Nonparametric Discriminant Analysis for Face Recognition

Zhifeng Li, *Member*, *IEEE*,
Dahua Lin, *Student Member*, *IEEE*, and
Xiaoou Tang, *Fellow*, *IEEE*

Abstract—In this paper, we develop a new framework for face recognition based on nonparametric discriminant analysis (NDA) and multiclassifier integration. Traditional LDA-based methods suffer a fundamental limitation originating from the parametric nature of scatter matrices, which are based on the Gaussian distribution assumption. The performance of these methods notably degrades when the actual distribution is non-Gaussian. To address this problem, we propose a new formulation of scatter matrices to extend the two-class NDA to multiclass cases. Then, in order to exploit the discriminant information in both the principal space and the null space of the intraclass scatter matrix, we develop two improved multiclass NDA-based algorithms (NSA and NFA) with each one having two complementary methods that are based on the principal space and the null space of the intraclass scatter matrix, respectively. Comparing to the NSA, the NFA is more effective in the utilization of the classification boundary information. In order to exploit the complementary nature of the two kinds of NFA (PNFA and NNFA), we finally develop a dual NFA-based multiclassifier fusion framework by employing the overcomplete Gabor representation for face images to boost the recognition performance. We show the improvements of the developed new algorithms over the traditional subspace methods through comparative experiments on two challenging face databases, the Purdue AR database and the XM2VTS database.

Index Terms—Face recognition, classifier design and evaluation, nonparametric discriminant analysis (NDA), multiclassifier fusion.

1 Introduction

Linear Discriminant Analysis (LDA) [6] is a representative subspace analysis method which has been extensively studied for face recognition. It aims to find the most discriminative features by maximizing the ratio of the determinant of the between-class scatter matrix to that of the within-class scatter matrix. To enhance the stability and performance of LDA, a number of improved LDA-based methods [1], [2], [4], [5], [11], [15], [18], [19], [20], [21], [24], [25], [26], [27], [28] have been proposed. However, most of the existing LDA-based methods inherit the parametric nature from the traditional LDA approach: The construction of the scatter matrices relies on the underlying assumption that the samples in each class satisfy the Gaussian distribution. Thus, they suffer performance degradation in cases of non-Gaussian distribution. In [7], a nonparametric technique is developed to overcome the problem by introducing a new definition for the between-class scatter matrix, which explicitly emphasizes the samples near boundary. Under the new formulation, by utilizing the whole training set, instead of merely the class centers, and weighting the sample pairs according to their contributions to discrimination, the learning algorithm generates a model more adapted to the sample space, especially in the non-Gaussian cases. However, this nonparametric definition is restricted to the two-class cases and cannot be applied in the multiclass classification such as face recognition. In this paper, we propose a new formulation of between-class scatter matrix by extending the definition of the original nonparametric between-class scatter matrix to the multi-class problem [9]. Based on this new formulation, a new method called multiclass nonparametric discriminant analysis (NDA) is proposed.

The well-known small sample size problem often arises when applying LDA in face recognition due to the high dimensionality of the sample space and the limited training set. In these cases, the within-class scatter matrix becomes nearly singular, which would incur serious instability and overfitting. In order to address this problem, we further propose a new method called principal nonparametric subspace analysis (PNSA) to extract nonparametric discriminating features within the principal subspace of within-class scatter matrix. This helps stabilize the transformation and thus improves the generalization performance.

A limitation of the PNSA method is that it only utilizes the principal subspace of the intrapersonal scatter with the whole null space discarded. It has been shown that the null space of within-class scatter also contains important discriminative information [3], [19], so we develop another method called null-space nonparametric subspace analysis (NNSA) to make use of the null space of the within-class scatter matrix.

However, the within-class scatter matrix in nonparametric subspace analysis (NSA) still has the parametric form, which may cause recognition performance degradation. In order to address this problem, we further propose a new formulation of scatter matrices in which both the within-class and between-class scatter matrices are redefined in nonparametric form to better exploit the discriminant information in training data. Based on this new formulation, an enhanced NSA algorithm called nonparametric feature analysis (NFA) is derived accordingly. Similar to NSA, we also derive two additional methods for the principal space and the null space: The principal space NFA (PNFA) is based on the principal space of the within-class scatter matrix and the null-space NFA (NNFA) is based on the null space of the within-class scatter matrix. Inspired by the dual-space LDA in [19], we can see that the two NFA-based approaches, PNFA and NNFA, are inherently complementary. Thus, it is desirable to combine the two types of classifiers.

