A COMPARATIVE STUDY OF DISTANCE METRICS USED IN FACE RECOGNITION

[1] Zunxiong Liu, [2]Zheng Xu, [1] Linfeng Hu

[1]Asstt Prof., School of Information Engineering, East China Jiaotong Uni., Nanchang, China-330013

2Assoc. Prof., School of Electronic & Electrical Engineering, , East China Jiaotong Uni., Nanchang, China-330013

E-mail: liuzx@ecjtu.jx.cn, zxu@4y.com.cn

ABSTRACT

Face recognition(FR) has become a specialized application area within the field of computer, among which the appearance-based approaches are popular. The appearance-based FR can be induced into the subspace projection step and a nearest neighbor classifier. Because face image data have the property of high dimension, the subspace projection methods should be employed for dimensionality reduction. They include Principal component analysis (PCA), Linear discriminant analysis (LDA), Independent component analysis (ICA) and Non-negative matrix factorization (NMF). In this paper, all the subspace projection algorithms are explored and face recognition system based on the four techniques presented. The simulation experiments are implemented with ORL face dataset, the results presented and analyzed. Also the four techniques are compared in recognition performance for FR.

Keywords: *Subspace Approaches(SA), Dimensionality Reduction(DR), Face Recognition(FR), Performance Comparison(PC)*

1. INTRODUCTION

The face plays a major role in our social intercourse in conveying identity and emotion. Automatic Face Recognition (FR) is a challenging task and has been one of the most successful applications of image analysis and understanding in many fields such as computer vision, pattern recognition. Image-based face recognition techniques can be divided into two groups according to the face representation which they use, which being the appearance-based approach and the feature-based approach, among which the appearance-based is more popular, that use holistic texture features[1].

With automatic face recognition there are many applications in human computer interaction, biometrics and security, etc. Over the decades, many computer systems that can recognize faces have been developed, some of which have been in commercial use. Generally, appearance-based face recognition techniques are finished with image matching in the space of compressed image. If image matching done in the original space, it will result in the curse of dimensionality in addition to the problems of large computational complexity and memory storage. It is necessary to pursue dimensionality reduction schemes in FR, which implemented by way of subspace projection popular in image representation.

Face recognition systems also employ a variety of techniques for selecting subspaces, and the common used subspace projection algorithms are: Primary component Analysis (PCA), Linear Discriminant Analysis (LDA), Independent Component Analysis (ICA) and Nonnegative Matrix Factorization (NMF). All the subspace algorithms can be viewed as calculating out a set of basic vectors according to different conditions and representing original data with their projection coefficients on the vectors. With all the subspace projection algorithms the basic vectors can be obtained by solving the optimization problems under different objective functions and the different constrains.

PCA [2,3]finds a set of the most representative projection vectors such that the projected samples retain the most information about original samples. It is widely used in FR, and the related projection vectors are called Eigenfaces. PCA deals with variance (second-order statistics) and solves the eigenvalue problems to the covariance matrix. Unlike PCA, ICA[4] captures both second and higher-order statistics and projects the input data onto the basis vectors that are as statistically independent as possible. LDA [5, 6] uses the class information and finds a set of vectors that maximize the between-class scatter while minimizing the within-class scatter. NMF is a linear projection technique that imposes non-negativity constraints on both basic vector and coefficient matrices during learning [7]. Although all the subspace projection methods can be used on face recognition, the claims about their performance are contradictory in the literature [9,10]. Usually, the comparative studies are implemented with PCA, LDA and ICA[11,12], and show that choice of appropriate algorithm-metric combination can only be made for a specific task and the performance also depends on the number of subspace dimension retained.

In this paper, we analyzed the comparative performance of PCA, LDA, ICA and NMF on face recognition problem. The rest of this paper is organized as follows. In Sect.2, the ideas of face space in comparison to image space and the general subspace appearance-based face recognition are given. Section 3 deals with subspace projection algorithm including five most popular methods in that area. The experiment with the ORL face dataset is presented In Sect.4. Following that conclusion presented in Sec.5 .

