Nonparametric Subspace Analysis for Face Recognition

Zhifeng Li[1], Wei Liu[1], Dahua Lin[1], and Xiaoou Tang[1,2]
[1]Department of Information Engineering
The Chinese University of Hong Kong, Shatin, Hong Kong
[2]Microsoft Research Asia, Beijing, China
{zli0, wliu, dhlin4, xtang}@ie.cuhk.edu.hk

Abstract

Linear discriminant analysis (LDA) is a popular face recognition technique. However, an inherent problem with this technique stems from the parametric nature of the scatter matrix, in which the sample distribution in each class is assumed to be normal distribution. So it tends to suffer in the case of non-normal distribution. In this paper a nonparametric scatter matrix is defined to replace the traditional parametric scatter matrix in order to overcome this problem. Two kinds of nonparametric subspace analysis (NSA): PNSA and NNSA are proposed for face recognition. The former is based on the principal space of intra-personal scatter matrix, while the latter is based on the null space. In addition, based on the complementary nature of PNSA and NNSA, we further develop a dual NSA-based classifier framework using Gabor images to further improve the recognition performance. Experiments achieve near perfect recognition accuracy (99.7%) *on the* **XM2VTS** *database.*

1. Introduction

Linear Discriminant Analysis (LDA), which is based on Fisher Linear Discriminant (FLD) [11], is a popular face recognition technique. It aims to find the most discriminative features maximizing the ratio of determinant of between-class variations to within-class variations. A number of LDA-based methods [1][4][8] have been proposed in face recognition. However, due to their parametric nature which assumes that the samples satisfy normal distribution, all these methods suffer from serious performance degeneration for cases of non-normal distribution. In [7], a nonparametric technique is proposed to overcome this problem for the case of two classes, in which a nonparametric between-class scatter is defined. However, it does not give a definition for multi-class problems. To apply it to face recognition, which is a typical multi-class recognition problem, we propose a novel algorithm called nonparametric discriminant analysis (NDA) which extends the definition of the nonparametric between-class scatter matrix to the multi-class problem. Similar to conventional LDA, the NDA features are obtained by computing the leading eigenvectors of $S_w^{-1}S_b^N$, where S_w is the within class scatter and S_b^N is the nonparametric between class scatter.

For high dimensional problems, there are often not enough training samples to guarantee the within class scatter matrix non-singularity. Inspired by the idea of the unified subspace [12][17], we propose a novel method called principal nonparametric subspace analysis (PNSA) to extract nonparametric discriminating features within the principal subspace of within class scatter matrix, This will help to stabilize the transformation and thus improve the recognition performance.

A limitation with the PNSA method is that it only utilizes the principal subspace of the intra-personal scatter with the whole null space discarded. Practically the null space of within class scatter also contains a great deal of discriminative information [13], so we develop another null-space method called null-space nonparametric subspace analysis (NNSA) that is based on the null space of the within-class scatter matrix. The two NSA based approaches: PNSA and NNSA are inherently complementary. PNSA utilizes the principal subspace of the intra-personal scatter matrix, while NNSA focuses on null space, thus it is desirable to combine the two types of classifiers. This strategy is similar to the one developed in [18] and [19].

Gabor wavelets have been shown to outperform original appearance features in recent study [6][15]. However, previous methods either downsample the Gabor responses or use only the Gabor responses at certain fiducial points. Apparently, the used Gabor features are not selected from the whole Gabor feature set in a statistically optimal way. This is mainly because the huge dimensions of the complete Gabor features prevent the application of traditional subspace methods. In this paper, we propose a multiple Gabor-based classifiers using the dual NSA framework to handle the high dimentionality of Gabor features. By using this new algorithm, our experiments yield very encouraging results: near perfect recognition accuracy (99.7%) are obtained on the **XM2VTS** database.

2. Parametric Discriminant Analysis

Linear Discriminant Analysis has been widely used for feature extraction in pattern recognition. It is also called the parametric discriminant analysis (PDA) in [7] since it uses the parametric form of the scatter matrix. In PDA, the within-class scatter matrix and the between-class scatter matrix are used to measure the class separability. They are defined as,

$$S_w = \sum_{i=1}^{c} \sum_{j \in C_i} (x_j - \mu_i)(x_j - \mu_i)^T \tag{1}$$

$$S_b = \sum_{i=1}^{c} N_i (\mu_i - \mu)(\mu_i - \mu)^T, \tag{2}$$

where μ_i denotes the mean of the class C_i and N_i denotes the number of samples in class C_i.

