ORIGINAL ARTICLE

Global–local fisher discriminant approach for face recognition

Qianqian Wang · Xiaolei Hu · Quanxue Gao ·
Bing Li · Yong Wang

Abstract   In this paper, we proposed a linear discriminant approach, namely global–local Fisher discriminant analysis (GLFDA) that explicitly considers both the local and global discriminant structures embedded in data. To be specific, GLFDA constructs two graphs to, respectively, model the global and local discriminant structures and then incorporates discriminant structures and local intrinsic structure, which characterizes the within-class compactness, into the objective function for dimensionality reduction. Thus, GLFDA well encodes the discriminant information, especially the local discriminant information of data. Experimental results on AR, YALE, and UMIST databases show the effectiveness of the proposed algorithm.

Keywords   Manifold learning · Discriminant analysis · Global and local geometrical structures · Face recognition

1 Introduction

Many previous works have demonstrated that manifold learning can effectively improve the image recognition accuracy [1–6]. Manifold learning technique constructs an adjacency graph over the training data to model the geometrical relationship of data and maps nearby data points in the high-dimensional data space to nearby data in the reduced space. It intuitively preserves the intrinsic structure of data and has been widely used in image recognition, clustering, image retrieval, and data classification. Locality Preserving Projection (LPP) [1, 7] is one of the prevalent approaches and has proven its effectiveness in face recognition, document clustering, and image retrieval. However, LPP neglects the label of samples when learning the intrinsic structure of data. This may impair the recognition accuracy.

Motivated by LPP and linear discriminant analysis (LDA) [8], many discriminant approaches based on manifold learning have been developed for linear dimensionality reduction [9–14], among which the representative approaches are Marginal Fisher Analysis (MFA) [2], Local Discriminant Embedding (LDE) [9], Local Fisher Discriminant Analysis (LFDA) [11, 12]. MFA, LFDA, and LDE represent the intraclass compactness by LPP [7] which maps nearby data points in the observed data space to nearby data with the low-dimensional representation. Different from MFA and LDE, which represent the between-class separability by maximizing the sum of distances among nearby data with different labels and characterize the local discriminant structure of data, LFDA represents the interclass separability by maximizing the sum of distances of data from different classes and detects the global discriminant structure of data. Moreover, the data pairs with large distance will dominate the global discriminant structure of data. This may impair the local discriminant structure.

Although the motivations of the above-mentioned discriminant approaches are different, they can be unified within the Graph Embedding Framework (GEF) that was proposed by Yan et al. [2]. GEF constructs two adjacency graphs, namely intrinsic graph and penalty graph, to model the intrinsic geometric structure and discriminant

Q. Wang · X. Hu · Q. Gao (✉) · B. Li
State Key Laboratory of Integrated Services Networks,
Xidian University, Xi'an 710071, China
e-mail: qxgao@xidian.edu.cn

Y. Wang
School of Electronic Engineering, Xidian University,
Xi'an 710071, China

geometric structure of the data, respectively. However, GEF only considers the local or global discriminant structure of data and does not simultaneously consider them for dimensionality reduction. In real cases, the intrinsic structure of data is unknown and complex, and testing data are usually different from the training data due to many factors. Thus, only local or global discriminant structure of data is not sufficient to guarantee the recognition accuracy [3, 15–17]. It indicates that the recognition accuracy of above-mentioned discriminant approaches is not good enough on testing data.

In this paper, motivated by [17], we propose a novel discriminant approach, namely global–local Fisher discriminate analysis (GLFDA), for dimensionality reduction. To be specific, GLFDA constructs two adjacency graphs to model the local intrinsic geometric structure and discriminant geometric structure of data, respectively, and then integrates the global discriminant structure that is learned by LDA, local discriminant structure and intrinsic structure into the objective function of dimensionality reduction. Experimental results on AR, YALE, and UMIST databases demonstrate the effectiveness of the proposed approach.

The rest of the paper is organized as follows. Section 2 introduces our GLFDA approach. Section 3 provides some experimental results. We conclude the paper in Sect. 4.

