Face Recognition Using Novel LDA-Based Algorithms

Guang Dai [1] and **Yuntao Qian** [1]

Abstract. Facial feature extraction with enhanced discriminatory power plays an important role in face recognition (FR) applications. Linear discriminant analysis (LDA) is a powerful tool used for dimensionality reduction and feature extraction in FR tasks. However, the classification performance of traditional LDA is often degraded, due to two factors: 1) their classification accuracies suffer from the small sample size problem (SSSP), which widely exists in FR; 2) their Fisher discriminant criteria are not directly related to the classification ability. Recently, so-called direct fractional-step LDA (DF-LDA) algorithm has been proposed to solve this problem. In this paper, the limitations of DF-LDA are discussed and a novel DF-LDA has been proposed to solve those problems. The novel DF-LDA has been tested, in terms of classification accuracy, on two face databases. Results reveal that the proposed method outperforms the existing methods, including: the Eigenfaces, Fisherfaces, D-LDA, previous DF-LDA, and EFM methods.

Keywords: Face recognition (FR), feature extraction, linear discriminant analysis (LDA), small sample size problem (SSSP).

1 INTRODUCTION

Face recognition (FR) technology has wide applications such as security, human-computer intelligent interaction, digital libraries and the web, and robotics [1]. Interests and research activities in FR have increased significantly over the past few years [1]. Feature extraction is one of the most popular and fundamental problems in FR tasks [1]. One of the popular approaches is principle component analysis (PCA) or Eigenfaces [2]. However, PCA is only the optimal representation criterion in the sense of mean-square error, and does not consider the classifier aspect. It should not be expected optimal performance for FR. Linear discriminant analysis (LDA), which defines a projection that makes the within-class scatter smaller and the between-class scatter larger, generally outperforms PCA in FR tasks [3,4,6]. However, the classification performance of classical LDA is often degraded by the fact that the Fisher discriminant criterion defined in the corresponding LDA is not directly related to classification accuracy in the output space. For effectively solving this problem, a weighted between-class scatter matrix is often constructed for the Fisher criterion, where classes, which are closer together in the input space, are more likely to result in misclassification and should therefore be more heavily weighted in the input space [8]. In addition, recently, Kothari et al [8] has proposed a fractional-step linear discriminant analysis (F-LDA) to extend this idea further by introducing the concept of fractional-step dimensionality reduction, wherein, the dimensionality is reduced in small fractional steps making the relevant distances be more correctly weighted. As same as classical LDA, the F-LDA can not be directly applied to solve the small sample size problem (SSSP), which widely exists in high-dimensional pattern recognition tasks (such as FR), where the number of training samples is smaller than the dimensionality of the samples, due to two factors: 1) the eigen-decomposition of the between-class scatter matrix is very difficult in the high-dimensional space; 2) the singular scatter matrices are caused by the SSSP.

In order to solve the SSSP, which widely exists in FR tasks, a very popular technique usually called PCA plus LDA has been proposed and verified to be effective by experience [3]. In this method, PCA is first utilized for dimensionality reduction before the application of LDA [3]. As PCA used in PCA plus LDA usually may be not compatible to the Fisher discriminant criterion defined in LDA, this technique will lose the null space of the within-class scatter matrix of the training samples, where may contain significant discriminatory information. To effectively avoid the loss due to PCA preprocessing step, some direct LDA (D-LDA) methods have been proposed and successfully applied to FR [4,5] recently. In addition, a more effective method, called DF-LDA, which combines the strengths of the D-LDA of Yu et al [4] and F-LDA of Kothari et al [8], has been proposed and successfully applied to FR. It not only overcomes the limitation of D-LDA [4] that the Fisher criterion function is not directly related to classification accuracy in the output space, but also can make the F-LDA [8] carried out in high-dimensional spaces. However, according to [5,11], it is obvious that the D-LDA algorithm [4], which is used in DF-LDA and first compute corresponding eigenvectors of the positive eigenvalues of the weighted between-class scatter, will discard the null space of the within-class scatter matrix of training set, where usually contains significant discriminatory information. In addition, the rank of the weighted between-class scatter matrix used in the DF-LDA is determined by $\min\{N, c-1\}$, as a result that the most number of optimal discriminant vectors calculated by DF-LDA is at most $c-2$, where N is the dimensionality of the training sample vectors and c is the number of the classes invoked.

