Is ICA Significantly Better than PCA for Face Recognition?

Jian Yang David Zhang
Biometrics Centre, Department of Computing, Hong Kong Polytechnic University, Kowloon, Hong Kong
{csjyang,csdzhang}@comp.polyu.edu.hk

Jing-yu Yang
Department of Computer Science, Nanjing University of Science and Technology, Nanjing 210094, P. R. China
yangjy@mail.njust.edu.cn

Abstract

The standard PCA was always used as baseline algorithm to evaluate ICA-based face recognition systems in the previous research. In this paper, we examine the two architectures of ICA for image representation and find that ICA Architecture I involves a PCA process by vertically centering (PCA I), while ICA Architecture II involves a whitened PCA process by horizontally centering (PCA II). So, it is reasonable to use these two PCA versions as baseline algorithms to revaluate the ICA-based face recognition systems. The experiments were performed on the FERET face database. The experimental results show there is no significant performance differences between ICA Architecture I (II) and PCA I (II), although ICA Architecture II significantly outperforms the standard PCA. It can be concluded that the performance of ICA strongly depends on its involved PCA process. The pure ICA projection has little effect on the performance of face recognition.

1. Introduction

Face recognition has received significant attention in the past decades due to its potential applications in biometrics, information security, law enforcement, etc. By far, numerous methods have been suggested to address this problem [1]. Among them, principal component analysis (PCA) turns out to be very effective [2]. Recently, a PCA closely-related method, independent component analysis (ICA) [3], has also been applied to face recognition. ICA can be viewed as a generalization of PCA since it concerns not only second-order dependencies but also high-order dependencies between variables. PCA makes the data uncorrelated while ICA makes the data as independent as possible. Bartlett et al. [4, 5], Yuen and Lai [6], and Liu and Wechsler [7] are among the first to apply ICA to face representation and recognition. They all claimed that ICA outperforms PCA for face recognition. Other researchers, however, reported different results on this subject. Baek et al [8] shown that PCA outperforms ICA while Moghaddam [9] and Jin [10] shown there is no significant difference in performance between the two methods. Socolinsky and Selinger [11] reported that ICA outperforms PCA on visible images but PCA outperforms ICA on infrared images. Recently, Draper et al [12] tried to explain why there exist such contradictory results. They re-tested ICA and PCA on the FERET database and made a comprehensive comparison between the performances of the two methods. They found the relative performance of ICA and PCA mainly depends on the ICA architecture and the distance metric.

The previous researchers [4-12], however, all use the standard PCA as the baseline algorithm to evaluate ICA-based face recognition systems. Now, a question is: is standard PCA a good choice to evaluate ICA? Actually, the ICA process, whether by Architecture I or II, contains more than a PCA process; see Figure 1. A whitening process exists between the standard PCA and ICA. After the sphering of data, we get the whitened PCA features, what about the performance of these features, in contrast to standard PCA features and the resulting ICA features? By far, no one has ever addressed this. The function of whitening process, particularly its potential effect on the recognition performance is still not clear. In the case where the performance of ICA is significantly different with that of the standard PCA, a critical problem is: what causes this difference, the whitening process or the subsequent pure ICA projection? If the whitened PCA features perform as powerful as ICA features, it is unnecessary to use a computationally expensive ICA projection for further processing. Therefore, the standard PCA seems not to be a good choice to evaluate ICA.

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**

IEEE COMPUTER SOCIETY

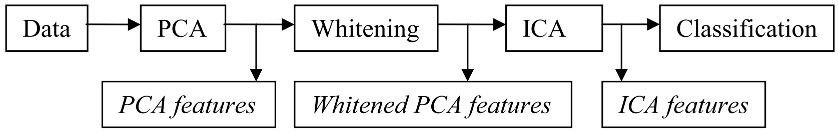


Figure 1 Illustration of ICA process for feature extraction and classification

Now, another question arises: what should be the PCA baseline algorithm to evaluate ICA? Roughly speaking, the "PCA+Whitening" (whitened PCA) can act as a baseline algorithm. In this paper, we will perform an in-depth analysis on the two architectures of ICA for image representation and find that ICA Architecture I involves a PCA process by vertically centering (PCA I), while ICA Architecture II involves a whitened PCA process by horizontally centering (PCA II). So, it is reasonable to use these two PCA versions as baseline algorithms to revaluate the ICA-based face recognition systems.

