# Face Recognition Based on ICA Combined with FLD

Juneho Yi, Jongsun Kim, Jongmoo Choi, Junghyun Han, Eunseok Lee

School of Electrical and Computer Engineering
Sungkyunkwan University
300, Chunchun-dong, Jangan-gu
Suwon 440-746, Korea
{jhyi, jskim, jmchoi}@ece.skku.ac.kr, eslee@seopo.skku.ac.kr,
han@graphics.skku.ac.kr

**Abstract.** Recently in face recognition, as opposed to our expectation, the performance of an ICA (Independent Component Analysis) method combined with LDA (Linear Discriminant Analysis) was reported as lower than an ICA only based method. This research points out that (ICA+LDA) methods have not got a fair comparison for evaluating its recognition performance. In order to incorporate class specific information into ICA, we have employed FLD (Fisher Linear Discriminant) and have proposed our (ICA+FLD) method. In the experimental results, we report that our (ICA+FLD) method has better performance than ICA only based methods as well as other representative methods such as Eigenface and Fisherface methods.

## 1. Introduction

Face recognition in real environments is a very difficult problem due to variations of facial expression, pose and illuminations. For face classification, employing features that best explain facial data is one of the most important tasks. This paper presents that (ICA+LDA) methods have better recognition performance than ICA only based methods as well as other representative face recognition methods such as Eigenfaces and Fisherfaces.

ICA is a technique that extracts statistically independent signals from mixed signals. An ICA based face recognition method assumes that a facial image can be represented by a linear combination of statistically independent sources. It is well known that ICA better represents a variety of data distributions than PCA [10]. Thus, ICA techniques have popularly been applied lately to the problem of face recognition [1][3][4][6], especially for face recognition under variations of illumination, pose and facial expression. However, ICA based methods do not consider class information. Hence an ICA only based method is appropriate for data representation, and is not tuned for classification of data.

Recently, Bartlett et al. who coined the ICA based face recognition method suggested that an (ICA+LDA) method might have better performance than ICA only based

M. Tistarelli, J. Bigun, A.K. Jain (Eds.): Biometric Authentication, LNCS 2359, pp. 10-18, 2002.

methods [2]. However, they did not provide any supporting experimental results. Among LDA techniques, FLD (Fisher Linear Discriminant) has often been used for face recognition. FLD is a classical statistical analysis method that finds a linear transformation matrix, which maximizes the ratio of between class scatter and within class scatter. Contrary to their expectation, C. Liu et al. reported the experimental result that their (ICA+FLD) method had lower performance than ICA only based methods [5]. In their experiment, however, the (ICA+FLD) methods did not get a fair comparison with ICA only based methods in that they used feature vectors of dimension less than $(C-1)$ although the proper use of FLD has to employ $(C-1)$ dimensional vectors. C denotes the number of classes.

In this paper, we propose our (ICA+FLD) method and report that the performance of face recognition for our (ICA+FLD) method is better than that for ICA only based methods. We present experimental results that compare in a fair manner the recognition performance of the (ICA+FLD) method with Eigenface, Fisherface and ICA only based methods. As can be seen in the experimental results, the (ICA+FLD) method performs better than the other methods.

The rest of this paper is organized as follows. Section 2 briefly reviews the Eigenface, Fisherface and ICA based methods that are most representative methods for face recognition. We describe the ICA method combined with FLD in section 3 and present experimental results in section 4.

## 2. Related Work

### 2.1 Eigenfaces and Fisherfaces

The Eigenface method is briefly described as follows. Let a face image be a two-dimensional M by N array of intensity values. This image can be represented a column vector $\overrightarrow{X_i}$ of dimension $M \times N$. Let $X = [\vec{X}_1, \vec{X}_2, ..., \vec{X}_T]$ be a sample set of the face images. $T$ denotes the total number of the face images. After subtracting the total mean from each face image, we get a new vector set $X' = [\overrightarrow{X_1'}, \overrightarrow{X_2'}, ..., \overrightarrow{X_T'}]$. Then the covariance matrix is defined as

$$S = \sum_{i=1}^{T} \overrightarrow{X_i'} \overrightarrow{X_i'}^T \\ = X'X'^T. \tag{1}$$

