# A Comparative Study of Linear Subspace Analysis Methods for Face Recognition

Wei Ge, Lijuan Cai, Chunling Han

School of Electronics and Information Engineering Changchun University of Science and Technology, Changchun, 130000, China

**Abstract.** Face recognition is a typical problem of pattern recognition and machine learning. Among these approaches to the problem of face recognition, subspace analysis gives the most promising results, and becomes one of the most popular methods. This paper researches typical subspace analysis approaches, based on the introduction of main approaches of linear subspace analysis, such as Principal Component Analysis (PCA) , Linear Discriminant Analysis(LDA) and Fast Independent Component Analysis (FastICA), the application of these approaches for face recognition by ORL database and YALE B database are investigated, and the advantages and disadvantages are compared. Experimental results show that the LDA approach leads to better classification performance than PCA approach, while the FastICA approach leads to the best classification performance with the improvement of nearly 3% compared with the LDA approach h.

**Keywords:** Face recognition, Principal Component Analysis, Linear Discriminant Analysis, Fast Independent Component Analysis.

## 1. Introduction

Face recognition is a typical issue of pattern analysis, classification and understanding for image. As one of the key technologies in biometrics, face recognition techniques are believed to have a great deal of potential applications in law enforcement, public security, financial security, and information security. In the past, most such research studies have been conducted using visible images, and a great variety of results are reported. In the past few decades, many face recognition methods have been developed [1-2]. Among the existing face recognition approaches, subspace analysis methods are widely used to reduce the high dimensionality of the raw face images. The first breakthrough production of the subspace techniques is Principal Component Analysis (PCA), it uses the KL transform to produce a most expressive subspace for face representation and recognition. Linear Discriminant Analysis (LDA) is an example of the most discrimination subspaces[3].The fast independent component analysis (Fast ICA) ,a fast converging algorithm based on the maximization of non-Gaussianity and implemented using an approximative Newton optimization method ,has become a standard for the separation of both sub- and super-Gaussian sources[4].These three methods are all linear and typical subspace methods.

In this paper, a comparative study of linear subspace analysis methods is researched by application for face recognition. Three linear subspace analysis methods: PCA, LDA and FastICA are investigated. This paper focuses on which linear subspace analysis method would outperform the others in the face recognition application. Results show that the best classification rate is obtained by FastICA approach and the descending order of classification rate by other methods is LDA, PCA.

## 2. Review of linear subspace analysis method

### 2.1. Principal Components Analysis Method

PCA tends to find such a subspace whose basis vectors correspond to the maximum variance direction in the original image space. New basis vector define a subspace of face images called face space. All training images are projected onto the subspace to find sets of weights, which describe the contribution of each vector [8]. Specifically, the PCA basis vectors are defined as eigenvectors $\mu_i$ of scatter matrix C:

$$C = \sum_{i=1}^{R} (x_i - m)(x_i - m)^T , \tag{1}$$

where $m$ is the mean of all images in the training set, $x_i$ is a $N \times N$ vector which represents each image and R is the number of faces in the training se. For identifying an unknown person, only a smaller number of eigenvectors $R_k$ corresponding to the largest eigenvalues is needed. Given its image x, we subtract the mean ( $x - m$ ) and compute the projection:

$$\tilde{\Phi} = \sum_{i=1}^{R_k} w_i \mu_i , \tag{2}$$

Where $w_i = \mu_i^T x$ is the weight vector of the projection $W_{pca}$. Then the weights are compared to the sets of weights of training images [9].

Fig.1 shows a few eigenvectors of ORL database images by using PCA method [5]. Since these eigenvectors look like some ghostly faces, they are conveniently named eigenfaces.

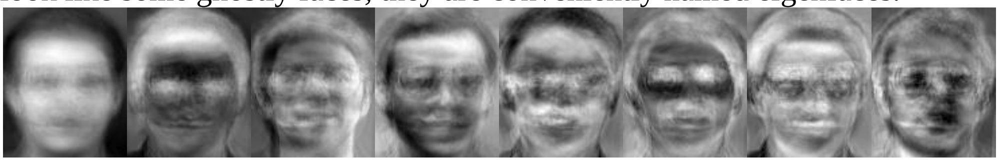

Fig.1 Five eigenvectors of PCA algorithm

**2.2 Linear Discriminant Method**

While PCA is oriented towards representing the data in their entirety, without paying any attention for the underlying structure, LDA finds the vectors in the underlying space that best discriminate among classes. LDA method tries to maximize the between-class differences and minimize the within-class ones.

