中正嶺學報 第三十三卷 第一期 民國 93.11.
JOURNAL OF C.C.I.T., VOL.33, NO.1, NOV., 2004.

The Evaluation of Noise Immunity for PCA and ICA in Face Recognition

Ping S. Huang, Chung-Shi Chiang, and Ying-Tsung Lin

Department of Electrical Engineering, Chung Cheng Institute of Technology, National Defense University

ABSTRACT

Principal Component Analysis (PCA) and Independent Component Analysis (ICA) have been successfully used for face recognition. PCA is computed from the global covariance matrix of the full set of image data, the obtained basis vectors are global representations that are not suitable for the recognition of non-aligned faces. Relatively, ICA basis vectors are more spatially local than PCA basis vectors and give better face representation. This paper addresses the evaluation results of noise immunity for PCA and ICA in face recognition. The recognition performance of PCA and ICA are compared and analyzed when the test images are contaminated by noises. Also, different methods for eigenvector selection and similarity measures are evaluated. ICA has achieved better recognition performance in noise immunity than PCA, as shown in the experimental results.

Keywords: biometrics, face recognition, template matching, principal component analysis, independent component analysis, noise immunity

PCA 與 ICA 方法用於人臉辨識之抗雜訊能力評估

黃炳森　江中熙　林穎聰

國防大學中正理工學院電機工程學系

摘　　要

主值分析(PCA)與獨立值分析(ICA)已經成功地被用在人臉辨識工作上。主值分析是藉由對整組影像資料的整體共異矩陣(global covariance matrix)執行計算所得，所獲得之特徵向量代表整體統計資訊，並不適合用於無對正之人臉影像上。相對地，獨立值分析之特徵向量較具有空間之區域特性，適合於人臉影像之表示。本篇論文著重在主值分析與獨立值分析用於人臉辨識時之抗雜訊能力評估。兩者之辨識效能在測試影像受雜訊干擾情況下作了比較與分析。同時，我們對於不同之特徵向量選取與相似度分析方法也作了評估。實驗結果顯示，獨立值分析法在影像受雜訊干擾情況下，對於人臉辨識所達成之效能優於主值分析法。

關鍵詞：生物認證，人臉辨識，樣板比對，主值分析，獨立值分析，抗雜訊力

文稿收件日期 92.12.4; 文稿修正後接受日期 93.2.24
Manuscript received December 4, 2003; revised February 24, 2004.

I. INTRODUCTION

Biometrics is the method to automatically recognize a person by his physiological or behavioral characteristics. Examples of human traits currently used for biometric recognition include fingerprints, speech and face. An important advantage of biometrics lies in the fact that physical or behavioral traits cannot be transferred to other individuals. Among those human traits mentioned previously, face recognition has been investigated and conducted on various aspects in psychophysics, neurosciences and engineering since the early 1960s. A variety of approaches ranging from Karhunen-Loeve expansion, feature matching and neural networks had been proposed for automatic face recognition [1]. Most of them can be divided into two methods: feature-based and template-based. Based on the statistical analysis in template matching, different template-based approaches including PCA (Principal Component Analysis) [2] and ICA (Independent Component Analysis) [3-5] are proposed.

Both of PCA and ICA include a subspace projection step that projects data from a high-dimensional space to a more meaningful, lower-dimensional space [6]. In PCA, the eigenvectors from the covariance matrix of training data constitute the eigenspace. The statistically independent basis vectors calculated by ICA create the independent component subspace. Face recognition is then implemented in this reduced space by using nearest-neighbor classification rule. PCA is optimal and useful only with respect to data compression and decorrelation of low (second) order statistics. ICA is a generalization of PCA that separates the high-order statistics of the input in addition to the second-order statistics. As PCA derives only the most expressive features for face reconstruction rather than face classification, ICA thus provides a more powerful data representation than PCA [7].

Face image representations based on PCA have been successfully used for recognition [2]. However, since PCA is computed from the global covariance matrix of the full set of image data, the obtained basis vectors are global representations that are not suitable for the recognition of non-aligned faces. To overcome this limitation, since ICA basis vectors are more spatially local than PCA basis vectors, ICA has been proposed for face recognition [3-5]. Using ICA, each face can be represented by the linear combination of extracted independent basis vectors. Face recognition can be achieved in the feature space spanned by their corresponding basis vectors.

This paper addresses the problem of face recognition using PCA and ICA in noise immunity. At first, two variations for PCA and ICA subspaces are examined. The first variation is to maintain the spectral energy needs of PCA and ICA for appropriate representation. This will affect the selection of eigenvectors. The second variation is the evaluation of similarity or distance measures in face recognition performed on a low-dimensional subspace. After the evaluation, suitable similarity or distance measures can be decided. After the examination of those two variations, the recognition performance of PCA and ICA are compared and analyzed when the test images are contaminated by noises.

The remaining of this article is organized as follows. Section 2 describes the method of PCA and ICA used in this work. Eigenvector Selection and Similarity and Distance Measures are presented in Section 3. The experimental results are demonstrated and discussed in Section 4, prior to Conclusions in Section 5.