We apply the developed NFA methods on Gabor features for face recognition. Gabor wavelets have been shown to outperform original appearance features [10], [22]. However, previous methods either downsample the Gabor responses [10] or use only the Gabor responses at certain fiducial points [22]. To fully utilize all the information embedded in the overcomplete Gabor representations without creating an extremely high-dimensional Gabor space, we use multiple classifiers in a dual NFA framework to handle the high dimensionality of Gabor features. Significant improvement over conventional subspace methods are achieved as demonstrated on two challenging face databases, the Purdue AR database and the XM2VTS database. We chose these two data sets because of their large variation of face samples over a reasonable data size.

- *Z. Li is with the Human-Computer Communications Laboratory, Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Hong Kong. E-mail: zfli@se.cuhk.edu.hk.*
- *D. Lin is with the Computer Science and Artificial Intelligence Laboratory, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA 02139. E-mail: dhlin@mit.edu.*
- *X. Tang is with the Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong. E-mail: xtang@ie.cuhk.edu.hk.*

Manuscript received 9 Nov. 2007; revised 16 Mar. 2008; accepted 5 May 2008; published online 27 June 2008.
Recommended for acceptance by M.-H. Yang.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2007-11-0758.
Digital Object Identifier no. 10.1109/TPAMI.2008.174.

2 Related Work

2.1 Parametric Discriminant Analysis

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

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

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

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 projection matrix $\mathbf{W}_{\mathrm{opt}}$, which maximizes the ratio of the determinant of between-class matrix to that of the within-class matrix,

$$\mathbf{W}_{\mathrm{opt}} = [\mathrm{w}_1 \mathrm{w}_2, \ldots \mathrm{w}_\mathrm{f}] = \operatorname{argmax} \frac{|\mathrm{W}^\mathrm{T} \mathrm{S}_\mathrm{b} \mathrm{W}|}{|\mathrm{W}^\mathrm{T} \mathrm{S}_\mathrm{w} \mathrm{W}|}, \tag{3}$$

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

From (1) to (3), we can see that the PDA has three disadvantages. First, the PDA algorithm is based on the assumption that all classes share the Gaussian distribution with the same covariance matrix. So, it cannot perform well in the cases of non-Gaussian distribution. Second, the number of the final LDA features has an upper limit $c-1$ 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 shown to be essential in classification [7].

2.2 Two-Class Nonparametric Discriminant Analysis

For a two-class problem, a nonparametric technique called NDA was proposed to solve the aforementioned problems in [7]. We call it two-class NDA. In two-class NDA, the within-class scatter matrix has the same form as the two-class PDA. The difference between them lies in the definition of the between-class scatter matrix. In [7], the two-class nonparametric between-class scatter matrix is defined as

$$\begin{aligned} S_b^N = &\sum_{l=1}^{N_1} w(1,l)\left(x_l^1 - m_2\left(x_l^1\right)\right)\left(x_l^1 - m_2\left(x_l^1\right)\right)^T \\ &+ \sum_{l=1}^{N_2} w(2,l)\left(x_l^2 - m_1\left(x_l^2\right)\right)\left(x_l^2 - m_1\left(x_l^2\right)\right)^T, \end{aligned} \tag{4}$$

where x_l^i denotes the lth face vector of class i and $m_j(x_l^i)$ is the local KNN mean, defined by

$$m_j\left(x_l^i\right) = \frac{1}{k} \sum_{p=1}^{k} NN_p\left(x_l^i, j\right), \tag{5}$$

where $NN_p(x_l^i, j)$ is the pth nearest neighbor from class j to the face vector x_l^i, and $w(i,l)$ is the value of the weighting function. Later, we will give an extended definition of $w(i,l)$ in (7) for multiclass problem and explain the advantage of the nonparametric between-class scatter matrix in detail.

