A Robust Algorithm for Generalized Orthonormal Discriminant Vectors

Wenming Zheng [1,2], Xiaoou Tang [2]
[1] Research Center for Learning Science, Southeast University, Nanjing, P.R. China
[2] Microsoft Research Asia, Beijing, P.R. China
wenming_zheng@seu.edu.cn; xitang@microsoft.com

Abstract

In this paper, we propose a robust and efficient algorithm for generalized orthonormal discriminant vectors (GODV). The major advantage of the proposed method is the use of the rank-one update technique, rather than the Lagrange multipliers method, to iteratively derive the formula of computing the discriminant vectors of GODV. By contrast with the previous algorithms of GODV, the proposed algorithm has the computational efficiency and the numerical stability because of the avoidance of solving the inverse computation of matrices. Moreover, the proposed algorithm can be easily extended to tackle the nonlinear problem via kernel trick. The performance of the proposed algorithm is tested on the Yale face database and ORL face database, respectively.

1. Introduction

Linear discriminant analysis (LDA) is one of the most popular feature extraction methods in pattern recognition [9][6]. It realizes an optimal set of discriminant vectors that maximize the between-class scatter distance and at the same time minimize the within-class scatter distance of the training data set. However, the classical LDA method is usually suffer from the rank limitation, which means that the available number of the discriminant vectors of LDA generally depends on the rank of the between-class scatter matrix. This problem occurs especially when the number of the data classes is small.

To overcome the rank limitation of LDA, Foley and Sammon firstly proposed the optimal orthonormal set of discriminant vectors (FSODV) in 1975 for two classes problem [2]. In 1985, Okada and Tomita extended Foley's method to multi-class problems, where they proposed a new algorithm based on an orthogonal subspace approach to solve the optimal discriminant vectors [10]. In 1988, Duchene and Leclercq proposed another algorithm based on Larange multipliers method for the same problem [11]. All the above three algorithms aim to overcome the rank limitation of LDA. However, Liu et al. pointed out in literature [1] that it is difficult to guarantee the property of maximum between-class scatter and minimum within-class scatter using the traditional Foley-Sammon orthogonal discriminant vectors, which is very desirable for designing a better classifier. Therefore, they proposed a generalized orthonormal set of discriminant vectors (GODV) to overcome this drawback of the FSODV method. The experiments in [1] showed that the GODV is superior to FSODV in terms of classification rates. However, Liu's algorithm may suffer from bosh computational and numerical problems because their algorithm involves many matrices computations including the inversion of matrices. To overcome this problem, we propose a new algorithm for GODV by using the rank one update technique. Moreover, consider that GODA is still a linear method, which is not appropriate to use for many nonlinear recognition problems, we had presented the kernel version of GODV (KGODV) in [4]. However, similar with the previous GODV algorithm, our algorithm still suffers from both computational and numerically instable problems. To overcome these drawbacks, in this paper we also propose another robust algorithm for KGODV based on the rank-one update technique and the singular value decomposition (SVD) approach.

2. Review of GODV

Let $\mathbf{X}=\{\mathbf{x}_i^j \mid j=1,\cdots,N_i; i=1,\cdots,c\}$ be an n-dimensional sample set with N elements, where $n >> N$, c is the number of the total classes and N_i is the number of the samples in ith class. The between-class scatter matrix $\mathbf{S}_B$ and the total-scatter matrix $\mathbf{S}_T$ are respectively defined as follows:

$$\mathbf{S}_B=\sum_{i=1}^{c} N_i(\mathbf{u}_i-\mathbf{u})(\mathbf{u}_i-\mathbf{u})^T \tag{1}$$

$$\mathbf{S}_T=\sum_{i=1}^{c}\sum_{j=1}^{N_i}(\mathbf{x}_i^j-\mathbf{u})(\mathbf{x}_i^j-\mathbf{u})^T \tag{2}$$

where $\mathbf{x}^T$ represents the transpose of the vector $\mathbf{x}$, $\mathbf{x}_i^j$ the jth sample in the ith class, $\mathbf{u}_i$ the mean of the ith class samples, and $\mathbf{u}$ the mean of all samples:

$$\mathbf{u}_i=\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbf{x}_i^j, \quad \mathbf{u}=\frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N_i}\mathbf{x}_i^j \tag{3}$$