II. METHOD

An image may be viewed as a vector of pixels where the value of each entry in the vector is the grayscale value of the corresponding pixel. The image is said to locate in a N-dimensional space, where N is the number of pixels. This vector representation of the image is considered to be the original space of the image. In this work, two specific subspaces are created. One is created from the eigenvectors of the covariance matrix of the training data computed by PCA. Another one is created from the basis vectors calculated by ICA. Although some of the details may vary, there is a basic algorithm to identify images by projecting them into a subspace. The first step is to select a subspace on which to project the images. When this subspace is selected, all training images are projected from their original image space into this subspace. Next step is to project each test image into this subspace. In this subspace, each test image is compared to all the training images by using a selected similarity or distance measure. The training image found to be most similar or closest to the test image is used to identify the test image. The fundamentals to create the subspace by PCA and ICA are described in Section 2.1 and Section 2.2.

2.1. Principal Component Analysis (PCA)

PCA is a standard decorrelation technique and one can derive an orthogonal projection basis that directly leads to dimensionality reduction, and possibly to feature selection [2]. Eigenspace is obtained by identifying the eigenvectors of the covariance matrix derived from a set of training images. The eigenvectors corresponding to non-zero eigenvalues of the covariance matrix form an orthonormal basis. Let $\boldsymbol{X} \in R^N$ be a random vector representing an image, where N is the dimensionality of the corresponding image space. By concatenating the image columns, each random vector can be represented by $\boldsymbol{X}_i$, $i = 1, 2, ..., P$, in which P is the number of images and N is the number of pixels. The images are mean-centered by subtracting the mean image from each image vector and combined to be a data matrix of size $N \times P$ given by

$$\boldsymbol{X} = [\boldsymbol{X}_1 \ \ \boldsymbol{X}_2 \ \ \cdots \ \ \boldsymbol{X}_P] \qquad (1)$$

The covariance matrix $\boldsymbol{C}$ of $\boldsymbol{X}$ is defined as follow:

$$\boldsymbol{C} = \boldsymbol{X}\boldsymbol{X}^T \qquad (2)$$

where $\boldsymbol{C} \in R^{N \times N}$. The PCA of a random vector $\boldsymbol{X}$ factorizes the covariance matrix $\boldsymbol{C}$ into the following form:

$$\boldsymbol{C} = \boldsymbol{\Phi}\boldsymbol{\Lambda}\boldsymbol{\Phi}^T \qquad (3)$$

where $\boldsymbol{\Phi} = [\boldsymbol{\Phi}_1, \boldsymbol{\Phi}_2, ..., \boldsymbol{\Phi}_N] \in R^{N \times N}$ is the orthonormal eigenvector matrix and $\boldsymbol{\Lambda} = diag\{\lambda_1, \lambda_2, ..., \lambda_N\} \in R^{N \times N}$ is the diagonal eigenvalue matrix with diagonal elements in decreasing order $(\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_N)$ This covariance matrix has up to P eigenvectors associated with non-zero eigenvalues, assuming $P < N$.

Assume that we have selected the number of eigenvectors that span an eigenspace, all image vectors can be projected into this eigenspace by

$$\boldsymbol{V} = \boldsymbol{\Phi}^T \boldsymbol{X} \tag{4}$$

where $\boldsymbol{V} = [v_1, v_2, \ldots, v_P]$ represents the P points with each point $v_i, i = 1, 2, \ldots, P$, corresponding to one training image. After any testing image is projected into the eigenspace, this projected point is compared to the P points of projected training images using a selected similarity measure. The training image that is found to be closest to the testing image is used to identify the testing image.

2.2. Independent component analysis (ICA)

Independent Component Analysis (ICA) [7, 8] has been used recently for the problem of blind source separation, while its possible use is also for face recognition [3-5]. ICA searches for a linear transformation to express a set of random variables as linear combinations of statistically independent source variables.

The search criterion involves the minimization of the mutual information expressed as a function of high-order cumulants. PCA considers only the second-order moments and uncorrelates data, while ICA accounts for higher order statistics and identifies the independent source components from their linear mixtures. ICA thus provides a more powerful data representation than PCA. As PCA derives only the most expressive features for face reconstruction rather than face classification, one would usually use some subsequent discriminant analysis to enhance PCA performance [9].

Here, based on [10], we only give a brief introduction on ICA needed for face image data in this paper. Using ICA, one attempts to represent each image as a linear combination of 'basis' patches. This can be represented by

$$\boldsymbol{Y} = \boldsymbol{A}s = \sum_{i=1}^{n} \boldsymbol{a}_i s_i \tag{5}$$

where the input image is denoted by $\boldsymbol{Y}$, and the basis images are the $\boldsymbol{a}_i$, the columns of $\boldsymbol{A}$. Fig. 1 shows an example that a face image can be represented by a linear combination of several image patches.

One then optimizes $\boldsymbol{a}_i$ such that for typical $\boldsymbol{Y}$, most of the s_i will be close to zero and only a few will have significantly non-zero values. This led to features qualitatively similar to simple cell receptive fields.

