Graph Embedding Discriminant Analysis and Semi-Supervised Extension for Face Recognition

Jiangshan Du[1], Liping Zheng[1], and Jun Shi[2(✉)]

[1] School of Computer Science and Information Engineering,
Hefei University of Technology, Hefei 230009, China
[2] School of Software, Hefei University of Technology, Hefei 230009, China
`juns@hfut.edu.cn`

Abstract. Considering graph embedding based dimension reduction methods are easily affected by outliers and the fact that there exist insufficient labeled samples in practical applications and thus degrades the discriminant performance, a novel graph embedding discriminant analysis method and its semi-supervised extension for face recognition is proposed in this paper. It uses the intraclass samples and their within-class mean to construct the intrinsic graph for describing the intraclass compactness, which can effectively avoid the influence of outliers to the within-class scatter. Meanwhile, we build the penalty graph through the local information of interclass samples to characterize the interclass separability, which considers the different contributions of interclass samples within the neighborhood to the between-class scatter. On the other hand, we apply the low-rank representation with sparsity constraint for semi-supervised learning, aiming to explore the global low-rank relationships of unlabeled samples. Experimental results on ORL, AR and FERET face datasets demonstrate the effectiveness and robustness of our method.

Keywords: Graph embedding · Low-rank representation · Semi-supervised learning · Dimension reduction · Face recognition

1 Introduction

Over the past decades, dimension reduction has been one research hotspot of face recognition, which aims to discover a low-dimensional representation from original high-dimensional data and thus capture the intrinsic structure of data and finally contribute to classification. Principal Component Analysis (PCA) [1] and Linear Discriminant Analysis (LDA) [2] are the most classic methods. PCA seeks a set of optimal projection directions to maximize the data covariance in an unsupervised way, while LDA uses the label information to construct the within-class scatter and between-class scatter and thus finds the discriminant projection directions. Therefore, LDA is better than PCA in face recognition and other recognition tasks. However, both of them fail to reveal the local structure due to their global linearity.

In order to capture the local information, the manifold learning-based methods have been proposed, which generally assume that the closely located samples are likely to

have the similar low-dimensional manifold embeddings. As the most representative manifold learning method, Locality Preserving Projections (LPP) [3] applies k-nearest-neighbor to explore the local structure of each sample. In contrast to LPP, Marginal Fisher Analysis (MFA) [4] uses k_1 nearest intraclass samples to build the intrinsic graph and simultaneously employs k_2 nearest interclass marginal samples to construct the penalty graph. Both of these two graphs can describe the intraclass compactness and the interclass separability respectively. It verifies that the manifold learning-based dimensionality reduction methods can be well explained by graph embedding framework [4]. However, compared with the k_1 nearest intraclass samples and k_2 nearest interclass samples, the distant intraclass and interclass samples are not considered during the process of graph construction. Therefore, it may cause the distant intraclass samples are far from each other and the distant interclass samples are close to each other in the projected feature space, which inevitably goes against classification. To address this problem, Huang et al. proposed the Graph Discriminant Embedding (GDE) [5] to focus on the importance of the distant intraclass and interclass samples in the graph embedding framework. It uses a strictly monotone decreasing function to describe the importance of nearby intraclass and interclass samples and simultaneously applies a strictly monotone increasing function to characterize the importance of distant intraclass and interclass samples. Consequently, the different contributions of nearby and distant samples to the within-class and between-class scatters are fully considered. For the most graph embedding methods, the importance of samples is directly based on the similarity between two samples. Therefore, outliers have the influence to the graph construction and especially affect the intraclass compactness.

On the other hand, considering there exist unlabeled samples in practical application, Cai et al. [6] proposed the Semi-Supervised Discriminant Analysis (SDA) which performs LDA on labeled samples and uses unlabeled samples to build k-nearest-neighbor graph for preserving local neighbor relationships. It effectively applies semi-supervised learning to achieve low dimensional embedding. However, the neighbor parameter k is difficult to define. Inappropriate k even influences the quality of graph construction and thus degrades the semi-supervised discriminant performance. In the past few years, Low Rank Representation (LRR) [7] has been proposed to capture the global structure of samples using low-rank and sparse matrix decomposition. The decomposed low-rank part reflects the intrinsic affinity of samples. In other words, the similar samples will be clustered into the same subspace. However, LRR is essentially unconstrained and is likely to be sensitive to noise for the corrupted data. Therefore, Guo et al. [8] performed the L_1-norm sparsity constraint on low-rank decomposition with the aiming of obtaining the more robust low-rank relationships.

