A Subspace Learning Based on a Rank Symmetric Relation for Fuzzy Kernel Discriminant Analysis

Minlun Yan
Department of Mathematics, Lianyungang Teachers' College, Lianyungang 222006, PR China
Email: yanminlun@yahoo.com.cn

Abstract—**Classification of nonlinear high-dimensional data is usually not amenable to standard pattern recognition techniques because of an underlying nonlinear small sample size conditions. To address the problem, a novel kernel fuzzy dual discriminant analysis learning based on a rank symmetric relation is developed in this paper. First, dual subspaces with rank symmetric relation on the discriminant analysis are established, by which a set of integrated subspaces of within-class and between-class scatter matrices are constructed, respectively. Second, a reformative fuzzy LDA algorithm is proposed to achieve the distribution information of each sample represented with fuzzy membership degree, which is incorporated into the redefinition of the scatter matrices. Third, considering the fact that the kernel Fisher discriminant is effective to extract nonlinear discriminative information of the input feature space by using kernel trick, a kernel algorithm based on the new discriminant analysis is presented subsequently, which has the potential to outperform the traditional subspace learning algorithms, especially in the cases of nonlinear small sample sizes. Experimental results conducted on the ORL and Yale face database demonstrate the effectiveness of the proposed method.**

Index Terms—**subspace learning; nonlinear small sample size problem; fuzzy set; rank symmetry**

I. INTRODUCTION

In recent years, subspace-based approaches have been widely studied as a viable solution to the challenging problem of face recognition across lighting conditions, facial poses and facial expressions, etc. Most traditional algorithms, such as traditional principal component analysis (PCA) [1-3] and linear discriminant analysis (LDA) [4-9] put an image object as a 1-D vector. However, for high-dimensional problem such as face identification, the traditional LDA still suffers from the so called small sample size (SSS) problem or "undersampled" problem which arises whenever the number of samples is smaller than the dimensionality of the samples [10]. In the past, many LDA extensions have been developed to deal with this problem. Briefly, there are four major extensions: pseudoinverse LDA (PLDA) [11], regularized LDA (R-LDA) [12-14], LDA/GSVD [15-16] and two-stage LDA [17-18]. Among these LDA approaches, a very popular technique usually called PCA plus LDA that belongs to the two-stage LDA is most frequently used. In this method, the PCA is first used for dimensionality reduction before the application of LDA, as it was done for the example in Fisherfaces [19] or in EFM [20]. Actually, it has been proved that the null space of S_w contains the most discriminant information when an SSS problem occurs. Based on this fact, Chen et al. [21] presented the null space LDA (NLDA) method, which only extracts the discriminatory information present in the null space of the S_w. Yu and Yang [22] proposed a direct-LDA (D-LDA) method, which takes the range space of the between-class scatter matrix as the intermediate subspace. Yang and Yang [17] proposed a complete LDA (C-LDA) framework, which searches the discriminant vectors both in the range space and in the null space of S_w. The random subspace [23-24] is an efficient technique to overcome the SSS problem, in which the dimensionality of the training data is reduced by random sampling on the facial features. On the basis of random subspace, Zhang and Jia [25] proposed a dual principal random discriminant analysis (RDA) algorithm, which combines the advantages of Fisherface and D-LDA. As analyzed above, it is observed that those classical subspace-based decomposition techniques were just carried out on the only one principal subspace from the within-class or between-class scatter matrix, which leads to a loss of some significant disciminant information in the high dimensional facial space since some potential subspaces are complementary in terms of the discriminative power. Recently, a novel approach called multilinear discriminant analysis (MDA) [26] has been presented to solve the supervised dimensionality reduction problem by encoding an image object as a general tensor of second or even higher order. MDA algorithm based on higher order tensors has the potential to outperform the traditional vector-based subspace learning algorithms, especially in the cases of small sample size.

Moreover, from another point of view, we know that the most well-known feature extraction techniques used for face recognition are those of Eigenface [3] and Fisherface [19]. The Eigenface method relies on a transformation of feature vectors by utilizing PCA. In essence, the PCA dwells on a linear projection of a high-

doi:10.4304/jcp.8.2.380-387

dimensional face image space into a new low-dimensional feature space [17-18]. The major problem of the Eigenface technique is that it can be affected by variations in lighting directions, different face poses and diversified face expressions. The second well-known approach of Fisherface is insensitive to large variation in the conditions we have enumerated above, which performs PCA on the training data and followed by Fisher's linear discriminant (FLD). It has been one of the effective algorithms due to its power of extracting the most discriminatory features. However, the most existing FLD based algorithms we have mentioned above are employed to dwell on the concept of a binary (yes/no) class assignment meaning that the samples are assigned to the given classes (categories) definitely. Evidently, as the samples are significantly affected by numerous environmental conditions (such as face images are affected by illumination, expression, etc.), it is advantageous to investigate these factors and quantify their impact on their "internal" class assignment [27-28]. Interestingly, the idea of such class assignment has been around for a long time and can be dated back to the results published by Keller et al. [29] coming under the notion of a fuzzy *k*-nearest neighbor classifier.