Recently, Yang et al [5] has proposed a novel D-LDA, which is a complete PCA plus LDA algorithm essentially, and the performance of this algorithm is superior to that of the previous D-LDA in FR, including D-LDA utilized in DF-LDA. Based on the D-LDA of Yang et al [5] and F-LDA [8], this paper introduces a novel FD-LDA for FR. This method combines the strengths of the F-LDA and D-LDA of Yang et al [5], which at the same time overcomes the limitations of the previous methods. As similar as

[1] College of Computer Science, Zhejiang University, Hangzhou, 310027, P.R.China. email: dai_guang@yahoo.com.cn, ytqian@zju.edu.cn.

the previous DF-LDA [6], the novel DF-LDA will first lower the dimensionality of the raw input space by using D-LDA of Yang at al [5] that leads to a low-dimensional subspace where the most discriminant features are preserved. In addition, a weighted between-class scatter also replaces the normal between-class scatter matrix in the algorithm of D-LDA of Yang et al [5], so that a subsequent F-LDA step can be directly applied to the low-dimensional subspace obtained by D-LDA of Yang et al [5] leading to a set of optimal discriminant features for face representation.

2 NOVEL DIRECT FRACTIONAL-STEP (DF-LDA)

Feature extraction in FR involves the derivation of salient features from the raw input data in order to reduce dimensionality of facial vectors for classification and simultaneously provide enhanced discriminatory power. LDA has been considered as one of the most effective approaches in the feature extraction of FR, and the problem of feature extraction by LDA in FR tasks can be described as follows. Its basic objective is to calculate the Fisher optimal discriminant vectors on the condition that the Fisher criterion function takes extremum, then project high-dimensional facial feature vectors on the obtained optimal discriminant vectors for constructing a low-dimensional facial feature representation.

2.1 Direct LDA (D-LDA)

Suppose there are c known pattern classes, S_b , S_w , and S_t denote the between-class scatter matrix, within-class scatter matrix, and popular scatter matrix, respectively. The traditional Fisher criterion function is generally defined by:

$$J(X)=\frac{X^T S_b X}{X^T S_w X} \quad \text{or} \quad J(X)=\frac{X^T S_b X}{X^T S_t X} \tag{1}$$

In the case of the SSSP, the latter one is usually adopted. The tradition solution to the SSSP, called PCA plus LDA, leads to loss of some significant discriminatory information [6]. Recently, Yu et al [4] has proposed a direct LDA algorithm that can effectively compensate this limitation at a certain extent, and has been successfully applied to FR. Their basic idea generally believes that the significant discriminatory information only exists in the intersection space $S_w(0)\cap S_b^{-1}(0)$, where $S_w(0)=\{x \mid S_w x=0, x\in R^n\}$, and $S_b^{-1}(0)=\{x \mid S_b x\neq 0, x\in R^n\}$. In order to obtain this intersection space, D-LDA of Yu et al [4] first calculated the corresponding eigenvectors of all positive eigenvalues of S_b to obtain the space $S_b^{-1}(0)$, and then calculated the corresponding eigenvectors of at most $c-1$ smallest eigenvalues of $\tilde{S}_W$, which is the projection of S_w in the space $S_b^{-1}(0)$, to obtain the space $S_w(0)$. In fact, according to the procedure of D-LDA of Yu et al [4], it is clear that $R(S_b^{-1}(0))=c-1$, and $R(S_w(0)\cap S_b^{-1}(0))<c-1$, where $R(\bullet)$ denotes the dimensionality of the space' • '. However, according to [5,11], we have $R(S_w(0)\cap S_b^{-1}(0))=c-1$ in the case of the SSSP. As a result, the algorithm of D-LDA of Yu et al [4] may discard some significant discriminatory information.

A novel D-LDA, which builds a theoretical foundation for the PCA plus LDA method in essence, has been proposed by Yang et al [5] recently. They believe that the optimal discriminant vectors only exist in $S_t^{-1}(0)=\{x \mid S_t x\neq 0, x\in R^n\}$; otherwise, if $\forall x\in S_t(0)=\{x \mid S_t x=0, x\in R^n\}$, then $x^T S_t x = x^T S_b x = 0$, i.e. the between-class distance equals 0, which means that the optimal discriminant vectors in $S_t(0)$ are meaningless for classification. The D-LDA of Yang et al [5] can be described as follows. The number of the optimal discriminant vectors is d .