2. OUTLINE OF FACE RECOGNITION

A two-dimensional face image with h (rows) by w (columns) pixels can be represented a vector in a d -dimensional image space ($\Re^{d=h\times w}$)after concatenating its rows or columns. Image space has huge dimensionality and recognition there would be computationally inefficient. However, the face image can also be viewed as a point in the image space, a set of n face images will be represented a set of points (samples of probability distribution) in the certain d -dimensional subspace [13]. This subspace can be modeled as a lower-dimensional manifold embedded in the image space. Subspace projection techniques devoted to extract the principle modes of the underlying manifold while retaining as much information from the original images as possible will derive lower dimensional spaces where face recognition can be done.

Figure.1 shows the block diagram of general subspace face recognition system. Given a set of face images, preprocessing and some other processes are performed on them. Then, the projection matrix $W \in \Re^{d\times p}$ can be obtained, whose columns spanning the feature space are basic vectors with a certain subspace projection techniques. G face images are selected as gallery images, and are projected on the feature space to form G sets of coefficients for later matching. That is implemented by $\left[\mathbf{p}_1,\mathbf{p}_2,\cdots\mathbf{p}_G\right]=W^T\tilde{X}$, $\tilde{X}=(\tilde{\mathbf{x}}_1,\cdots\tilde{\mathbf{x}}_G)$. In testing, the probe image $\mathbf{x}_i$ will be processed as the above mentioned and also be represented by the projection coefficient vector $\mathbf{p}_i = W^T\tilde{\mathbf{x}}_i$. In this research all the subspace projection techniques are combined with nearest neighbor classifier, the most similar gallery image to $\mathbf{x}_i$ is found with the minimum distance between $\tilde{\mathbf{x}}_i$ and $\tilde{X}$, the identity of $\mathbf{x}_i$ will be the same as the most similar gallery image.

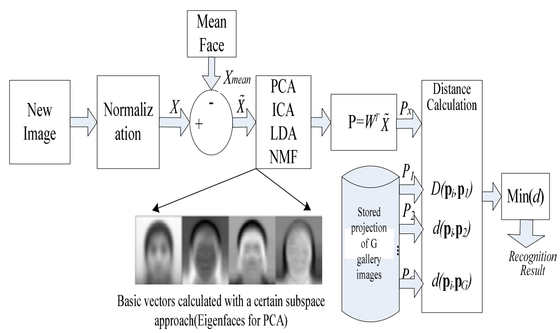

Fig. 1. block diagram of general subspace face recognition system

3. PROJECTION TECHNIQUES FOR DIMENSION REDUCTION

Each one of the four appearance-based subspace projection approaches has its own set of basic functions which are derived based on different statistic viewpoints. In the following these algorithms PCA, LDA, ICA and NMF will be explored.

3.1 PRINCIPAL COMPONENT ANALYSIS (PCA)

Principal Component analysis (PCA) is a popular dimensionality reduction technique, seeking an orthonormal set of principal axes, that is to say, a set of subspace basic vectors correspond to the maximum variance direction in the original image space. In face recognition, the basic vectors can be displayed like some ghostly faces, also called eigenfaces. Let the scatter matrix S_T is

$$S_T = \sum_{i=1}^{N} (\mathbf{x}_i - \boldsymbol{\mu}) \cdot (\mathbf{x}_i - \boldsymbol{\mu})^T \quad (1)$$

Where $\boldsymbol{\mu} = \left(\frac{1}{N}\right)\sum_{i=1}^{N} \mathbf{x}_i$, referred to mean face, and $\mathbf{x}_i$ is i-th face image. The projection matrix W_{PCA} is chosen to maximize the determinant of the total scatter matrix of the projected samples,

$$W_{PCA} = \arg\max_W \left|W^T S_T W\right| = [\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_p] \quad (2)$$

W_{PCA} is composed of p eigenvectors corresponding to p largest eigenvalues. The amount of overall variation that one eigenface counts for, is actually known by the eigenvalue associated with the corresponding eigenvector. If the eigenface with small eigenvalues are neglected, then an image can be a linear combination of a reduced number of these eigenfaces.