The respective eigenvector and eigenvalue matrices, $\Psi$ and $\Lambda$, are computed from Eq. (2).

$$S\Psi = \Psi\Lambda. \tag{2}$$

The size of the matrix, $S$ is MN MN and determining the MN eigenvectors and eigenvalues is an intractable task for typical image sizes. A computationally feasible method that employs the eigen-analysis of $X'^T X'$ instead of $X'X'^T$ is used. The size of $X'^T X'$ is T T [9]. Let $V$ and $\Lambda'$ the eigenvector and eigenvalue matrices of $X'^T X'$, respectively. Then we have the following equation.

$$(X'^T X')V = V\Lambda' \tag{3}$$

Pre-multiplying $X'$ on both sides, we have

$$X'(X'^T X')V = (X'X'^T)(X'V) = (X'V)\Lambda' \tag{4}$$

$V = [\vec{V}_1, \vec{V}_2, ..., \vec{V}_T]$ and $\Lambda' = diag(\lambda_1, \lambda_2, ..., \lambda_T)$. Assuming $\lambda_i$'s are sorted as $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_T$, we obtain eigenvectors of $X'X'^T$ corresponding to the first $m$ largest eigenvalues, and the projection matrix $W_{PCA}$ is obtained as follows:

$$W_{PCA} = [X'\overrightarrow{V_1}, X'\overrightarrow{V_2}, ..., X'\overrightarrow{V_m}]. \tag{5}$$

$X'\overrightarrow{V_1}, X'\overrightarrow{V_2}, ..., X'\overrightarrow{V_m}$ represent the first $m$ eigenfaces. A vector $\overrightarrow{X_i'}$ that represents a face image is projected to a vector $\vec{Y}_i$ using the projection matrix $W_{PCA}$ as in the following equation.

$$\vec{Y}_i = W_{PCA}^T \overrightarrow{X_i'} \tag{6}$$

At online recognition, a new face image $\vec{X}_i$ is subtracted by the mean face and projected to $\vec{Y}_i$ using Eq. (6). It is recognized by comparison of $\vec{Y}_i$ with the projected vectors of the training face images that are computed off-line. Since PCA maximizes for all the scatter, it is more appropriate for signal representation rather than for recognition purpose.

The idea of the Fisherfaces method is that one can perform dimensionality reduction using PCA and still preserve class separablility. It applies FLD to the reduced PCA subspace to achieve more reliability for classification purpose. The Fisherfaces method is briefly described as follows. Let $C$ and $N_1, N_2, ..., N_c$ denote the number of classes and the number of face images in each class, respectively. Let $\overrightarrow{M_1}, \overrightarrow{M_2}, ..., \overrightarrow{M_c}$ and $\vec{M}$ be the means of each class and the total mean in the reduced PCA subspace. Since $\overrightarrow{Y_{ij}} = W_{PCA}^T \overrightarrow{X_{ij}'}$, we can then have $\overrightarrow{M_i} = \frac{1}{N_i}\sum_{j=1}^{N_i}\overrightarrow{Y_{ij}} = W_{PCA}^T(\frac{1}{N}\sum_{j=1}^{N_i}\overrightarrow{X_{ij}'})$. $\overrightarrow{X_{ij}'}$ denotes the $j^{th}$ face image vector belonging to the $i^{th}$ class (or subject). The between-class scatter and within-class scatter matrices $S_b'$ and $S_w'$ of $\overrightarrow{Y_{ij}}$ are expressed as follows.