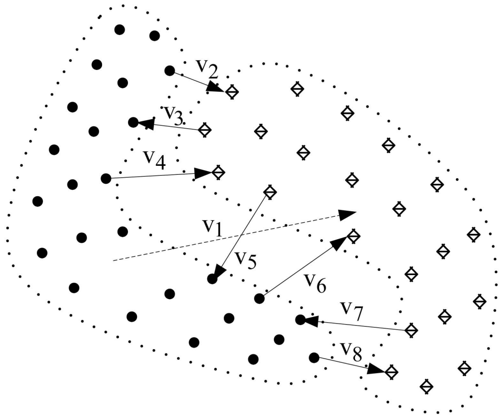

Fig. 1. Nonparametric between-class scatter and parametric between-class scatter. v_1: difference vector of the centers of the two classes; $\{v_2, \ldots, v_8\}$: difference vectors from the samples located at the classification boundary.

3 Nonparametric Discriminant Analysis-Based Methods

3.1 Multiclass Nonparametric Discriminant Analysis

The original nonparametric between-class matrix definition, as shown in (4), is only available for two-class cases. For face recognition, which is a typical multiclass recognition problem, we propose to generalize (4) to a multiclass form. We define the nonparametric between-class scatter matrix for multiclass problem as follows:

$$S_b^N = \sum_{i=1}^{c} \sum_{\substack{j=1 \\ j \neq i}}^{c} \sum_{l=1}^{N_i} w(i,j,l)\left(x_l^i - m_j\left(x_l^i\right)\right)\left(x_l^i - m_j\left(x_l^i\right)\right)^T, \tag{6}$$

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

$$w(i,j,l) = \frac{\min\left\{d^\alpha\left(x_l^i, NN_k\left(x_l^i, i\right)\right), d^\alpha\left(x_l^i, NN_k\left(x_l^i, j\right)\right)\right\}}{d^\alpha\left(x_l^i, NN_k\left(x_l^i, i\right)\right) + d^\alpha\left(x_l^i, NN_k\left(x_l^i, j\right)\right)}, \tag{7}$$

where α is a parameter ranging from zero to infinity which controls the changing speed of the weight with respect to the distance ratio. $d(v_1, v_2)$ is the Euclidean 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 applied before hand.

From (6), we have the following observations. First, if we select $k = N_i$ and set all of the values of the weighting function to be one, $m_j(x_l^i)$ becomes μ_j, the center of class j. It means the NDA is essentially a generalization of the PDA.

Second, in contrast to the PDA, which can only extract at most $c-1$ discriminant features, the NDA inherently breaks such limitation by making use of all the samples in the construction of S_b instead of merely using the class centers. Accordingly, many more features can be extracted for discrimination and thus enhance the classification performance with more information utilized.

Third, the NDA 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 $(x_l^i - m_j(x_l^i))$. As illustrated in Fig. 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. On the one hand, the nonparametric between-class scatter matrix spans a space involving the subspace spanned by the vectors $\{v_2, \ldots, v_8\}$, where boundary structure is embedded. Therefore, the boundary information can be effectively utilized. On the other hand, as mentioned before, the weighting function in (6) can help emphasize 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 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.

One thing to note is that the number of neighbors, k, would affect the recognition performance to some extent. Hence, it is desirable to discuss how to choose an appropriate k for multiclass NDA. As discussed above, if we select $k = N_i$, which is actually the largest possible value for k, and set all the values of the weighting function to be 1, $m_j(x_l^i)$ would become μ_j, which is the center of class j. As a result, multiclass NDA would perform the same as PDA. That means k should not be set too large. Otherwise, multiclass NDA would approach PDA and, thus, may lose the advantages of NDA. On the contrary, if we set k too small, such as 1, that means only a very small amount of training sample pairs are utilized in the learning procedure of multiclass NDA, which may lead to suboptimal performance due to the loss of much information. Therefore, in our experiments, k is chosen as the median of the sample number for each training class.

3.2 Nonparametric Subspace Analysis

Further, considering that S_w may become singular when the number of samples of each class is small, directly solving eigenvectors of $S_w^{-1}S_b$ is infeasible. Inspired by the unified subspace [18] and the dual-space method [19], we propose two kinds of NSA. One is based on the principal space of intraclass scatter and the other is based on the null space of intraclass scatter. We call them PNSA and NNSA, respectively.