Assume that a random vector $\boldsymbol{Y}$ can be represented by a linear combination of image

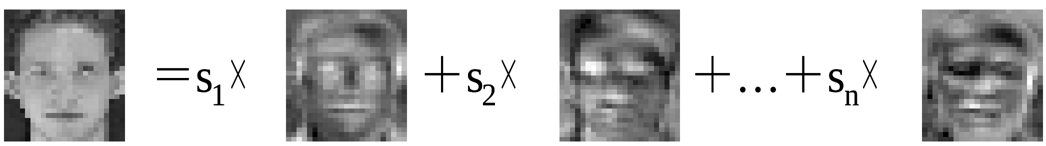

Fig. 1. Using ICA, each image is represented as a linear combination of basis patches.

patches shown in Fig. 1. Then, the goal of ICA is to express the data by a linear model where the stochastic sources (s_i) are as mutually independent as possible. Basically, there are two standard preprocessing steps in ICA [10]. First, the mean of the data $\boldsymbol{Y}$ is subtracted to center the data on the origin. This does not alter the ICA model in Eq. (5) unless we have zero-mean sources. The second step is to whiten the data so that the components are uncorrelated and have unit variance. The whitening can be done by

$$\boldsymbol{z} = \boldsymbol{W}\boldsymbol{Y} \tag{6}$$

where $\boldsymbol{W}$ is the whitening matrix and $\boldsymbol{z}$ the whitened data.

There are two commonly used and analytically available solutions [11] for image data. The first one is the symmetric whitening matrix $\boldsymbol{W}_{swm} = E\{\boldsymbol{Y}\boldsymbol{Y}^T\}^{-1/2}$. Second, there is the PCA solution $\boldsymbol{W}_{PCA} = \boldsymbol{D}^{-1/2}\boldsymbol{E}^T$, where $\boldsymbol{E}\boldsymbol{D}\boldsymbol{E}^T = E\{\boldsymbol{Y}\boldsymbol{Y}^T\}$ is the eigensystem of the correlation matrix of $\boldsymbol{Y}$. Eigenvectors are placed in the columns of $\boldsymbol{E}$ and their corresponding eigenvalues are in the diagonal elements of $\boldsymbol{D}$. For image data, PCA allows one to optimally reduce the dimension by selecting only a subset of the components of $\boldsymbol{z} = \boldsymbol{W}_{PCA}\boldsymbol{Y}$.

After the data is preprocessed and whitened, the goal of ICA is to find a transform $\boldsymbol{U}$ that minimizes the statistical dependencies between the estimated sources [10]

$$\hat{s} = \boldsymbol{U}\boldsymbol{z} = \boldsymbol{U}\boldsymbol{W}_{PCA}\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{D}_n^{-1/2}\boldsymbol{E}_n^T\boldsymbol{Y} \tag{7}$$

where $\boldsymbol{D}_n$ denotes the diagonal matrix containing the n largest eigenvalues (of the correlation matrix $E\{\boldsymbol{Y}\boldsymbol{Y}^T\}$) and E_n is the matrix with corresponding eigenvectors as columns. Optimally, the dependencies are measured by the mutual information of the sources [12]. However, for a fast and simple algorithm, $\boldsymbol{U}$ is constrained to be orthogonal (i.e. the estimated sources are constrained to be uncorrelated), and the mutual information is approximated [12]. Based on the FastICA algorithm [12], we use the algorithm proposed in [10] to calculate the estimated basis. The algorithm starts from a random orthogonal matrix $\boldsymbol{U}$ and iterates to update each row $\boldsymbol{u}_i^T$ of $\boldsymbol{U}$ until converged. After the algorithm is converged, the estimated basis can be constructed by

$$\boldsymbol{A} = \boldsymbol{E}_n\boldsymbol{D}_n^{1/2}\boldsymbol{U}^T \tag{8}$$

and each column $\boldsymbol{a}_i$ of $\boldsymbol{A}$ can be identified as the basis patch of one independent component. The transformation matrix $\boldsymbol{T}$ of ICA is now defined by

$$\boldsymbol{T} = \boldsymbol{U}\boldsymbol{D}_n^{-1/2}\boldsymbol{E}_n^T \tag{9}$$

The new feature vector $\hat{s}$ of the image is given by

$$\hat{s} = \boldsymbol{T}\boldsymbol{Y} \tag{10}$$

III. EIGENVECTOR SELECTION AND SIMILARITY MEASURES

Since the computation of PCA and ICA algorithms is originated from the eigensystem, the recognition performance is affected by the selection of eigenvectors. Furthermore, the performance is also affected by the selected similarity measures while calculating the distances between training and testing images in the subspace of PCA or ICA. Therefore, the evaluation of the methods for Eigenvector Selection and Similarity Measures are presented in this section. The evaluation techniques are originated from [13] which is used to compare the face recognition performance of PCA-Based and Fisher Discriminant-Based techniques. Here, they are adopted to compare PCA and ICA algorithms in face recognition.

3.1. Eigenvector Selection

Until now, for PCA or ICA, all eigenvectors associated with non-zero eigenvalues are used to create a subspace. The computation time of PCA and ICA are directly proportional to the number of eigenvectors used to create the subspace. Therefore, removing some portion of the eigenvectors can reduce the computation time. Furthermore, as the small trailing eigenvalues tend to capture noise, by removing additional eigenvectors that do not contribute to the classification of the face images, recognition performance can be improved and more robust. Many approaches for eigenvector selection have been considered, here, five of them are described [13]:

A. Standard eigenspace projection: All eigenvectors corresponding to non-zero eigenvalues are used to create the subspace.