$$S'_w = W_{PCA}^T S_w W_{PCA}, \quad S'_b = W_{PCA}^T S_b W_{PCA} \tag{7}$$

$S'_b$ and $S'_w$ are the between-class scatter and within-class scatter matrices of $\overrightarrow{X'_{ij}}$ 's, respectively. The projection matrix that maximizes the ratio of the determinant $\frac{\left|W_{fld}^T S'_b W_{fld}\right|}{\left|W_{fld}^T S'_w W_{fld}\right|}$ is chosen as the optimal projection, $W_{fld}$. The rank of $\left(S'_w\right)^{-1} S'_b$ is $(C-1)$. Thus the columns of $W_{fld}$ are computed as the $(C-1)$ leading eigenvectors of the matrix $\left(S'_w\right)^{-1} S'_b$ [11]. Therefore, given an input face image $\overrightarrow{X'_{ij}}$, it is projected to $\overrightarrow{\Omega'_{ij}} = W_{fld}^T W_{PCA}^T \overrightarrow{X'_{ij}}$ and classified by comparison with the vectors $\overrightarrow{\Omega'_{ij}}$ 's that were computed off-line from a set of training face images [13].

### 2.2 ICA Based Face Recognition

Face recognition using ICA is briefly described as follows. We observe L linear mixtures $x_1, x_2, ..., x_L$, which are assumed to be linear combinations of N unknown zero mean mutually statistically independent components $s_1, s_2, ..., s_N$. The above mixing model is written as:

$$x = As. \tag{8}$$

$A$ is an unknown mixing matrix. The task is to estimate the independent component $s$ by computing the matrix $W$ as in the following equation that corresponds to the mixing matrix $A$.

$$U = Wx \text{ and } x = W^{-1}U \tag{9}$$

On the other hand, a face image $\vec{X}_i$ is represented by a linear combination of the independent components $s_1, s_2, ..., s_n$ as in Eq. (10) whose coefficients $a_{i1}, a_{i2}, ..., a_{in}$ are the column vectors of the matrix $A$ [7].

$$\vec{X}_i = a_{i1}s_1 + a_{i2}s_2 + ... + a_{in}s_n \tag{10}$$

Let $X'$ be the set of column vectors where each vector is obtained by subtracting the total mean from $\vec{X}_i$. The mean value of the column vectors of $X'$ becomes 0. The matrix $W$ is a de-correlating matrix when the covariance matrix of the output vector $U$ becomes a diagonal matrix [12]. That is, in the case of Eq. (9), when $UU^T = I$, then $W$ can be written as:

$$W^T W = (X'X'^T)^{-1}. \tag{11}$$

Then, the independent components $U$ become uncorrelated by the whitening process. For each face image in the training set, we can calculate its feature vector $U$ and a weight matrix $W$. Using Eq. (11), we get the following equation,

$$EDE^{-1} = (X'X'^{T}) \tag{12}$$

$E$ and $D$ are the eigenvector and eigenvalue matrices of the covariance matrix $X'X'^{T}$, respectively. Substituting Eq. (12) into Eq. (11), we get the whitening matrix $W_p$ as follows:

$$W_p = D^{-\frac{1}{2}} E^{T} . \tag{13}$$

We transform $X'$ to $V$ using the matrix $W_p$ (M×M) as follows.

$$V = W_p X' \tag{14}$$

Thus, substituting Eq. (14) into Eq. (15) gives $V$ as follows:

$$V = D'^{-\frac{1}{2}} E'^{T} X' . \tag{15}$$

Therefore, the independent components $U$ is obtained as:

$$U = W^{T} V . \tag{16}$$

Kurtosis could be used to estimate mutually statistically independent components from observed images. The kurtosis measures non-Gaussianity of data distribution [7]. The kurtosis of $U$ is classically computed as follows:

$$kurt(U) = \sum_i \left| E\{U_i^{\,4}\} - 3(E\{U_i^{\,2}\})^2 \right| . \tag{17}$$

A linear transformation matrix $W^{T}$ could be found by the kurtosis maximization. This ICA algorithm using the kurtosis method of Eq. (17) estimates the independent components one at a time [10], while Bell and Sejnowski's algorithm simultaneously estimates all the components [8].

## 3. (ICA+FLD) Method

Our (ICA+LDA) method that is evaluated is shown in Fig. 1. In off line training, the method is divided into two stages. First, we get statistically independent basis vectors by applying ICA to all training patterns. Then, in order to incorporate class specific information into ICA, we apply FLD to the subspace that the basis vectors span.

