of experiments

The experimental face database consists of 3175 images of 635 persons (5 images of each person), which is the data set adapted for the MPEG-7 VCE-4 (face descriptor) standardization effort [14]. 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 be alike duplicate our results easily. The images in the database are manually cropped and normalized to 46x56 pixels^2 for the holistic representation and 128x128 pixels^2 for the component-based representation 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). An example of
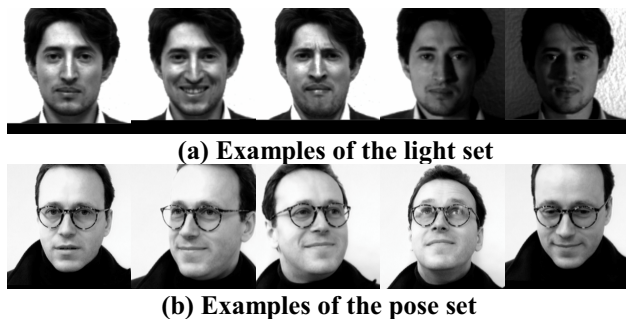
(a) Examples of the light set

(b) Examples of the pose set

Figure 6. Examples of the face database

Table 1. The data set for the experiments

(Unit : # of images)	Training set		Test set	
	Illumination set	Pose set	Illumination set	Pose set
Experiment 1	250		490	
Experiment 2		250		710
Experiment 3	500		1200	
Experiment 4	1685		1490	

data-set is shown in Figure 6. Four different experiments in Table 1 were performed.

6.2. Feature Selection Scheme

The class discriminability of basis vectors, defined as

$$r = \frac{\sigma_B}{\sigma_W} \qquad (8)$$

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

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, μ_i is the ith class mean and μ is the global mean. Figure 7 (a) shows the class discriminability of the basis vectors of PCA, ICA, 2nd 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 greater class discriminability than the PCA and this is consistent with the result of Bartlett [1]. 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.

6.3. Results

Five methods - PCA, ICA, the 2nd-order PCA, "ICA in residual space" and "ICA in local residual space" - were
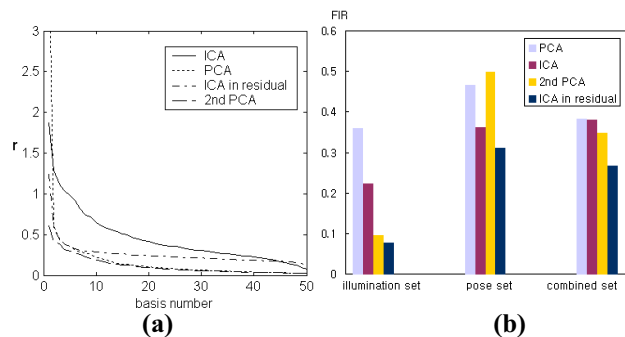

(a)   **(b)**

Figure 7. (a) Class discriminability of basis vectors of PCA, ICA, 2nd-order PCA and "ICA in residual" (b) FIR for the same number (bit-rate) of components

Table 2. Best results of the Experiment 1 and 2

Method	Experiment 1 (light set)		Experiment 2 (pose set)	
	# of bases	FIR	# of bases	FIR
PCA	40	0.1530	130	0.4549
ICA	80	0.1795	60	0.3605
2nd PCA	220	0.0612	130	0.4760
ICA in residual	220	0.0530	60	0.3112
ICA in local residual	*8×100	0.0448	*10×40	0.1845
ICA in local residual (+manual alignment)	*8×80	0.0489	*10×60	0.0943

*(# of components)×(# of component bases)

performed on the training sets of the experiment to extract basis vectors. The residual space was obtained by removing 10 leading eigenfaces. The number, 10, was arbitrarily chosen as an inflection point of eigenvalue plot. However, changing this number did not affect the recognition result significantly.

Table 2 shows the best FIR (False Identification Rate) of each method for the light set. In the best case, ICA was inferior to PCA. As mentioned earlier, it is difficult to argue that ICA is always better than PCA in face recognition. Both 2nd-order PCA and "ICA in residual space" were much better than PCA and ICA by removing illumination factors. The statistical characteristics of face residual space were such that ICA was more suitable to encode them rather than PCA. The ICA in local residual space also enhanced the result of the residual ICA of a holistic face. Figure 7 (b) 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 the accuracy gets worse with the increasing number of components, with the exception of this case the results in the figure are similar to the best case of each experiment.