The detailed algorithm of the PNSA is given as follows:

1. Project a face vector V to its PCA subspace established by training samples and then adjust the PCA dimension to better reduce noise.
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 better reduce the intrapersonal variations.
3. Calculate the nonparametric between-class scatter matrix S_b^N in the whitened intrapersonal subspace according to (6) and then determine the dominant eigenvectors of S_b^N to obtain the PNSA subspace transformation matrix T_{PNSA}. The final PNSA transformation is formulated as

$$V_{PNSA} = T_{PNSA}^T V. \tag{8}$$

The difference between multiclass NDA and PNSA is that the feature dimension of each step above is variant instead of fixed. This will not only help reduce the feature dimension but also make the transformation more stable and, hence, increase the generalization ability.

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

1. Compute the within-class scatter matrix from the training data and then calculate the null space projection of the within-class scatter matrix.
2. Project the sample space to the null space and compute the nonparametric between-class scatter matrix S_b^N in null space according to (6).
3. Calculate the dominant eigenvectors of S_b^N to obtain the NNSA subspace transformation matrix T_{NNSA}. For any given face vector V, the NNSA transformation is formulated as

$$V_{NNSA} = T_{NNSA}^T V. \tag{9}$$

3.3 Nonparametric Feature Analysis

However, the NSA algorithm still has some limitations. First, as mentioned before, the within-class scatter matrix in NSA still has the same form as PDA. This may affect the recognition performance. Second, the NSA algorithm uses the simple local mean instead of all the selected KNN samples to compute the between-class vectors for the calculation of between-class scatter matrix without considering the fact that different KNN points contribute differently to the construction of between-class scatter matrix.

In order to address these problems, we further develop an enhanced NSA algorithm called NFA for face recognition. In NFA, the new nonparametric within-class scatter matrix and between-class scatter matrix are defined as

$$S_w^{NFA} = \sum_{i=1}^{c}\sum_{p=1}^{k_1}\sum_{l=1}^{N_i}\left(x_l^i - NN_p\left(x_l^i, i\right)\right)\left(x_l^i - NN_p\left(x_l^i, i\right)\right)^T, \tag{10}$$

$$\begin{aligned} S_b^{NFA} = \sum_{i=1}^{c}\sum_{\substack{j=1\\ j\neq i}}^{c}\sum_{p=1}^{k_2}\sum_{l=1}^{N_i} & w(i,j,p,l)\left(x_l^i - NN_p\left(x_l^i, j\right)\right) \\ & \left(x_l^i - NN_p\left(x_l^i, j\right)\right)^T, \end{aligned} \tag{11}$$

where the weighting function in (9) is defined as

$$w(i,j,p,l) = \frac{\min\left\{d^\alpha\left(x_l^i, NN_p\left(x_l^i, i\right)\right), d^\alpha\left(x_l^i, NN_p\left(x_l^i, j\right)\right)\right\}}{d^\alpha\left(x_l^i, NN_p\left(x_l^i, i\right)\right) + d^\alpha\left(x_l^i, NN_p\left(x_l^i, j\right)\right)}. \tag{12}$$

Compared with the NSA, the within-class scatter matrix of NFA has the nonparametric form. Moreover, instead of using the simple local mean to estimate the between-class scatter matrix in NSA, the NFA estimates the contribution of the KNN points, respectively, for the calculation of the between-class scatter matrix. This leads to a more flexible and accurate estimation of the between-class scatter matrix. The experimental results given in Section 5 clearly show the considerable recognition performance improvement of NFA over the NSA.

In order to fully utilize the discriminant information contained in the principal space and the null space of the intraclass scatter matrix, similar to NSA, we also propose two kinds of NFA methods: PNFA and NNFA. The former is based on the principal space of intrapersonal scatter, and the latter is based on the null space of intrapersonal scatter.

