

Dual Adaptive K-SVD Algorithm Based on a Rank Symmetrical Relationship

Minlun Yan

Department of Mathematics, Lianyungang Teachers' College, Lianyungang 222006, PR China

Abstract—Applications that use sparse representation are many and include compression, regularization in inverse problems, feature extraction, and more. Recent activity in this field has concentrated mainly on the study of pursuit algorithms that decompose signals with respect to a given dictionary. The K-SVD algorithm is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the data. However, the existing K-SVD algorithm is employed to a single feature space meaning that the pursuit algorithms are assigned to the given subspace definitely. The work proposed in this paper provides a novel adaptive way to adapting dictionaries in order to achieve the dual subspace sparse signal representations, the update of the dictionary is combined with a rank symmetrical relationship of the proposed dual subspace by incorporated a new mechanism of matrix transform, which is called dual K-SVD. Experimental results conducted on the ORL and Yale face databases demonstrate the effectiveness of the proposed method.

Index Terms—subspace learning; sparse representation; K-SVD; 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 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 approaches 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 methods, 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 [22] proposed a direct-LDA (D-LDA) method, which takes the range space of the between-class scatter matrix as the intermediate subspace. 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 [25] proposed a dual principal random discriminant analysis (RDA) algorithm, which combines the advantages of Fisherface and D-LDA. As discussed 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 discriminant information in the high dimensional facial space since some potential subspaces are complementary in terms of the discriminative power. Recently, Song [26] presented a novel fuzzy supervised learning method with dynamical parameter estimation for discriminant analysis. By this means, it dynamical achieves the distribution information of each sample of images that represented with fuzzy membership degree.

Moreover, sparse signal representations using over-complete dictionaries are used in a variety of fields such as pattern recognition, image and video coding [27-28]. Over-completeness of a set means that it has more members than the dimensionality of the members. Given an over-complete set of basis signals, called the dictionary, the goal is to express input signals as sparse linear combinations of the dictionary members. The advantage of over-completeness of a dictionary is its robustness in case of noisy or degraded signals. Also, it introduces greater variety of shapes in the dictionary, thus leading to sparser representations of a variety of input signals. Extraction of the sparsest representation is a hard problem that has been extensively investigated in the past few years. The K-SVD algorithm is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the data. However, the existing K-SVD algorithm [29-30] is employed to a

single feature space meaning that the pursuit algorithms are assigned to the given subspace definitely.

In this paper, the objective of our study is to establish a dual adaptive K-SVD algorithm, which is based on a rank symmetrical relationship to solve supervised dimensionality reduction problem by unfolding the feature vectors along different projection directions. The work proposed in this paper provides a novel adaptive way to adapting dictionaries in order to achieve the dual subspace sparse signal representations, the update of the dictionary is combined with a rank symmetrical relationship of the proposed dual subspace by incorporated a new mechanism of matrix transform, which is called dual K-SVD (DK-SVD). Therefore, the DK-SVD approach has the potential to outperform the traditional feature extraction algorithms, especially in the cases of 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 dimensional 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 $A \in M_{m,n}(F)$ 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 $A \in M_{m,n}(F)$ 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, $P \in M_{m,n}(F)$, $Q \in M_{n,n}(F)$, E_r is identity matrix, I_r is equivalent standard form of matrix A .

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

Obviously, $\text{rank}(B_{n-r}^*) = n - r$

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

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

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

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 . \square

As analyzed above, by Theorem 1, we may further deduce the dual subspace learning with respect to the Fisher discriminant analysis.

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 . 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 | $O(M^3)$ | $O(C^3)$ | $O(2M^3)$ | $O(M^2d)$ | $O(2M^3)$ |

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 proposed algorithm can be more effective than other ones for classification, it needs more CPU time for whole process because it costs computation using more feature vectors.