The PDA features are the optimal projections W_{opt}, which maximizes the ratio of the determinant of between-class matrix to that of the within-class matrix,

$$W_{opt} = [w_1, w_2, \cdots, w_f] = arg \max_{W} \frac{|W^T S_b W|}{|W^T S_w W|}, \tag{3}$$

and mathematically it is equivalent to the leading eigenvectors of $S_w^{-1} S_b$.

From Eq. (1) to (3), we can see that the PDA has at least three disadvantages. Firstly, the PDA algorithm is based on the assumption that all classes share the same normal distribution. So it cannot perform well in the cases of non-normal distribution. Practically, the real samples often are not normally distributed. Secondly, the number of the final LDA features has an upper limit because the rank of the between-class matrix is at most $c - 1$. However, it is often insufficient to separate the classes well with only $c - 1$ features, especially in high dimensional spaces. Third, with only the centers of classes taken into account for computing between class scatter matrix, it fails to capture the boundary structure of classes effectively, which has been proven to be essential in classification.

3. Nonparametric Discriminant Analysis

For the two-class problem, a nonparametric technique called nonparmetric discriminant analysis (NDA) was proposed to solve above problems [7]. In NDA, the calculation of within-class scatter matrix keeps the same form as PDA. The difference between NDA and PDA is in the definition of the between-class scatter matrix. In [7], the two-class nonparametric between-class scatter matrix is defined as,

$$S_b^N = \sum_{t=1}^{N_1} W(t)(x_t^1 - \mu_2(x_t^1))(x_t^1 - \mu_2(x_t^1))^T + \sum_{t=1}^{N_2} W(t)(x_t^2 - \mu_1(x_t^2))(x_t^2 - \mu_1(x_t^2))^T, \tag{4}$$

where x_t^i denotes the t-th face vector of class i, and $\mu_j(x_t^i)$ is the local K-NN mean, defined by

$$\mu_j(x_t^i) = \frac{1}{k} \sum_{p=1}^{k} NN_p(x_t^i, j), \tag{5}$$

where $NN_p(x_t^i, j)$ is the p-th nearest neighbor from class j to the face vector x_t^i. $W(t)$ is the value of the weighting function. Later we will give an extended definition of $W(t)$ for multi-class problem.

However, such nonparametric between-class matrix definition is only available for two-class cases. For face recognition, which is a typical multi-class recognition problem, we need to generalize Eq. (4) to its multi-class form. We propose a new definition of the nonparametric between-class scatter matrix for multi-class problem is as follows:

$$S_b^N = \sum_{i=1}^{c} \sum_{j=1, j \neq i}^{c} \sum_{t=1}^{N_i} W(i,j,t)(x_t^i - \mu_j(x_t^i))(x_t^i - \mu_j(x_t^i))^T, \tag{6}$$

where $w(i, j, t)$ is defined as

$$W(i,j,t) = \frac{\min\{d^\alpha(x_t^i, NN_k(x_t^i, i)), d^\alpha(x_t^i, NN_k(x_t^i, j))\}}{d^\alpha(x_t^i, NN_k(x_t^i, i)) + d^\alpha(x_t^i, NN_k(x_t^i, j))} \tag{7}$$

Here α is a control parameter that can be selected between zero and infinity, $d(v_1, v_2)$ is the distance between two vectors v_1 and v_2. The weighting function has the property that for samples near the classification boundary it approaches 0.5 and drops off to zero if the samples are far away from the classification boundary. By using such a weighting function, the boundary information contained in the training set is emphasized.

After computing S_w and S_b^N, the final NDA features are the eigenvectors of the matrix $S_w^{-1} S_b^N$. To overcome the singularity problem, PCA is required before hand.

From Eq. (6) we have the following observations. Firstly, if we select $k = N_i$ and set all the values of the weighting function equal one, $\mu_j(x_t^i)$ will become μ_j, the center of the class j. It means the nonparametric discriminant analysis is indeed a generalized version of the parametric discriminant analysis (PDA). So it is expected to inherit the advantage of PDA and furthermore provide more flexibility.

Secondly, compared to the PDA, which can at most extract $c-1$ discriminant features, the nonparametric discriminant does not have such limitation. Thus more features can be extracted for discrimination, and accordingly enhance the classification performance with more information utilized.