In this paper, the objective of our study is first to establish a novel kernel fuzzy dual subspaces learning based on a rank symmetric relation to solve supervised dimensionality reduction problem by unfolding the feature vectors along different projection directions. Under the Fisher's discriminant criterion, a set of solution spaces including within-class scatter matrix, between-class scatter matrix and their corresponding transformed complements have been partitioned, respectively. Second, the remaining problem of our framework is how to incorporate a new mechanism of fuzzy set into the proposed dual subspaces model. After review the fuzzy Fisherface feature extraction algorithm, we augmented it by some improved mechanisms of fuzzy set. Compared with the fuzzy Fisherface algorithm, the presented fuzzy algorithm computes the discriminant vectors associated with the membership grade from each training sample, which is theoretically effective to overcome the classification limitation originated from the imprecise samples. Third, different from the traditional LDA subspace learning criterion which derives only one principal subspace, in our approach two kernel subspaces and their transformed complements were respectively obtained through the optimization of fuzzy dual discriminant analysis. Therefore, the KFDDA approach has the potential to outperform the traditional LDA algorithms, especially in the cases of nonlinear small sample sizes. Experimental results conducted on two face image databases demonstrate the effectiveness of the proposed method.

II. Theoretical Analysis On A New Dual Subspaces Learning Model

Most previous approaches to subspace learning, such as Fisherface, D-LDA and C-LDA, are performed on the only one principal subspace from the within-class or between-class scatter matrix. In this work, we study how to conduct discriminant analysis in high dimensonal space by unfolding the feature vectors along different projection directions. Also, we explore the characteristics of the dual discriminant analysis based algorithm in theoretical aspect.

A. A New Dual Subspaces Learning Model

Theorem 1: Suppose $A \in M_{n,n}(F)$ is a matrix with $m \times n$ dimension in the field of F, and its rank is r. Then a matrix product can be defined as $AB = 0$ if and only if $B \in M_{n,n}(F)$ and its rank is $n - r$.

Proof: Since $A \in M_{n,n}(F)$ and its rank is r, by the theory of matrix analysis, a matrix transformation can be attained by two invertible matrices P and Q as follows,

$$PAQ = \begin{bmatrix} E_r & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & 0_{(n-r)\times(n-r)} \end{bmatrix} \triangleq I_r$$

where, $P \in M_{n,n}(F)$, $Q \in M_{n,n}(F)$, E_r is identity matrix, I_r is equivalent standard form of matrix A.

Let $Q^{-1}B = \begin{bmatrix} 0_{r\times r} & 0_{r\times(n-r)} \\ 0_{(n-r)\times r} & E_{(n-r)\times(n-r)} \end{bmatrix} \triangleq B^*_{n-r}$

Obviously, $rank(B^*_{n-r}) = n - r$

Thus, $PAQB^*_{n-r} = PAQQ^{-1}B = PAB = 0$

Since P, Q are invertible matrices, a conclusion can be reached

$$AB = 0 \text{ and } rank(B) = rank(B^*_{n-r}) = n - r$$

Hence, $B = QB^*_{n-r}$ is a concise representation of a dual subspace with a rank symmetrical relationship to A, which is complementary to the original feature space of A. □

As analyzed above, by Theorem 1, we may further deduce the fuzzy dual subspace learning with respect to the Fisher discriminant analysis, by which fourfold subspaces originated from with-class scatter matrix and between-class scattermatrix can be obtained, respectively.

B. Discussions

As described previously, most existing approaches to subspace learning are performed on the only one principal subspace from the within-class or between-class scatter matrix. Specifically, Fisherface is implemented in the principal subspace of S_w, D-LDA is carried out in the principal subspace of S_b, C-LDA is conducted by splitting the $\tilde{S}_w$ into its null space and its orthogonal complement. Subsequently, Zhang [25] proposed a RDA algorithm which combines the advantages of Fisherface and D-LDA. In this method, Fisherface and D-LDA are respectively applied to the two principal subspaces of S_w and S_b for simultaneous discriminant analysis. However, due to the defects of Fisherface and D-LDA, some potential and valuable discriminatory information is also lost in the space of S_w and S_b. The MDA [26] is a recently proposed algorithm for dimensionality reduction, it utilizes the higher order statistics for data analysis. The general idea of Yan's work is to explore the characteristics of the high order tensor-based discriminant analysis and search for the k-mode optimization. It is different from the dual discriminant analysis algorithm in the following aspects: 1) In MDA, an image object is directly treated as a matrix and the whole data set is encoded as a third-order tensor. Whereas in the dual discriminant analysis, an image object is still treated as a vector. 2) The semantics of the learned subspaces are different. In MDA, these subspaces originated from Fisher's maximal criterion characterize the discriminating information from internal factors such as row and column directions. While in dual discriminant analysis, the learned subspaces characterize another kind of integrated discriminatory information which is obtained from two proposed subpace matrices originated from the Fisher's within-class scatter matrix and between-class one. 3) In Yan's work, the projection matrices are optimized iteratively. While in dual discriminant analysis, the projection matrices are obtained respectively. Therefore, as the above differences between the MDA and dual discriminant analysis, we do not further compare them in the experiment section.

Also, the computational complexities of Fisherface [19], R-LDA [12-14], D-LDA [22], C-LDA [17] and the proposed dual discriminant analysis are listed in the Table 1.