1) In order to obtain $S_t^{-1}(0)$, the PCA is carried out in essence. $\tilde{S}_b$, $\tilde{S}_w$, and $\tilde{S}_t$ are the corresponding projection of S_b , S_w , and S_t in this PCA transformed space R^m , respectively;
2) Calculate the $\tilde{S}_w$'s orthonormal eigenvectors $\gamma_1,\ldots,\gamma_m$, suppose the first q ones are corresponding to positive eigenvalues, then the PCA transformed space R^m can be divided into two sub -spaces: $\tilde{S}_w(0)=\{x \mid \tilde{S}_w x=0, x\in R^m\}$ and $\tilde{S}_w^{-1}(0)=\{x \mid \tilde{S}_w x\neq 0, x\in R^m\}$;
3) Let $P_1=(\gamma_{q+1},\ldots,\gamma_m)$ and $\overline{S}_b=P_1^T\tilde{S}_b P_1$, calculate $\overline{S}_b$'s orthonormal eigenvectors $Z_1,\ldots,Z_l$, then, the first l optimal discriminant eigenvectors contained in $\tilde{S}_w^{-1}(0)$ are $Y_j=P_1 Z_j$, $(j=1,\ldots,l)$. In the case of the SSSP, $l=c-1$, c is the number of classes;
4) Let $P_2=(\gamma_1,\ldots,\gamma_q)$ and $\breve{S}_b=P_2^T\tilde{S}_b P_2$, $\breve{S}_t=P_2^T\tilde{S}_t P_2$, calculate corresponding $d-l$ eigenvectors $Z_{l+1},\ldots,Z_d$ of the $d-l$ leading eigenvalues of $(\breve{S}_b)^{-1}\breve{S}_t$. Then, the remaining $d-l$ optimal discriminant vectors contained in $\tilde{S}_w^{-1}(0)$ are $Y_j=P_2 Z_j$, $(j=l+1,\ldots,d)$;
5) The $Y_j,(j=1,\ldots,d)$ constitute the d optimal discriminant vectors in the PCA transformed space R^m .

In fact, this whole intersection space $S_w(0)\cap S_b^{-1}(0)$ can be obtained in the step (3) above, and the space $S_w^{-1}(0)\cap S_b^{-1}(0)$, where some discriminatory information may exist, can be obtained in the step (4) above too. Yang et al has successfully applied this algorithm to FR tasks, and the comparative results in [5] have shown that the performance of this D-LDA is more effective than that of the D-LDA of Yu et al [4]. Hence, it is reasonable to assume that a novel DF-LDA that combines the strengths of D-LDA [5] and F-LDA [8] can effectively overcome the limitations of previous DF-LDA [6].

2.2 Novel Direct Fractional-Step LDA (DF-LDA)

As same as the classical LDA, the Fisher discriminant criterion defined in the D-LDA above is not directly linked to classification accuracy. A weighted Fisher discriminant criterion, where a weighted between-class scatter matrix replaces a conventional between-class scatter matrix, can effectively avoid those classes that close in input space and can potentially result in misclassification in the output space. According to [9], a weighted between-class scatter matrix can be defined:

$$\hat{S}_b=\sum_{i=1}^{c-1}\sum_{j=i+1}^{c} P_i P_j w(d_{i,j})(m_i-m_j)(m_i-m_j)^T , \tag{2}$$

where $d_{i,j}$ is the Euclidean distance between the means of class i and class j , the weighting function $w(d_{ij})$ is generally a monotonically decreasing function, P_i and P_j are the prior probabilities of class i and class j respectively, m_i and m_j are the means of class i and class j respectively. According to [8], the weight should drop faster than the Euclidean distance between the means of class i and class j ; as a result only constraint of the weighting function is $w(d)=d^{-p}$, where p is integer and $p\geq 3$.