2 Global–local fisher discriminant analysis

2.1 Motivation

According to above-mentioned analysis, graph embedding framework only encodes the local or global discriminating information, which will reduce the flexible of the algorithm and impair the recognition accuracy. In this section, we learn a novel dimensionality reduction approach, namely GLFDA, which explicitly considers both the global and local discriminating information of data. To be specific, we construct three adjacency graphs, namely intrinsic graph, global penalty graph, and local penalty graph, to model the intrinsic and discriminant geometrical structures of the data manifold. The global penalty graph encodes the global discriminating information of data and maps the large distance pairs from different classes to be faraway in the reduced space. The local penalty graph encodes the local discriminating information and maps nearby points from different classes to be faraway in the reduced space. Thus, the discriminant structure of data can be effectively characterized by local and global penalty graph. The intrinsic graph maps nearby data points from the same class to nearby points with the low-dimensional representation. In the following section, we first describe the intrinsic graph of data.

2.2 Intrinsic graph

Suppose, we have a set of p-dimensional samples $X = \{x_1, x_2, ..., x_n\}$, belonging to C classes. We construct an adjacency graph $G_w = \{X, W\}$, namely intrinsic graph, with a vertex set X and weight matrix W to learn the similarity of data from the same class, which characterizes the intrinsic structure. The elements W_{ij} in W are defined as follows:

$$W_{ij} = \begin{cases} A_{ij}/n_c, & \text{if } \tau_i = \tau_j = c; \\ 0, & \text{if } \tau_i \neq \tau_j \end{cases} \quad (1)$$

where the elements A_{ij} are 1 if x_j is the k_1-nearest neighbor of x_i or vice versa; otherwise $A_{ij} = 0$, n_c denotes the number of samples in class c, τ_i and τ_j denote the class label of data x_i and x_j, respectively.

Intrinsic graph aims to map nearby data points from the same class in the observed space to nearby data in the reduced space. As it happens, a reasonable criterion for choosing a good map is to optimize the following objective function

$$\min \sum_{ij} (y_i - y_j)^2 W_{ij} \quad (2)$$

where $y_i = \alpha^T x_i$ denotes the one-dimensional representation of x_i, α is the transformation vector.

The objective function (2) with our choice of symmetric weight matrix W_{ij} incurs a heavy penalty, if nearby points x_i and x_j, which are close or very close to each other and share the same label, are mapped far apart. Therefore, minimizing Eq. (2) is an attempt to ensure that if x_i and x_j are close, then y_i and y_j are also close. Thus, the intrinsic structure of the data manifold, which characterizes the similarity of data, is well preserved by Eq. (2).

2.3 Local penalty graph

Likewise, we construct an adjacency graph $G_B = \{X, B\}$, namely local penalty graph, with a vertex set X and weight matrix B to encode the discriminating information embedded in nearby data from different classes. The elements B_{ij} in B can be defined as follows:

$$B_{ij} = \begin{cases} \dfrac{1}{n_c} & \text{if } x_i \in N_{k_2}(x_j) \text{ or } x_j \in N_{k_2}(x_i) \text{ and } \tau_i \neq \tau_j, \\ 0 & \text{else} \end{cases} \quad (3)$$

where $N_{k_2}(x_i)$ denotes the set of k_2 nearest neighbors of x_i.

Local penalty aims to find a map so that the connected data in the adjacency graph G_B stay as distant as possible. Suppose y_i is a map of training data x_i ($i = 1, 2, ..., n$), a reasonable map is to optimize the following objective function:

$$\max \quad \sum_{ij} (y_i - y_j)^2 B_{ij} \tag{4}$$

The objective function (4) on the graph G_B incurs a heavy penalty if nearby two points x_i and x_j are mapped close together while they actually belong to different classes. So, maximizing Eq. (4) is an attempt to ensure that if x_i and x_j are close but have different labels, then y_i and y_j are also far apart. Thus, we can well detect the local discriminant geometrical structure of data by Eq. (4) which encodes the local discriminating information.

Obviously, the objective function (4) does not explicitly consider the data points which are not in the local neighborhood and have different labels. In other words, these data points may be mapped to be close to each other in the reduced space. This impairs the global discriminant structure of data, resulting in not better recognition accuracy for the algorithm. In Sect. 2.4, we construct a global graph, namely global penalty graph, to address this problem.