The detailed algorithms of the PNFA and NNFA are given as follows:

1. Project a face vector V to its PCA subspace established by training samples and then adjust the PCA dimension to better reduce the noise. Compute the nonparametric within-class scatter matrix S_w^{NFA} in the reduced PCA subspace.
2. Apply PCA to S_w^{NFA} and calculate the principal space F and its complementary subspace $\overline{F}$.
3. In F, compute the whitened intrapersonal subspace and then adjust the dimension to better reduce the intrapersonal variations. Calculate the nonparametric between-class scatter matrix S_b^{NFA} in the reduced intrapersonal principal subspace and then determine the dominant eigenvectors of S_b^{NFA} to obtain the PNFA subspace transformation T_{PNFA}.

This serves as a PNFA classifier, and the transformation is formulated as

$$V_{PNFA} = T_{PNFA}^T V. \tag{13}$$

4. In $\overline{F}$, compute the nonparametric between-class scatter matrix S_b^{NFA} and then determine the dominant eigenvectors of S_b^{NFA} to obtain the NNFA subspace transformation T_{NNFA}. This serves as an NNFA classifier and the transformation is

$$V_{NNFA} = T_{NNFA}^T V. \tag{14}$$

4 Dual-NFA for Gabor Feature Extraction

As discussed before, the PNFA and the NNFA are complementary of each other. The former preserves the principal subspace of the within-class scatter matrix with the information in the null space of the within-class scatter matrix discarded. The latter preserves the null space while discards the information in the principal subspace. It is therefore desirable to integrate them together to fully utilize the discriminative information in the whole space. We call this integrated method the dual NFA.

To demonstrate the advantage of the developed NFA methods, we apply them on Gabor features for face recognition [10], [22]. By applying Gabor wavelet transform, we acquire a set of Gabor-based images for each face image. Contrary to the traditional Gabor wavelet representation, where only the Gabor wavelet coefficients around some fiducial points [22] are computed, we extract the Gabor wavelet features based on the whole image and generate a complete sequence. Therefore, much more information is available for further analysis. Nonetheless, such an approach improves the utilization of information at the expense of increasing processing complexity. For example, in our experiments, we have 40 Gabor images of size 61×41 for each sequence; thus, the feature dimension is 100,040. Such a huge amount of data is difficult to process directly.

In order to handle these data efficiently without notably compromising the utilization of information, inspired by the fusion framework developed for face video sequence in [16], a multiple classifier fusion framework is developed. We first apply the appropriate classifier to process each individual Gabor image. Then, all of the classifiers are integrated via a fusion rule to obtain the final decision.

A variety of methods on combining multiple classifiers have been proposed in [8], [23] such as majority voting and sum rule. In this paper, we use two simple fusion rules to combine the frame-based classifiers: majority voting and sum rule. More sophisticated combination algorithms may further improve the recognition accuracy. By incorporating all these strategies, a multiclassifier framework integrating both PNFA and NNFA on Gabor image representation is developed, which is called dual NFA-based multiple classifier fusion method. The procedure of the algorithm is illustrated in Fig. 2.

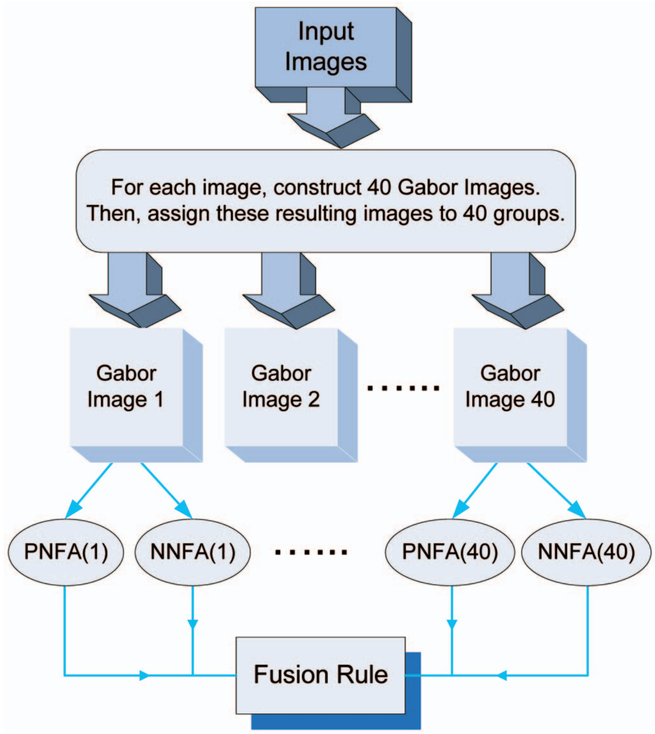

Fig. 2. The procedure of dual NFA-based multiple classifier framework. $PNFA(i)$ and $NNFA(i)$ mean performing PNFA and NNFA on the ith Gabor image.