TABLE I.
THE COMPUTATIONAL COMPLEXITIES OF FISHERFACE, D-LDA, C-LDA, R-DA AND PROPOSED METHOD

Method	Fisherface	D-LDA	C-LDA	R-DA	Proposed method
Complexity	$O(M^3)$	$O(C^3)$	$O(2M^3)$	$O(M^2d)$	$O(4M^3)$

Obviously, the computation requirement of Fisherface increase cubically with the increase of the training sample size M , while the complexity of R-LDA depends on the sample size M and data dimensionality d , therefore, for high-dimensional data where d is larger than M . Moreover, the computation requirement of D-LDA does with the increase of the number of classes C and the computation scales of C-LDA and the proposed dual discriminant analysis depend on the number of reduced subspaces. As analyzed above, although the dual discriminant analysis can be more effective than other ones for classification, it still needs more CPU time for whole process (training and testing) because it costs more computation using feature vectors which are twice as many as those in C-LDA for classification.

III. HOW TO INCORPORATE A NOVEL MECHANISM OF FUZZY SETS INTO DUAL DISCRIMINANT ANALYSIS

A. Why Introduce Fuzzy Set

According to the study in Ref. [27-28], face recognition is a very difficult problem due to a substantial variation in light direction, different face, poses, and diversified facial expressions. By taking advantage of the technology of fuzzy sets [30], a number of studies have been carried out for fuzzy image recognition, fuzzy image filtering, fuzzy image segmentation, and fuzzy edge detection with an ultimate objective to cope with the factor of uncertainty being inherently present in many problems of image processing and pattern recognition [31]. From this point of view, we address the uncertainty associated with a significant variation in illumination, viewing directions, and facial expression in the face images. Moreover, in the feature space of face images samples, unclassifiable regions still remain by those conventional methods. The shaded region in Fig.1 illustrates the problem, it seems that the data in the shaded region is difficult to be classified.

Specifically, by analyzing the existing fisherface, we note that the algorithm dwells on the concept of a binary (yes–no) class assignment meaning that the face images samples come fully assigned to the given classes (categories). Evidently, as the faces are significantly affected by numerous environmental conditions (including illumination, poses, etc.), it is advantageous to investigate these factors and quantify their impact on their "internal" (viz. algorithm-driven) class assignment. In essence, the intent is to reflect all these factors in a "soft" viz. fuzzy class allocation to the individual faces under consideration. Interestingly, the idea of such "fuzzification" of class assignment has been around for a long time and can be dated back to the results published by Keller et al. [29].

In this section, we are concerned with face recognition using a novel fuzzy LDA algorithm. According to the studies in Ref. [29], we conclude that the traditional fuzzy mechanism attempts to "fuzzify" or refine the membership degrees of the labeled patterns only by fuzzifying the each class center. How can we make full use of the distribution information of each sample to the redefinition of scatter matrices? In addition, we note that the performance of those hard feature extraction methods will degenerate whenever the outlier samples exist in the patterns. How can we appropriately represent the membership degrees of these special samples before computing its discriminant vectors? In this section, a complete fuzzy discriminant analysis approach based on the relaxed normalized condition will be proposed and established.

B. An Improved Fuzzy LDA

In the method, the first key step of reformative fuzzy LDA (RFDA) method is how to address the problem coming under the influence of the outlier samples in the patterns. As shown there, the membership matrix denoted by $U=\left[\mu_{ij}\right]$ for $i=1,2,\cdots,C$ and $j=1,2,\cdots,N$ satisfies the obvious property [27]:

$$\sum_{i=1}^{C} \mu_{ij}=1 \qquad (1)$$

In particular, it is worth stressing that the condition helps us to assure the mathematical tractability.

Regretfully, the misclassification results often occur due to the existence of those outliers in the patterns. As usual, since the outliers are far from the class center of each pattern, it is disadvantageous to obtain the exact membership degrees in the setting of fuzzy sets. However, under the restriction of condition (1), the relatively large membership degrees of the patterns to those outliers will be achieved instead. For instance, in a two classes outlier problem, the membership degrees close to 0.5 indicate that the outliers exhibit unprecise membership degrees to the several classes.

Taking into account the fact that the outliers may have some adverse influence to the performance of conventional fuzzy LDA approach, a new relaxed normalized condition in the fuzzy membership degrees is proposed as follows:

$$\sum_{i=1}^{C}\sum_{j=1}^{N}\mu_i(x_j)=N \tag{2}$$

By the new condition (2), we can redefine the fuzzy membership as follows:

$$\mu_{ij}=\begin{cases}\frac{\gamma}{N}+\frac{1-\gamma}{N}\cdot\left(n_{ij}/k\right) \\ \frac{1-\gamma}{N}\cdot\left(n_{ij}/k\right) & otherwise\end{cases} \tag{3}$$

$$\gamma=\frac{N-C}{2^m\cdot N} \tag{4}$$

where m and γ are constants which ultimately control the values of μ_{ij} and satisfy the constraints $m\in(0,1),\gamma\in(0,1)$. Since the normalized condition of the membership degree has been relaxed, the all samples coming with use of membership matrix U are insensitive to the class center of each pattern. Thus, it shows high stability of this improved normalized condition compared to others due to its power of regulating the fuzzy membership grades.

Now, the second key step of our strategy is how to incorporate the contribution of each training sample into the redefinition of scatter matrices. Under the direction of fuzzy set theory [27], each sample can be classified into multi-classes under the membership degrees of the labeled patterns. Thus, the contribution of every sample to classification can be represented by its corresponding fuzzy membership degree.