In FR tasks, where the SSSP widely exists, the weighted Fisher criterion function can be expressed as:

$$\hat{J}(Y) = \frac{Y^T \hat{S}_b Y}{Y^T \hat{S}_t Y} \tag{3}$$

where $\hat{S}_t = S_w + \hat{S}_b$, and $\hat{S}_b$ is the weighted between-class scatter matrix defined in (2). According to the algorithm of the D-LDA of Yang et al [5], we start by calculating the orthonormal bases of $\hat{S}_t^{-1}(0) = \{x \mid \hat{S}_t x \neq 0, x \in R^n\}$. However, it is intractable to directly calculate the orthonormal bases of $\hat{S}_t^{-1}(0)$, which is a large matrix in FR tasks. For example, a typical face pattern of (112x92) used in this paper leads to a scatter matrix of size 10304x10304. Fortunately, we can proof the following lemma to solve this problem (The procedure of the proof of this lemma is very intricate, and it is omitted in this paper, due to space limitations.):

Lemma 1 *Suppose* $\varphi_1, \ldots, \varphi_k$ *are a set of orthonormal bases of* $span\{x_1, \ldots, x_N\}$, *then* $\hat{S}_t^{-1}(0) = \{Qy \mid Q^T \hat{S}_t Q y \neq 0, y \in R^m\}$, *where* $\hat{S}_t$ *is defined in (3),* $Q = (\varphi_1, \ldots, \varphi_k)$, *and* $x_1, \ldots, x_N$ *are the training samples.*

From this lemma, we can see that not only the orthonormal bases of $\hat{S}_t^{-1}(0)$ can be obtained, but also the number of these orthonormal bases is no more than the number of the training samples. As a result, the SSSP can be effectively solved. Suppose $Q = [\varphi_1, \ldots, \varphi_m]$ is all orthonormal bases of $\hat{S}_t^{-1}(0)$. Let all raw input vectors project on $\hat{S}_t^{-1}(0)$ and obtain the m -dimensional transformed space R^m. Suppose the $\tilde{S}_w$, $\tilde{S}_b$ and $\tilde{S}_t$ are the corresponding projection of S_w, $\hat{S}_b$ and $\hat{S}_t$ in the transformed space R^m. In the transformed space R^m, split the within-class scatter matrix $\tilde{S}_w = Q^T S_w Q$ into its null space $\tilde{S}_w(0) = span\{\gamma_{q+1}, \ \ldots, \gamma_m\}$ and its orthonormal complementary space $\tilde{S}_w^{-1}(0) = span\{\gamma_1, \ldots, \gamma_q\}$, where $\gamma_1, \ \ldots, \gamma_m$ are all orthonormal eigenvectors of $\tilde{S}_w$ and the first q ones are corresponding to positive eigenvalues. According to [5], it is clear that the Fisher criterion function $\hat{J}(Y)$ can be replaced by $\hat{J}_b(Y) = Y^T \tilde{S}_b Y$ in $\tilde{S}_w(0)$. So the first l optimal discriminant vectors in $\tilde{S}_w(0)$ can be obtained by calculating $\bar{S}_b$'s orthonormal eigenvectors Z_i and the first l optimal discriminant vectors contained in raw input space are $QP_1Z_i (i = 1, \ldots, l)$, where $\bar{S}_b = P_1^T \tilde{S}_b P_1$ is the corresponding projection of $\tilde{S}_b$ in $\tilde{S}_w(0)$, and $P_1 = (\gamma_{q+1}, \cdots, \gamma_m)$. Generally, $c = l - 1$, c is the number of classes. The reminding $d - l$ optimal discriminant vectors in the raw input space can be obtained by calculating the $d - l$ optimal discriminant vectors in $\tilde{S}_w^{-1}(0)$. It is clear that $\breve{S}_t$ is nonsingular in $\tilde{S}_w^{-1}(0)$, so the $d - l$ optimal discriminant vectors in the space $\tilde{S}_w^{-1}(0)$ can be obtained by calculating the corresponding eigenvectors of the $d - l$ leading eigenvalues of $(\breve{S}_t)^{-1} \breve{S}_b$, where $P_2 = (\gamma_1, \ldots, \gamma_q)$, $\breve{S}_b = P_2^T \tilde{S}_b P_2$, and $\breve{S}_t = P_2^T \tilde{S}_t P_2$. As a result, the $QP_2Z_i (i = l+1, \ldots, d)$ constitute the remaining $d - l$ optimal discriminant vectors in raw space.