The between-class and within-class difference are represented by the corresponding scatter matrices $S_b$ and $S_w$.Considering $X_c$ , $c = 1,....N_c$ as subsets of X containing $N_i$ images of the same subject:

$$S_w = \sum_{i=1}^{c} \sum_{x_k \in X_c} (x_k - m_i)(x_k - m_i)^T ,$$

$$S_b = \sum_{i=1}^{c} N_i (m_i - m)(m_i - m)^T \tag{3}$$

Where $m_i = \dfrac{1}{N_i} \sum_{k=1}^{N_i} x_k$ is the mean vector of samples belonging to class i and m is the mean of all images .If $S_w$ is not singular, the goal is to find a projection $W_{opt} = (w_1, w_2, ..., w_l)$ that satisfies the Fisher criterion

$$W_{opt} = \arg\max w \frac{\left|W^T S_b W\right|}{\left|W^T S_w W\right|} , \tag{4}$$

where $w_1, w_2, ..., w_l$ are the eigenvectors of $S_w^{-1} S_b$ corresponding to $l$ $(\le c - 1)$ largest eigenvalues. So LDA is also known as Fisher linear discriminant. Using this algorithm, it must be paid attention that the precondition of this algorithm is that $S_w$ is not singular.

Practically, $S_w$ is usually singular, that is the small sample size problem of LDA. And the inverse of $S_w$ does not exist. So PCA plus LDA method (Fisherface) is adopted usually. This method makes use of PCA to project the image set to a lower dimensional space, so the new within-class scatter matrix $\widehat{S}_w$ is nonsingular, and then applies the standard LDA[8][10]. Spesifically,

$$\widehat{S}_w = W_{pca}^T S_w W_{pca} , \ \widehat{S}_b = W_{pca}^T S_b W_{pca} ,$$

$$W_{fld} = \arg\max_{W} \frac{\left|W^T \widehat{S}_b W\right|}{\left|W^T \widehat{S}_w W\right|} \quad W_{opt}^T = W_{fld}^T W_{pca}^T \tag{5}$$

Some middle results are shown. Fig. 2 gives the pictorial examples of $S_w$ and $S_b$ projected onto PCA space(based on ORL database).

(a) images of $S_w$ projected onto PCA space

(b) images of $S_b$ projected onto PCA space

Fig.2 Images of $S_w$ and $S_b$ projected onto PCA space

**2.3 FastICA Method**

ICA is related to principal component analysis and factor analysis superficially. While PCA minimizes second-order statistics, ICA minimizes both second-order and higher-order dependencies in the input data and attempts to find the basis along which the projected data are statistically independent. The simplest form of ICA is followed:

$$X = AS, \tag{6}$$

We denote the observed variables $x_i$ as a observed vector $X = (x_1, x_2, ..., x_m)^T$ .The observed variables are assumed to be linear combinations of n ( $n \le m$ ) independent components $s_1, s_2, ..., s_n$ , the mean value of $s_i$ is zero and the variance of $s_i$ is 1. A is an unknown $m \times n$ matrix of full rank, called the mixing matrix. S is assumed to be non-gaussian distribution. ICA is simply obtain the sources by using inverse matrix W of A ( $W = A^{-1}$ ). The goal of ICA is to estimate A and $s_i$ by using X[11].

In this paper, FastICA is used to experiment face recognition for its good performance. FastICA method is based on fixed-point iteration scheme. Fixed point method for performing ICA derives from the entropy optimization methods and its speed is about second-order function[4][11].Generally, the mutual information that is a natural measure of the dependence between random variables ,can be defined by negentropy. The definition of negentropy $N_g(y)$ for a random vector Y whose function is f(y) is given as follows:

$$N_g(y) = H(y_{gauss}) - H(y), \tag{7}$$

Where $H(y) = -\int f(y)\log f(y)dy$ , is the entropy of random variable and $y_{gauss}$ is Gaussian random variables of the same covariance matrix as y. The process of the FastICA algorithm in face recognition is shown in[11].

Fig. 3 shows some basis images of Fast ICA method(based on ORL database).