B. Remove the last 40% of the eigenvectors: Since the eigenvectors are sorted by the corresponding descending eigenvalues, 40% of the eigenvectors with the least amount of eigenvalues are removed.

C. Energy dimension: This method uses the minimum number of eigenvectors to guarantee that the accumulated energy is greater than a threshold value. A typical threshold value is 0.9. The accumulated energy of the first i eigenvectors is the ratio of the sum of the first i eigenvalues over the sum of all the eigenvalues:

$$e_i = \frac{\sum_{j=1}^{i} \lambda_j}{\sum_{j=1}^{k} \lambda_j} \tag{11}$$

where k is the number of eigenvectors with non-zero eigenvalues.

D. Stretching dimension: Another method for selecting eigenvectors is to calculate the stretch of an eigenvector. The stretch of the i^{th} eigenvector is the ratio of the i^{th} eigenvalue (λ_i) over the maximum eigenvalue (λ_i). A common threshold for the stretching dimension is 0.01.

$$S_i = \frac{\lambda_j}{\lambda_1} \tag{12}$$

E. Removing the first eigenvector: The three methods from Eqs. (2)-(4) assume that the information in the last eigenvectors relate to noises and work against classification. However, this method assumes that information in the first eigenvector is affected by lighting conditions and works against classification. For example, lighting can cause considerable variation for identical images. Hence, this method is presented to remove the first eigenvector and improve the classification accuracy.

3.2. Similarity and Distance Measures

For PCA or ICA, to improve the classification performance, the number of eigenvectors has to be determined for subsequent subspace projection. After images are projected into the subspace, they are converted into associated feature vectors. The recognition task is to determine which images are most like one another. In general, there are two ways to achieve this task. One is to measure the distance between two associated feature vectors in subspace. The second way is to measure how similar two associated feature vectors are. The goal in measuring distance is to minimize the distance; so two images that are alike produce a small distance between their associated feature vectors. When measuring similarity, one wishes to maximize similarity, so that two like images produce a high similarity value between their associated feature vectors. However, there are many possible similarity and distance measures. Here, four of them are described [13]:

A. *L1* norm: The *L1* norm is a distance measure that is also known as the city block norm. It sums up the absolute difference between feature vectors. The *L1* norm of feature vector A and feature vector B is:

$$D_{L1}(A, B) = \sum_{i=1}^{n} \left| A_i - B_i \right| \tag{13}$$

where n is the number of elements in vector A and vector B.

B. *L2* norm: The *L2* norm is a distance measure that is also known as the Euclidean norm or the Euclidean distance when its square root is calculated. It sums up the squared difference between feature vectors. The *L2* norm of feature vector A and feature vector B is:

$$D_{L2}(A, B) = \sum_{i=1}^{n} (A_i - B_i)^2 \tag{14}$$

where n is the number of elements in vector A and vector B.

C. *Covariance:* Covariance is also known as the angle measure. It calculates the angle between two normalized vectors. Taking the dot product of the normalized vectors performs the calculation. The covariance between feature vectors A and B is:

$$D_{Cov}(A, B) = \frac{A}{\|A\|} \times \frac{B}{\|B\|} \tag{15}$$

Covariance is a similarity measure. By negating the covariance value, it becomes a distance measure.

D. Mahalanobis distance: The Mahalanobis distance calculates the product of the vector elements and the eigenvalue of a specific dimension and sums all these products. The Mahalanobis distance between feature vectors A and B is:

$$D_{Md}(A, B) = \sum_{i=1}^{n} A_i B_i C_i \tag{16}$$

where

$$C_i = \frac{1}{\sqrt{\lambda_i}} \tag{17}$$

and n is the number of elements in vectors A, B and C. Mahalanobis distance is a distance measure.

In this work, once the similarity measure is determined, classification is achieved by using Nearest Neighbor Classifiers [9]. Let $M_k, k = 1, 2, ..., L$, be the training samples for class w_k after the PCA or ICA transformation. The nearest neighbor rule for classification uses a similarity (distance) measure δ. The image feature vector is classified as belonging to the class of the closest M_k using the similarity (distance) measure δ.

IV. EXPERIMENTAL RESULTS

To perform experiments, we use a face database came from the Olivetti Research Laboratory in Cambridge, UK. The Olivetti Research Laboratory (ORL) face database is a small database of 40 subjects with 10 images each showing some pose, lighting and expression variation. The 10 images of each subject are quite similar to each other, only allowing some changes to hairstyle, background and the wearing of glasses. Therefore, the ORL is the most useful standard database for our purposes. The experiments involve 400 face images corresponding to 40 subjects such that each subject has ten images of size 112×92 with 256 gray scale levels. Some of the face images used are shown in Fig. 2.

For different purposes, three experiments are conducted for face recognition in this work. First experiment is conducted to evaluate the recognition performance achieved by PCA and ICA using all eigenvectors and different similarity measures. Second experiment is conducted to take only a portion of eigenvectors. Third one is conducted to test the recognition robustness of PCA and ICA for noise immunity. The details of three experiments are described below.