A low-dimensional SSSP-free subspace has been obtained by the enhanced D-LDA step of Yang et al [5] discussed above, without losing the significant discriminatory information. Then, an F-LDA step will be directly applied to further reduce the dimensionality of SSSP-free subspace from d to d'. In fact, when a pair of classes are well separated in the input space, it is possible that the weight of the pair of classes in computing weighted between-class scatter matrix $\hat{S}_b$ is so small that they would heavily overlap in the lower-dimensional space (The weighting function defined in (2) is the monotonically decreasing function.). The F-LDA can effectively solve this problem [8]. F-LDA, which substantially improves the robustness of the choice of the weighting function, reduces the dimensionality in small fractional steps making the relevant distances be more correctly weighted.

From the statements above, a novel DF-LDA, where the enhanced D-LDA step of Yang et al [5] replaces the enhanced D-LDA step of DF-LDA of [6] in fact, can be constructed, and the pseudocode implementation of the novel DF-LDA for selecting the facial feature has been depicted in Figure.1.

Input: A set of training face images $\{x_i\}_{i=1}^N$

Output: A low-dimensional feature y of the face image x with enhanced discriminatory power.

Algorithm:

Step 1. According to lemma 1, calculate the orthonormal bases of $\hat{S}_t^{-1}(0)$: $Q = [\varphi_1, \ldots, \varphi_m]$, where $m \leq N$.

Step 2. A transformed space R^m can be constituted by projecting all training face images $\{x_i\}_{i=1}^N$ to $\hat{S}_t^{-1}(0)$, and $\tilde{S}_b$, $\tilde{S}_w$ and $\tilde{S}_t$ are the corresponding projection of $\hat{S}_b$, S_w and $\hat{S}_t$ in this transformed space R^m, respectively.

Step 3. Calculate the $\tilde{S}_w$'s orthonormal eigenvectors $\gamma_1, \ldots, \gamma_m$, suppose the first q ones are corresponding to positive eigenvalues.

Step 4. Let $P_1 = (\gamma_{q+1}, \ldots, \gamma_m)$ and $\bar{S}_b = P_1^T \tilde{S}_b P_1$, calculate $\bar{S}_b$'s orthonormal eigenvectors $Z_1, \ldots, Z_l$, then the first l optimal discriminant eigenvectors contained in $\tilde{S}_w(0)$ are $V_j = P_1 Z_j (j = 1, \ldots, l)$.

Step 5. Let $P_2 = (\gamma_1, \ldots, \gamma_q)$ and $\breve{S}_b = P_2^T \tilde{S}_b P_2$, $\breve{S}_t = P_2^T \tilde{S}_t P_2$, calculate the corresponding $d - l$ generalized eigenvectors $Z_{l+1}, \ldots, Z_d$ of the $d - l$ leading eigenvalues of $(\breve{S}_b)^{-1} \breve{S}_t$. Then, the remaining optimal discriminant vectors are $V_j = P_2 Z_j (j = l+1, \ldots, d)$.

Step 6. $Y_j = QV_j = QP_1Z_j (j = 1, \ldots, l)$ and $Y_j = QV_j = QP_2Z_j (j = l+1, \ldots, d)$ can constitute the d optimal discriminant vectors in raw input space, let $Y_j (j = 1, \ldots, d)$ act as projection axes to form the feature extractor $\Theta = [Y_1, \ldots, Y_l, Y_{l+1}, \ldots, Y_d]$.

Step 7.Project all face images $\{x_i\}_{i=1}^N$ to the d -dimensional SSSP-free subspace by the optimal discriminant vectors $\Theta = [Y_1, \ldots, Y_d]$ obtained in step 6, and result in $\{z_i\}_{i=1}^N$, where $z_i = \Theta^T x_i$.

Step 8.Further reduce the dimensionality of z_i from d to d' by directly applying the F-LDA to $\{z_i\}_{i=1}^N$ and let W be the basis of the output space.

Step 9. The final optimal discriminant feature representation of the face image x can be expressed by $y = (\Theta W)^T x$.

Figure 1. Pseudocode for the computation of the novel DF-LDA algorithm.