In the redefinition of fuzzy within-class scatter matrix, samples that are more close to the class center have more contributions to classification. Then, the membership degree of each sample should be considered and the corresponding fuzzy within-class scatter matrix can be remodified as follows:

$$RFS_W=\sum_{i=1}^{C}\left(\sum_{x_j\in w_i}u_{ij}{}^{p}\left(x_j-fm_i\right)\left(x_j-fm_i\right)^T\right) \tag{5}$$

where p is a constant which controls the influence of fuzzy membership degree.

In the redefinition of fuzzy between-class scatter matrix, in contrast to the redescription of the within-class scatter matrix, samples that are far from the center of total classes will have more contribution to classification. Thus, the fuzzy between-class scatter matrix can be remodified as follows:

$$RFS_b=\sum_{i=1}^{C}\left(\left(1-\sum_{x_j\in w_i}u_{ij}^{p}\Big/\sum_{j=1}^{N}u_{ij}^{p}\right)\left(fm_i-\overline{m}\right)\left(fm_i-\overline{m}\right)^T\right) \tag{6}$$

where p is the constant chosen above which controls the influence of fuzzy membership degree, $\overline{m}$ stands for the mean of all vectors (images).

Similarly, the fuzzy total scatter matrix can be achieved as follows:

$$RFS_t=RFS_b+RFS_W \tag{7}$$

Consequently, with the contribution of each sample incorporated, all scatter matrices coming under the notion of complete fuzzy sets are redefined, respectively.

IV. OUTLINE OF KFD (KPCA+LDA)

A. Fundamental

For a given nonlinear mapping Φ, the input data space R^n can be mapped into the feature space H:

$$\Phi: R^n \to H, \quad x \to \Phi(x). \tag{8}$$

Correspondingly, a pattern in the original input space R^n is mapped into a potentially much higher dimensional feature vector in the feature space H.

The idea of KFD [18] is to solve the problem of LDA in the feature space H, thereby yielding a set of nonlinear discriminant vectors in input space. This can be achieved by maximizing the following Fisher criterion:

$$J^{\phi}(\varphi)=\frac{\varphi^T S_b^{\phi}\varphi}{\varphi^T S_t^{\phi}\varphi}, \quad \varphi\neq 0 \tag{9}$$

where S_b^{ϕ} and S_t^{ϕ} are the between-class and total scatter matrices defined in feature space H:

$$S_b^{\phi}=\frac{1}{M}\sum_{i=1}^{C}l_i(m_i^{\phi}-m_0^{\phi})(m_i^{\phi}-m_0^{\phi})^T \tag{10}$$

$$S_t^{\phi}=\frac{1}{M}\sum_{i=1}^{C}(\Phi(x_i)-m_0^{\phi})(\Phi(x_i)-m_0^{\phi})^T \tag{11}$$

Here, $x_1,x_2,...,x_M$ is a set of M training samples in input space; l_i is the number of training samples of class i and satisfies $\sum_{i=1}^{C}l_i=M$; m_i^{ϕ} is the mean vector of the mapped training samples of class i; m_0^{ϕ} is the mean vector across all mapped training samples.

B. KFD Algorithm

The optimal discriminant vectors with respect to the Fisher criterion are actually the eigenvectors of the generalized equation $S_b^{\phi}\varphi=\lambda S_t^{\phi}\varphi$. Since any of its eigenvectors can be expressed by a linear combination of the observations in feature space, we have

$$\varphi=\sum_{j=1}^{M}a_j\Phi(x_j)=Q\alpha \tag{12}$$

where $\mathrm{Q}=\left[\Phi(x_1),\cdots,\Phi(x_M)\right]$ and $\alpha=(a_1,\cdots,a_M)^{\mathrm{T}}$.

Substituting Eq. (12) into Eq. (9), the Fisher criterion in space H is converted to

$$J^{\mathrm{K}}(\alpha)=\frac{\alpha^{\mathrm{T}}(\mathrm{KWK})\alpha}{\alpha^{\mathrm{T}}(\mathrm{KK})\alpha} \tag{13}$$

where the matrix K is defined as

$$\mathrm{K}=\tilde{\mathrm{K}}\text{-}\mathrm{I}_M\tilde{\mathrm{K}}\text{-}\tilde{\mathrm{K}}\mathrm{I}_M+\mathrm{I}_M\tilde{\mathrm{K}}\mathrm{I}_M \tag{14}$$

Here, $\mathrm{I}_M=(1/M)_{M*M}$ is a scaled identity matrix; $\tilde{\mathrm{K}}=\mathrm{Q}^{\mathrm{T}}\mathrm{Q}$ is an $M\times M$ matrix and its elements are determined by

$$\tilde{\mathrm{K}}_{ij}=\Phi(x_i)^{\mathrm{T}}\Phi(x_j)=\left(\Phi(x_i)\cdot\Phi(x_j)\right)=\mathrm{k}(x_i\cdot x_j) \tag{15}$$

where $\mathrm{k}(x_i\cdot x_j)$ is the kernel function corresponding to the given nonlinear mapping function Φ. And $\mathrm{W}=\left(\mathrm{W}_1,\cdots,\mathrm{W}_C\right)$, where W_j is a $l_j\times l_j$ matrix with terms all equal to $1/l_i$. Thereby, W is an $M\times M$ block diagonal matrix.