4.1. First experiment

The goal of the first experiment is to evaluate the recognition performance resulted by 4 different similarity and distance measures described in Eqs. (13)-(16). They are *L1* norm, *L2* norm, Covariance (*Cov*) and Mahalanobis distance (Md), respectively. In the first experiment, performing PCA and ICA in separate, those 4 similarity and distance measures are adopted and achieved results are used to compare their performance in face recognition. As such, we can identify the most adequate measure resulting the best performance in PCA or ICA. Nearest neighbor rule is adopted to measure the nearness between two samples.

For the eigenvector selection, we select the

Fig. 2. Some example images from the Olivetti database.

first 40 eigenvectors that accumulate most of the energy. To each similarity or distance measure, nine cases are tested. In the first case, 40 images of one from each subject are used as training samples; the remaining 360 images are used as testing images. In the remaining cases, training samples from each subject are increased from 2 to 9 and the testing samples decreasing from 320 to 40 images.

Using PCA and the training (testing) conditions mentioned above, Fig. 3 shows the recognition performance achieved by using *L1*, *L2*, *Cov* and Md, respectively. As shown in Fig. 3, although the performance achieved by Mahalanobis distance measure degraded when the number of training or testing images from each subject is 8 or 2, however, in average, the Mahalanobis distance measure achieves the best performance when using PCA. The anomaly condition is resulted by the insufficient number of testing images from each subject. Even only one misclassification will reduce the recognition performance to a noticeable level.

Using ICA and the same training (testing) conditions as using PCA, Fig. 4 shows the recognition performance achieved by using similarity and distance measures described in Eqs. (13)-(16). In Fig. 4, although the performance achieved by the similarity measure of Covariance starts to degrade when the number of training (testing) images from each subject increases (decreases) to 8 (2), however, in average, the similarity measure of Covariance achieves the best performance when using ICA.

To compare PCA with ICA, we use the best performance achieved in Fig. 3 and Fig. 4. As shown in Fig. 5, when the number of training (testing) images from each subject increases (decreases) from 3 (7) to 8 (2), ICA has a better performance than PCA when ICA uses the similarity measure of Covariance and PCA adopts the Mahalanobis distance measure. When the number of training or testing images from each subject reaches to 9 or 1, ICA and PCA accomplish the same recognition performance.

Therefore, using the first 40 eigenvectors as the transformation basis and different number of training images in the first experiment, PCA achieves its best performance when using the Mahalanobis distance measure and ICA uses the similarity measure of Covariance to reach its optimal case. Furthermore, in average, ICA accomplishes better performance than PCA when using their individual best cases.

4.2. Second Experiment

To evaluate the performance of face recognition by different number of eigenvectors, the second experiment is conducted. The similarity or distance measures used are the same as in the first experiment. PCA and ICA are performed separately and different numbers of eigenvectors are used to compare their recognition performance. Nearest neighbor rule is adopted to measure the distance between two samples. In each case, every five of ten images from each subject are used as the training samples (200 samples), the other five images as testing images (200 samples). Using PCA and the training (testing) conditions mentioned above, Fig. 6 has shown the recognition results achieved by using *L1*, *L2*, *Cov* and Md, respectively.

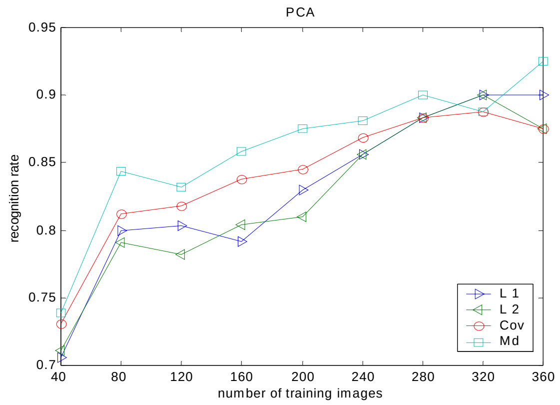

Fig. 3. Face recognition results resulted by 4 different similarity measures when using PCA in the first experiment.

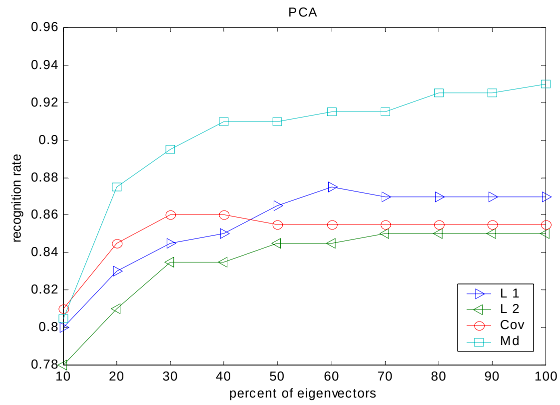

Fig. 6. Face recognition results achieved by 4 different similarity measures and different number of eigenvectors when using PCA in the second experiment.