Define the Fisher criterion $J_F(\boldsymbol{\omega})$ by

$$J_F(\boldsymbol{\omega})=\frac{\boldsymbol{\omega}^T\mathbf{S}_B\boldsymbol{\omega}}{\boldsymbol{\omega}^T\mathbf{S}_T\boldsymbol{\omega}} \tag{4}$$

The 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Then, according to literature [1], the optimal orthonormal vectors of GODV can be obtained step by step using the following procedures:

(1) *The first orthonormal vector, denoted by* $\boldsymbol{\omega}_1$, *is the unit vector that maximizes* $J_F(\boldsymbol{\omega})$;

(2) *Suppose that we have obtain the first* i *optimal orthonormal vectors of GODV, then the* $(i+1)$ *th orthonormal vector, denoted by* $\boldsymbol{\omega}_{i+1}$, *is defined as the unit vector that maximizes the new criterion* $J_{i+1}(\boldsymbol{\omega})$ *under the following orthogonal constraints:*

$$\boldsymbol{\omega}_{i+1}^T \boldsymbol{\omega}_j = 0 \ (j = 1,2,\cdots,i) \tag{5}$$

where $J_{i+1}(\boldsymbol{\omega})$ *is defined as:*

$$J_{i+1}(\boldsymbol{\omega}) = \frac{\sum_{j=1}^{i} \boldsymbol{\omega}_j^T \mathbf{S}_B \boldsymbol{\omega}_j \big/ (\boldsymbol{\omega}_j^T \boldsymbol{\omega}_j) + \boldsymbol{\omega}^T \mathbf{S}_B \boldsymbol{\omega} \big/ \|\boldsymbol{\omega}\|^2}{\sum_{j=1}^{i} \boldsymbol{\omega}_j^T \mathbf{S}_T \boldsymbol{\omega}_j \big/ (\boldsymbol{\omega}_j^T \boldsymbol{\omega}_j) + \boldsymbol{\omega}^T \mathbf{S}_T \boldsymbol{\omega} \big/ \|\boldsymbol{\omega}\|^2} = \frac{\boldsymbol{\omega}^T (a_i \mathbf{I} + \mathbf{S}_B) \boldsymbol{\omega}}{\boldsymbol{\omega}^T (b_i \mathbf{I} + \mathbf{S}_T) \boldsymbol{\omega}} \tag{6}$$

where $a_i = \sum_{j=1}^{i} \boldsymbol{\omega}_j^T \mathbf{S}_B \boldsymbol{\omega}_j \big/ (\boldsymbol{\omega}_j^T \boldsymbol{\omega}_j)$, $b_i = \sum_{j=1}^{i} \boldsymbol{\omega}_j^T \mathbf{S}_T \boldsymbol{\omega}_j \big/ (\boldsymbol{\omega}_j^T \boldsymbol{\omega}_j)$, and $\mathbf{I}$ is the n by n identity matrix.

To solve the optimal discriminant vectors of GODV, Liu et al. proposed the following algorithm [1]:

(1) *The first optimal vector* $\boldsymbol{\omega}_1$ *is the eigenvector of the matrix* $\mathbf{S}_T^{-1}\mathbf{S}_B$ *corresponding to the largest eigenvalue;*

(2) *Suppose that we have obtained the first* i *discriminant vectors of GODV, denoted by* $\boldsymbol{\omega}_1,\cdots,\boldsymbol{\omega}_i$, *then the* $(i+1)$ *th optimal discriminant vector can be calculated by using the following procedures:*