2.4 Global penalty graph

Motivated by Fisher discriminant analysis, we construct a global graph, namely global penalty graph, $G_g = \{X, F\}$ with a vertex set X and weight matrix F to model the global discriminant structure that encodes the global discriminating information of data. The elements F_{ij} in F can be defined as follows:

$$F_{ij} = \begin{cases} 1/n, & \text{if } \tau_i \neq \tau_j \\ 0, & \text{Otherwise} \end{cases} \tag{5}$$

The global penalty graph seeks to find a map so that the data points, which have different labels, can be mapped to be faraway in the reduced space. Suppose y_i is a map of training data x_i ($i = 1, 2, \ldots, n$), a reasonable map is to optimize the following objective function:

$$\max \quad \sum_{ij} (y_i - y_j)^2 F_{ij} \tag{6}$$

The objective function (6) on the graph G_g incurs a heavy penalty if two points x_i and x_j are mapped close together while they actually belong to different classes. So, maximizing Eq. (6) is an attempt to ensure that if x_i and x_j have different labels, then y_i and y_j are far apart. Thus, we can well detect the global discriminant geometrical structure of data by Eq. (6).

Moreover, the objective function (6) emphasizes the large distance pairs and guarantees that the larger the distance between two data points is, the farther they are embedded in the reduced space. It means that Eq. (6) maps data points, which are in the un-local neighborhoods and have different labels, to be still distant in the reduced space. This effectively avoids the disadvantage caused by Eq. (4).

2.5 Objective function

In this section, we describe the GLFDA approach that solves the objective functions (2), (4), and (6). Suppose α is a projection vector, substituting $y_i = \alpha^T x_i$ into the objective function (2), and following some simple algebraic steps, we can see that

$$\begin{aligned}
\sum_{ij} (y_i - y_j)^2 W_{ij} &= \sum_{ij} (\alpha^T x_i - \alpha^T x_j)^2 W_{ij} \\
&= 2\alpha^T \left(\sum_j x_i x_i^T W_{ii} - \sum_{ij} x_i x_j^T W_{ij} \right) \alpha \\
&= 2\alpha^T X (D_w - W) X^T \alpha \\
&= 2\alpha^T X L_w X^T \alpha
\end{aligned} \tag{7}$$

where $L_w = D_w - W$, D_w is a diagonal matrix whose entries are column (or row, since W is symmetric) sum of W, i.e., $D_w(i, i) = \sum_j W_{ij}$.

Substituting $y_i = \alpha^T x_i$ into the objective function (4), we see that

$$\begin{aligned}
\sum_{ij} (y_i - y_j)^2 B_{ij} &= \sum_{ij} (\alpha^T x_i - \alpha^T x_j)^2 B_{ij} \\
&= 2\alpha^T \left(\sum_j x_i x_i^T B_{ii} - \sum_{ij} x_i x_j^T B_{ij} \right) \alpha \\
&= 2\alpha^T X (D_b - B) X^T \alpha \\
&= 2\alpha^T X L_b X^T \alpha
\end{aligned} \tag{8}$$

where $L_b = D_b - B$, D_b is a diagonal matrix whose entries are column (or row, since B is symmetric) sum of B, i.e., $D_b(i, i) = \sum_j B_{ij}$.

Likewise, substituting $y_i = \alpha^T x_i$ into the objective function (6), we see that

$$\begin{aligned}
\sum_{ij} (y_i - y_j)^2 F_{ij} &= \sum_{ij} (\alpha^T x_i - \alpha^T x_j)^2 F_{ij} \\
&= 2\alpha^T \left(\sum_j x_i x_i^T F_{ii} - \sum_{ij} x_i x_j^T F_{ij} \right) \alpha \\
&= 2\alpha^T X (D_f - F) X^T \alpha \\
&= 2\alpha^T X L_f X^T \alpha
\end{aligned} \tag{9}$$

where $L_f = D_f - F$, D_f is a diagonal matrix whose entries are column (or row, since F is symmetric) sum of F, i.e., $D_f(i, i) = \sum_j F_{ij}$.