5 Experiments

In this section, we conduct experiments on two standard face data sets, the AR database [12] and the XM2VTS database [13]. Comparing to other standard data sets, these two data sets have large within-class variations for a relatively large number of people, thus showing a higher degree of non-Gaussian distribution. To better evaluate the recognition performance with geometric and photometric interferences filtered out, we preprocess the face images through the following steps:

1. Rotate the face images to align the vertical face orientation.
2. Scale the face images so that the distances between the two eyes are the same for all images.
3. Crop the face images to remove the background and the hair region.
4. Apply histogram equalization to the face images for photometric normalization.

5.1 Experiment on the AR Face Database

The AR face database contains 126 different persons (70 males and 56 females). Each person has 26 frontal face images, which are divided into two sessions with different expression and different lighting and occlusion. For this database, there are, in total, 90 people who have complete face sequences from both sessions. Here, we select the training data and the testing data from the face images of these 90 persons. For the training data, we select $90 * 7$ nonoccluded face images of 90 persons from the first session. The testing data are composed of a gallery set and a probe set. The gallery set consists of 90 normal face images of 90 persons from the first session while the probe set consists of $90 * 7$ nonoccluded face images of 90 persons from the second session. The face images in the data set are subject to significant illumination variation and exaggerated expression. This makes the recognition task very challenging. The poor recognition accuracies of the traditional subspace methods in Table 1 clearly show this point.

Using the gray level features, the first experiment is to compare the three proposed nonparametric methods: NDA, NSA, and NFA with several popular subspace methods: PCA [17], LDA [1], Bayesian method [14], Kernel LDA [24], where we use the popular polynomial kernel with degree 2, and Local Discriminant Embedding (LDE) [2], which is an improved LDA-based method that relies on manifold learning to learn the class statistics. The comparative recognition results are summarized in Table 1. For the proposed nonparametric methods, we set the number of neighbors to 4 and, finally, extract 89 nonparametric discriminant features for classification. From these results, we can see that the

TABLE 1
Comparison of the Nonparametric Methods with Other Subspace Methods Using the Gray Level Features on the AR Face Database

Methods		Recognition Accuracy (%)
PCA		51.4
LDA		74.6
Bayesian method		74.6
Kernel LDA		75.6
LDE		77.1
Multi-class NDA		78.9
NSA	PNSA	80.0
	NNSA	81.3
NFA	PNFA	82.1
	NNFA	83.8

TABLE 2
The Recognition Results of the NFA-Based Multiclassifier Fusion Framework on the Gabor Wavelet Images on the AR Face Database

Methods	Recognition Accuracy (%)
Sum rule using PNFA	88.9
Voting rule using PNFA	90.0
Sum rule using NNFA	88.3
Voting rule using NNFA	89.8
Sum rule using PNFA and NNFA	91.3
Voting rule using PNFA and NNFA	91.9

multiclass NDA method, which is the original nonparametric method, outperforms all five of the popular subspace methods. In addition, further improvement on the recognition accuracy is achieved by the improved multiclass NDA-based method: NSA (PNSA and NNSA). Furthermore, when using the enhanced NSA method, NFA (PNFA and NNFA), we achieve the best recognition accuracy. This shows the advantage of the nonparametric techniques.

In the second experiment, we investigate the performance of the NFA-based multiclassifier fusion framework on Gabor wavelet images. As mentioned above, 40 Gabor wavelet images are obtained for each face; accordingly, 40 PNFA-based classifiers and 40 NNFA-based classifiers are constructed, with each one corresponding to the image acquired by a certain Gabor kernel. Two popular fusion methods (sum rule and voting rule) are applied to combine the classifiers. The results of the experiment are reported in Table 2. From the results, we have the following observations: 1) By fusing the classifiers based on 40 different Gabor wavelet images, we achieve much better performance than single appearance models and 2) combining the PNFA and NNFA models leads to higher accuracy than combining only the PNFA models or NNFA models, confirming that the PNFA models and NNFA models are mutually complementary to each other.