3 EXPERIMENTS

The comparative experiments are performed using the ORL [12] and the UMIST [13] databases, which are the popular testbed for FR technologies. The ORL database contains 40 different persons and each person has 10 different images, including variations in pose, face expression (open or closed eyes, smiling or non-smiling) and with glasses or no-glasses. All images were taken a dark homogenous background with the subjects in an upright frontal position, with tolerance for some tilting and rotation of up to about

20 degrees. The UMIST database is a multiview database, which consists of 575 gray-scale images of 20 persons, each covering a wide range of poses from profile to frontal views as well as face gender and appearance. All original face images in both databases are sized into 92x112 with 256-level gray scale. Figure.2 shows some images from the two databases.

Figure 2. Some sample images from two databases. (a):ORL. (b):UMIST.

In the following experiments, each one of the two databases can be divided into a training set and a testing set with no overlapping between the two sets. For the ORL database, we select five training images and five testing images per person from this database, and a training set of 200 images and a testing set of 200 images will be created for the following experiments. For the UMIST database, we select five training set of images per person from this database, and a training set of 100 images and a testing set of 475 images will be created for the following experiments. The number of fractional-steps used in DF-LDA is 30 , and the Nearest Neighbor Classifier (NNC) rule is used for classification (In fact, the classification accuracy of the following experiments will lead to being improved if a more sophisticated classifier (such as SVM) is used to instead of the NNC [14]; however, this experiment is beyond the scope of this paper). In addition, since the recognition performance will be affected by the selection of training images, we do each experiment on 10 times and the results reported in this paper are an average of them.

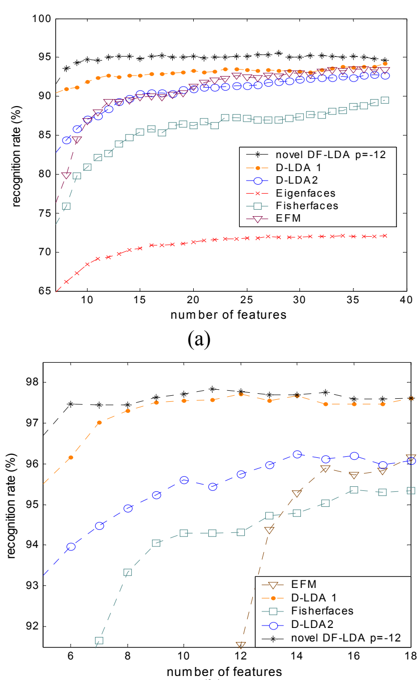

Figure 3. Comparison of recognition rate obtained by the six methods as functions of the feature vectors, where $w(d) = d^{-12}$ is used in the novel DF-DLA (D-LDA 1 and D-LDA 2 denote the D-LDA of Yang et al [5] and the D-LDA of Yu et al [4], respectively). (a):ORL. (b):UMIST[1]

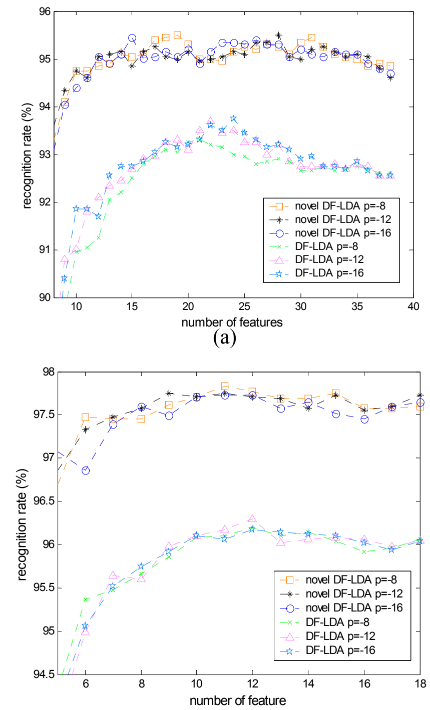

Figure 4. FR performance of the novel DF-LDA method and the previous DF-LDA method with the different weighting functions. (a): ORL. (b):UMIST.