Motivated by the above discussion, we introduce a novel graph embedding discriminant analysis method and its semi-supervised extension for face recognition in this paper. For the labeled samples, we redefine the weights of the intraclass graph through the mean of intraclass samples. In other words, the similarity between two intraclass samples is based on the similarities between them and the mean of intraclass samples, instead of directly using the distance of them. It effectively avoids the influence of outliers to the within-class scatter. Meanwhile we concentrate the different contributions of interclass samples to between-class scatter in view of the idea of GDE. Consequently, we build the improved discriminant criterion to make the intraclass

samples as close to each other as possible and interclass samples far from each other. More critically, the constructed graph embedding framework can relieve the influence of outliers to the intraclass compactness. On the other hand, considering there may exist a large number of unlabeled samples in practical application, we apply the low rank representation constrained by L_1-norm sparsity to capture their global low-rank relationships, and then minimizes the error function through preserving the low-rank structure of samples. Finally, we integrate the improved graph embedding discriminant criterion and LRR error function for achieving semi-supervised learning. Experimental results on ORL, AR and FERET face datasets demonstrate the effectiveness and robustness of our method.

The remainder of this paper is organized as follows. Section 2 briefly reviews MFA and GDE. Section 3 introduces the proposed method. Section 4 shows the experimental results and the conclusion is given in Sect. 5.

2 Related Works

Given a set of n training samples $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n]$, containing C classes: $\mathbf{x}_i \in l_c$, $c \in \{1, 2, ..., C\}$. By using the projection matrix $\mathbf{A}$, the representation of the data set $\mathbf{X}$ in the low-dimensional space can be represented as $\mathbf{y}_i = \mathbf{A}^T \mathbf{x}_i$. In this section, we will briefly review MFA and GDE.

2.1 Marginal Fisher Analysis (MFA)

As the most representative graph embedding method, MFA [4] builds the intrinsic graph to characterize the compactness of intraclass samples, and constructs the penalty graph to describe the separability of interclass samples. Let $G_{intrinsic} = \{\mathbf{V}, \mathbf{E}, \mathbf{W}^{MFA}\}$ denotes the intrinsic graph, and $G_{penalty} = \{\mathbf{V}, \mathbf{E}, \mathbf{B}^{MFA}\}$ denotes the penalty graph, where $\mathbf{V}$ is the samples, and $\mathbf{E}$ is the set of edges connecting samples. $\mathbf{W}^{MFA}$ and $\mathbf{B}^{MFA}$ are two adjacency matrices identifying the similarity between samples. The weights of these two matrices are defined [4]:

$$\mathbf{W}_{ij}^{MFA} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in N_{k_1}^+(j) \text{ or } \mathbf{x}_j \in N_{k_1}^+(i) \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

$$\mathbf{B}_{ij}^{MFA} = \begin{cases} 1, & \text{if } \mathbf{x}_i \in P_{k_2}(j) \text{ or } \mathbf{x}_j \in P_{k_2}(i) \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $N_{k_1}^+(i)$ indicates the k_1 nearest intraclass samples of $\mathbf{x}_i$, and $P_{k_2}(i)$ indicates the k_2 nearest interclass marginal samples of $\mathbf{x}_i$. The intraclass scatter $\mathbf{S}_w^{MFA}$ and interclass scatter $\mathbf{S}_b^{MFA}$ can be formulated [4]:

$$\mathbf{S}_w^{MFA} = \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 \mathbf{W}_{ij}^{MFA} = 2tr(\mathbf{A}^T \mathbf{X} \mathbf{L}^{MFA} \mathbf{X}^T \mathbf{A}) \tag{3}$$

$$\mathbf{S}_b^{MFA} = \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 \mathbf{B}_{ij}^{MFA} = 2tr\left(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}^{MFA'}\mathbf{X}^{\mathrm{T}}\mathbf{A}\right) \qquad (4)$$

where $\mathbf{L}^{MFA} = \mathbf{D}^{MFA} - \mathbf{W}^{MFA}$ and $\mathbf{L}^{MFA'} = \mathbf{D}^{MFA'} - \mathbf{B}^{MFA}$ are the Laplacian matrices; $\mathbf{D}_{ii}^{MFA} = \sum_j \mathbf{W}_{ij}^{MFA}$ and $\mathbf{D}_{ii}^{MFA'} = \sum_j \mathbf{B}_{ij}^{MFA}$ are the diagonal matrices. The discriminant criterion function of MFA can be expressed as [4]:

$$J_{MFA}(\mathbf{A}) = \min_{\mathbf{A}} \frac{\mathbf{S}_w^{MFA}}{\mathbf{S}_b^{MFA}} = \min_{\mathbf{A}} \frac{tr(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}^{MFA}\mathbf{X}^{\mathrm{T}}\mathbf{A})}{tr(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}^{MFA'}\mathbf{X}^{\mathrm{T}}\mathbf{A})} \qquad (5)$$