Substituting Eqs. (7), (8), and (9) into Eqs. (2), (4), and (6), respectively, we have several ways to build the objective function of GLFDA by integrating Eqs. (2), (4), and (6). The representative one is the Rayleigh quotients form that is formally similar to LDA [8]. Thus, we can write the objective function of GLFDA as follows:

Fig. 1 Some images of one subject from the Yale database

Table 1 The top recognition accuracies (%) of the six approaches and the corresponding number of features on the Yale database (shown in parentheses)

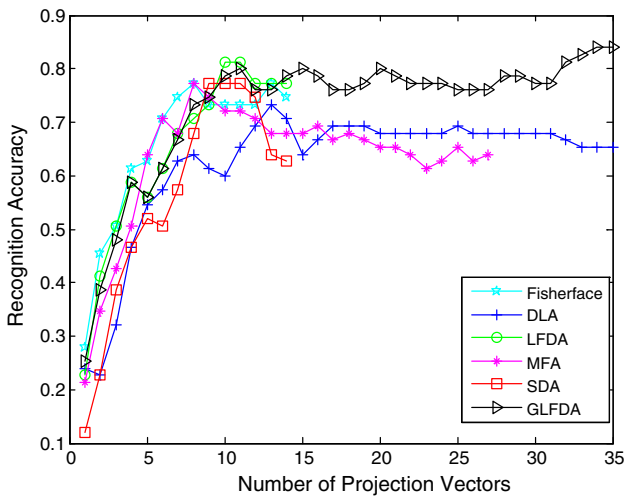| Training/ testing numbers	Fisherface	DLA	LFDA	MFA	SDA	GLFDA
45/120	61.67	66.67	70.00	65.83	65.00	72.50
	(6)	(15)	(14)	(8)	(14)	(13)
90/75	77.33	73.33	81.33	77.33	77.33	84.00
	(8)	(13)	(10)	(8)	(9)	(34)
135/30	93.33	96.67	96.67	96.67	96.67	100.00
	(12)	(13)	(12)	(18)	(11)	(57)

Fig. 2 The recognition accuracy versus number of projection vector on the YALE database

$$\arg\max_{\alpha} \frac{\alpha^T \left(S_b \cdot t + S_f(1-t)\right)\alpha}{\alpha^T S_w \alpha} \qquad (10)$$

where $t \in [0, 1]$ is a parameter and balances the global and local discriminant information of data. $S_b = XL_bX^T$ is the global between-class scatter matrix, $S_f = XL_fX^T$ is the local between-class scatter matrix, and $S_w = XL_wX^T$ is the local within-class scatter matrix.

The optimal projection vector α that maximizes Eq. (10) is given by the maximum eigenvalue solution to the generalized eigenvalue problem:

$$\left(t \cdot S_b + (1-t)S_f\right)\alpha = \lambda S_w \alpha \qquad (11)$$

Let the column vector $\alpha_1, \alpha_2, ..., \alpha_d$ be the solution of (11), ordered according to its eigenvalues, $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_d > 0$, then the projection matrix P can be written as $P = [\alpha_1, \alpha_2, ..., \alpha_d]$.

Note that, in face recognition, S_w is singular, and the optimal projection matrix is not directly calculated from Eq. (11). In this case, many approaches have been developed to solve the optimal projection vector α. In the experiments, we simply choose the same way as in Fisherface [8] for its simplicity, i.e., PCA is first used to reduce dimension such that XL_wX is nonsingular, then solving the Eq. (11) in the PCA space.

2.6 Algorithm

Given training data set $X = [x_1 x_2 ... x_n]$ including C classes. The algorithmic procedure of GLFDA is summarized as follows.

2.6.1 Construct intrinsic and penalty graphs

Suppose that $G_w = \{X, W\}$, $G_B = \{X, B\}$, and $G_g = \{X, F\}$ denote local intrinsic graph, local penalty graph, and global penalty graph, respectively. The weighted matrices W, B, and F are calculated by Eqs. (1), (3), and (5), respectively.

2.6.2 Calculate scatter matrices

In this step, we calculate scatter matrices S_w, S_b, and S_f by Eqs. (7), (8), and (9), respectively.