For comparison purpose, we first carry out Eigenfaces [2], Fisherfaces [3], EFM [10], D-LDA [4], D-LDA [5], and novel DF-LDA of this article. The weighting function utilized in the novel DF-LDA is $w(d) = d^{-12}$. The recognition rate curves obtained for the six methods are described in Figure.3 as functions of the number of feature vectors. From Figure.3, it can be seen that the performance of the novel DF-LDA is overall superior to those of the other five methods on the different number of the feature vectors. As the Fisher discriminant criterion defined in the novel DF-LDA is directly related to classification accuracy, the novel DF-LDA can outperform the D-LDA of Yang et al [4] on the different number of the feature vectors. In addition, the D-LDA of Yang et al [5] effectively avoid the loss of discriminatory information in D-LDA of Yu et al [4], so the D-LDA of Yang et al [5] can be overall superior to the D-LDA of Yu et al [4] on both databases. It is not surprising that the D-LDA can not outperform the EFM on the ORL database, because the EFM method can effectively improve the generalization ability of the standard LDA based classifiers as they overfit to the training samples.

The following experiment will exploit the performance of the novel DF-LDA of this article and the previous DF-LDA [6] with the influences of the different weighting function. The comparative results are obtained in Figure.4 as functions of the number of feature vectors. From Figure.4, it can be seen that the performance of the novel DF-LDA can be overall superior to that of the previous DF-LDA with the different weighting functions $w(d) = \{d^{-8}, d^{-12}, d^{-16}\}$ recommended in [8]. In fact, for the ORL database, the maximum recognition rate of the novel DF-DLA in

[1] The recognition rate of the Eigenfaces method is so low on the UMIST database that it is not depicted in the Figure.3 (b).

one of the testing groups can reach 99.5% only using a weighting function of $w(d) = d^{-12}$ and a set of $N = 12$ feature vectors; for the UMIST database, the maximum recognition rate of the novel DF-DLA in one of the testing groups can reach 99.6% only using a weighting function of $w(d) = d^{-12}$ and a set of $N = 8$ feature vectors. It is a result comparable to the best results reported previously in many literatures. It is clear that the feature extraction of face images by the novel DF-LDA is more effective than feature extraction of face images by the previous DF-LDA. In fact, the classification performance of the DF-LDA [6] is degraded by the fact that the algorithm of D-LDA of Yu et al [4] used in the DF-LDA will discard some significant discriminatory information, and from the experimental results above, we can see that the novel DF-LDA can effectively compensate this shortcoming. In order to sufficiently illustrate the effectiveness of the novel DF-LDA method, we also calculate the average percentages of the error rate of the novel DF-LDA over those of other methods above on both databases. Let α_i and β_i be the average error rates of the novel DF-LDA and one of the remaining methods above respectively, then the average percentages of the error rate of the novel DF-LDA over those of the others are obtained by $\sum_{i=5}^{38} \alpha_i / \beta_i$ for the ORL database and $\sum_{i=5}^{18} \alpha_i / \beta_i$ for the UMIST database, where i is the number of feature vectors. The results summarized in Table.1 show that the novel DF-LDA is more effective than the existing FR methods above. From all conclusions above, we can see that the novel DF-LDA can effectively overcome the limitations and shortcomings of those various methods.

Table 1. Average percentages of the error rate of the novel DF-LDA over the others on both databases (O and U denote the ORL database and the UMIST database respectively).

The weighting functions	$w(d) = d^{-8}$			$w(d) = d^{-12}$			$w(d) = d^{-16}$		
Methods	O	U	O+U	O	U	O+U	O	U	O+U
Eigenfaces [2]	17.77%	5.39%	11.58%	18.09%	5.37%	11.73%	18.12%	5.53%	11.83%
Fisherfaces [3]	35.61%	40.85%	38.23%	36.12%	40.71%	38.42%	36.12%	41.92%	39.02%
EFM [10]	56.48%	32.86%	44.67%	57.13%	32.81%	44.97%	57.09%	33.78%	45.44%
D-LDA [4]	54.36%	53.64%	54.00%	55.16%	53.49%	54.33%	55.17%	55.07%	55.12%
D-LDA [5]	73.39%	90.71%	82.05%	74.62%	90.42%	82.52%	74.70%	93.07%	83.89%
Previous DF-LDA [6]	65.20%	58.18%	61.69%	67.96%	57.92%	62.94%	67.82%	59.68%	63.75%

4 CONCLUSIONS

Feature extraction is one of the most significant and fundamental problems in FR tasks, and extracting efficient feature is always the key to solving a problem in FR. One of the most effective tools is the DF-LDA, which has been proposed and successfully extracted the facial feature recently. This paper analyzes the limitations of the DF-LDA and introduces a novel DF-DLA, which combines of the strengths of the F-LDA and D-LDA of Yang et al [5] to effectively compensate those limitations. The efficiency of the novel DF-LDA has been tested through experimentation on the ORL and UMIST face databases.