III. THE K-SVD ALGORITHM

A different update rule for the dictionary can be proposed, in which the atoms (i.e., columns) in dictionary are handled sequentially. This leads to the K-SVD algorithm, as developed by Aharon et al [29-30]. Keeping all the columns fixed apart from the j_0 -th one, a_{j_0} , this column can be updated along with the coefficients that multiply it in X . We isolate the dependency on a_{j_0} by rewriting minimization as

$$\|Y - AX\|_F^2 = \left\| Y - \sum_{j=1}^m a_j x_j^T \right\|_F^2 = \left\| \left(Y - \sum_{j \neq j_0} a_j x_j^T \right) - a_{j_0} x_{j_0}^T \right\|_F^2 \quad (1)$$

In this description, x_j^T stands for the j -th row of X . The update step targets both a_{j_0} and $x_{j_0}^T$, and refers to the term in parentheses,

$$E_{j_0} = Y - \sum_{j \neq j_0} a_j x_j^T \quad (2)$$

as a known pre-computed error matrix.

The optimal a_{j_0} and $x_{j_0}^T$ minimizing Equation (1) are the rank-1 approximation of E_{j_0} , and can be obtained via an SVD, but this typically would yield a dense vector $x_{j_0}^T$, implying that we increase the number of non-zeros in the representations in X .

In order to minimize this term while keeping the cardinalities of all the representations fixed, a subset of the columns of E_{j_0} should be taken – those that correspond to the signals from the example-set that are using the j_0 -th atom, namely those columns where the entries in the row $x_{j_0}^T$ are non-zero. This way, we allow only the existing non-zero coefficients in $x_{j_0}^T$ to vary, and the cardinalities are preserved. Therefore, we define a restriction operator, P_{j_0} , that multiplies E_{j_0} from the right to remove the non-relevant columns. The matrix P_{j_0} has M rows (the number of overall examples), and M_{j_0} columns (the number of examples using the j_0 -th atom). We define $(x_{j_0}^R)^T = x_{j_0}^T P_{j_0}$ as the restriction on the row $x_{j_0}^T$, choosing the non-zero entries only.

For the sub-matrix, E_{j_0} , P_{j_0} , a rank-1 approximation via SVD can be applied, updating both the atom a_{j_0} and the corresponding coefficients in the sparse representations, $x_{j_0}^R$. This simultaneous update may lead to a substantial speedup in the convergence of the training algorithm. The K-SVD algorithm is described in detail as follows.

K-SVD Task: Train a dictionary A to sparsely represent the data $\{y_i\}_{i=1}^M$, by approximating the solution to the optimization problem.

Step1. Initialization: Initialize $k = 0$, and initialize dictionary: Build $A_0 = R^{m \times n}$, either by using random entries, or using m randomly chosen examples;

Step2. Normalization: Normalize the columns of A_0 .

Main Iteration: Increment k by 1, and apply

Step3. Sparse Coding Stage: Use a pursuit algorithm to approximate the solution of

$$\hat{x}_i = \arg \min_x \|y_i - A_{k-1} x\|_2^2 \text{ subject to } \|x\|_0 \leq k_0$$

obtaining sparse representations \hat{x}_i for $1 \leq i \leq M$. These form the matrix X_k .

Step4. K-SVD Dictionary-Update Stage: Use the following procedure to update the columns of the dictionary and obtain A_k : Repeat for $j_0 = 1, 2, \dots, m$

Define the group of examples that use the atom a_{j_0} ,

$$\Omega_{j_0} = \{i | 1 \leq i \leq M, X_k[j_0, i] \neq 0\}$$

Compute the residual matrix

$$E_{j_0} = Y - \sum_{j \neq j_0} a_j x_j^T,$$

where x_j^T are the j 'th rows in the matrix X_k .

Restrict E_{j_0} by choosing only the columns corresponding to Ω_{j_0} , and obtain $E_{j_0}^R$.

Apply SVD decomposition $E_{j_0}^R = U \Delta V^T$. Update the dictionary atom $a_{j_0} = u_1$, and the representations by $x_{j_0}^R = \Delta[1, 1] \cdot v_1$.

Step5. Stopping Rule: If the change in $\|Y - A_k X_k\|_F^2$ is small enough, stop. Otherwise, apply another iteration.
Output: The desired result is A_k .

IV. DUAL ADAPTIVE K-SVD ALGORITHM

In this section, the detailed algorithm of dual adaptive K-SVD is described as follows:

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

Step2. In PCA transformed space R^m , work out the \tilde{S}_w 's orthogonal eigenvectors $\theta_1, \theta_2, \dots, \theta_m$, and the first h ones are corresponding to positive eigenvalues. Let $P_1 = (\theta_1, \theta_2, \dots, \theta_h)$ and $\hat{S}_b = P_1^T \tilde{S}_b P_1$, $\hat{S}_t = P_1^T \tilde{S}_t P_1$, work out orthonormal eigenvectors $\mu_1, \mu_2, \dots, \mu_h$ ($h \leq C-1$) 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 $\alpha_j = P_1 \mu_j$ ($j=1, 2, \dots, h$).