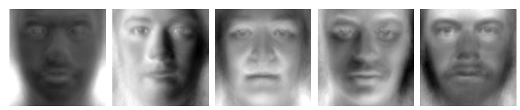
In this paper, our main goal is to investigate (1) what role the PCA whitening step and centering mode play in the ICA-based face recognition system; (2) what effect the pure ICA projection takes on the performance of face recognition, and further to reveal how the performance of two ICA architectures depends on their related PCA versions. By virtue of these investigations, we try to find some inherent reasons to explain why ICA performs better than PCA in some cases and why not in other cases.

2. Two architectures of ICA-based image representation and their corresponding PCA baseline algorithms

Regardless of what algorithm is used to perform ICA, there are generally two different architectures to be followed to apply ICA to face recognition. In Architecture I, the observed face images are viewed as a linear mixture of a set of statistically independent basis images. ICA is used to recover the set of statistically independent basis images. Then, the projection coefficients of a face image onto these basis images are used as features to represent the image for recognition. Although the basic basis images obtained by Architecture I are independent, the coefficients that code each image are not necessarily independent. Architecture II tries to use ICA to find a set of statistically independent coefficients to represent an image for recognition. Figure 2 shows some basis images corresponding to ICA Architecture I, ICA Architecture II and PCA when they are applied to face representation. From Figure 2, it can be seen that ICA Architecture I provides a more localized representation for faces, while ICA Architecture II, like PCA in a sense, provides a more holistic representation.

(a) Basis Images corresponding to ICA Architecture I

(b) Basis Images corresponding to ICA Architecture II



(c) Basis Images (eigenfaces) corresponding to PCA

Figure 2. Basis images corresponding to ICA Architecture I, ICA Architecture II and PCA

In the follow subsections, we will perform analysis on these two architectures and build the corresponding PCA baseline algorithms.

2.1. ICA Architecture I: statistically independent basis images

Given a set of M training samples (image column vectors) $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M$ in $\mathbb{R}^N$. Let us form the *image column data matrix* $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M)$ and its transpose (*image row data matrix*) $\mathbf{Y} = \mathbf{X}^{\mathrm{T}}$.

In Architecture I, the face images are viewed as random variables and the pixel values provide observations for these variables. This means that ICA is performed on the *image row data matrix* $\mathbf{Y} = \mathbf{X}^{\mathrm{T}}$. Denote $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N)$. $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N$ are used as observation vectors to evaluate the unmixing matrix of ICA model.

Centering data

Let us center the data in the observation vector space $\mathbb{R}^M$. The mean vector $\mathbf{m}_1 = E\{\mathbf{y}\} = \frac{1}{N}\sum_{j=1}^{N}\mathbf{y}_j$.

Denote $\mathbf{m}_{\mathrm{I}} = (m_1, m_2, \cdots, m_M)^{\mathrm{T}}$. Actually, $m_j = E\{\mathbf{x}_j\}$, i.e., the mean of all pixel values of image *j*. Every observation is subtracted by the mean vector $\mathbf{m}_{\mathrm{I}}$, i.e., $\mathbf{y}_j \leftarrow (\mathbf{y}_j - \mathbf{m}_{\mathrm{I}})$. Then, we get the centered *image row data matrix* $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N)$. Let us denote $\mathbf{X}_v = \mathbf{Y}^{\mathrm{T}}$. $\mathbf{X}_v$ is called the vertically-centered *image column data matrix,* in which every column is an zero-mean image, i.e., the image that has been removed the mean of pixel values.