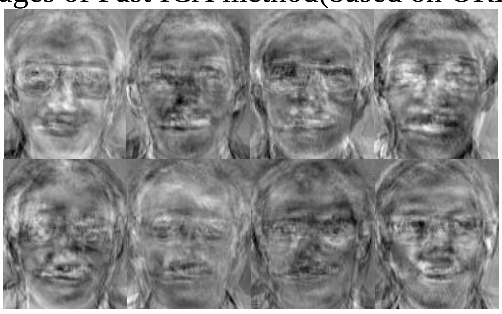

Fig.3 Basis images of FastICA algorithm

## 3. Experiment Results

The recognition algorithm is executed by the nearest algorithm. In this paper Cosine similarity measures are used, which are previous found to be effective for face recognition. For two vectors, the distance measures for x and y are defined as:

$$d_{COS(x,y)} = 1 - \frac{x^T \cdot y}{\|x\| \cdot \|y\|}, \tag{8}$$

Two recognition experiments are conducted based on ORL database and Yale database B[5- 6]. In the ORL database, there are 400 images with c=10 classes(different persons).Each class contains a different number of persons. Several groups are used in the experiment. Train sets are 2 images per person and 5 images per person respectively. 200 images are used to test. The experimental results can be seen in Table 1 .In the Yale database B, there are 5760 images with 10 classes. Two groups are used in the experiment. Train sets are 2 images per person and 5 images per person respectively. 50 images are used to test. The experimental results can be seen in Table 2.

Table 1.Recognition rate based on ORL database

| train umbers / Algorithms | 80 | 200 |
|---|---|---|
| PCA | 86.5% | 90.5% |
| LDA | 84.5% | 92.5% |
| FastICA | 88.5% | 93.5% |

Table 2.Recognition rate based on Yale

| train numbers / Algorithms | 20 | 50 |
|---|---|---|
| PCA | 84% | 88% |
| LDA | 84% | 90% |
| FastICA | 86% | 90% |

According to the results, we can see that recognition rate is gradually improving when the train samples is more. The LDA approach is better than PCA. FastICA always obtained good results. Nearly 3% of the recognition rate is improved by FastICA. Combination with the experiment results, some discussions are as follows:

1. PCA approach is the most typical subspace method, which has the best representation ability but not the best classification ability. Its main drawbacks are that the extracted features are necessarily orthogonal and variant under transformation.

2. LDA approach is good at discriminating different classes. But LDA always suffers from a small sample size problem. The problem will happen when the number of training samples is less than the total number of pixels in an image. It is worse than PCA when the train samples number is not more.

3. ICA is the proper method which is sensitive to high-order statistics. FastICA is the improved method with a lot of merits: it is parallel, distributed, computationally, simple, and requires little space. In our study, FastICA obtained the best recognition performance.

## 4. Conclusion

In this paper, a comparative study of linear subspace analysis methods are researched by application for face recognition. Three linear subspace methods (PCA, LDA and FastICA) are investigated. Our study shows that FastICA obtains the best classification performance (93.5%).LDA method leads to better results than PCA method (when the train samples are more enough).Results also show that FastICA method outperforms the other two linear subspace methods in the task of face recognition.

## References

[1] Rabia Jafri* and Hamid R. Arabnia. A survey of face recognition techniques. Journal of Information Processing Systems, 2(5),41-68(2009).

[2] V.Vijayakumari.Face recognition techniques: a survey. World Journal of Computer Application and Technology, 1(2):41-50(2013).

[3] Xiaogang Wang and Xiaoou Tang. Unified subspace analysis for face recognition.Proceedings of the Ninth IEEE International Conference on Compute Vision(ICCV 2003),2,(2013)

[4] Soroush Javidi, Clive Cheong Took, and Danilo P. Mandic. Fast independent component analysis algorithm for quaternion valued signals. IEEE Transaction on Neural Networks, 12(22), 1967-1978(2011)

[5] http://www.cam-orl.co.uk.

[6] Georghiades, A.S. and Belhumeur, P.N. and Kriegman, D.J. From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose, IEEE Trans. Pattern Anal, Mach. Intelligence, 23(6), 643-660(2001).

[7] Gonzalo Carvajal, Waldo Valenzuela and Miguel Figueroa, Subspace-based face recognition in analog VLSI, In Neural Information Processing Systems 20,(2007).

[8] Borut Batagelj and Franc Solina., Face recognition in different subspaces:a comparative study, PRIS 2006 in conjunction with ICEIS, (2006).

[9] Saurabh Singh, Aglika Gyaourova, George Bebis and Ioannis Pavlidis, Infrared and visible image fusion for face recognition, Proc. SPIE 5404,585-596(2004).

[10] Weishi Zheng, Jianhuang Lai, and Pong C. Yuen. GA-Fisher: a new LDA-based face recognition algorithm with selection of principal components, IEEE transactions on systems, man and cybernetics—part B: Cybernetics., 35(5), 1065-1078(2005).

[11] A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Networks, 10(3), 626-634(1999).