Step3. Since $\tilde{S}_w \in M_{n,n}(F)$ and its rank is r , where, $r = \min\{n, N-C\}$, 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 $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_{n-r}$, suppose the first t ones corresponding

to the positive eigenvalues. Let $P_2 = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_t)$ and $\bar{S}_b = P_2^T \tilde{S}_b P_2$, $\bar{S}_t = P_2^T \tilde{S}_t P_2$, work out the orthogonal eigenvectors v_1, v_2, \dots, v_t ($t \leq C-1$) of \bar{S}_b and \bar{S}_t corresponding to the first t largest eigenvalues. Then, the optimal discriminant vectors derived from the dual space of \tilde{S}_w^* are $\beta_j = P_2 v_j$ ($j=1, 2, \dots, t$).

Step4. Work out the between-class scatter matrix \tilde{S}_b 's orthogonal eigenvectors $\vartheta_1, \vartheta_2, \dots, \vartheta_n$, suppose the first l ones are corresponding to positive eigenvalues. Let $Q_1 = (\vartheta_1, \vartheta_2, \dots, \vartheta_l)$ 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 $\omega_1, \omega_2, \dots, \omega_l$ ($l \leq C-1$) 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 $\xi_j = Q_1 \omega_j$ ($j=1, 2, \dots, l$).

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 $\sigma_1, \sigma_2, \dots, \sigma_{n-d}$, suppose the first z ones corresponding to the positive eigenvalues. Let $Q_2 = (\sigma_1, \sigma_2, \dots, \sigma_z)$ and $\bar{S}_w = Q_2^T \tilde{S}_w Q_2$, $\bar{S}_t = Q_2^T \tilde{S}_t Q_2$, work out the orthogonal eigenvectors $\kappa_1, \kappa_2, \dots, \kappa_z$ ($z \leq C-1$) of \bar{S}_w and \bar{S}_t corresponding to the first z largest eigenvalues. Then, the optimal discriminant vectors derived from the dual space of \tilde{S}_b^* are $\delta_j = Q_2 \kappa_j$ ($j=1, 2, \dots, z$).

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

Step7. Apply K-SVD algorithm to update the dictionary atom of feature extractor φ , and the corresponding coefficients in the dual sparse representations can be obtained.

V. EXPERIMENTAL RESULTS

The proposed method is used for face recognition and tested on the ORL [31] and Yale [32] face image database. To evaluate the proposed method properly, we also include experimental result for the D-LDA [22], C-LDA [17], LPP [33], NPE [34]. For its simplicity, the k nearest neighbor (k -NN) [35] 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 upright, 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 θ ($\theta = 3, 4, 5$) images per individual for training, and the rest images are used for testing. To make full use of the available data and to evaluate the generalization power of algorithms more accurately, ten experiments were performed. The final result was the average recognition rate over the ten random training sets.

Table I shows the average recognition accuracies of D-LDA, C-LDA, LPP, NPE 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 | Number of training samples | | |
|-----------------|----------------------------|-------|-------|
| | 3 | 4 | 5 |
| D-LDA | 88.19 | 90.46 | 92.81 |
| C-LDA | 88.78 | 91.72 | 94.23 |
| LPP | 89.21 | 92.02 | 94.46 |
| NPE | 88.69 | 91.62 | 94.11 |
| Proposed method | 90.58 | 93.21 | 95.19 |

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. Table 2 presents the recognition accuracies of D-LDA, C-LDA, LPP, NPE 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

| Methods | D-LDA | C-LDA | LPP | NPE | Proposed method |
|--------------|-------|-------|-------|-------|-----------------|
| Accuracy | 97.78 | 98.42 | 98.09 | 97.91 | 98.97 |
| CPU Time (s) | 3.09 | 5.53 | 4.28 | 4.47 | 5.80 |

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 (i.e. dual space for K-SVD) because it costs more computation for classification.