Thirdly, the nonparametric discriminant analysis is more effective in capturing the boundary structural information for different classes compared with the PDA algorithm. This can be explained by examining the vectors $(x_t^i - \mu_j(x_t^i))$. As illustrated in Figure 1, where k is set to 1 and some of these vectors are visualized, NDA has two advantages over PDA in utilization of boundary information. Firstly, the non-parametric between-class scatter matrix spans a space involving the subspace spanned by the vectors $\{v_2, \cdots, v_8\}$ where boundary structure is embedded. Therefore, the boundary information can be fully utilized. Secondly, as mentioned before, the weighting function in Eq. (6) can help to emphasis the samples near the boundary, and thus capture the boundary structure information more effectively. For PDA, it computes the between-class scatter matrix only using the vector v_1, which is merely the difference between the centers of the two classes. It is obvious that v_1 fails to capture the boundary structure information effectively.

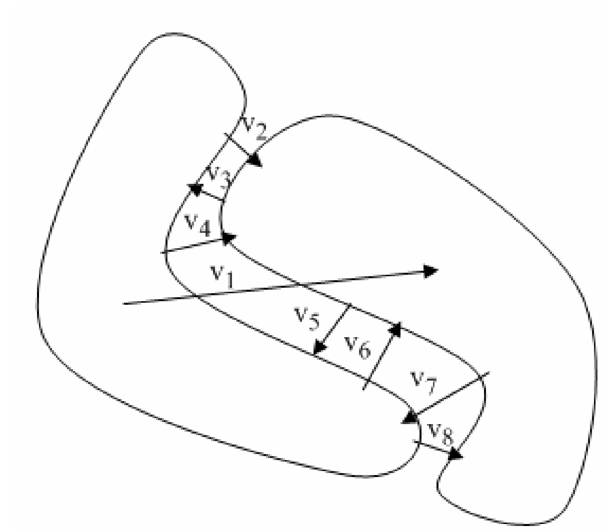

Figure 1. Nonparametric between-class scatter and parametric between-class scatter. v_1: Difference vector of the centers of two classes; $v_2, \cdots, v_8$: Difference vectors from the samples located in the classification boundary.

Furthermore, inspired by the idea of the unified subspace [12][17] and the dual-space LDA [18][19], we propose two kinds of nonparametric subspace analysis (NSA). One is based on the principal space of the intra-personal scatter, and the other is based on the null space of the intra-personal scatter. We call them principal nonparametric subspace analysis (PNSA) and nullspace nonparametric subspace analysis (NNSA) respectively.

The detailed algorithm of the PNSA is as follows:

Step 1. Project a face vector to its PCA subspace established by training samples and then adjust the PCA dimension to better reduce noise.

Step 2. Compute the whitened intrapersonal subspace using the within-class scatter matrix in the reduced PCA subspace and adjust the dimension of the whitened intrapersonal subspace to reduce the dimension of intrapersonal variation.

Step 3. Determine the nonparametric between-class scatter matrix in the intra-personal subspace and then apply PCA to obtain the final nonparametric discriminant features.

The difference between standard NDA and PNSA is that the dimension of the whitened space is variant instead of fixed. This will not only help to reduce the feature dimension but can also make the transform more stable and hence increase the generalization ability.

NNSA is another NSA based technique. As opposed to PNSA that is based on the principal space of the within-class scatter matrix, NNSA focuses on the null-space.

Step 1. Compute the within-class scatter matrix from the training data.

Step 2. Calculate the null space projection of within-class scatter matrix.

Step 3. Project the sample space to the null space and compute the nonparametric between-class scatter matrix in null space.

Step 4. Project the sample to the space of the null space nonparametric inter-personal scatter matrix.

4. Dual NSA-based Classifiers Framework

As discussed before, the PNSA and the NNSA are complementary of each other. The former preserve the principal space of the within-class scatter matrix with the information residing in the null space of the within-class scatter matrix discarded. On the contrary, the latter preserve the null space while discard the information in the principal space. Motivated by the complementary nature of the two kinds of classifiers, it is desirable to integrate them together to fully utilize the discriminative information in the whole space, thus further boost the recognition performance.

Moreover, the property of the Gabor wavelet feature makes it attractive in face representation [6]. The Gabor images obtained by different kernels is similar to multiple frames in a video sequence, thus the fusion technique for video sequence [14] can be applied here to improve the recognition accuracy and efficiency.