Table 2 also shows the result on the pose set. ICA was much better than PCA due to non-Gaussian distribution of

Table 3. Best results on the combined set

The type of Experiment	Method	# of basis vectors	FIR
Experiment 3	PCA	80	0.37
	ICA	40	0.3625
	2nd PCA	160	0.320833
	ICA in residual	60	0.265833
Experiment 4	2nd PCA	50	0.306
	ICA in residual	50	0.205
	ICA in local residual	14×50	0.112
	ICA in local residual (64x64 pixels^2 image)	14×50	0.145

Figure 8. Top FIR for the number of basis vectors

rotated face images. While 2nd-order PCA lost its merits compared with PCA, ICA in residual was still better than ICA. ICA in local residual space provided the best recognition result. Alignment of the components improved the result of ICA in local residual space. This is because the pose set includes much geometrical variation of faces and the component alignment could compensate it.

To check that the proposed ICA in the residual space conserves the benefits of residual space and 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 3. Figure 8 shows the recognition result as a function of the number of basis vectors. The result of superiority of ICA and inferiority of holistic representations for the combined set is consistent with all dimensionality of the representation space.

60 independent components in the holistic face residual space were utilized and each coefficient for an independent component 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 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 non-linear classifiers.

7. Conclusion

In this paper, "ICA in residual space" and "ICA in 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 residual space gives the same principal components as those of conventional PCA, "ICA in residual images" provides a new set of independent components and this new feature set exhibits to more robust face recognition performance to illumination variation. This is achieved by utilizing high-order statistics of face residual space. It was observed that the method based on ICA outperformed
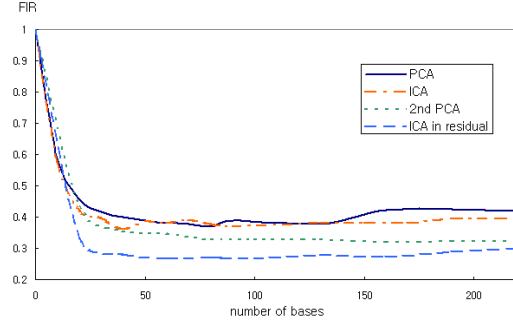
PCA in terms of recognition of face images including large pose variations, which make the face space non-Gaussian. The experimental results show that the proposed description based on ICA in facial residual space is better in face recognition as compared with the conventional ICA/PCA and 2nd-order PCA methods. Moreover, the recognition result was further enhanced by splitting a face space into several local spaces and combining them. The lower statistical complexity of local region makes the linear encoding like 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 zero mean input matrix such that $\mathbf{X} = [\mathbf{x_1}, \mathbf{x_2}, ..., \mathbf{x_m}]$ where $\mathbf{x}_i$ is a column vector and $\mathbf{x}_i \in \mathbf{R^N}$ with $\mathbf{m} < \mathbf{N}$. Eigenvectors $\mathbf{v_i}$ $^{\forall}\mathbf{i}$ of the covariance matrix $\mathbf{XX^T}$ are defined as:

$$\mathbf{XX^T v_i} = \lambda_i \mathbf{v_i} \quad \text{s.t.} \quad \mathbf{v_i} \perp \mathbf{v_j} \ (\mathbf{i} \neq \mathbf{j}), \quad (1)$$

where λ_i is an eigenvalue. Let $\mathbf{\Phi}$ be the diagonal matrix whose elements are eigenvalues and let the eigenmatrix $\mathbf{V}$ be defined by $\mathbf{XX^T V} = \mathbf{\Phi V}$. The projection and reconstruction using $\mathbf{m_1} \leq \mathbf{m}$ eigenvectors are obtained by $\mathbf{R_{m_1}} = \mathbf{V_{m_1}^T X}$ and $\mathbf{\hat{X}} = \mathbf{V_{m_1} R_{m_1}}$ respectively where $\mathbf{V_{m_1}} = [\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_{m_1}}]$. By subtracting the reconstruction from the original, we have a $N \times m$ matrix $\mathbf{\overline{X}}$ whose columns are residual vectors defined by

$$\mathbf{\overline{x}_i} = \mathbf{x_i} - \mathbf{V_{m_1} V_{m_1}^T x_i} = (\mathbf{I} - \mathbf{V_{m_1} V_{m_1}^T})\mathbf{x_i}. \quad (2)$$