Suppose that $\mathbf{V}_i = span\{\boldsymbol{\omega}_1,\cdots,\boldsymbol{\omega}_i\}$ *be the subspace spanned by the first* i *discriminant vectors of GODV. Construct the complementary subspace* $\overline{\mathbf{V}}_i$ of $\mathbf{V}_i$. Let $\overline{\mathbf{V}}_i = span\{\boldsymbol{\varphi}_1,\cdots,\boldsymbol{\varphi}_{n-i}\}$, where $\boldsymbol{\varphi}_1,\cdots,\boldsymbol{\varphi}_{n-i}$ are orthonormal vectors. Let $\mathbf{P}_i = [\boldsymbol{\varphi}_1,\cdots,\boldsymbol{\varphi}_{n-i}]$, and let $\boldsymbol{\xi}_i$ be *the eigenvector of* $(a_i\mathbf{I} + \mathbf{P}_i^T\mathbf{S}_T\mathbf{P}_i)^{-1}(b_i\mathbf{I} + \mathbf{P}_i^T\mathbf{S}_B\mathbf{P}_i)$ *corresponding to the largest eigenvalue. Then the* $(i+1)$ *th optimal discriminant vector of GODV can be written as:* $\boldsymbol{\omega}_{i+1} = \boldsymbol{\xi}_{i+1} \big/ \|\boldsymbol{\xi}_{i+1}\|$.

From the above algorithm, we can see that Liu's algorithm may suffer from the computational problem and the numerical stability problem. This is because their algorithm needs to compute the inverse of matrices in each step, and also needs to construct an orthonormal basis of a new subspace in each step.

3. Robust Algorithm for GODV

In this section, we will propose another algorithm for GODV, which is computationally more efficient and numerically more stable than Liu's algorithm. Moreover, our proposed GODV method can be easily extended to the nonlinear version of GODV via the kernel trick [3][8].

It is easy to know that any optimal discriminant vector of GODV, denote by $\boldsymbol{\omega}$, can be written as the span of the columns of $\mathbf{I}$. For simplicity of our derivation, we make the following notations

$$\mathbf{I} = \mathbf{I}_{(1)}, \quad \mathbf{S}_B^{(1)} = \mathbf{S}_B, \quad \mathbf{S}_T^{(1)} = \mathbf{S}_T$$

Then, the first discriminant vector $\boldsymbol{\omega}_1$ can be written as the following form:

$$\boldsymbol{\omega}_1 = \mathbf{I}_{(1)}\boldsymbol{\alpha}_1 \tag{7}$$

where $\boldsymbol{\alpha}_1$ is an n dimensional vector.

Suppose that we have obtained the ith orthonormal discriminant vector $\boldsymbol{\omega}_i$. Then the $(i+1)$th orthonormal discriminant vector $\boldsymbol{\omega}_{i+1}$ lies in the subspace spanned by the columns of the matrix $\mathbf{I}_{(i+1)}$, where $\mathbf{I}_{(i+1)}$ is obtained using the following rank one update:

$$\mathbf{I}_{(i+1)} = \mathbf{I}_{(i)} - \boldsymbol{\omega}_i\boldsymbol{\omega}_i^T\mathbf{I}_{(i)} \tag{8}$$

Thus, finding $\boldsymbol{\omega}_{i+1}$ is equivalent to finding the coefficient vector $\boldsymbol{\alpha}_{i+1}$ such that

$$\boldsymbol{\omega}_{i+1} = \mathbf{I}_{(i+1)}\boldsymbol{\alpha}_{i+1} \tag{9}$$

In this case, the criterion $J_{i+1}(\boldsymbol{\omega})$ can be rewritten as:

$$J_{i+1}(\boldsymbol{\alpha}) = \frac{\boldsymbol{\alpha}^T\mathbf{I}_{(i+1)}^T(a_i\mathbf{I} + \mathbf{S}_B)\mathbf{I}_{(i+1)}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T\mathbf{I}_{(i+1)}^T(a_i\mathbf{I} + \mathbf{S}_T)\mathbf{I}_{(i+1)}\boldsymbol{\alpha}} = \frac{\boldsymbol{\alpha}^T\mathbf{B}_{(i+1)}\boldsymbol{\alpha}}{\boldsymbol{\alpha}^T\mathbf{W}_{(i+1)}\boldsymbol{\alpha}} \tag{10}$$