Based on the analysis above, we develop a multiple NSA-based classifiers framework to further improve the recognition performance, in which the two types of complementary classifiers: PNSA and NNSA based on Gabor images are combined.

4.1. Gabor Wavelet Representation

We use the method in [6] to extract the Gabor wavelet features of an image by convolving the image with a family of Gabor kernels (wavelets, filters) , which is the product of a Gaussian envelope and a plane wave, defined as follows

$$\psi_{\vec{k}}(\vec{z}) = \frac{\|\vec{k}\|^2}{\sigma^2} \cdot e^{-\frac{\|\vec{k}\|^2 \cdot \|\vec{z}\|^2}{2\sigma^2}} \cdot [e^{i\vec{k}\cdot\vec{z}} - e^{-\frac{\sigma^2}{2}}], \qquad (8)$$

where $\vec{z} = (x, y)$ is the variable in the spatial domain, and $\|\cdot\|$ denotes the norm operator. $\vec{k}$ is the frequency (wave) vector, which determines the scale and the orientation of Gabor kernels

$$\vec{k} = k_s e^{i\phi_d}, \qquad (9)$$

where

$$k_s = \frac{k_{max}}{f^s},\ k_{max} = \frac{\pi}{2},\ f = \sqrt{2},\ s = 0, 1, 2, 3, 4$$

and

$$\phi_d = \frac{\pi d}{8},\ for\ d = 0, 1, 2, 3, 4, 5, 6, 7.$$

Let s and d denote the scale and orientation of the Gabor kernels, here we use Gabor kernels of five different scales, $s \in \{0, \cdots, 4\}$, and eight different directions, $d \in \{0, \cdots, 7\}$. For simplicity, we express the Gabor kernels $\psi_{\vec{k}}(\vec{z})$ as $\psi_{s,d}(\vec{z})$.

Let $I(x, y) = I(\vec{z})$ be the gray level distribution of an image, the convolution of image I and a Gabor kernel $\psi_{s,d}(\vec{z})$ is defined as follow

$$W_{s,d}(\vec{z}) = I(\vec{z}) * \psi_{s,d}(\vec{z}), \qquad (10)$$

where $*$ denotes the convolution operator, and $W_{s,d}(\vec{z})$ is the convolution result corresponding to the Gabor kernel at orientation d and scale s. Hence, the set $\mathcal{S} = \{W_{s,d}(x, y) \mid s \in \{0, \cdots, 4\},\ d \in \{0, \cdots, 7\}\}$ forms the Gabor wavelet representation of the image $I(x, y)$. As in [6], $W_{s,d}(x, y)$ can be computed efficiently via the Fast Fourier Transform (FFT).

Different from the feature extraction in [6], we extend the Gabor representation $W_{s,d}(x, y)$ to float valued images, called Gabor-based images or Gabor images, instead of concatenating all Gabor representation results to derive an augmented feature vector. Therefore, we can obtain an image sequence composed of 40 float valued images, which is similar to an extended video, for each facial image. In addition, we will not conduct arbitrary down sample to reduce the data size. In stead we use efficient multiple subspace analysis algorithms to optimally extract Gabor features from the full set of Gabor responses.

4.2. Integration of Dual NSA-based Classifiers

By applying Gabor wavelet transform, for each face, we can finally acquire a sequence composing of a set of Gabor-based images. Contrary to the traditional Gabor wavelet representation where only the Gabor wavelet coefficients around some discrete fiducial points [15] are computed, we extract the Gabor wavelet features based on the whole image. Therefore, much more information is available for further analysis. Nonetheless, such approach improves the utilization of information at the expense of increasing processing complexity. For example, in our experiments we have 40 images of size 61x41 for each sequence, thus the feature dimension is 100040. Such a huge amount of data is infeasible to manipulate directly.

In order to handle these data efficiently and extract the discriminative feature effectively, a multiple classifiers fusion framework is developed. We first break the whole sequence into slices, with features from each Gabor image as a slice. Then, we apply the appropriate classifier to process each individual slice. Finally all the slice-based classifiers are integrated via a fusion rule to obtain the final decision.

A variety of methods on combining multiple classifiers have been proposed [9][10]. In this paper, we use the majority voting rule to combine the frame-based classifiers: majority voting and sum rule. More sophisticated combination algorithms may further improve the recognition accuracy.