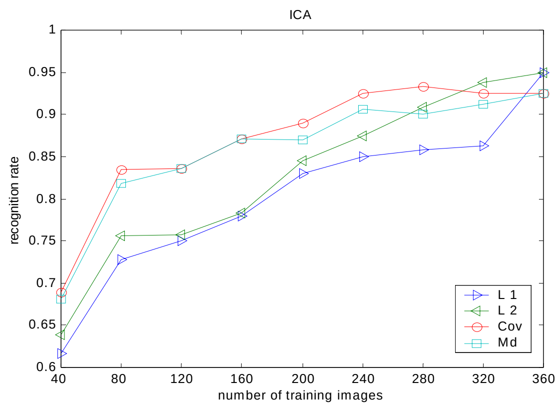

Fig. 4. Face recognition results resulted by 4 different similarity measures when using ICA in the first experiment.

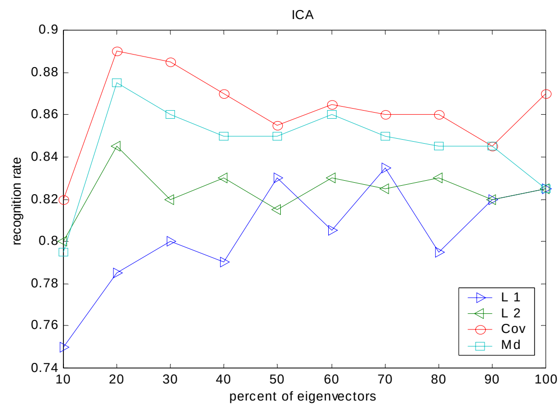

Fig. 7. Face recognition results achieved by 4 different similarity measures and different number of eigenvectors when using ICA in the second experiment.

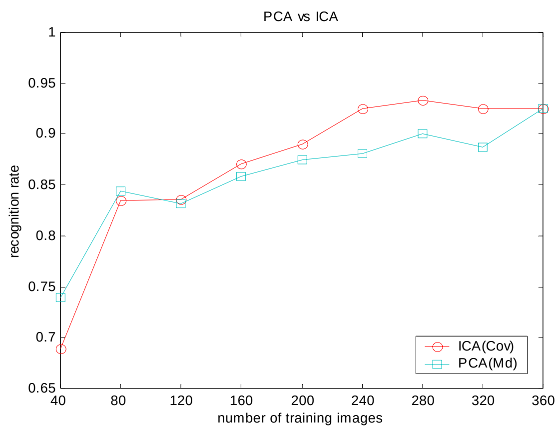

Fig. 5. PCA vs. ICA when using the Mahalanobis and Covariance measures for PCA and ICA respectively.

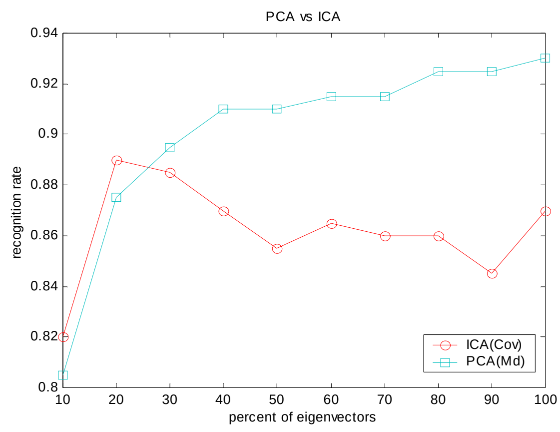

Fig. 8. PCA vs. ICA when using the Mahalanobis and Covariance measures for PCA and ICA in the second experiment.

In Fig. 6, when 100% of eigenvectors are used, the best performance is achieved by using Mahalanobis distance measure for PCA. When the eigenvectors are adopted from 10% up to 40%, the Covariance similarity measure accomplishes better results than *L1* and *L2* measures. However, when the eigenvectors are increased over 50%, the Covariance similarity measure reduces its performance and achieves worse results than *L1* measure. In the other hand, for *L1* measure, when the number of eigenvectors is increased over 70%, the recognition performance is reduced. Therefore, we can conclude that the performance is not promised to be better by increasing the number of eigenvectors for certain measures.

Using ICA and the same training (testing) conditions as for PCA, Fig. 7 shows the recognition results achieved by using *L1*, *L2*, *Cov* and Md, respectively.

In Fig. 7, when 100% of eigenvectors are used, the best performance is achieved by using Covariance measure for ICA. Using *L2*, Covariance and Mahalanobis measures, Fig. 7 shows that the best recognition performance of those measures can be achieved by adopting only small number (20%) of eigenvectors when using ICA. However, their results degrade when the number of eigenvectors is increased from 30% to 100%.

The comparison of PCA (using Mahalanobis distance measure) and ICA (using the Covariance similarity measure) is shown in Fig. 8. PCA achieves better performance than ICA when the number of eigenvectors is increased from 30% to 100%.

Therefore, for the second experiment, the summary is that Mahalanobis distance measure and Covariance similarity measure are the most suitable measures for PCA and ICA when using different number of eigenvectors. Furthermore, the best recognition performance achieved by PCA when using Mahalanobis distance measure is better than ICA when using Covariance similarity measure.

4.3. Third Experiment

The goal of the third experiment is to evaluate the noise immunity power of PCA and ICA for face recognition by adding Gaussian noise into testing images. Again, Covariance (*Cov*) similarity measure, *L1* norm, *L2* norm and Mahalanobis distance (Md) measures in Eqs. (13)-(16) are used in this experiment.