2.6.3 PCA projection

In this step, we project the data points x_i $(i = 1, ..., n)$ into the PCA subspace. Suppose that the projection matrix of PCA is $W_{PCA} \in R^{p \times l}$, whose columns consist of the l leading eigenvectors of the covariance matrix of data; the scatter matrixes S_w, S_b, and S_f become $\tilde{S}_w = W_{PCA}^T XL_wX^T W_{PCA}$, $\tilde{S}_b = W_{PCA}^T XL_bX^T W_{PCA}$, and $\tilde{S}_f = W_{PCA}^T XL_fX^T W_{PCA}$, respectively, after PCA projection.
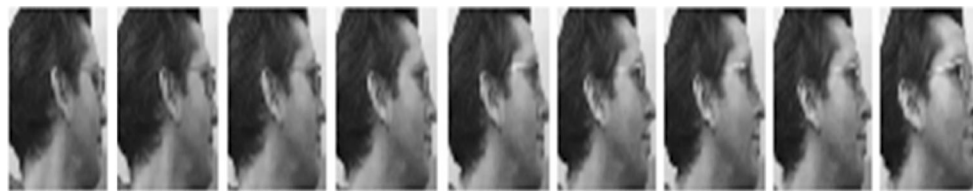
Fig. 3 Some sample images of one subject in the UMIST database

Table 2 The top classification accuracies (%) of the six approaches and the corresponding number of dimensions on the UMIST database (shown in parentheses)

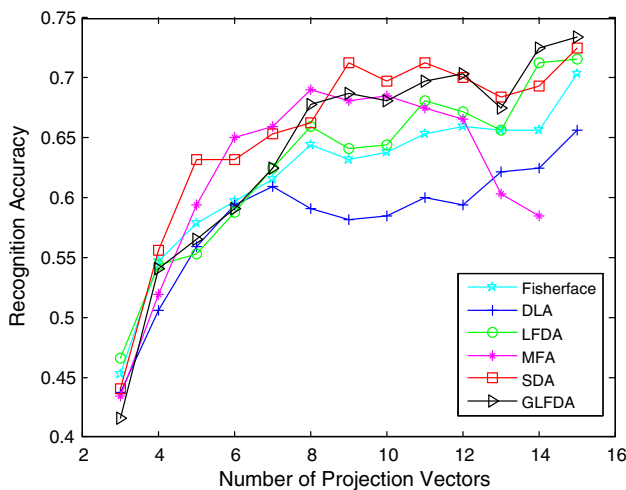| Training/ testing numbers	Fisherface	DLA	LFDA	MFA	SDA	GLFDA
45/120	70.31	65.63	71.56	69.06	72.50	73.44
	(15)	(15)	(15)	(8)	(15)	(15)
90/75	85.38	73.46	87.31	84.23	85.77	87.69
	(9)	(7)	(11)	(13)	(8)	(11)
135/30	89.50	77.00	90.50	90.50	90.00	92.00
	(7)	(13)	(7)	(8)	(7)	(7)

Fig. 4 The recognition accuracy versus number of projection vector on the UMIST database

2.7 Compute the projection matrix

Solve the following generalized eigenvector problem:

$$\left(t \cdot \tilde{S}_b + (1-t)\tilde{S}_f\right)\alpha = \lambda \tilde{S}_w \alpha$$

Let $\alpha_1, \alpha_2, \ldots, \alpha_d$ be the solutions of the above equation, ordered according to their eigenvalues, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d$. Thus, the projection matrix of GLFDA is $P_S = [\alpha_1 \alpha_2 \ldots \alpha_d]$.

2.7.1 Linear embedding

In this step, we compute the linear projections. Given an arbitrary training vector x_i, the linear embedding is as follows:

$$x_i \rightarrow y_i = P^T x_i$$

$$P = W_{\text{PCA}} P_S$$

3 Experiments

In this section, we evaluate the proposed approach on three well-known face databases (YALE, UMIST, and AR), and compare its performance with Fisherface [8], MFA [2], SDA [17], DLA [18], and LFDA [11]. In the experiments, we use Euclidean metric and nearest classifier for classification due to the simplicity. In real cases, it is very difficult to select a suitable neighbor parameter k for MFA, LFDA, DLA, and GLFDA approaches due to the complex distribution of training data. Note that the same problem is often encountered in many manifold learning approaches. In experiments, we empirically determined a proper neighbor parameter k within the interval $[1, n-1]$ for all approaches, where n denotes the number of training samples. The parameter t depends on the distribution of data. In our approaches, we determined t as 0.5, 0.1, and 0.4 for YALE, UMIST, and AR databases, respectively.