However, all methods compared in this article can not reflect the correlation of facial feature well under variations due to facial expression and pose changes. The Gabor wavelet representation of face images can capture the local structure corresponding to spatial frequency (scale), spatial location, and orientation selectivity. As a result, the Gabor wavelet representation of face images should be robust to variations due to facial expression and changes. Our next goal is to further search for a FR system, which combines the Gabor wavelet representation of face images and the novel DF-LDA of this article, to further improve recognition rate.

ACKNOWLEDGMENTS

The authors would like to thank Dr.D.Graham and Dr.N.Allison, for providing the UMIST face database, and thank AT&T Laboratories Cambridge for providing the ORL face database. This work is supported by the National Natural Science Foundation of P.R.China (No.60103018).

REFERENCES

[1] R.Chellappa, C.L.Wilson, S.Sirohey, 'Human and machine recognition of faces: a survey', *Proceedings of the IEEE*, **83**, 705-741, (1995).

[2] M.Turk, A.Pentland, 'Eigenfaces for recognition', *J.Cogn. Neurosci.,* **3**, 71–86, (1991).

[3] P.N.Belhumeur, J.P.Hespanha, D.J.Kriegman, 'Eigenfaces vs.Fisherfaces: Recognition using class specific linear projection', *IEEE Trans. Pattern Anal. Machine Intell.,* **19**, 711–720, (1997).

[4] H.Yu, J.Yang, 'A direct LDA algorithm for high-dimensional data with application to face recognition', *Pattern Recognition*, **34**, 2067–2070, (2001).

[5] J.Yang, J.Y.Yang, 'Why can LDA be performed in PCA transformed space', *Pattern recognition*, **36**, 563-566, (2003).

[6] J.Lu, K.N.Plataniotis, A.N.Venetsanopoulos, 'Face recognition using LDA-based algorithms', *IEEE trans on neural networks,* **14**, 195-200, (2003).

[7] G.Dai, H.Zhou, L.Zhang, C.Zhou, 'Feature extraction method based on the Generalized weighted Foley-Sammon transform in high dimensional spaces', *Proceedings of the Indian International Conference on Artificial Intelligence,* (2003).

[8] R.Lotlikar, R.Kothari, 'Fractional-Step dimensionality reduction', *IEEE Trans. Pattern Anal. Machine Intell.,* **22**, 623-627, (2000).

[9] M.Loog, P.R.W.Duin, R.Haeb-Umbach, 'Multiclass linear dimension reduction by weighted pairwise Fisher criteria', *IEEE Trans. Pattern Anal. Machine Intell.,* **23**, 762-766, (2001).

[10] C.J.Liu, H.Wechsler, 'Robust coding schemes for indexing and retrieval from large face recognition', *IEEE Trans. Image Processing*, **9**, 132-137, (2000).

[11] Y.Guo, L.Wu, 'A novel optimal discriminant principle in high dimensional spaces', *Proceedings of the 2nd IEEE International Conference on Development and Learning*, (2002).

[12] [Online]. Available: http://www.cam-orl.co.uk/faceda tabase.html

[13] D.B.Graham, N.M.Allinson, 'Characterizing Virtual Eigen-signatures for General Purpose Face Recognition', (in) Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences, **163**. H. Wechsler, P. J. Phillips, V. Bruce, F. Fogelman -Soulie and T. S. Huang (eds), (446-456), 1998. Available: http://images. ee.umist.ac.uk/danny/database.html.

[14] O.Deniz, M.Castrillon, M.Hernandez, **'**Face recognition using independent component analysis and support vector machines', *Pattern Recognition Letters*, **24**, 2153-2157, (2003).