The proposed multiclassifier fusion framework can also be used for other algorithms such as PCA and LDA to boost the performance of these methods. To verify this point, we conduct an additional experiment using this framework combined with PCA and LDA. The results are reported in Table 3, from which we have the following observations: On the one hand, we can clearly see that the results in Table 3 are much better than those in Table 1. This shows the advantage of this multiclassifier fusion framework again. On the other hand, the results in Table 3 are still lower than Table 2. This is because we use the traditional subspace methods (PCA and LDA) to replace the NFA in the framework in Table 3.

5.2 Experiment on the XM2VTS Database

For the *XM2VTS* database, we select all 295 people with four face images from four different sessions for each person. For the training data, we select $295 * 3$ images of 295 people from the first three sessions. The gallery set is composed of 295 images of 295 people from the first session. The probe set is composed of 295 images of 295 people from the fourth session.

We implement the comparative experiment similarly to the Purdue AR face database experiment. The comparative results are reported in Tables 4, 5, and 6. For the proposed nonparametric methods, we set the number of neighbors to 2 and, finally, extract 294 nonparametric discriminant features for classification. The results further confirm our observations in the Purdue AR face database.

TABLE 3
The Recognition Results of the Multiclassifier Fusion Framework Using the Traditional Subspace Methods on the Gabor Wavelet Images on the AR Face Database

Methods	Recognition Accuracy (%)
Sum rule using PCA	63.5
Voting rule using PCA	62.2
Sum rule using LDA	87.8
Voting rule using LDA	89.4

TABLE 4
Comparison of the Nonparametric Methods with Other Subspace Methods Using the Gray Level Features on the XM2VTS Face Database

Methods		Recognition Accuracy (%)
PCA		66.1
LDA		88.1
Bayesian method		88.5
Kernel LDA		89.8
LDE		90.2
Multi-class NDA		91.5
NSA	PNSA	93.9
	NNSA	93.6
NFA	PNFA	94.9
	NNFA	94.9

TABLE 5
The Recognition Results of the NFA-Based Multiclassifier Fusion Framework on the Gabor Wavelet Images on the XM2VTS Face Database

Methods	Recognition Accuracy (%)
Sum rule using PNFA	99.0
Voting rule using PNFA	99.0
Sum rule using NNFA	98.6
Voting rule using NNFA	98.0
Sum rule using PNFA and NNFA	99.3
Voting rule using PNFA and NNFA	99.7

6 CONCLUSION

Linear discriminant analysis (LDA) is a popular face recognition method. However, conventional LDA faces difficulty in addressing the non-Gaussian aspects of sample distributions due to its parametric nature of scatter matrices. In this paper, a nonparametric formulation of scatter matrices has been proposed to overcome this problem. Using this new formulation, we have proposed two kinds of nonparametric methods: PNSA and NNSA. The former is based on the principal space of intraclass scatter, while the latter is based on the null space of intraclass scatter. Further, to achieve better stability and generalization performance, an enhanced NSA algorithm called NFA (PNFA and NNFA) is derived. Finally, based on the complementary nature of PNFA and NNFA and the Gabor feature representation, we develop a dual NFA-based classifier fusion framework to boost the recognition performance. Experiments show the effectiveness of our framework on the challenging AR and XM2VTS face databases.

ACKNOWLEDGMENTS

The authors would like to thank Professor Helen Meng for her help and support. The work described in this paper was supported by grants from the Research Grants Council of the Hong Kong Special Administrative Region (Project CUHK 4190/01E, Project CUHK 4224/03E, and Project CUHK1/02C).