where $a_i = \sum_{j=1}^{i} \boldsymbol{\omega}_j^T\mathbf{S}_B\boldsymbol{\omega}_j$, $b_i = \sum_{j=1}^{i} \boldsymbol{\omega}_j^T\mathbf{S}_T\boldsymbol{\omega}_j$, and

$$\mathbf{B}_{(i+1)} = \mathbf{S}_B^{(i+1)} + a_i\Delta_{(i+1)} \tag{11}$$

$$\mathbf{W}_{(i+1)} = \mathbf{S}_T^{(i+1)} + b_i\Delta_{(i+1)} \tag{12}$$

where

$$\mathbf{S}_B^{(i+1)} = \mathbf{S}_B^{(i)} - \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_B^{(i)}) - \mathbf{S}_B^{(i)}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T + \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_B^{(i)}\boldsymbol{\omega}_i)\boldsymbol{\omega}_i^T \tag{13}$$

$$\mathbf{S}_T^{(i+1)} = \mathbf{S}_T^{(i)} - \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_T^{(i)}) - \mathbf{S}_T^{(i)}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T + \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_T^{(i)}\boldsymbol{\omega}_i)\boldsymbol{\omega}_i^T \tag{14}$$

$$\Delta_{(i+1)} = \Delta_{(i)} - (\mathbf{I}_{(i)}^T\boldsymbol{\omega}_i)(\boldsymbol{\omega}_i^T\mathbf{I}_{(i)}) \ \text{ where } \ \Delta_{(1)} = \mathbf{I} \tag{15}$$

Suppose that $\boldsymbol{\omega}_1,\cdots,\boldsymbol{\omega}_K$ are the optimal orthonormal discriminant vectors of GODV. According to the above analysis, we have the following algorithm for solving the optimal discriminant vectors of GODV, where $a_1 = b_1 = 0$, $\mathbf{B}_{(1)} = \mathbf{S}_B^{(1)}$, and $\mathbf{W}_{(1)} = \mathbf{S}_T^{(1)}$:

GODV Algorithm: *Repeat for* $i = 1,\cdots,K$

1. *Let* $\mathbf{W}_{(i)} = \mathbf{U}_{(i)}\mathbf{D}_{(i)}\mathbf{U}_{(i)}^T$ *be the singular value decomposition (SVD) of* $\mathbf{W}_{(i)}$ and let $\widetilde{\mathbf{B}}_{(i)} = \mathbf{D}_{(i)}^{-1/2}\mathbf{U}_{(i)}^T\mathbf{B}_{(i)}\mathbf{U}_{(i)}\mathbf{D}_{(i)}^{-1/2}$;

2. *Compute the eigenvector, denoted by* $\boldsymbol{\alpha}$, *corresponding to the largest eigenvalue of the symmetric matrix* $\widetilde{\mathbf{B}}_{(i)}$;
3. *Let* $\boldsymbol{\omega}_i = \mathbf{I}_{(i)}\mathbf{U}_{(i)}\mathbf{D}_{(i)}^{-1/2}\boldsymbol{\alpha}$, *and let* $\boldsymbol{\omega}_i = \boldsymbol{\omega}_i / \|\boldsymbol{\omega}_i\|$;
4. *Update the following matrices:*

$a_i = a_{i-1} + \boldsymbol{\omega}_i^T \mathbf{S}_B \boldsymbol{\omega}_i$;

$b_i = b_{i-1} + \boldsymbol{\omega}_i^T \mathbf{S}_T \boldsymbol{\omega}_i$;

$$\mathbf{S}_B^{(i+1)} = \mathbf{S}_B^{(i)} - \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_B^{(i)}) - \mathbf{S}_B^{(i)}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T + \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_B^{(i)}\boldsymbol{\omega}_i)\boldsymbol{\omega}_i^T$$

$$\mathbf{S}_T^{(i+1)} = \mathbf{S}_T^{(i)} - \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_T^{(i)}) - \mathbf{S}_T^{(i)}\boldsymbol{\omega}_i\boldsymbol{\omega}_i^T + \boldsymbol{\omega}_i(\boldsymbol{\omega}_i^T\mathbf{S}_T^{(i)}\boldsymbol{\omega}_i)\boldsymbol{\omega}_i^T$$

$$\Delta_{(i+1)} = \Delta_{(i)} - (\mathbf{I}_{(i)}^T\boldsymbol{\omega}_i)(\boldsymbol{\omega}_i^T\mathbf{I}_{(i)})$$