Whitening data

We will perform PCA based on the centered observation vectors $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N$ to whiten the data. The covariance matrix is

$$\Sigma_{\mathrm{I}} = \frac{1}{N} \sum_{j=1}^{N} \mathbf{y}_j \mathbf{y}_j^{\mathrm{T}} = \frac{1}{N} \mathbf{Y}\mathbf{Y}^{\mathrm{T}} \tag{1}$$

Let us denote $\mathbf{G}_v = \mathbf{Y}\mathbf{Y}^{\mathrm{T}}$. Calculate the orthonormal eigenvectors $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m$ of $\mathbf{G}_v$ corresponding to *m* largest positive eigenvlaues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. Then, the *m* largest positive eigenvlaues of Σ_{I} are $\frac{\lambda_1}{N}, \frac{\lambda_2}{N}, \cdots, \frac{\lambda_m}{N}$, and the associated orthonormal eigenvectors are $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m$.

Denoting $\mathbf{V} = (\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m)$ and $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m)$, we obtain the whitening matrix

$\mathbf{P} = \mathbf{V}(\frac{1}{N}\Lambda)^{-\frac{1}{2}} = \sqrt{N}\,\mathbf{V}\Lambda^{-\frac{1}{2}}$, such that

$$\mathbf{P}^{\mathrm{T}} \Sigma_{\mathrm{I}} \mathbf{P} = \mathbf{I} \tag{2}$$

The data matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_N)$ can be whitened by the following transformation

$$\mathbf{R} = \mathbf{P}^{\mathrm{T}} \mathbf{Y} \tag{3}$$

ICA Processing

We perform ICA on **R**, producing a matrix $\mathbf{U}_{\mathrm{I}}$ with *m* indepdent basis images in its rows (see Figure 2 (a)), that is,

$$\mathbf{U}_{\mathrm{I}} = \mathbf{W}_{\mathrm{I}} \mathbf{R} \tag{4}$$

where $\mathbf{W}_{\mathrm{I}}$ is the unmixing matrix of ICA.

For a given image **x** in a column vector, after vertically-centering and projecting it onto these *m* indepdent basis images, we have

$$\mathbf{z} = \mathbf{U}_{\mathrm{I}} \mathbf{x} = \mathbf{W}_{\mathrm{I}} \mathbf{R}\, \mathbf{x} \tag{5}$$

This transformation can be decomposed into two ones:

$$\mathbf{y} = \mathbf{R}\,\mathbf{x} = \mathbf{P}^{\mathrm{T}} \mathbf{Y}\,\mathbf{x} = \left(\sqrt{N}\,\mathbf{V}\Lambda^{-\frac{1}{2}}\right)^{\mathrm{T}} \mathbf{Y}\,\mathbf{x} = \sqrt{N}\,\Lambda^{-\frac{1}{2}} \mathbf{V}^{\mathrm{T}} \mathbf{Y}\,\mathbf{x} \tag{6}$$

$$\mathbf{z} = \mathbf{W}_{\mathrm{I}}\, \mathbf{y} \tag{7}$$

Note that here, we use a projection-based method [10] to get the representation coefficients of an image, that is, projecting the image directly on the axes formed by the *m* indepdent source images resulting from ICA. This is different from the ways used in [5], where a minimum squared error (MSE) based representation was suggested. It is not hard to show the two kinds of representation methods are actually equivalent when the unmixing matrix $\mathbf{W}_{\mathrm{I}}$ is orthogonal ($\mathbf{W}_{\mathrm{I}}$ should be orthogonal in theory; actually, we can obtain an orthogonal unmixing matrix by FastICA algorithm [3]).

2.2. PCA baseline algorithm I for ICA Architecture I

PCA by Vertically Centering

Recall the vertically-centered *image column data matrix* $\mathbf{X}_v = \mathbf{Y}^{\mathrm{T}}$, based on which we can construct the following covariance matrix:

$$\Sigma_v = \frac{1}{M} \mathbf{X}_v \mathbf{X}_v^{\mathrm{T}} = \frac{1}{M} \mathbf{Y}^{\mathrm{T}} \mathbf{Y} \tag{8}$$