TABLE 6
The Recognition Results of the Multiclassifier Fusion Framework Using the Traditional Subspace Methods on the Gabor Wavelet Images on the XM2VTS Face Database

Methods	Recognition Accuracy (%)
Sum rule using PCA	85.4
Voting rule using PCA	84.1
Sum rule using LDA	98.0
Voting rule using LDA	98.0

REFERENCES

[1] P. Belhumeur, J. Hespanda, and D. Kiregeman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.
[2] H. Chen, H. Chang, and T. Liu, "Local Discriminant Embedding and Its Variants," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 2005.
[3] 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, 2000.
[4] K. Etemad and R. Chellappa, "Face Recognition Using Discriminant Eigenvectors," *Proc. IEEE Int'l Conf. Acoustics, Speech, and Signal Processing,* vol. 4, pp. 2148-2151, 1996.
[5] K. Etemad and R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *J. Optical Soc. Am. A,* vol. 14, no. 8, pp. 1724-1733, 1997.
[6] R. Fisher, "The Use of Multiple Measures in Taxonomic Problems," *Annals of Eugenics,* vol. 7, pp. 179-188, 1936.
[7] K. Fukunaga, *Statistical Pattern Recognition.* Academic Press, 1990.
[8] T.K. Ho, J. Hull, and S. Srihari, "Decision Combination in Multiple Classifier Systems," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 1, pp. 66-75, Jan. 1994.
[9] Z. Li, W. Liu, D. Lin, and X. Tang, "Nonparametric Subspace Analysis for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 2005.
[10] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing,* vol. 11, no. 4, pp. 467-476, 2002.
[11] M. Loog and R.P.W. Duin, "Linear Dimensionality Reduction via a Heteroscedastic Extension of LDA: The Chernoff Criterion," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 6, pp. 732-739, June 2004.
[12] A.M. Martinez and R. Benavente, "The AR Face Database," CVC Technical Report 24, Purdue Univ., June 1998.
[13] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Matitre, "XM2VTSDB: The Extended M2VTS Database," *Proc. Second Int'l Conf. Audio- and Video-Based Biometric Person Authentication,* Mar. 1999.
[14] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian Face Recognition," *Pattern Recognition,* vol. 33, pp. 1771-1782, 2000.
[15] D.L. Swets and J. Weng, "Using Discriminant Eigenfeatures for Image Retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, no. 8, pp. 831-836, Aug. 1996.
[16] X. Tang and Z. Li, "Frame Synchronization and Multi-Level Subspace Analysis for Video Based Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* 2004.
[17] M. Turk and A. Pentland, "Face Recognition Using Eigenfaces," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,* pp. 586-591, 1991.
[18] X. Wang and X. Tang, "A Unified Framework for Subspace Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 9, pp. 1222-1228, Sept. 2004.
[19] X. Wang and X. Tang, "Dual-Space Linear Discriminant Analysis for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 2004.
[20] X. Wang and X. Tang, "Random Sampling LDA for Face Recognition," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 2004.
[21] X. Wang and X. Tang, "Random Sampling for Subspace Face Recognition," *Int'l J. Computer Vision,* vol. 70, no. 1, pp. 91-104, 2006.
[22] L. Wiskott, J.M. Fellous, N. Krüger, and C. von der Malsburg, "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 775-779, July 1997.
[23] L. Xu, A. Krzyzak, and C.Y. Suen, "Method of Combining Multiple Classifiers and Their Applications to Handwriting Recognition," *IEEE Trans. Systems, Man, and Cybernetics,* vol. 22, no. 3, pp. 418-435, 1992.
[24] M. Yang, "Kernel Eigenface versus Kernel Fisherface: Face Recognition Using Kernel Methods," *Proc. Fifth Int'l Conf. Automatic Face and Gesture Recognition,* 2002.
[25] W. Zhao and R. Chellappa, "Discriminant Analysis of Principal Components for Face Recognition," *Proc. Third IEEE Conf. Automatic Face and Gesture Recognition,* pp. 336-341, 1998.
[26] W. Zhao, "Discriminant Component Analysis for Face Recognition," *Proc. Int'l Conf. Pattern Recognition,* vol. 2, pp. 818-821, 2000.
[27] W. Zhao, R. Chellappa, and N. Nandhakumar, "Empirical Performance Analysis of Linear Discriminant Classifiers," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* pp. 164-169, 1998.
[28] W. Zheng, C. Zou, and L. Zhao, "Real-Time Face Recognition Using Gram-Schmidt Orthogonalization for LDA," *Proc. IEEE Conf. Pattern Recognition,* pp. 403-406, 2004.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/publications/dlib.