The Yale face database [19] contains 165 images of 15 individuals (each person providing 11 different images) under various facial expressions and lighting conditions. In our experiments, each image is manually cropped and resized to 32×32 pixels. Figure 1 shows some images of one person. In the experiments, the first 3, 6, and 9 images per person are selected for training, respectively, and the corresponding rest images for testing. Thus, we have three training sets that include 45, 90, and 135 images, and the corresponding three testing sets including 120, 75, and 30 images. The top recognition accuracies of the six approaches are shown in Table 1. Figure 2 plots the curve of recognition accuracy versus number of projection vectors when the first 6 images per person are selected for training.
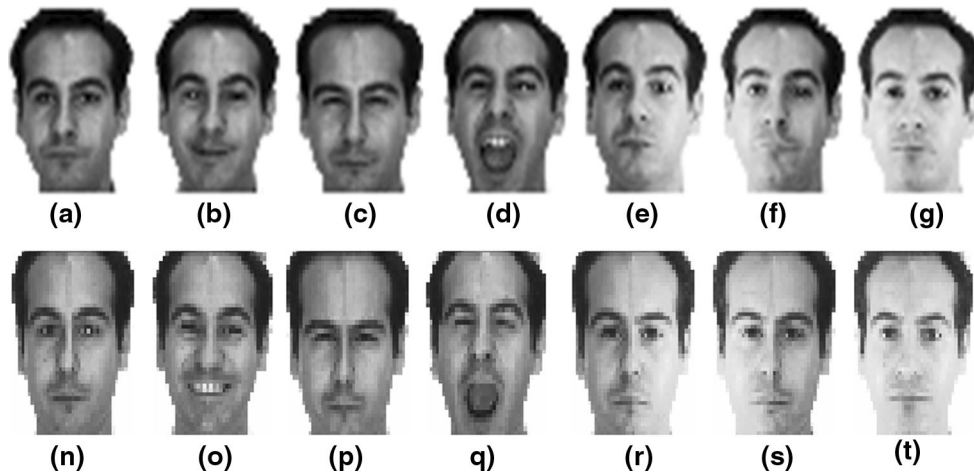
Fig. 5 Some sample images of one subject in the AR database

Table 3 The top classification accuracies (%) of the six approaches and the corresponding number of dimensions on the AR database (shown in parentheses)

Training/ testing numbers	Fisherface	DLA	LFDA	MFA	SDA	GLFDA
840/840	66.90 (66)	67.50 (66)	69.52 (119)	69.17 (125)	65.59 (72)	70.24 (124)
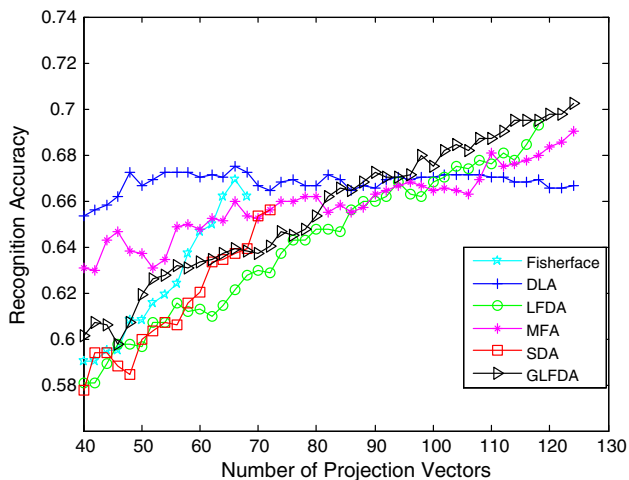
Fig. 6 The recognition accuracy versus number of projection vector on the AR database

The UMIST database [20] is a multi-view database, consisting of 564 images from 20 individuals under various poses. Each image is of the 112 × 92 size. In the experiments, the first 19 images per image are selected as gallery. Figure 3 shows some images of one subject. In experiments, the first 3, 6, and 9 images per person are, respectively, selected as training images and the remaining

images for the corresponding testing images. Thus, we have three training sets that include 60, 120, and 180 images, and the corresponding three testing sets including 320, 260, and 200 images, respectively. The top recognition accuracies of the six approaches are shown in Table 2. Figure 4 plots the curve of the recognition accuracy versus number of projection vectors of six approaches when the first 3 images per person are selected for training.