From the Singular value decomposition (SVD) theorem [13], we know $\mathbf{G}_v = \mathbf{Y}\mathbf{Y}^{\mathrm{T}}$ and $\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$ have the same nonzero eigenvalues. Suppose the orthonormal eigenvectors $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, \cdots, \boldsymbol{\gamma}_m$ of $\mathbf{G}_v$ corresponding to *m* largest positive eigenvlaues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. Then, we obtain the corresponding orthonormal eigenvectors of $\mathbf{Y}^{\mathrm{T}}\mathbf{Y}$:

$$\boldsymbol{\beta}_j = \frac{1}{\sqrt{\lambda_j}} \mathbf{Y}^{\mathrm{T}} \boldsymbol{\gamma}_j, \quad j = 1, \cdots, m. \tag{9}$$

Thus, the *m* largest positive eigenvlaues of Σ_v are $\frac{\lambda_1}{M}, \frac{\lambda_2}{M}, \cdots, \frac{\lambda_m}{M}$, and the associated orthonormal eigenvectors are $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m$.

Let $\mathbf{P}_v = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m) = \mathbf{Y}^{\mathrm{T}}\mathbf{V}\Lambda^{-\frac{1}{2}} = \mathbf{X}_v \mathbf{V}\Lambda^{-\frac{1}{2}}$. The corresponding PCA transformation (*by vertically centering*) is

$$\mathbf{y} = \mathbf{P}_v^{\mathrm{T}}\, \mathbf{x} = \left(\mathbf{Y}^{\mathrm{T}} \mathbf{V}\Lambda^{-\frac{1}{2}}\right)^{\mathrm{T}} \mathbf{x} = \Lambda^{-\frac{1}{2}} \mathbf{V}^{\mathrm{T}} \mathbf{Y}\,\mathbf{x} \tag{10}$$

Comparing Eq. (10) with the Eq. (6), we find that the first transformation in ICA Artchetecture I is exactly the PCA transformation *by vertically centering* (the constant $\sqrt{N}$ has no effect on the results in the sense of feature extraction). The transform matrix $\mathbf{R} = \sqrt{N}\,\mathbf{P}_v^{\mathrm{T}}$. That is, its rows are principal eigenvectors of Σ_v, i.e., a set of eigen-images. The followed ICA is to recover a set of independent basic images from these eigen-images.

Since the ICA Artchetecture I involves a PCA process *by vertically centering*, this kind of PCA should be taken as a baseline algorithm to evaluate

IEEE COMPUTER SOCIETY

ICA Artchetecture I. This PCA algorithm (*by vertically centering*) is named *PCA Baseline Algorithm I* (PCA I).

2.3. ICA Architecture II: statistically independent representation coefficients

Given a set of M training samples (image column vectors) $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M$ in $\mathbb{R}^N$. Let us form the *image column data matrix* $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M)$.

The goal of ICA Architecture II is to find statistically independent coefficients for input image data. In this architecture, the face images are viewed as observations and the pixel values are random variables. ICA is performed directly on the *image column data matrix* $\mathbf{X}$. In other words, $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M$ are used as observation vectors to evaluate the unmixing matrix of ICA model.

Centering data

Let us center the data in the observation vector space $\mathbb{R}^N$. The mean vector $\mathbf{m}_{\mathrm{II}} = E\{\mathbf{x}\} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{x}_j$. Every observation is subtracted by the mean vector $\mathbf{m}_{\mathrm{II}}$, i.e., $\mathbf{x}_j \leftarrow (\mathbf{x}_j - \mathbf{m}_{\mathrm{II}})$, then, we get the centered *image row data matrix* $\mathbf{X}_h = (\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M)$. This way of centering is named *horizontally centering*, in contrast with the *vertically centering* of data in Section 2.2.

Whitening data

PCA is first performed based on the centered observation vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M$ to whiten the data. This process will be detailed in the next subsection. Suppose the PCA whitening matrix is $\mathbf{P}_w$. We obtain the whitened data matrix $\mathbf{Y} = (\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_M)$ after the following transformation

$$\mathbf{Y} = \mathbf{P}_w^{\mathrm{T}}\mathbf{X}_h \tag{16}$$

ICA Processing

Then, we perform ICA based on the sphered data $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_M$. Suppose the resulting unmixing matrix is $\mathbf{W}_{\mathrm{II}}$. The whole transform matrix $\mathbf{U}_{\mathrm{II}}$ of ICA Architecture II is

$$\mathbf{U}_{\mathrm{II}} = \mathbf{W}_{\mathrm{II}}\mathbf{P}_w^{\mathrm{T}} \tag{17}$$

The row vectors of $\mathbf{U}_{\mathrm{II}}$ are corresponding to the basic images of ICA Architecture II (see Figure 2 (b)). These images are not necessarily independent. They, however, form the coordinate axes to make the projection coefficients of each sample as independent as possible.