Now, let us consider the QR decomposition of the matrix K. Suppose $\gamma_1,\gamma_2,\cdots,\gamma_m$ are K's orthonormal eigenvectors corresponding to the m nonzero eigenvalues $\lambda_1\geq\lambda_2\geq\cdots\geq\lambda_m$. Then, K can be expressed by $\mathrm{K}=\mathrm{P}\Lambda\mathrm{P}^{\mathrm{T}}$, where $\mathrm{P}=\left(\gamma_1,\gamma_2,\cdots,\gamma_m\right)$ and $\Lambda=\mathrm{diag}\left(\lambda_1,\lambda_2,\cdots,\lambda_m\right)$. Obviously, $\mathrm{P}^{\mathrm{T}}\mathrm{P}=\mathrm{I}$, where I is the identity matrix.

Substituting $\mathrm{K}=\mathrm{P}\Lambda\mathrm{P}^{\mathrm{T}}$ into Eq. (13), we have

$$J^{\mathrm{K}}(\alpha)=\frac{\left(\Lambda^{1/2}\mathrm{P}^{\mathrm{T}}\alpha\right)^{\mathrm{T}}\left(\Lambda^{1/2}\mathrm{P}^{\mathrm{T}}\mathrm{WP}\Lambda^{1/2}\right)\left(\Lambda^{1/2}\mathrm{P}^{\mathrm{T}}\alpha\right)}{\left(\Lambda^{1/2}\mathrm{P}^{\mathrm{T}}\alpha\right)^{\mathrm{T}}\Lambda\left(\Lambda^{1/2}\mathrm{P}^{\mathrm{T}}\alpha\right)} \tag{16}$$

Let

$$\beta=\Lambda^{1/2}\mathrm{P}^{\mathrm{T}}\alpha \tag{17}$$

Then, Eq. (16) becomes

$$J(\beta)=\frac{\beta^{\mathrm{T}}S_b\beta}{\beta^{\mathrm{T}}S_t\beta} \tag{18}$$

where, $S_b=\Lambda^{1/2}\mathrm{P}^{\mathrm{T}}\mathrm{WP}\Lambda^{1/2}$ and $S_t=\Lambda$.

It is easy to know that S_t is positive definite and S_b is positive semi-definite. So, Eq. (18) is a standard generalized Rayleigh quotient. By maximizing this Rayleigh quotient, we can obtain a set of optimal solutions $\beta_1,\beta_2,\cdots,\beta_d$, which are actually the eigenvectors of $S_t^{-1}S_b$ corresponding to $d(d\leq C-1)$ largest eigenvalues.

From Eq. (17), we know that for a given β, there exists at least one α satisfying $\alpha=\mathrm{P}\Lambda^{1/2}\beta$. Thus, after determining $\beta_1,\beta_2,\cdots,\beta_d$, we can obtain a set of optimal solutions $\alpha_j=\mathrm{P}\Lambda^{1/2}\beta_j\left(j=1,...,d\right)$ with respect to the criterion in Eq. (13). Thereby, the optimal discriminant vectors with respect to the Fisher criterion in feature space are

$$\varphi_j=\mathrm{Q}\alpha_j=\mathrm{QP}\Lambda^{1/2}\beta_j,\qquad j=1,...,d. \tag{19}$$

Up to now, however, the KFD algorithm described above is somewhat complicated. According to the study in Ref. [18], Yang revealed the essence of KFD and proposed a concise framework: KPCA plus LDA. In Yang's opinion, Eq. (19) can be divided into two steps as follows.

Step1. KPCA transformation from feature space H into Euclidean space R^m, i.e.,

$$\begin{aligned} y&=\left(\frac{\gamma_1}{\sqrt{\lambda_1}},\cdots,\frac{\gamma_m}{\sqrt{\lambda_m}}\right)^{\mathrm{T}}\left(\Phi(x_1),\cdots,\Phi(x_M)\right)^{\mathrm{T}}\Phi(x)\\ &=\left(\frac{\gamma_1}{\sqrt{\lambda_1}},\cdots,\frac{\gamma_m}{\sqrt{\lambda_m}}\right)^{\mathrm{T}}\left[\mathrm{k}(x_1,x),\cdots,\mathrm{k}(x_M,x)\right] \end{aligned} \tag{20}$$

Step2. LDA transformation in the KPCA transformed space R^m, i.e.,

$$\varphi=G^{\mathrm{T}}y \tag{21}$$

where $G=\left(\beta_1,\cdots,\beta_d\right)$.

V. KENEL FUZZY DUAL DISCRIMINANT ANALYSIS BASED ALGORITHM

In this section, the detailed algorithm of kernel fuzzy dual discriminant analysis is described as follows:

Step1. Use KPCA to transform the input space R^n into an m-dimensional space R^m, where m=rank(R). Pattern x in R^n is transformed to be KPCA-based feature vector y in R^m.