VI. CONCLUSIONS

In this paper, we presented a novel dual K-SVD algorithm based on a rank symmetrical relationship to accomplish the mission of feature extraction and recognition. In particular, it is worth stressing that the method which is developed in the feature extraction approach revealed more robust characteristics as far as the relationship between the potential subspaces of scatter matrices and the novel mechanism of rank symmetrical relationship is concerned. The reason why the presented method yields a better performance can be attributed to the fact that the proposed DK-SVD can efficiently manage the sparse representations of different face subspaces being degraded by poor illumination component.

ACKNOWLEDGMENT

This work is supported by the Natural Science Foundation of Jiangsu Province (Grant no. BK2012700), the Natural Science Foundation of the Jiangsu Higher Education Institutions of China (Grant no. 10KJB520006), the China Postdoctoral Science Foundation (Grant no. 2011M500926), the Jiangsu Province Postdoctoral Science Foundation (Grant no. 1102063C) and the Foundation of Artificial Intelligence Key Laboratory of Sichuan Province (Grant no. 2012RZY02).

REFERENCES

- [1] T. Shakhnazarov, 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] B. Yang, Y.Y. Bu, "A Comparative Study on Vector-based and Matrix-based Linear Discriminant Analysis," *Journal of Computer*, vol. 6, no. 4, pp. 818-824, 2011.
- [5] D.Y. Zhang, B.Q. Liu, C.J. Sun, X.L. Wang, "Learning the Classifier Combination for Image Classification," *Journal of Computer*, vol. 6, no. 8, pp. 1756-1763, 2011.
- [6] Z.Q. Wang, X. Sun, "Optimal Kernel Marginal Fisher Analysis for Face Recognition," *Journal of Computer*, vol. 7, no. 9, pp. 2298-2305, 2012.
- [7] X.N. Song, J.Y. Yang, X.J. Wu, X.B. Yang, "An Optimal Symmetrical Null Space Criterion of Fisher Discriminant for Feature Extraction and Recognition," *Soft Computing*, vol. 15, pp. 281-293, 2011.
- [8] 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.
- [9] X.N. Song, X.B. Yang, J.Y. Yang, X.J. Wu, Y.J. Zheng, "Discriminant Analysis Approach Using Fuzzy Fourfold Subspaces Model," *Neurocomputing*, vol. 73, pp. 2255-2265, 2010.
- [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] X.N. Song, X.B. Yang, Z. Liu, X.N. Ding. "A novel fuzzy supervised learning method with dynamical parameter estimation for discriminant analysis", In: *9th International Conference on Fuzzy Systems and Knowledge Discovery*, 2012.
- [27] R. Mazhar, P.D. Gader, J.N. Wilson, "A Matching Pursuit Based Similarity Measure for Fuzzy Clustering and Classification of Signals," In: *International Conference on Fuzzy Systems*, Hong Kong, 2008.
- [28] A.E. Moghadam, S. Shirani, "Matching Pursuit-Based Region-of-Interest Image Coding," *IEEE Trans. Image Processing*, vol. 16, no.10, pp. 406-415, 2007.
- [29] M. Aharon, M. Elad, A. Bruckstein, "K-SVD: An Algorithm for Designing Overcomplete Dictionaries for Sparse Representation," *IEEE Trans. Signal Processing*, vol.54, no.11, pp. 4311-4322, 2006.
- [30] M. Elad. "The Quest for a Dictionary", *Sparse and Redundant Representations*, Springer New York, pp. 227-246, 2010.
- [31] The ORL face database at the AT&T (Olivetti) research laboratory. Available from: <<http://www.uk.research.att.com/facedatabase.html>>.
- [32] The Yale database. Available From: <[Http://cvc.yale.edu/projects/yalefaces/yalefaces.html](http://cvc.yale.edu/projects/yalefaces/yalefaces.html)>.
- [33] X. He, S. Yan, Y. Hu, P. Niyogi, H. Zhang, "Face recognition using laplacianfaces," *IEEE Transaction on pattern analysis and machine intelligence*, vol.27, pp.328-40, 2005.
- [34] Q.B. You, N.N. Zheng, S.Y. Du, Y. Wu, "Neighborhood Discriminant Projection for Face Recognition," *J. Pattern Recognition Letters*, vol.28, pp.1156-1163, 2007.
- [35] 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.