3.2 LINEAR DISCRIMINANT ANALYSIS (LDA)

LDA, also known as Fisher's Discriminant Analysis, deals directly with discrimination between classes and finds the projection axes that best discriminate among classes. With the information of all samples and their classes, the between-class scatter matrix S_B and the within-class scatter matrix S_w are defined as $S_B = \sum_{i=1}^{c} N_i \cdot (\mathbf{x}_i - \boldsymbol{\mu}) \cdot (\mathbf{x}_i - \boldsymbol{\mu})^T$ and $S_w = \sum_{i=1}^{c} \sum_{\mathbf{x}_k \in X_i} (\mathbf{x}_k - \boldsymbol{\mu}_i) \cdot (\mathbf{x}_k - \boldsymbol{\mu}_i)^T$. N_i is the number of training samples in class i, c is the number of classes, $\boldsymbol{\mu}_i$ is the centroid of class i. x_i is the samples set which are attached to class i. $\boldsymbol{\mu}$ represents the mean of all classes.

LDA finds a set of vectors W_{LDA} such that Fisher Discriminant Criterion is maximized:

$$W_{LDA} = \arg\max_W \left|\frac{W^T \cdot S_B \cdot W}{W^T \cdot S_W \cdot W}\right| \quad (3)$$

The optimization problem can be solved when the column vectors of projection matrix(W_{LDA}) are the eigenvectors of $S_W^{-1} \cdot S_B$. In general PCA is used as a preprocessing step to prevent S_W from becoming singular, so the final transformation is $W_{OPT}^T = W_{LDA}^T \cdot W_{PCA}^T$.

3.3 INDEPENDENT COMPONENT ANALYSIS (ICA)

ICA tries to explain the original data using statistically independent random vectors, the observed signal matrix $\mathbf{X}$ can be modeled as $\mathbf{X} = \mathbf{AS}$. $\mathbf{S}$ is the source signal matrix, and the source signals are independent of each other. The mixing matrix $\mathbf{A}$ is invertible. Based on the observed signals, ICA algorithms will find $\mathbf{A}$ such that $\mathbf{U} = \mathbf{WX} = \mathbf{WAS}$ is an estimation of the independent source signals[14]. There are many algorithms that perform ICA (InfoMax [14], JADE [15], FastICA [16]) but they all seem to converge to the same solution for any given data set.

There are two fundamentally different ways to apply ICA to face recognition, respectively referenced to as Architecture I and Architecture II. However, PCA is always employed to reduce dimensionality prior to performing ICA.

In Architecture I, the images in $\mathbf{X}$ are considered to be a linear combination of statistically independent basis images $\mathbf{S}$, which estimated as the learned ICA output $\mathbf{U}$. Let $\mathbf{R}$ be a $d \times m$ matrix containing the first m eigenvectors of a set of n face image matrix $\mathbf{X}$ with the dimension of $d \times n$, and ICA be performed on $\mathbf{R}^T$. The m independent basis images in the rows of $\mathbf{U}$ are computed as $\mathbf{U} = \mathbf{W} \cdot \mathbf{R}^T$. Because of the assumption that $\mathbf{W}$ is invertible, there is

$$\mathbf{R}^T = \mathbf{W}^{-1} \cdot \mathbf{U} \quad (4)$$

Let $\mathbf{C}$ be the $n \times m$ matrix of PCA coefficients, so

$$\mathbf{C} = \mathbf{X} \cdot \mathbf{R} \text{ and } \mathbf{X} = \mathbf{C} \cdot \mathbf{R}^T \quad (5)$$

Substituting (4) into (5), therefore

$$\mathbf{X} = \left(\mathbf{C} \cdot \mathbf{W}^{-1}\right) \cdot \mathbf{U} = \mathbf{B} \cdot \mathbf{U} \quad (6)$$

where each row of $\mathbf{B}$ contains the coefficients for linearly combining the basis images to comprise the image in the corresponding row of $\mathbf{X}$.