$\mathbf{B}_{(i+1)} = \mathbf{S}_B^{(i+1)} + a_i\Delta_{(i+1)}$;

$\mathbf{W}_{(i+1)} = \mathbf{S}_T^{(i+1)} + b_i\Delta_{(i+1)}$;

After performing the above procedures, we obtain that the projection matrix of GODV, denoted by $\mathbf{W}_{GODV}$, where

$$\mathbf{W}_{GODV} = [\boldsymbol{\omega}_1, \cdots, \boldsymbol{\omega}_K] \tag{16}$$

Then the projection of a test point $\mathbf{x}_{test}$ onto $\mathbf{W}_{GODV}$ can be computed by

$$\mathbf{z}_{test} = \mathbf{W}_{GODV}^T \mathbf{x}_{test} \tag{17}$$

The classification is obtained by minimizing the distance

$$d = \left\|\mathbf{z}_i^j - \mathbf{z}_{test}\right\|^2 \tag{18}$$

where $\mathbf{z}_i^j$ are the projections of $\mathbf{x}_i^j$ onto $\mathbf{W}_{GODV}$.

4. Robust GODV Algorithm with Kernel

Let $\mathbf{X}$ is mapped into a Hilbert space F through a nonlinear mapping kernel function Φ, i.e., $\Phi: \mathbf{X} \to F$, $\mathbf{x} \to \Phi(\mathbf{x})$. Then $\mathbf{S}_B$ and $\mathbf{S}_T$ in F can be respectively rewritten as:

$$\mathbf{S}_B^{\Phi} = \sum_{i=1}^{c} N_i(\mathbf{u}_i^{\Phi} - \mathbf{u}^{\Phi})(\mathbf{u}_i^{\Phi} - \mathbf{u}^{\Phi})^T \tag{19}$$

$$\mathbf{S}_T^{\Phi} = \sum_{i=1}^{c}\sum_{j=1}^{N_i}(\Phi(\mathbf{x}_i^j) - \mathbf{u}^{\Phi})(\Phi(\mathbf{x}_i^j) - \mathbf{u}^{\Phi})^T \tag{20}$$

$$\mathbf{u}_i^{\Phi} = \frac{1}{N_i}\sum_{j=1}^{N_i}\Phi(\mathbf{x}_i^j), \quad \mathbf{u}^{\Phi} = \frac{1}{N}\sum_{i=1}^{c}\sum_{j=1}^{N_i}\Phi(\mathbf{x}_i^j) \tag{21}$$

The Fisher discriminant analysis can be expressed as:

$$J_F(\boldsymbol{\omega}) = \frac{\boldsymbol{\omega}^T\mathbf{S}_B^{\Phi}\boldsymbol{\omega}}{\boldsymbol{\omega}^T\mathbf{S}_T^{\Phi}\boldsymbol{\omega}} \tag{22}$$

And the expression of $J_{i+1}(\boldsymbol{\omega})$ in equation (6) can be rewritten by

$$J_{i+1}(\boldsymbol{\omega}) = \frac{\sum_{j=1}^{i}\boldsymbol{\omega}_j^T\mathbf{S}_B^{\Phi}\boldsymbol{\omega}_j / (\boldsymbol{\omega}_j^T\boldsymbol{\omega}_j) + \boldsymbol{\omega}^T\mathbf{S}_B^{\Phi}\boldsymbol{\omega} / \|\boldsymbol{\omega}\|^2}{\sum_{j=1}^{i}\boldsymbol{\omega}_j^T\mathbf{S}_T^{\Phi}\boldsymbol{\omega}_j / (\boldsymbol{\omega}_j^T\boldsymbol{\omega}_j) + \boldsymbol{\omega}^T\mathbf{S}_T^{\Phi}\boldsymbol{\omega} / \|\boldsymbol{\omega}\|^2} \tag{23}$$