Let $W_I$ and $W_{fld}$ be the linear transformation matrices that are computed from ICA and FLD, respectively. Then, given an input face image $\vec{X}_{ij}$, it is projected to $\vec{\Omega}_{ij} = W_{fld}^T W_I^T \vec{X}_{ij}$ and classified by comparison with the vectors $\vec{\Omega}_{ij}$'s that were computed off-line from a set of training face images. The fundamental difference between C. Liu et al.'s and our experiments is that we use the full $(C-1)$ dimensional feature vectors for classification while they employ feature vectors of reduced dimension.

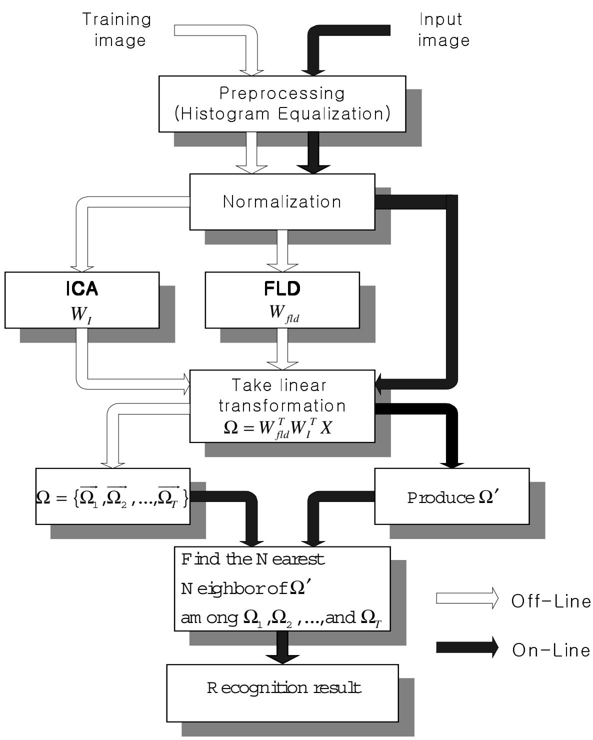


**Fig. 1:** System overview

## 4. Experiments

We have used several facial image databases such as SKKU [15], Yale [16] and ORL [17] in order to compare the recognition performance of Eigenfaces [9], Fisherfaces [14], ICA, and (ICA+FLD) methods. SKKU database contains ten different images each for ten different subjects. The size of the image is 50×40. For a subject, five images out of ten images were taken first and the rest five images at a different time. All the images are frontal views of upright faces with changes in illumination, facial expression (open/closed eyes, smiling/non-smiling/surprised), facial details (glasses/no glasses) and hair-style. Fig. 2 shows a part of images of a subject in SKKU facial database. In Yale facial database, each of sixteen different subjects has ten images, which consist of three images under illumination changes, six with changes in facial expression and one with glasses worn. The ORL (Olivetti Research Laboratory) face image database are 10 different images of 40 distinct subjects [17], All the images were taken against a dark homogeneous background and the subjects are in up-right, frontal position with tolerance for some side movement. Fig. 3 shows a set of images of a subject in Yale and ORL facial database.

**Fig. 2:** Example images from SKKU facial database.

**Fig. 3:** Left: Example images from Yale facial database. Right: Example images from ORL facial database.

We have compared the recognition rate of our (ICA+FLD) method with that of Eigenface [9], Fisherfaces [14] and ICA based face recognition [7] using SKKU, Yale and ORL facial databases. In order to determine the recognition rates, the half of the entire database is used for training and the other half for testing. The classification was performed using the nearest neighbor classifier.

Table 1 shows the relative performance of the algorithms when applied to SKKU, Yale and ORL facial databases, respectively. As can be seen in Table 1, the performance of the (ICA+FLD) method is better than that of Eigenfaces [9], Fisherfaces [14] and ICA based face recognition [7]. Interestingly, however, the performance of Fisherfaces turns out to be better than ICA only based method. We think that why the performance of ICA is lower than that of Fisherfaces is that the Fisherfaces method uses class information while ICA does not. This experimental results proves our claim that the (ICA+FLD) method is more appropriate for classification than the other methods. The application of FLD yields the better discriminating power for pattern classification, and the recognition accuracy has been improved.