The goal of ICA in architecture II is to find statistically independent coefficients for input data, where the pixels are variables and the images are observation. The source separation is performed on the pixels, and each row of the learned weight matrix W is an image. **A**, the inverse matrix of W, contains the basis images in its columns. The statistically independent source coefficients in **S** that comprise the input images are recovered in the columns of **U.**

Same as architecture I, ICA is also performed on the PCA coefficients in architecture II. The statistically independent coefficients are computed as $\mathbf{U} = \mathbf{W} \cdot \mathbf{C}^T$. As the above definition of $\mathbf{C} = \mathbf{X} \cdot \mathbf{R}$, it can be obtained that

$$\mathbf{X} = \mathbf{R} \cdot \mathbf{C}^T = \left(\mathbf{R} \cdot \mathbf{W}^{-1}\right) \cdot \mathbf{U} \qquad (7)$$

and the actual basis images are obtained from the columns of $\mathbf{R} \cdot \mathbf{W}^{-1}$.

3.4 NON-NEGATIVE MATRIX FACTORIZATION (NMF)

NMF applies non-negativity constraints to the basis vectors and the corresponding coefficients when representing a vector in data matrix as combination of the basis vectors, therefore it leads to a part-based representation because the only additive combinations of the basis vectors are allowed [7]. Given a non-negative face image matrix $\mathbf{X}$ with the dimension of $d \times n$ (n image vectors) , NMF finds the $d \times p$ basic vector matrix $\mathbf{W}$ and the $p \times n$ coefficient matrix $\mathbf{H}$ in the condition of $\mathbf{X} \approx \mathbf{W} \times \mathbf{H}$, here all the elements are not allowed to be negative in the factorized matrices $\mathbf{W}$ and $\mathbf{H}$, p is selected meeting the requirement $\left(d+n\right) \times p < d \times n$ to accomplish dimension reduction. Each column of $\mathbf{W}$ contains a basic vector while each of column of $\mathbf{H}$ contains the weights needed to approximate the corresponding face image in $\mathbf{X}$. In order to estimate the factorization matrices, the objective functions need defining, it can be defined as

$$F = \sum_{i=1}^{d} \sum_{\mu=1}^{n} \left[\mathbf{X}_{i\mu} \log(\mathbf{WH})_{i\mu} - (\mathbf{WH})_{i\mu} \right] \qquad (8)$$

This function can be related to the likelihood of generating the images in $\mathbf{X}$ from the bases $\mathbf{W}$ and encodings $\mathbf{H}$. An iterative approach can be employed to reach a local maximum of the equation (8) as follows:

$$\mathbf{W}_{ia} \leftarrow \mathbf{W}_{ia} \sum_{\mu} \frac{\mathbf{X}_{i\mu}}{\left(\mathbf{WH}\right)_{i\mu}} \mathbf{H}_{a\mu} \qquad (9)$$

$$\mathbf{W}_{ia} \leftarrow \frac{\mathbf{W}_{ia}}{\sum_{j} \mathbf{W}_{ja}} \qquad (10)$$

$$\mathbf{H}_{a\mu} \leftarrow \mathbf{H}_{a\mu} \sum_{i} \mathbf{W}_{ia} \frac{\mathbf{X}_{i\mu}}{\left(\mathbf{WH}\right)_{i\mu}} \qquad (11)$$

After learning the NMF basis functions, new data are projected into p dimensional space by fixing $\mathbf{W}$, randomly initializing $\mathbf{H}$, the iterative algorithm is also implemented until convergence. In this way the encoding coefficients for probe images are obtained.

4. EXPERIMENTAL RESULTS

The experiments are performed on the ORL dataset constructed by AT&T Laboratories at Cambridge[17], being composed of 400 images of size 112×92. There are 40 persons, with 10 images for each person. The images are taken at different times, lighting and facial expressions. The faces are in up-right position of frontal view, with slight leftright rotation. So we have an input dimensionality of $d = 10304$. Fig.2 depicts some samples.