Step2. In KPCA transformed space R^m, work out the $\tilde{S}_w$'s orthogonal eigenvectors $\theta_1,\theta_2,\cdots,\theta_m$, and the first h ones are corresponding to positive eigenvalues. Let $P_1=\left(\theta_1,\theta_2,\cdots,\theta_h\right)$ and $\hat{S}_b=P_1^T\tilde{S}_bP_1$, $\hat{S}_t=P_1^T\tilde{S}_tP_1$, work out orthonormal eigenvectors $\mu_1,\mu_2,\cdots,\mu_h$ ($h\leq C-1$) of $\hat{S}_b$ and $\hat{S}_t$ corresponding to the first h largest eigenvalues. Then, the optimal discriminant vectors derived from the range space of $\tilde{S}_w$ are $\alpha_j=P_1\mu_j\left(j=1,2,\cdots,h\right)$.

Step3. Since $\tilde{S}_w\in M_{n,n}(F)$ and its rank is r, where, $r=\min\left\{n,N-C\right\}$, n is the dimension of samples, N is the total number of training samples, and C is the number of classes. By Theorem 1, we may obtain a dual matrix $\tilde{S}_w^*\in M_{n,n}(F)$, where, the rank of $\tilde{S}_w^*$ is $n-r$. Work out the $\tilde{S}_w^*$'s orthogonal eigenvectors $\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{n-r}$, suppose the first t ones corresponding to the positive eigenvalues. Let $P_2=\left(\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_t\right)$ and $\overline{S}_b=P_2^T\tilde{S}_bP_2$, $\overline{S}_t=P_2^T\tilde{S}_tP_2$, work out the orthogonal eigenvectors $v_1,v_2,\cdots,v_t$ ($t\leq C-1$) of $\overline{S}_b$ and $\overline{S}_t$

corresponding to the first t largest eigenvalues. Then, the optimal discriminant vectors derived from the dual space of $\tilde{S}_w^*$ are $\beta_j = P_2\nu_j\left(j=1,2,\cdots,t\right)$.

Step4. Work out the between-class scatter matrix $\tilde{S}_b$'s orthogonal eigenvectors $\vartheta_1, \vartheta_2, \cdots, \vartheta_n$, suppose the first l ones are corresponding to positive eigenvalues. Let $Q_1 = \left(\vartheta_1, \vartheta_2, \cdots, \vartheta_l\right)$ and $\hat{S}_w = Q_1^T \tilde{S}_w Q_1$, $\hat{S}_t = Q_1^T \tilde{S}_t Q_1$, work out orthonormal eigenvectors $\omega_1, \omega_2, \cdots, \omega_l$ ($l \le C-1$) of $\hat{S}_w$ and $\hat{S}_t$ corresponding to the first l largest eigenvalues. Then, the optimal discriminant vectors derived from the range space of $\tilde{S}_b$ are $\xi_j = Q_1\omega_j\left(j=1,2,\cdots,l\right)$.

Step5. Since $\tilde{S}_b \in M_{n,n}(F)$ and its rank is d, we may also obtain another dual matrix $\tilde{S}_b^* \in M_{n,n}(F)$, where, the rank of $\tilde{S}_b^*$ is $n-d$. Work out the $\tilde{S}_b^*$'s orthogonal eigenvectors $\sigma_1, \sigma_2, \cdots, \sigma_{n-d}$, suppose the first z ones corresponding to the positive eigenvalues. Let $Q_2 = \left(\sigma_1, \sigma_2, \cdots, \sigma_z\right)$ and $\overline{S}_w = Q_2^T \tilde{S}_w Q_2$, $\overline{S}_t = Q_2^T \tilde{S}_t Q_2$, work out the orthogonal eigenvectors $\kappa_1, \kappa_2, \cdots, \kappa_z$ ($z \le C-1$) of $\overline{S}_w$ and $\overline{S}_t$ corresponding to the first z largest eigenvalues. Then, the optimal discriminant vectors derived from the dual space of $\tilde{S}_b^*$ are $\delta_j = Q_2\kappa_j\left(j=1,2,\cdots,z\right)$.

Step6. Let $\alpha_j = P_1\mu_j\left(j=1,\cdots,h\right)$, $\beta_j = P_2\nu_j\left(j=1,\cdots,t\right)$, $\xi_j = Q_1\omega_j\left(j=1,\cdots,l\right)$ and $\delta_j = Q_2\kappa_j\left(j=1,\cdots,z\right)$ act as projection axes to form the feature extractor $\varphi = \left(\alpha_j, \beta_j, \xi_l, \delta_z\right)$.

VI. EXPERIMENTAL RESULTS

The proposed method is used for face recognition and tested on the ORL [32] and Yale [33] face image database. To evaluate the proposed method properly, we also list experimental results for the kernel C-LDA (KC-LDA) algorithm. For its simplicity, the k nearest neighbor (k-NN) [34] classifier with Euclidean distance is employed for the classification. The parameter of k-NN is fixed as $k=3$.

A. On ORL Database

The ORL contains a set of faces taken between April 1992 and April 1994 at the Olivetti Research Laboratory in Cambridge. It contains 40 distinct persons with 10 images per subject. The images were taken at different time instances, with varying lighting conditions, facial expressions, and facial details. All persons are in the up-right, frontal position, with tolerance for some side movement. In this experiment, each image is normalized and presented by a 23×28 pixel array whose gray levels ranged between 0 and 255. Some sample images from the ORL database are shown in Figure 1.