**Table 1:** Recognition rates of Eigenfaces, Fisherfaces, ICA and our (ICA+FLD) method for SKKU, Yale and ORL facial images.

| | | Eigenfaces | Fisherfaces | ICA | (ICA+ FLD) |
|---|---|---|---|---|---|
| correct retrieval rate | SKKU | 0.86 | 0.94 | 0.90 | 0.94 |
| | Yale | 0.829 | 0.988 | 0.829 | 1.000 |
| | ORL | 0.925 | 0.950 | 0.925 | 0.965 |

## 5. Conclusions

In this paper, we have achieved better classification performance by combining ICA with FLD method. Although the ICA method shows a good performance under variations of illumination, pose, and facial expression, it is not tuned for classification of data in that it does not consider class information. In order to incorporate class specific information into ICA, we have employed FLD after ICA has been applied. The experimental results using several facial databases have shown that the (ICA+FLD) method has better performance than Eigenface, Fisherface, and ICA only based methods.

**Acknowledgements**
This work was supported by grant number R01-1999-00339 from the Basic Research program of the Korea Science and Engineering Foundation.

## References

1. Bartlett, M. Stewart, Lades, H. Martin and Sejnowski, T.J., Independent component representations for face recognition, Proceedings of the SPIE, Vol. 3299: Conference on Human Vision and Electronic Imaging III, pp. 528-539, 1998.
2. Bartlett, M. Stewart and Sejnowski, T. J., Independent components of face images: A representation for face recognition, Proceedings of the 4th Annual Joint Symposium on Neural Computation, Pasadena, CA, May 17, 1997.
3. Bartlett, M. Stewart and Sejnowski, T.J., Viewpoint invariant face recognition using independent component analysis and attractor networks, In M. Mozer, M. Jordan, T. Petsche (Eds.), Advances in Neural Information Processing Systems 9, MIT Press, Cambridge, MA. pp. 817-823, 1997.
4. Guan, A.X. and Szu, H.H.A, A local face statistics recognition methodology beyond ICA and/or PCA, Neural Networks, IJCNN '99. International Joint Conference on , Vol. 2 1999.
5. C.Liu and H.Wechsler, Comparative assessment of independents component analysis(ICA) for face recognition, Second Int. Conf. Audio and Video-based Biometric person Authentication, pp.22-24, March 1999.

6. O. Déniz, M. Castrillón and M. Hernández, Face recognition using independent component analysis and support vector machines, IX Spanish Symposium on Pattern Recognition and Image Analysis, Castellón (Spain), May 2001.
7. Aapo Hyvarinen and Erki Oja, Independent component analysis: a tutorial, http://www.cis.hut.fi/ ~aapo/papers/IJCNN99_tutorialweb/1999.
8. A. Bell and Sejnowski, An information maximization approach to blind separation and blind deconvolution, Neural Computation 7, pp. 1129-1159, 1995.
9. M. A. Turk and A.P Pentland, Eigenfaces for recognition, Cognitive Neuroscience, vol.3, no.1, pp. 71-86, 1991, ftp://whitechapel.media.mit.edu/pub/ images/.
10. Oja, E., Karhunen, J.,Wang, L. and Vigario, R. Principal and independent components in neural networks-recent developments, Proc. Italian Workshop on Neural Networks, May 1995.
11. K. Fukunaga, Introduction to statistical pattern recognition, Academic Press, second edition, 1991.
12. A. Bell and Sejnowski, The independent components of natural scenes are edge filters, Advance in Neural Information Processing Systems 9, 1997.
13. J. Yi, H. Yang and Y. Kim, Enhanced fisherfaces for robust face recognition, Lecture Notes in Computer Science, Vol, 1811, pp. 502-511, May 2000.
14. P. Belhumeur, J. Hespanha, and D. Kriegman, Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, IEEE Trans. on PAMI, Vol. 19, no. 7, pp. 711-720, 1997.
15. http://vision.skku.ac.kr/research/skkuface.html.
16. http://cvc.yale.edu/projects/yalefaces/yalefaces.html
17. ftp://ftp.uk.research.att.com:pub/data/att_faces.zip