The residual vector $\mathbf{\overline{x}_i}$ is orthogonal to the $\mathbf{m_1}$ eigen-

vectors, $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{m_1}$ as can be seen by multiplying the above equation by $\mathbf{v}_j^T$:

$$
\begin{aligned}
\mathbf{v}_j^T \overline{\mathbf{x}}_i &= \mathbf{v}_j^T \mathbf{x}_i - \mathbf{v}_j^T \mathbf{V}_{m_1} \mathbf{V}_{m_1}^T \mathbf{x}_i \\
&= \mathbf{v}_j^T \mathbf{x}_i - [\mathbf{0}, ..., \mathbf{v}_j^T \mathbf{v}_j, ..., \mathbf{0}][\mathbf{v}_1, ..., \mathbf{v}_j, ..., \mathbf{v}_{m1}]^T \mathbf{x}_i \\
&= \mathbf{v}_j^T \mathbf{x}_i - \mathbf{v}_j^T \mathbf{v}_j \mathbf{v}_j^T \mathbf{x}_i = \mathbf{0}
\end{aligned}
\tag{3}
$$

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 $\overline{\mathbf{X}} \overline{\mathbf{X}}^T \mathbf{u}_i = \lambda_i \mathbf{u}_i$. It is noted that $\mathbf{u}_i$ is also orthogonal to $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_{m_1}$ as:

$$
\begin{aligned}
\mathbf{v}_j^T \overline{\mathbf{X}} \overline{\mathbf{X}}^T \mathbf{u}_i &= \mathbf{v}_j^T \lambda_i \mathbf{u}_i \\
[\mathbf{v}_j^T \overline{\mathbf{x}}_1, \mathbf{v}_j^T \overline{\mathbf{x}}_2, ..., \mathbf{v}_j^T \overline{\mathbf{x}}_m] \overline{\mathbf{X}}^T \mathbf{u}_i &= \lambda_i \mathbf{v}_j^T \mathbf{u}_i \\
\mathbf{0} &= \lambda_i \mathbf{v}_j^T \mathbf{u}_i \quad \therefore \mathbf{v}_j \perp \mathbf{u}_i
\end{aligned}
\tag{4}
$$

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

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

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

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

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

References

[1] M.S. Bartlett and T.J. Sejnowski, "Independent component representations for face recognition", *SPIE Conf. On Human Vision and Electronic Imaging III*, vol. 3299, pp. 528-539, San Jose, Jan. 1998. Recently appeared in M.S. Bartlett, J.R. Movellan and T.J. Sejnowski, "Face recognition by ICA", *IEEE Trans. on Neural Networks*, vol. 13, no. 6, Nov. 2002.

[2] T.W. Lee, M. Girolami and T.J. Sejnowski, "Independent Component Analysis using an Extended Infomax Algorithm for Mixed Sub-Gaussian and Super-Gaussian Sources", *Neural Computation*, vol. 11(2), pp. 417-441, 1999.

[3] A. X. Guan and H. H. Szu, "Local Face Statistics Recognition Methodology beyond ICA and/or PCA", *International Joint Conference on Neural Networks*, vol. 2, pp. 1016-1021, Washington, DC, 1999.

[4] N. Kwak, C. Choi and N. Ahuja, "Face recognition using feature extraction based on independent component analysis", *International Conference on Image Processing*, Rochester, New York, Sept. 2002.

[5] B. Fasel and J. Luettin, " Recognition of Asymmetric Facial Action Unit Activities and Intensities", *International Conference on Pattern Recognition*, vol. 1, pp. 1100-1103, Barcelona, Spain, 2000.

[6] B. Moghaddam, " Principal Manifolds and Bayesian Subspaces for Visual Recognition , *IEEE International Conference on Computer Vision*, vol. 2, pp. 1131-1136, Kerkyra, Greece, Sept. 1999.

[7] B. Moghaddam, W. Wahid and A. Pentland, "Beyond Eigenfaces: Probabilistic Matching for Face Recgonition", *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 30-35, Nara, Japan, 1998.

[8] B. Moghaddam, and A. Pentland, "Probabilistic visual learning for object representation", *IEEE Trans. on PAMI*, vol. 19, no. 7, pp. 696-710, July 1997.

[9] 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, Sept. 2002.

[10] L. Wang, and T. K. Tan, "Experimental Results of Face Description Based on the 2nd-order Eigenface Method", ISO/IEC JTC1/SC21/WG11/M6001, Geneva, May 2000.

[11] Kah-Kay Sung and Tomaso Poggio, "Example-based learning for view-based human face detection", *IEEE Trans. on PAMI*, vol. 20, no. 1, pp. 39-51, Jan. 1998.

[12] A. Nefian and M. Hayes. "An embedded hmm-based approach for face detection and recognition," *In Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3553–3556, 1999.

[13] B. Heisele, P. Ho, and T. Poggio, "Face Recognition with Support Vector Machines: Global versus Component-based Approach," *In Proc. IEEE International Conference on Computer Vision*, 2001.

[14] T. Kamei, "The MPEG-7 Face Dataset," TM-TK-MPEG001, Jan 18, 2002.