Given an image $\mathbf{x}$ in a column vector, after horizontally-centering and projecting it onto these basic images, we have

$$\mathbf{z} = \mathbf{U}_{\mathrm{II}}\mathbf{x} = \mathbf{W}_{\mathrm{II}}\mathbf{P}_w^{\mathrm{T}}\mathbf{x} \tag{18}$$

The vector $\mathbf{z}$, containing a set of independent coefficients, is used to represent the image $\mathbf{x}$ for recognition.

It is obvious that the transformation in Eq.(18) can be decomposed into two ones: a whitened PCA transform $\mathbf{y} = \mathbf{P}_w^{\mathrm{T}}\mathbf{x}$ followed by an ICA transform $\mathbf{z} = \mathbf{W}_{\mathrm{II}}\mathbf{y}$.

2.4. PCA baseline algorithm II for ICA Architecture II

PCA by Horizontally Centering (Standard PCA)

Let us construct the covariance matrix based on the horizontally-centered observation vectors $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_M$:

$$\Sigma_{\mathrm{II}} = \frac{1}{M}\sum_{j=1}^{M}\mathbf{x}_j\mathbf{x}_j^{\mathrm{T}} = \frac{1}{M}\mathbf{X}_h\mathbf{X}_h^{\mathrm{T}} \tag{19}$$

Suppose $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m$ are the orthonormal eigenvectors of Σ_{II} corresponding to m largest positive eigenvlaues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_m$. Then, the standard PCA transform is

$$\mathbf{y} = \mathbf{P}_h^{\mathrm{T}}\mathbf{x}\text{, where } \mathbf{P}_h = (\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \cdots, \boldsymbol{\beta}_m) \tag{20}$$

Whitened PCA

Denoting $\mathbf{P}_w = \mathbf{P}_h\Lambda^{-\frac{1}{2}}$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \cdots, \lambda_m)$, we have

$$\mathbf{P}_w^{\mathrm{T}}\Sigma_{\mathrm{II}}\mathbf{P}_w = \mathbf{I} \tag{21}$$

Thus, we obtain the whitening matrix $\mathbf{P}_w$. The data can be sphered by the following *whitened PCA* transform

$$\mathbf{y} = \mathbf{P}_w^{\mathrm{T}}\mathbf{x} \tag{22}$$

The *whitened PCA* transform not only eliminates the correlation betweens variables but also normalizes the deviation of each variable, that is, makes the deviation to be one.

Since the ICA Artchetecture II involves not only a standard PCA (*by horizontally centering*) but also a whitened PCA process, the two PCA versions should be taken as baseline algorithms to evaluate ICA Artchetecture II. The whitened PCA algorithm (*by horizontally centering*) is named *PCA Baseline Algorithm II* (PCA II).

2.5. Summary

The two architectures of ICA involve two different versions of PCA: ICA Architecture I includes a PCA by vertically centering (PCA I), while ICA

1550-5499/05 $20.00 © 2005 IEEE

Architecture II includes a whitened PCA by horizontally centering (PCA II), that is

$$\boxed{\text{ICA Architecture I}} \Leftrightarrow \boxed{\text{PCA I}} + \boxed{\text{ICA}} \quad (23)$$

$$\boxed{\text{ICA Architecture II}} \Leftrightarrow \boxed{\text{PCA II}} + \boxed{\text{ICA}} \quad (24)$$

In contrast to standard PCA, PCA I removed the mean of each image while standard PCA removes the mean image of all training samples. PCA II is the whiten version of standard PCA. It can normalize the deviation of each coefficient as well as make the coefficients uncorrelated.