In the experiments, the different numbers of subject can be chosen to evaluate the subspace models. For each of the selected subject number, 6 images per person are randomly chosen to constitute the stored gallery images, and the left images will be taken as the probe images.

Fig. 2. Some sample images of two persons with ORL

After the gallery images are set, the different subspace approaches can be used to obtain a series of rotation basis vectors for dimensionality reduction. They are PCA-based subspace, LDA-based subspace, two ICA-based subspaces and NMF-based subspace, and eight basis vectors for each technique are illustrated in Fig.3, where the rows from top down are for eigenfaces with the highest eigenvalues(PCA), LDA, architecture Ⅰ with ICA, architecture Ⅱ with ICA and NMF respectively. Among them, the basis vectors with architectureⅠwith ICA and NMF show localized features to some degrees. All the images are projected the different subspaces or obtained by calculation, and the encodings are fed into a simple nearest neighbor classifier, where L2 (euclidean) distance metrics are used.

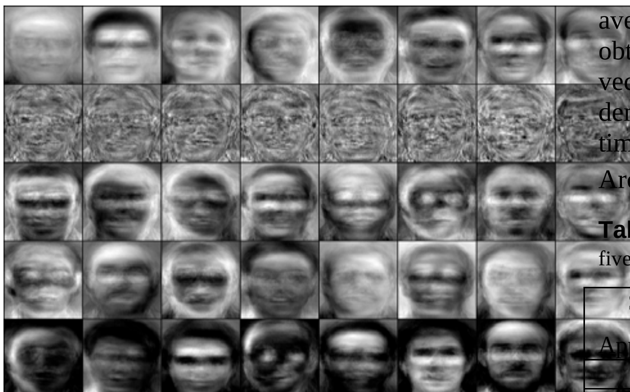

Fig. 3. Eight basis vectors for each technique for the subspace approaches

According to the above dataset partition, if 40 subjects are selected, the training dataset with 240 samples and the testing dataset with 160 samples are produced. In the following experiments, the figures of merit are recognition rates averaged over 6 runs. The recognition rate curves obtained for the five methods are shown in Fig. 4 as functions of the number of feature vectors, which ranges for 4 to 24 at the interval of 2.

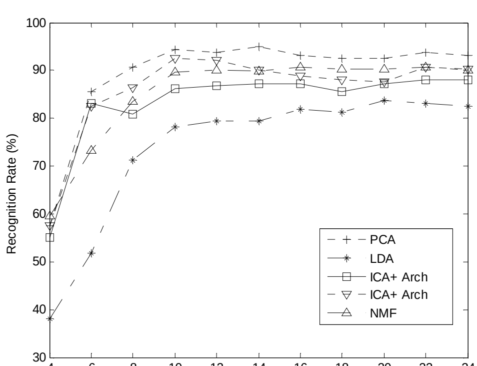

Fig. 4. Comparison of recognition rates obtained by the five FR methods as functions of the number of feature vectors

From Fig. 4, it can be seen that the recognition performances with the five methods tend to be stationary when the number of basis vector is more than 9. Face recognition based on PCA holds top in recognition rate compared with others. What's more, the LDA approach is in the opposite position. These may be in contradiction to the allegations from other voices, because face recognition rates vary with the task and the related dispositions. The five algorithms with a different number of subjects are also tested, the dataset partition is the same, and the number of basis vector is fixed in 9. Tab. 1 presents the average recognition rate data with 6 runs, obtained for the five methods with 9 basis vectors for different subject number. It also demonstrates that PCA behaves best almost all time in the experiment, and that PCA and ICA Arch II are leading in recognition performance.