For each similarity or distance measure, 400 images from 40 subjects are used as training samples; all training images degraded by Gaussian noise are used as testing images.

For the eigenvector selection, we select the first 40 eigenvectors that accumulate most of the energy. Each testing image is tested by PCA and ICA when Gaussian noise is added by mean=0.0 and variance=0.01, 0.05, 0.1, 0.5 and 1, respectively. That is, there are 400 testing images for each case of different variance value. The image examples contaminated by different degree of Gaussian noises are shown in Fig. 9.

Using PCA and the training (testing) conditions mentioned above, Fig. 10 shows the recognition performance achieved by adding different degree of Gaussian noises to training images.

As shown in Fig. 10, for each similarity or distance measure, the recognition performance is degraded when the variance of Gaussian noise is increased. Using PCA, for the four different measures, *L1* norm has the most robust recognition power in noise immunity and *L2* norm accomplishes the worst performance.

Using PCA and the training (testing) conditions mentioned above, Fig. 10 shows the recognition performance achieved by adding different degree of Gaussian noises to training images. As shown in Fig. 10, for each similarity or distance measure, the recognition performance is degraded when the variance of Gaussian noise is increased. Using PCA, for the four different measures, *L1* norm has the most robust recognition power in noise immunity and *L2* norm accomplishes the worst performance.

Using ICA and the training (testing) conditions as in PCA, Fig. 11 shows the recognition performance achieved by using four different measures and adding different degree of Gaussian noise to training images.

Same as the results shown in Fig. 10 when using PCA, the recognition performance achieved by ICA is degraded when the variance of Gaussian noise is increased. As shown in Fig. 11, for the four different similarity measures, in average, Covariance measure achieves the most robust recognition power in noise immunity when using ICA under five noise conditions. The performance achieved by using Covariance measure is only degraded to the second when the noise variance value equals 0.05.

By using *L1* norm for PCA and Covariance measure for ICA, the performance comparison of PCA and ICA under different noise conditions is shown in Fig. 12. Fig. 12 shows that ICA has accomplished better noise immunity power than PCA.

Therefore, for the third experiment, the summary is that the recognition performance is degraded in each similarity or distance measure when the variance of Gaussian noise is increased. PCA achieves its best recognition performance in

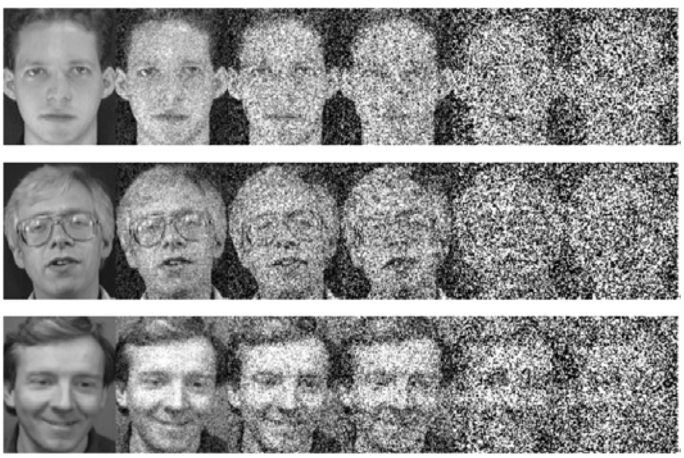

Fig. 9. Testing images of three persons in three rows, from left to right, when Gaussian noises are added by mean=0.0,variance=0, 0.01, 0.05, 0.1, 0.5 and 1, respectively.

noise immunity when the distance measure of *L1* norm is used. However, the best performance of ICA is accomplished by using Covariance similarity measure. ICA has better noise immunity power than PCA.

V. CONCLUSIONS

PCA is computed from the global covariance matrix of the full set of image data, the obtained basis vectors are global representations that are not suitable for the recognition of non-aligned faces. Relatively, ICA basis vectors are more spatially local than PCA basis vectors and give better face representation. In the experimental results, the recognition performance of PCA and ICA are compared and analyzed when the test images are contaminated by noises. Also, different methods for eigenvector selection and similarity measures are evaluated.

Using part of eigenvectors that accumulate most of energy as the transformation basis and different number of training images in the first experiment, ICA accomplishes better performance than PCA in each individual case. However, when using 50% of face dataset as the training samples and the other 50% as testing images, Mahalanobis distance measure and Covariance similarity measure are the most suitable measures for PCA and ICA when different number of eigenvectors are adopted as the transformation basis. Furthermore, the best recognition performance achieved by PCA when using Mahalanobis distance measure is better than ICA when using Covariance similarity measure. When testing images are contaminated by Gaussian noises, the recognition performance is degraded in each similarity or distance measure. PCA achieves its best recognition performance in noise immunity when the distance measure of *L1* norm is used. However, the best performance of ICA is accomplished by using Covariance similarity measure. By selecting the best results achieved by PCA and ICA, ICA accomplishes better noise immunity power than PCA.

In this work, we have obtained promising results in face recognition when using ICA in the Olivetti dataset contaminated by Gaussian noises. The characteristics of ICA with spatially local basis vectors not only give better face representation than PCA, but also make ICA perform better performance for face recognition in noisy conditions.