To evaluate the performance of the two architectures of ICA for image representation and recognition, it is necessary to compare them with the two different versions of PCA, as well as the standard PCA. In other words, PCA I, PCA II and standard PCA should be selected as baseline algorithms to evaluate ICA.

3. Experiments and Analysis

The experiments are performed using the FERET 1996 standard subset. In this subset, the basic gallery contains 1,196 face images. There are four sets of probe images compared to this gallery: *fafb*, *fafc* Duplicates I and II [14]. In our experiments, the face portion of each original image is automatically cropped based on the location of eyes and resized to an image of 80×80 pixels. The resulting image is then pre-processed by a histogram equalization algorithm.

The purpose of our experiments is to compare the performance of the standard PCA, PCA I, PCA II, and ICA Architectures I and II based face recognition systems. In order to reduce the effect that might be induced by the choice of the training sample set, we run each system ten times. In each time, the training sample set that containing 500 images is randomly selected from the gallery so that the training sample sets are different for ten tests.

By training, the projector of each method is obtained and, for each face image, 200 features are extracted for representation and recognition purpose. Note that the PC (principle component) number is selected as 205 in the pre-processing phase (i.e., PCA I or II) of ICA Architecture I or II. We adopt a *kurtosis* closely related contrast function $G(u) = \frac{1}{4}u^4$ in ICA model and use the FastICA codes that are publicly available at the website http://www.cis.hut.fi/projects/ica/fastica to calculate the projector of ICA. After feature extraction, a near neighbor classifier with different distance metrics is employed for classification. Three distance metrics: L2 (Euclidean) distance metric, L1 (city-block) distance metric, and cosine distance, are used in standard PCA. Only the cosine distance is adopted in ICA (Architectures I and II), PCA I and PCA II because this metric was demonstrated most effective for ICA [5, 12] and, PCA I and PCA II need the same distance metric to evaluate ICA. For each method and each probe set, the average recognition rate across ten tests is listed in Table 1. Taking all four probe sets as a whole, the total recognition rate of each method is also calculated and listed in this table.

Table 1 The average recognition rates of the standard PCA, PCA I, PCA II, and ICA Architectures I and II with different distance metrics on the FERET database

Probe set	ICA Arch. I	PCA I	Standard PCA			PCA II	ICA Arch. II
	Cosine	Cosine	L2	L1	Cosine	Cosine	Cosine
fafb	77.66	77.74	77.18	76.49	76.67	81.66	81.23
fafc	14.21	14.32	14.84	38.42	11.06	63.91	63.81
dup. I	32.32	32.35	32.06	33.89	33.80	46.37	46.33
Dup. II	10.46	10.41	10.15	13.03	12.81	26.37	26.41
Total	51.74	51.79	51.44	53.89	51.67	63.80	63.57

Table 1 shows us that **i**) ICA Architecture II (with cosine distance) significantly outperforms the standard PCA no matter what distance metric is used by PCA. **ii**) PCA with L1 distance is slightly better than ICA Architecture I in terms of the total recognition rate. All of these results are consistent with Draper et al.'s studies [12]. Based on these studies, Draper et al. concluded that ICA Architecture II is better than PCA for identifying faces.

However, if we use the proposed PCA baseline algorithms to reevaluate the performance of ICA for face recognition, we will draw absolutely different conclusions. As shown in Table 1, PCA II (I) can perform as well as (even slightly better than) ICA Architecture II (I). There is no significant performance difference between ICA Architecture I (II) and PCA I (II). It seems that *the pure ICA projection has trivial effect on the performance of face recognition*.