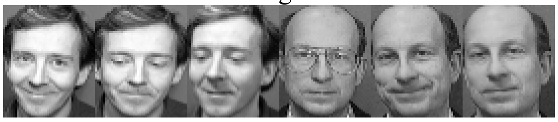

Figure 1. Some sample images from the ORL face image database

We randomly choose $\theta(\theta = 3,4,5)$ images per individual for training, and the rest are used for testing. To make full use of the available data and to evaluate the generalization power of algorithms more accurately, ten experiments are performed. The final result is the average recognition rate over the ten random training sets. In this experiment, two popular kernel functions are employed respectively. One is the second-order Polynomial kernel function $\mathrm{k}(x,y) = (x\cdot y+1)^2$ and the other is Sigmoid function $\mathrm{k}(x,y) = \tanh\left[q(x\cdot y)+\Theta\right]$, where, the parameters of Sigmoid function are chosen by cross-validation technique as $q=0.01$ and $\Theta = 4$.

Table1 shows the average recognition accuracies of KC-LDA and the proposed method under a varying number of the training samples per individual on the ORL face image database. As shown in Table 2, it is therefore reasonable to believe that the proposed method is the most effective one no matter what kind of kernel function is employed.

TABLE II.

COMPARISON OF RECOGNITION RATES ON THE ORL DATABASE IN THE KERNEL SPACE

Methods/Kernel function	Number of training samples/ CPU Time (s)		
	3	4	5
KC-LDA (Polynomial)	89.29/9.27	93.33/9.73	95.54/9.81
Proposed method (Polynomial)	90.23/17.59	94.35/17.71	96.24/17.82
KC-LDA (Sigmoid)	88.93/9.39	93.33/9.56	95.51/9.67
Proposed method (Sigmoid)	90.64/17.84	94.67/17.89	96.36/17.93

B. On Yale Database

The Yale face image database contains 165 grayscale images of 15 individuals. There are 11 images per subject, one per different facial expression or configuration. We manually crop the facial portion of each face image. The each cropped face is resized to 40×50 pixels. Some sample images from the Yale database are shown in Figure 2.

Figure 2. Some sample images from the Yale face image database

We randomly choose the former 5 images per individual for training, and the rest images are used for testing. Similarly, ten experiments were performed to obtain the average recognition rate. In the experiment, the $\mathrm{k}(x,y)=\tanh\left[q(x\cdot y)+\Theta\right]$ is employed, where, the parameters of the Sigmoid function are chosen by cross-validation technique as $q=0.01$ and $\Theta=4$.

Table 2 presents the recognition accuracies of Fisherface [19], D-LDA [22], C-LDA [17] and the proposed method. For all methods, the corresponding dimensionality of the reduced subspace is also given in Table 3.

TABLE III.

THE RECOGNITION RATES (%) AND CORRESPONDING DIMENSIONALITY ON THE YALE DATABASE

Results/ Methods	Fisherface	D-LDA	C-LDA	Proposed method
Accuracy	97.28	97.78	98.42	99.07
Dimensionality	14	14	28	56
CPU Time (s)	3.82	3.09	5.53	9.24

Again, the recognition accuracy of each method listed in Table 3 indicates that the proposed method is still the most effective one among the other traditional approaches. However, it is worth stressing that the proposed method needs more CPU time for whole process (training and testing) because it costs more computation using 56 features for classification

VII. CONCLUSIONS

In this paper, we present a novel kernel fuzzy dual discriminant analysis learning based on a rank symmetric relationship to accomplish the mission of feature extraction and recognition. In particular, it is worth stressing that the method which is developed in the kernel fuzzy feature extraction approach revealed more robust characteristics as far as the relationship between the potential subspaces of scatter matrices and the novel mechanism of fuzzy sets is concerned. The reason why the presented method yields a better performance can be attributed to the fact that the proposed kernel fuzzy dual subspace learning can efficiently manage the vagueness and ambiguity of different face subspaces being degraded by poor illumination component.

ACKNOWLEDGMENT

This work is supported by the Natural Science Foundation of the Jiangsu Higher Education Institutions of China(Grant no. 10KJB520006).

REFERENCES

[1] T. Shakunaga, K. Shigenari, “Decomposed eigenface for face recognition under various lighting conditions,” in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 864–871, 2001.

[2] J.Tenenbaum, W. Freeman, “Separating style and content with bilinear models,” Neural Comput., vol.12, no. 6, pp. 1247–1283, 2000.

[3] M. Turk, A. Pentland, “Face recognition using eigenfaces,” in: Proceeding of IEEE Conference on Computer Vision and Pattern Recognition, Maui, HI, USA, pp. 586–591, June 1991.

[4] N. Kwak, J. Oh, “Feature extraction for one-class classification problems: Enhancements to biased discriminant analysis,” Pattern Recognition, vol. 42, pp. 17–16, 2009.

[5] Z.Z. Liang, Y.F. Li, P.F. Shi, “A note on two-dimensional linear discriminant analysis,” Pattern Recognition Letters, vol. 29, pp. 2122–2128, 2008.

[6] L. Rueda, M. Herrera, “Linear dimensionality reduction by maximizing the Chernoff distance in the transformed space,” Pattern Recognition, vol. 41, pp. 3138–3152, 2008.

[7] Y.Yan, Y.J. Zhang, “A novel class-dependence feature analysis method for face recognition,” Pattern Recognition Letters, vol. 29, pp. 1907–1914, 2008.