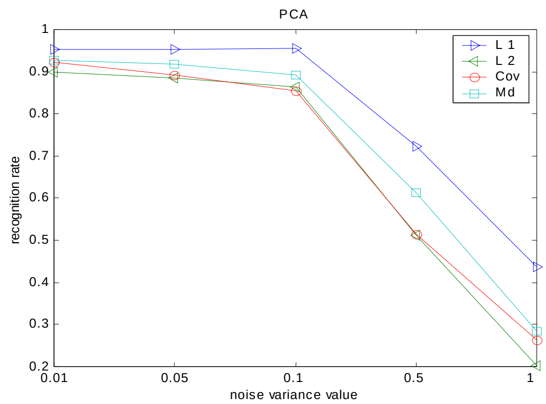

Fig. 10. Face recognition results by using PCA and different Gaussian noises.

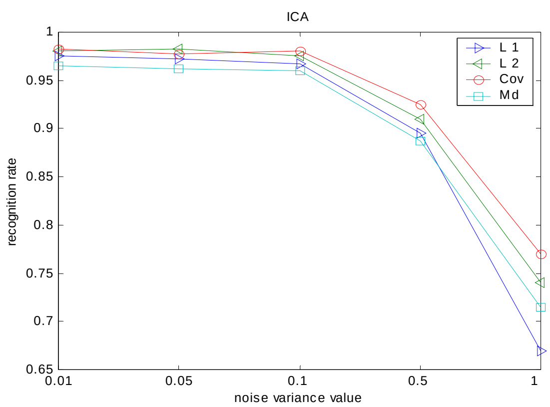

Fig. 11. Face recognition results by using ICA and different Gaussian noises.

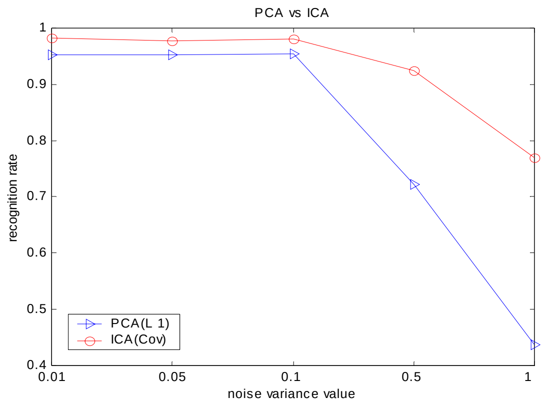

Fig. 12. The performance comparison of PCA (using *L1* measure) and ICA (using *Cov* measure) in noise immunity.

Acknowledgment

This project was support by the National Science Council, Taiwan, R.O.C., under Grant NSC 90-2213-E-014-006.

References

[1] Chellappa, R., Wilson, C. L., and Sirohey, S., "Human and machine recognition of faces: a survey," Proceedings of the IEEE, Vol. 83, No. 5, pp. 705-740, 1995.

[2] Turk, M. and Pentland, A., "Eigenfaces for recognition," Journal of Cognitive Neuroscience, Vol. 13, No. 1, pp. 71–86, 1991.

[3] Bartlett, M. and Sejnowski, T., "Independent components of face images: A representation for face recognition," in Proceedings of the 4th Annual Joint Symposium on Neural Computation, pp 3-10, 1997.

[4] Liu, C. and Wechsler, H., "Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition," in Proceedings of International Conference on Audio and Video Based Biometric Person Authentication, 1999.

[5] Yuen, P. C. and Lai, J. H., "Face representation using independent component analysis," Pattern Recognition, Vol. 35, No. 6, pp. 1247-1257, 2002.

[6] Baek, K., Draper, B. A., Beveridge, J. R., and She, K., "PCA vs. ICA: A comparison on the FERET data set," in Proceedings of Joint Conference on Information Sciences (JCIS '02), pp. 824-827, 2002.

[7] Karhunen, J., Oja, E., Wang, L., Vigario, R., and Joutsensalo, J., "A class of neural networks for independent component analysis," IEEE Transactions on Neural Networks, Vol. 8, No. 3, pp. 486– 504, 1997.

[8] Comon, P., "Independent Component Analysis – A New Concept?," Signal Processing, Vol. 36, No. 3, pp. 287-314, 1994.

[9] Belhumeur, P., Hespanha, J., and Kriegman, D., "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 7, pp. 711–720, 1997.

[10] Hoyer, P. O., and Hyvärinen, A., "Independent Component Analysis Applied to Feature Extraction from Colour and Stereo Images," Network: Computation in Neural Systems, Vol. 11, No. 3, pp. 191-210, 2000.

[11] Bell, A. J., and Sejnowski, T. J., "The 'Independent Components' of Natural Scenes are Edge Filters," Vision Research, Vol. 37, No. 23, pp. 3327-3338, 1997.

[12] Hyvärinen, A., "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis," IEEE Transactions on Neural Networks, Vol 10, No. 3, pp. 626-634, 1999.

[13] Yambor, W. S., "Analysis of PCA-Based and Fisher Discriminant-Based Image Recognition Algorithms," Technical report CS-00-013, Computer Science Department, Colorado State University, Fort Collins, July 2000.