To solve the optimal discriminant vector of KGODV, we firstly divide the whole data space into two orthogonal subspaces, i.e., the null space of $\mathbf{S}_T^{\Phi}$, denoted by $\mathbf{S}_T^{\Phi}(0)$, and its complementary subspace, denoted by $\overline{\mathbf{S}_T^{\Phi}(0)}$. It has been proven that the subspace $\mathbf{S}_T^{\Phi}(0)$ carries no discriminant information [3], thus we only limit our attention to the subspace $\overline{\mathbf{S}_T^{\Phi}(0)}$ to find the optimal discriminant vectors of KGODV.

We use the kernel principal component analysis (KPCA) [8] to remove the null space of $\mathbf{S}_T^{\Phi}$. Let $\mathbf{W}_{KPCA}$ be the transform matrix of KPCA, and let $\mathbf{y}_i^j$ be the projections of $\mathbf{x}_i^j$ onto $\mathbf{W}_{KPCA}$, i.e.,

$$\mathbf{y}_i^j = \mathbf{W}_{KPCA}^T\mathbf{x}_i^j \tag{24}$$

Then the between-class scatter matrix in the projection space, denoted by $\widetilde{\mathbf{S}}_B$, and the total scatter matrix in the projection space, denoted by $\widetilde{\mathbf{S}}_T$, in the projection space can be respectively expressed as

$$\widetilde{\mathbf{S}}_B = \sum_{i=1}^{c} N_i(\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{u}})(\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{u}})^T \tag{25}$$

$$\widetilde{\mathbf{S}}_T = \sum_{i=1}^{c}\sum_{j=1}^{N_i}(\mathbf{y}_i^j - \widetilde{\mathbf{u}})(\mathbf{y}_i^j - \widetilde{\mathbf{u}})^T \tag{26}$$

where $\widetilde{\mathbf{u}}_i$ the mean of the ith class samples and $\widetilde{\mathbf{u}}$ the mean of all samples. Thus, our goal turns to solve the GODV problem of the projection sample set $\{\mathbf{y}_i^j\}$, which can be easily implemented according to section 3. We omit the derivation of this algorithm due to the space limitation.

5. Experiments

We will test the proposed algorithm on the Yale face database and ORL face database. The nearest neighbor is used over the experiments.

5.1 Experiment on Yale face database

The Yale face database contains 15 subjects, each of which contains 11 face images with variations in both facial expression and lighting condition. The original face images are sized 243×320 pixels with a 256-level gray scale. Figure 1 shows all the 11 images of one subject. To reduce the background part of each full image, each image is manually centered and cropped into a size of 190×170 pixels, and then down-sampled into the size of 95×85. Moreover, note that the normal facial expression image and the without glasses image (or with glasses if subject normally wears glasses) in the face subjects numbered 2, 3, 6, 7, 8, 12 and 14 are copies of each other (see Figure 1). We remove the normal facial expression image. Consider that the size of this database is relatively small, we adopt the "leave-one-out" cross validation strategy to test the proposed method. Moreover, for the comparison purpose, we also perform the same experiments using the PCA method, the Fisherfaces method, the direct LDA (D-LDA) method [5], and the null space method [13]. Table 1 shows the experimental results of various systems. From Table 1, we can see that the GODV achieve best recognition rate among the various methods.

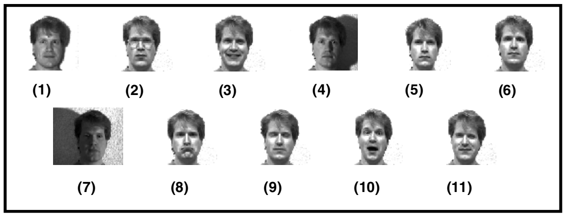

Figure 1 Ten images for one subject in Yale face database

Table 1 Comparison of error rate on Yale face database

Methods	Reduced Space	Error Rate (%)
PCA	40	23.33 (35/150)
D-LDA	14	11.33 (17/150)
Fisherfaces	14	4.00 (6/150)
Null Space	14	4.00 (6/150)
GODV	50	3.33 (5/150)