This framework has obvious advantages both in efficiency and effectiveness. Firstly, this fusion framework is composed of several parallel classifiers with each one processing a portion of data. So, efficient parallel computation is enabled. Secondly, in our framework only the synchronized data are contained in the same classifier. It has been shown that the synchronization information is conductive to increase the recognition accuracy [14]. So this framework is also quite effective in extracting the discriminating features. Another advantage of this framework is the functional agility. We can apply different algorithms to this framework respectively and then combine them easily.

By incorporating all these strategies, a novel framework integrating both PNSA and NNSA on Gabor image representation is developed. It is called dual NSA-based classifiers fusion method.

The detailed procedure of the whole framework is summarized as below:

In the training stage

Step 1. For each Gabor Kernel, train PCA model based on the slices obtained by filtering with that kernel.

Step 2. Construct PNSA and NNSA classifiers based on slices obtained using different Gabor kernels.

In the testing stage, for a new testing image

Step 1. Obtain the 40 Gabor images by convolution with Gabor kernels.

Step 2. Use each PNSA or NNSA classifier to determine the classification respectively based on the corresponding Gabor image.

Step 3. Combine the decisions made by slice-based classifiers with the majority voting rule.

The procedure of the algorithm is clearly illustrated in Figure 2.

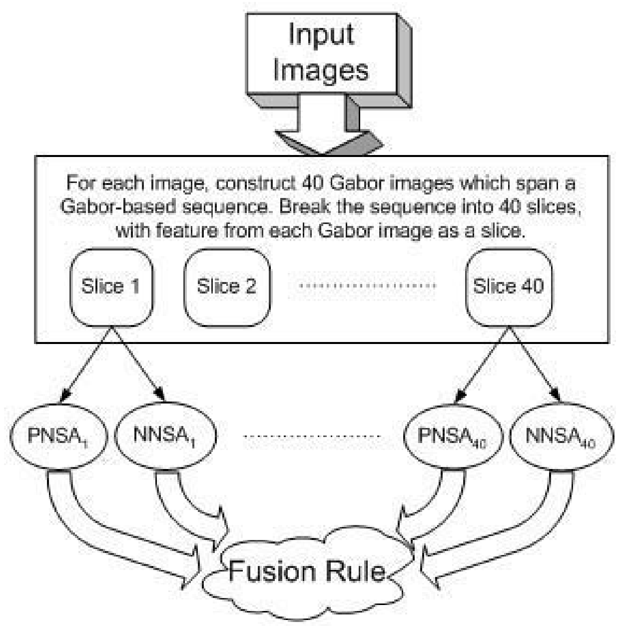

Figure 2. The procedure of dual NSA-based classifiers framework. $PNSA_i(i = 1, 2, \cdots, 40)$**,** $NNSA_i(i = 1, 2, \cdots, 40)$ **means performing PNSA and NNSA on the** i**-th slice**

5. Experiments

In this section, we conduct experiment on the **XM2VTS** face database [5]. We select 293*4 face images of 293 distinct persons captured in four different sessions. For training, we select the 293*3 face images of 293 people captured in the first three sessions. For testing, we use the 293 face images captured in the first session as gallery samples and the 293 face images captured in the fourth session as probe samples.

The first experiment is to compare the three proposed nonparametric methods, NDA, PNSA and NNSA, with all the three traditional subspace methods, PCA [3], LDA [1], and Bayesian method [2]. The comparative recognition results are summarized in Table 1. From these results we can see that the NDA method, which is the original nonparametric method, outperforms all the three traditional subspace methods, PCA, LDA, and Bayesian analysis. In addition, further improvement on the recognition accuracy is achieved by the two improved nonparametric algorithms PNSA and NNSA. This clearly shows the superiority of the nonparametric technique. Comparing to the best results of the conventional subspace methods, the error rate is reduced at least by 45%.

In the second experiment, we investigate the performance of the NSA-based multi-classifier fusion framework on Gabor wavelet images. As mentioned above, 40 Gabor wavelet images are obtained for each face, accordingly 40 PNSA based classifiers and 40 NNSA based classifiers are constructed with each one corresponding to the images acquired by a certain Gabor kernel. Three different fusion schemes are tested in the experiments. The first fusion scheme combines the 40 PNSA classifiers, while the second fusion scheme combines the 40 NNSA classifiers. And the third fusion scheme integrates all PNSA and NNSA classifiers together.

Table 2 reports the results in this experiment. Compared with the results in experiment 1, the classifier fusion strategy does remarkably boost the recognition performance. This clearly shows the superiority of NSA-based multiple-classifier framework. In addition, it is encouraging to see that the third scheme which combines both PNSA and NNSA classifiers achieves a near perfect recognition accuracy (99.7%).