[8] Q.B.You, N.N. Zheng, S.Y. Du, Y. Wu, “Neighborhood discriminant projection for face recognition,” Pattern Recognition Letters, vol. 28, pp. 1156–1163, 2007.

[9] J. Yu, Q. Tian, T. Rui, T.S. Huang, “Integrating discriminant and descriptive information for dimension reduction and classification,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 17, no. 3, pp. 372–377, 2007.

[10] K. Fukunaga, “Introduction to Statistical Pattern Recognition,” second ed., Academic Press, Boston, MA, 1990.

[11] R.P.W. Duin, “Small sample size generalization,” in: Proceedings of Ninth Scandinavian Conference on Image Analysis, pp. 957–964, June 1995.

[12] S.W. Ji, J.P. Ye, “Generalized linear discriminant analysis: A unified framework and efficient model selection,” IEEE Transactions on Neural Networks, vol. 19, no. 10, pp. 1768–1782, 2008.

[13] J.W. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, “Regularized discriminant analysis for the small sample size problem in face recognition,” Pattern Recognition Letters, vol. 24, no. 16, pp. 3079–3087, 2003.

[14] J.W. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, “Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition,” Pattern Recognition Letters, vol. 26, no. 2, pp. 181–191, 2005.

[15] P. Howland, H. Park, “Generalizing discriminant analysis using the generalized singular value decomposition,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp. 995–1006, 2004.

[16] P. Howland, J.L. Wang, H. Park, “Solving the small sample size problem in face recognition using generalized discriminant analysis,” Pattern Recognition, vol. 39, no. 2, pp. 277–287, 2006.

[17] J. Yang, J.Y. Yang, “Why can LDA be performed in PCA transformed space,” Pattern Recognition, vol.36, no.2, pp. 563–566, 2003.

[18] J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, Z, “Jin. KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol.27, no.2, pp.230–244, 2005.

[19] P.N. Belhumeur, J.P. Hespanha, D.J. Kriegman, “Eigenfaces vs. Fisherfaces: recognition using class specific linear projection,” IEEE Trans. Pattern Anal. Mach. Intell., vol.19, no.7, pp.711–720, 1997.

[20] C. Liu, H. Wechsler, “A shape- and texture-based enhanced Fisher classifier for face recognition,” IEEE

Transactions on Image Process, vol. 10, no. 4, pp. 598–608, 2001.
[21] L.F. Chen, H.Y. Liao, J.C. Lin, M.T. Ko, G.J. 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.
[22] H. Yu, J. Yang, “A direct LDA algorithm for high-dimensional data—with application to face recognition,” Pattern Recognition, vol.34, no.10, pp. 2067–2070, 2001.
[23] M.Skurichina, R.P.W. Duin, “Bagging, boosting, and the random subspace method for linear classifiers,” Pattern Analysis and Applications, vol. 5, pp. 121–135, 2002.
[24] X. Wang, X.Tang, “Subspace analysis using random mixture models,” in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, pp. 574–580, June 2005.
[25] X.X. Zhang, Y.D. Jia, “A linear discriminant analysis framework based on random subspace for face recognition,” Pattern Recognition, vol.40, pp.2585–2591, 2007.
[26] S.C. Yan, D. Xu, Q. Yang, X.O. Tang, H.J. Zhang, “Multilinear discriminant analysis for face recognition,” IEEE Transactions on Image Processing, vol.16, no.1, pp.212–220, 2007.
[27] K.C. Kwak, W. Pedrycz, “Face recognition using a fuzzy fisherface classifier,” Pattern Recognition, vol. 38, no. 10, pp. 1717–1732, 2005.
[28] X.N. Song, Y.J. Zheng, X.J. Wu, X.B. Yang, J.Y. Yang, “A complete fuzzy discriminant analysis approach for face recognition,” Applied Soft Computing, vol. 10, pp. 208-214, 2010.
[29] J.M. Keller, M.R. Gray, J.A. Givens, “A fuzzy k-nearest neighbor algorithm,” IEEE Transactions on Systems, Man and Cybernetics, vol. 15, no. 4, pp. 580–585, 1985.
[30] L.A. Zadeh, “Fuzzy sets,” Info. Control, vol. 8, pp. 338–353, 1965.
[31] J.C. Bezdek, J. Keller, R. Krishnapuram, Fuzzy Models and Algorithms for Pattern Recognition and Image Processing, Kluwer Academic Publishers, Dordrecht, 1999.
[32] The ORL face database at the AT&T (Olivetti) research laboratory. Available from: <http://www.uk.research.att.com/facedatabase.html>.
[33] The Yale database. Available From: <Http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
[34] N.García-Pedrajas, D. Ortiz-Boyer, “Boosting k-nearest neighbor classifier by means of input space projection,” Expert Systems with Applications, vol. 36, pp.10570–10582, 2009.

Minlun Yan was born in October 1966, he received his B.S. degree in Mathematics and Information Technology from Soochow University, Suzhou, PR China, in 1988. Since 1989, he has been teaching in the Department of Mathematics, Lianyungang Teachers’ College, Lianyungang, PR China, where he is a vice professor. Now, he is also a visiting scholar in Soochow University on the subject of applied mathematics and information technology. His current research interests include applied mathematics, feature analysis.