The AR face database [21] is established by American Purdue University, which contains over 4,000 color face images of 126 people (70 men and 56 women) including frontal views of faces with different facial expressions, lighting conditions, and occlusions. The pictures of most persons are taken in two sessions (separated by 2 weeks). Each session contains 13 color images and 120 individuals (65 men and 55 women) participate in both sessions. In the experiments, the facial portion of each image is manually cropped and then normalized to the size of 50 × 40. The images from the first session with (a) "neutral expression", (b) "smile", (c) "anger", (d) "scream", (e) "left light on", (f) "right light on", and (g) "both side light on" are selected for training, and the corresponding second session images, i.e., from (n)–(t), are selected for testing. Thus, training set and testing set both include 840 images. Figure 5 shows some sample images of one subject. Table 3 lists the top recognition accuracies of the six approaches. Figure 6 plots the curve of the recognition accuracy versus number of projection vectors of the six approaches.

From Tables 1, 2 and 3, Figs. 2, 4, and 6, we can see that:

1. Our GLFDA approach is superior to Fisherface. This is probably because that Fisherface does not well encode the discriminative information embedded in nearby data from different classes and also does not preserve the local intrinsic geometrical structure embedded in

Fig. 7 **a** Two-dimensional data and one-dimensional projection direction obtained by GLFDA, Fisherface, and MFA, respectively; **b** One-dimensional embedded results obtained by GLFDA; **c** One-dimensional embedded results obtained by Fisherface; **d** One-dimensional embedded results obtained by MFA
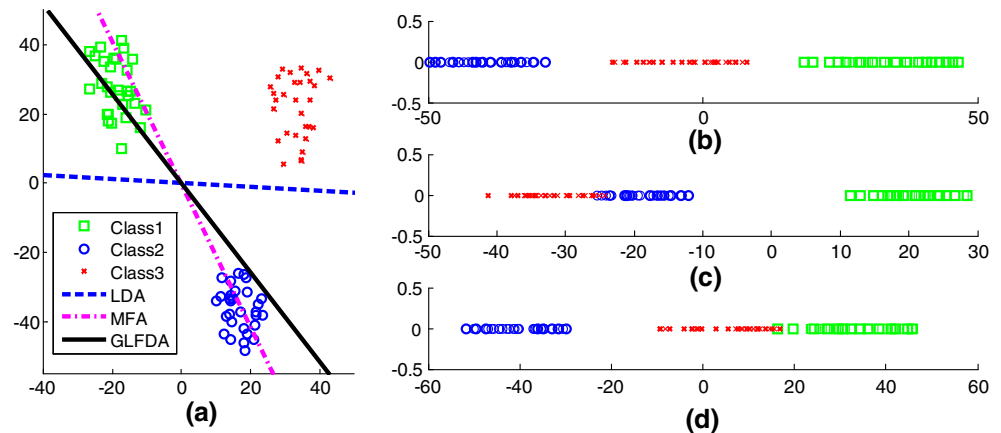
nearby data from the same class. This impairs the recognition accuracy of the algorithm.

2. The top recognition accuracy of GLFDA is also superior to MFA and DLA. This is probably because that MFA and DLA only encode the discriminative information embedded in the nearby data points from different classes and do not explicitly consider the global discriminative information embedded in data points having different labels. This will impair the recognition accuracy.

3. Global–local Fisher discriminant analysis is also superior to LFDA. This is probably due to the fact that LFDA only well detects the global discriminant structure, which maps data points from different classes to be faraway in the reduced space by the quadratic function. However, this quadratic function emphasizes the large distance pairs and may impair the local discriminant information of data. This leads to inseparability among nearby data from different classes.

4. SDA is inferior to GLFDA. Although SDA considers both the global and local geometrical structures of data, it ignores the local discriminant geometrical structure. This may impair the local discriminating information of data and lead to inseparability among nearby data having different labels. In Fig. 4, the recognition accuracy of GLFDA is not obvious better than SDA. This is probably because that GLFDA ignores the un-local intrinsic structure of data.