Now, a question is: what on earth causes the remarkable performance difference between ICA Architecture I (II) and the standard PCA? In our opinion, the underlying reason is the inherent difference between the involved PCA I (II) and the standard PCA, rather than the involvement of a *pure ICA projection*. The difference between PCA I and the standard PCA is the centering mode, PCA I centers the data by removing the mean of each image (i.e., vertical centering) while the standard PCA by removing the mean image of all training samples (i.e., horizontal centering). Vertical centering may benefit for the recognition in varying illumination, especially that caused by lighting intensity. This leads to the result that PCA I performs better than standard PCA with the same distance metric (cosine distance) on Probe set *fafc*, as shown in Table 1. In such a case, ICA Architecture I performs better than standard PCA. However, the centering mode does not affect the recognition accuracy much in the average sense. So, there is no significant performance difference between PCA I and standard PCA with the same distance metric in terms of the total recognition rate. This gives rise to the similar results between ICA Architecture I and standard PCA with the same distance metric.

PCA II is the whiten version of standard PCA, i.e., standard PCA plus a whitening step. The whitening step is really helpful for improving the recognition rate on this database. It causes the significant performance difference between PCA II and standard PCA, as shown in Table 1. Further, this whitening step makes ICA Architecture II significantly outperform standard PCA.

In a word, it is the *centering mode* (not the *pure ICA projection*) that causes the remarkable performance difference between ICA Architecture I and the standard PCA (if there exists such a performance difference between them). It is the *whitening step* (not the *pure ICA projection*) that causes the significant performance difference between ICA Architecture II and the standard PCA

4. Conclusions

In this paper, we examine the two architectures of ICA for image representation and find that ICA Architecture I involves a PCA process by vertically centering (PCA I), while ICA Architecture II involves a whitened PCA process by horizontally centering (PCA II). These two PCA versions are used as baseline algorithms to revaluate the ICA-based face recognition systems. The experimental results show there is no significant performance differences between ICA Architecture I (II) and PCA I (II), although ICA Architecture II significantly outperforms the standard PCA on the FERET database. The recognition performance of ICA, whether using Architecture I or II, strongly depends on its involved PCA process (PCA I or II). The pure ICA projection seems to have little effect on the performance of face recognition.

References

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, "Face recognition: A literature survey." *Technical Report CAR-TR-948, UMD CS-TR-4167R*, August, 2002.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition", *J. Cognitive Neuroscience*, 1991, 3(1), pp. 71-86.

[3] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.

[4] M. S. Bartlett and T. J. Sejnowski, "Independent components of face images: a representation for face recognition", *Proceedings of the 4th Annual Jount Symposium on Neural Computation*, Pasadena, CA, May 17, 1997.

[5] M.S. Bartlett, J.R. Movellan and T.J. Sejnowski, "Face recognition by independent component analysis", *IEEE Trans. Neural Networks*, 2002, 13(6), pp. 1450-1464.

[6] Pong C. Yuen, and J.H. Lai, "Face representation using independent component analysis", *Pattern Recognition*, 2002, 35 (6), 1247-1257.

[7] C. Liu and H. Wechsler: "Independent component analysis of Gabor features for face recognition", *IEEE Trans. Neural Networks*, 2003, 14 (4), pp. 919-928.

[8] K. Baek, B. A. Draper, J. R. Beveridge, and K. She, "PCA vs ICA: A comparison on the FERET data set", *Joint Conference on Information Sciences*, Durham, N.C., 2002.

[9] B. Moghaddam, "Principal manifolds and probabilistic subspaces for visual recognition*", IEEE Trans. Pattern Anal. Machine Intell.*, 2002, June, 24(6), pp.780-788.

[10] Z. Jin, F. Davoine, "Orthogonal ICA representation of images", *Proceedings of 8th International Conference on Control, Automation, Robotics and Vision*, pp.369-374, Kunming, China, 6-9th, December 2004.

[11] D. Socolinsky and A. Selinger, "A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery", *International Conference on Pattern Recognition*, Quebec City, 2002.

[12] B.A. Draper, K. Baek, M.S. Bartlett, J.R. Beveridge, "Recognizing faces with PCA and ICA", *Computer vision and image understanding*, Special issue on Face Recognition, 2003, July, 91(1/2), 115-137.

[13] G. H. Golub and C. F. Van Loan, Matrix Computations, The Johns Hopkins University Press, Baltimore and London, Third edition, 1996.

[14] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms", *IEEE Trans. Pattern Anal. Machine Intell.*, 2000, 22 (10), pp.1090-1104.