Table 1. Comparison of recognition rates obtained by the five FR methods with different subject number (%)

Subject Num / Approach	10	20	30	40
PCA	97.50	95.00	92.50	93.13
LDA	97.50	86.25	80.83	76.25
ICA+Arch I	97.92	89.58	86.53	85.31
ICA+Arch II	97.50	93.75	92.64	90.63
NMF	95.83	90.21	89.03	86.15

5. CONCLUSION

In this paper, the five subspace projection techniques, which are PCA, LDA, ICA (two architectures) and NMF are studied. They are nice approaches for dimensionality reduction. As known, face images are with high dimension and the face recognition system generally consists of two steps of feature extraction and classification following. The experiments are done wiith ORL face dataset, recognition performance for all the 5 algorithms is evaluated . The experimental results tell that PCA is a nice dimensionality reduction technique, and can be employed for face recognition very well.

6. ACKNOWLEGEMENT

This work was supported by the Scientific Research Foundation with Jiangxi Provincial Department of Education in China(Grant No: GJJ08237).

REFRENCES

[1]. ZHAO Wenyi, CHELLAPPA R, PHILLIPS J, et al. Face recognition in still and video images:a literature surrey[J].ACM Computing Surveys,2003,35(4)：399－458.

[2]. M. Turk. A Random Walk through Eigenspace[J]. IEICE Trans. Inf. & Syst., Vol. E84-D, No. 12, December 2001, pp. 1586-1595

[3]. M. Turk, A. Pentland. Eigenfaces for Recognition[J]. Journal of Cognitive Neuroscience, Vol. 3, No. 1, 1991, pp. 71-86.

[4]. M.S. Bartlett, J.R. Movellan, T.J. Sejnowski. Face Recognition by Independent Component Analysis[J]. *IEEE Trans. on Neural Networks*, Vol. 13, No. 6, November 2002, pp. 1450-1464.

[5]. P. Belhumeur, J. Hespanha, D. Kriegman. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection. *Proc. of the Fourth European Conference on Computer Vision*, Vol. 1, 14-18 April 1996, Cambridge, UK, pp. 45-58.

[6]. A. Martinez, A. Kak. PCA versus LDA[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2001, 23 (2) : 228-233.

[7]. D. Lee and H. Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401:788-791, 1999.

[8]. Guillamet, D., Vitrià, J. Classifying faces with non-negative matrix factorization. In: Proc. 5th Catalan Conf. for Artif. Intell., (2002), pp.24-31.

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

[10]. M. S. Bartlett, H. M. Lades, and T. J. Sejnowski, Independent component representations for face recognition, presented at SPIE Symposium on Electronic Imaging: Science and Technology; Conference on Human Vision and Electronic Imaging III, San Jose, CA, 1998.

[11]. K. Delac, M. Grgic, and S. Grgic, Independent comparative study of PCA, ICA, and LDA on the FERET data set［J］, Int. J. Imaging Systems and Technology, 2005, 15 (5) : pp.252–260.

[12]. DRAPER B A, BAEK K, BARTLETT M S, et al. Recognizing faces with PCA and ICA［J］. Computer Vision and Image Understanding, 2003, 91(1/2): 115-137.

[13]. G. Shakhnarovich and B.Moghaddam, Face recognition in Subspaces, Handbook of Face Recognition, Springer-Verlag, 2004.

[14]. J.-F. Cardoso, Infomax and Maximum Likelihood for Source Separation［J］, *IEEE Letters on Signal Processing*, 1997 , 4: 112-114.

[15]. M. Zibulevsky and B. A. Pearlmutter, "Blind Separation of Sources with Sparse Representations in a Given Signal Dictionary", *International Workshop on Independent Component Analysis and Blind Signal Separation*, 19-22 June 2000, Helsinki, Finland.

[16]. A. Hyvarinen, "The Fixed-point Algorithm and Maximum Likelihood Estimation for Independent Component Analysis", *Neural Processing Letters*, Vol. 10, Issue 1, August 1999, pp. 1-5.

[17]. ORL face database.. AT&T Laboratories, Cambridge, U.K.. [Online]. Available: http://www.cam-orl.co.uk/facedatabase.html