5. Global–local Fisher discriminant analysis is superior to the other discriminant approaches. The reason is probably because that the other discriminant approaches only encode the global or local discriminant information of data. In real applications, the intrinsic structure of data is unknown and complex, only single structure is not suitable to guarantee the discrimination ability in the reduced space. Different from these approaches, GLFDA explicitly considers both the global and local discriminant information of

data. Thus, GLFDA can well separate the data having different class labels in the reduced space. We randomly generate some two-dimensional data points using MATLAB and show the projection directions of GLFDA, Fisherface, and MFA in Fig. 7a. We show the corresponding one-dimensional embedded results in Fig. 7b–d, respectively. It is easy to see that GLFDA well encodes the more discriminant information embedded in data than Fisherface and MFA.

4 Conclusion

In this paper, we propose a novel dimensionality reduction approach, namely GLFDA. GLFDA explicitly considers the local intrinsic structure, global discriminant structure, and local discriminant structure of the data manifold by constructing three adjacency graphs and then incorporates them into one objective function for dimension reduction. In this way, GLFDA well encodes the discriminative information. Experiments on the YALE, UMIST, and AR databases show the effectiveness of the proposed GLFDA approach.

References

1. He X, Yan S, Hu Y, Niyogi P, Zhang HJ (2005) Face recognition using Laplacianfaces. IEEE Trans Pattern Anal Mach Intell 27(3):328–340
2. Yan S, Xu D, Zhang B, Zhang H, Yang Q, Lin S (2007) Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans Pattern Anal Mach Intell 29(1):40–51

3. Gao Q, Xu H, Li Y, Xie D (2010) Two-dimensional supervised local similarity and diversity projection. Pattern Recognit 43(10):3359–3363

4. Wang J, Yang W, Yang J (2013) Face recognition using fuzzy maximum scatter discriminant analysis. Neural Comput Appl 23:957–964

5. Xu Y, Feng G, Zhao Y (2009) One improvement to two-dimensional locality preserving projection method for use with face recognition. Neurocomputing 73(1–3):245–249

6. Xu Y, Zhong A, Yang J, Zhang D (2010) LPP solution schemes for use with face recognition. Pattern Recognit 43(12):4165–4176

7. He X, Niyogi P (2003) Locality preserving projections. In: Proceedings of NIPS, vol 16, pp 153–160

8. Belhumeur PN, Hepanha JP, Kriegman DJ (1997) Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans Pattern Anal Mach Intell 19(7):711–720

9. Chen HT, Chang HW, Liu TL (2005) Local discriminant embedding and its variants. IEEE Conf Comput Visual Pattern Recognit (CVPR) 2:846–853

10. Fan Z, Xu Y, Zhang D (2011) Local linear discriminant analysis framework using sample neighbors. IEEE Trans Neural Netw 22(7):1119–1132

11. Sugiyama M (2006) Local fisher discriminant analysis for supervised dimensionality reduction. In: Proceedings of international conference on machine learning (ICML), vol 23, pp 905–912

12. Sugiyama M, Roweis S (2007) Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. J Mach Learn Res 8:1027–1061

13. Cai D, He X, Zhou K, Han J, Bao H (2007) Locality sensitive discriminant analysis. In: Proceedings of joint conference on artificial intelligence (IJCAI), vol 20, pp 708–713

14. Lin F, Jiao LC, Shang F, Wang S, Hou B (2013) Fast Fisher sparsity preserving projections. Neural Comput Appl 23:691–705

15. Gao Q, Liu J, Zhang H, Hou J, Yang X (2012) Enhanced fisher discriminant criterion for image recognition. Pattern Recognit 45:3717–3724

16. Chen J, Ye J, Li Q (2007) Integrating global and local structures: a least squares framework for dimensionality reduction. IEEE Conf Comput Visual Pattern Recognit (CVPR) 1:1–8

17. Cai D, He X, Han J (2007) Semi-supervised discriminant analysis. IEEE Conf Comput Vision (ICCV) 11:1–8

18. Zhang T, Tao D, Li X, Yang J (2009) Patch alignment for dimensionality reduction. IEEE Trans Knowl Data Eng 21(9):1299–1313

19. The Yale Face Database, http://cvc.yale.edu/projects/yalefaces/yalefaces.html

20. The UMIST database, http://images.ee.umist.ac.uk/danny/database.html

21. The AR face data base, http://rvll.ecn.purdue.edu/~aleix/aleix_face_DB.html