5.2 Experiment on ORL face database

In this example, we use the ORL face database to test the performance of the proposed algorithm. The ORL database contains 40 distinct subjects, where each one contains 10 different images taken at different times, varying lighting slightly. All the images are taken against a dark homogeneous background and the persons are in upright, frontal position, with tolerance for some tilting and rotation. The original face images are all sized 112×92 pixels with a 256-level gray scale. We use the two-fold cross validation to test the performance of the various systems. Table 2 shows the experimental results. It can be clearly seen from Table 2 that the proposed method still achieves the best recognition result.

Table 2 Comparison of error rate on ORL face database

Methods	Reduced Space	Error Rate (%)
PCA	80	9.75
D-LDA	39	10.50
Fisherfaces	39	14.00
Null Space	39	8.50
GODV	80	8.25

6. Discussions and Conclusions

We have presented a robust algorithm for GODV in this paper. By using the rank one update technique, we can greatly save the computation of update the matrices of the Fisher criterion. On the other hand, by only using the SVD operator rather than computing the inverse of matrices, our algorithm will be more numerically stable than the previous algorithms. Moreover, by using the kernel trick, our algorithm can be easily extended to handle the nonlinear version of GODV, i.e., the KGODV algorithm. The experiments on face recognition also show the better performance of the proposed algorithm in terms of recognition rates.

Acknowledges

This work was partly supported by the national natural science foundations of China (60503023), and partly by the natural science foundations of Jiangsu province (BK2005407), and partly by Program for New Century Excellent Talents in University (NCET-05-0467).

References

[1] K. Liu, Y.Q. Cheng, J.Y. Yang, "A Generalized Optimal Set of Discriminant Vectors", Pattern Recognition, 25(7), pp.731-739, 1992.

[2] D.H., Foley, J.W. Sammon, "An optimal set of discriminant vectors", IEEE Transactions on Computers, Vol. C-24, No.3, pp.139-144, 1985.

[3] W. Zheng, L. Zhao, and C. Zou, "Foley-Sammon optimal discriminant vectors using kernel approach", IEEE Transactions on Neural Networks, Vol.16, No.1, pp.1-9, 2005.

[4] W. Zheng, L. Zhao, and C. Zou, "Generalized Optimal Set of Discriminant Vectors Using Kernel Approach", Proceedings of JCIS CVPRIP2003, Cary, North Carolina, Sept. 26-30, pp.777-780, 2003.

[5] H. Yu and J. Yang, "A direct LDA algorithm for high-dimensional data - with application to face recognition", Pattern Recognition, 34, pp. 2067-2070, 2001.

[6] R.O. Duda, & P.E. Hart, "Pattern Classification and Scene Analysis", New York: John Wiley & Sons, Inc., 1973.

[7] Peter N. Belhumeur, Joao P. Hespanha, and David J.Kriegman, "Eigenfaces VS. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Trans. On PAMI, Vol.19, No.7, pp.711-720 1997.

[8] B. Schölkopf, A.J. Smola, and K.-R. Müller, "Nonlinear component analysis as a kernel eigenvalue problem", Neural Computation, Vol.10, pp.1299-1319, 1998.

[9] R.A. Fisher, "The use of multiple measurements in taxonomic problems", Annu. Eugenics, Vol.7, pp.179-188, 1936.

[10] T. Okada, S. Tomita, "An optimal orthonormal system for discriminant analysis", Pattern Recognition, Vol.18, No.2, pp.139-144, 1985.

[11] J. Duchene, S. Leclercq, "An optimal transformation for discriminant and principal component analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.10, pp.978-983, 1988.

[12] P.N. Belhumer, J. Hespanda, and D. Kiregeman, "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.19, No.7, pp.711-720, 1997.

[13] L.F. Chen, H.Y.M. Liao, M.T. Ko, J.C. Lin, and G.J. Yu, "A New LDA-based face recognition system which can solve the small sample size problem", Pattern Recognition, Vol.33, pp.1713-1726, 2000.