Table 1. The comparison of the nonparametric methods with all the three traditional subspace methods.

Method	Recognition Accuracy
PCA	66.6%
LDA	88.7%
Bayesian method	89.1%
NDA	92.2%
PNSA	94.5%
NNSA	94.2%

Table 2. The recognition results of the NSA-based multi-classifier fusion framework on the Gabor wavelet images.

Method	Recognition Accuracy
The first fusion scheme using PNSA	99.0%
The second fusion scheme using NNSA	98.0%
The third fusion scheme using PNSA and NNSA	99.7%

6. Conclusions

Conventional LDA has an inherent problem due to its parametric nature of the scatter matrix. In this paper, a novel algorithm based on the definition of nonparametric between-class scatter matrix is proposed. Inspired by the idea of the unified subspace and null space, two kinds of nonparametric subspace analysis method PNSA and NNSA are presented to achieve better stability and generalization performance. Moreover, based on the complementary nature of PNSA and NNSA, and the Gabor feature representation, we further develop a dual NSA-based classifiers fusion framework to further boost the recognition performance. Experiments show the effectiveness of our framework with near perfect recognition accuracy (99.7%) achieved on the **XM2VTS** face database.

Acknowledgement

The work described in this paper was fully supported by grants from the Research Grants Council of the Hong Kong SAR and a Hong Kong RGC/NSFC joint project N_CUHK409/03.

References

[1] P. Belhumeur, J. Hespanda, and D. Kiregeman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. on PAMI*, Vol. 19, No. 7, pp. 711-720, July 1997.

[2] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian Face Recognition," *Pattern Recognition*, Vol. 33, pp. 1771-1782, 2000.

[3] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," in *Proc. of CVPR*, pp. 586-591, 1991.

[4] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical Performance Analysis of Linear Discriminant Classifiers," in *Proc. of CVPR*, pp. 164-169, 1998.

[5] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Matitre, "XM2VTSDB: The extended M2VTS database," *Second International Conference on AVBPA*, March 1999.

[6] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. on Image Processing*, Vol. 11, No. 4, pp. 467-476, 2002.

[7] K. Fukunaga, *Statistical Pattern Recognition*, Academic Press, 1990.

[8] D. L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. on PAMI*, Vol. 18, pp. 831-836, Aug. 1996.

[9] L. Xu, A. Krzyzak, and C. Y. Suen, "Method of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," *IEEE Trans. on System, Man, and Cybernetics*, Vol. 22, No. 3, pp. 418-435, 1992.

[10] T. K. Ho, J. Hull, and S. Srihari, "Decision Combination in Multiple Classifier Systems," *IEEE Trans. on PAMI*, Vol. 16, No. 1, pp. 66-75, Jan. 1994.

[11] R. Fisher, "The Use of Multiple Measures in Taxonomic Problems," *Ann. Eugenics*, Vol. 7, pp. 179-188, 1936.

[12] X. Wang and X. Tang, "Unified Subspace Analysis for Face Recognition," in *Proc. of ICCV*, Vol. 1, pp. 679-686, 2003.

[13] L. Chen, H. Liao, J. Lin, M. Ko, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognition*, Vol. 33, No. 10, pp. 1713-1726, 2000.

[14] X. Tang and Z. Li, "Frame Synchronization and Multi-Level Subspace Analysis for Video Based Face Recognition," in *Proc. of CVPR*, Vol. 2, pp. 902-907, 2004.

[15] L. Wiskott, J. M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. on PAMI*, Vol. 19, No. 7, pp. 775-779, 1997.

[16] W. Zhao, and R. Chellappa, "Discriminant Analysis of Principal Components for Face Recognition," in *Proc. of Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 336-341, 1998.

[17] X. Wang and X. Tang, "A Unified Framework for Subspace Face Recognition," *IEEE Trans. on PAMI*, Vol. 26, No. 9, pp. 1222-1228, 2004.

[18] X. Wang and X. Tang, "Dual-Space Linear Discriminant Analysis for Face Recognition," in *Proc. of CVPR*, Vol. 2, pp. 564-569, 2004.

[19] X. Wang and X. Tang, "Random Sampling LDA for Face Recognition," in *Proc. of CVPR*, Vol. 2, pp. 259-265, 2004.