2.2   Graph Discriminant Embedding (GDE)

The purpose of MFA is to construct two graphs by comprehensively considering the local structure and discriminant structure of samples. However, MFA neglects the importance of the distant samples which may influence the discriminant performance. Therefore, GDE redefines the intrinsic graph $G_{intrinsic}$ and the penalty graph $G_{penalty}$. The weight matrices $\mathbf{W}^{GDE}$ and $\mathbf{B}^{GDE}$ in these two graphs are redefined [5]:

$$\mathbf{W}_{ij}^{GDE} = \begin{cases} e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{t}}, & \text{if } \mathbf{x}_j \in N_{k1}^+(i) \text{ or } \mathbf{x}_i \in N_{k1}^+(j), l_i = l_j \\ e^{-\frac{t}{||\mathbf{x}_i - \mathbf{x}_j||^2}}, & \text{if } \mathbf{x}_j \notin N_{k1}^+(i) \text{ and } \mathbf{x}_i \notin N_{k1}^+(j), l_i = l_j \\ 0, & \text{if } l_i \neq l_j \end{cases} \qquad (7)$$

$$\mathbf{B}_{ij}^{GDE} = \begin{cases} e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{t}}, & \text{if } \mathbf{x}_j \in P_{k2}(i) \text{ or } \mathbf{x}_i \in P_{k2}(j), l_i \neq l_j \\ e^{-\frac{t}{||\mathbf{x}_i - \mathbf{x}_j||^2}}, & \text{if } \mathbf{x}_j \notin P_{k2}(i) \text{ and } \mathbf{x}_i \notin P_{k2}(j), l_i \neq l_j \\ 0, & \text{if } l_i = l_j \end{cases} \qquad (8)$$

where the parameter t $(t > 0)$ is used to tune the weight of samples. The intraclass scatter $\mathbf{S}_w^{GDE}$ and the interclass scatter $\mathbf{S}_b^{GDE}$ can be computed [5]:

$$\mathbf{S}_w^{GDE} = \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 \mathbf{W}_{ij}^{GDE} = 2tr\left(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}^{GDE}\mathbf{X}^{\mathrm{T}}\mathbf{A}\right) \qquad (9)$$

$$\mathbf{S}_b^{GDE} = \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 \mathbf{B}_{ij}^{GDE} = 2tr\left(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}^{GDE'}\mathbf{X}^{\mathrm{T}}\mathbf{A}\right) \qquad (10)$$

where $\mathbf{L}^{GDE} = \mathbf{D}^{GDE} - \mathbf{W}^{GDE}$ and $\mathbf{L}^{GDE'} = \mathbf{D}^{GDE'} - \mathbf{B}^{GDE}$ are the Laplacian matrices; $\mathbf{D}_{ij}^{GDE} = \sum_j \mathbf{W}_{ij}^{GDE}$ and $\mathbf{D}_{ij}^{GDE'} = \sum_j \mathbf{B}_{ij}^{GDE}$ are the diagonal matrices. Therefore, the discriminant criterion function of GDE can be computed [5]:

$$J_{GDE}(\mathbf{A}) = \min_{\mathbf{A}} \frac{\mathbf{S}_w^{GDE}}{\mathbf{S}_b^{GDE}} = \min_{\mathbf{A}} \frac{tr(\mathbf{A}^{T}\mathbf{X}\mathbf{L}^{GDE}\mathbf{X}^{T}\mathbf{A})}{tr(\mathbf{A}^{T}\mathbf{X}\mathbf{L}^{GDE'}\mathbf{X}^{T}\mathbf{A})} \qquad (11)$$

In GDE method, the similarity defined by monotonic functions can effectively solve the effect of the distant samples on recognition results. However, the similarity of samples is directly defined by the distance of samples, and outliers will have a certain impact on the discriminant performance. Furthermore, GDE computes the similarity for labeled samples without taking unlabeled samples into account.

3   Methodology

3.1   The Proposed Method

As described in Sect. 2.2, GDE does not study the influence of outliers and unlabeled samples on the recognition performance. In this paper, we propose a novel graph embedding discriminant analysis method where the similarity of samples is based on the similarities between them and the mean of intraclass samples. It intends to eliminate the impact of outliers on the within-class scatters and improves the recognition performance.

Let $G_{intrinsic} = \{V, E, W\}$ denotes the redefined intrinsic graph, and $G_{penalty} = \{V, E, B\}$ denotes the penalty graph, and the weight matrices $\mathbf{W}$ and $\mathbf{B}$ are given as follows:

$$\mathbf{W}_{ij} = \begin{cases} e^{-\frac{||\mathbf{x}_i - \tilde{\mathbf{x}}_i||^2 + ||\mathbf{x}_j - \tilde{\mathbf{x}}_i||^2}{2t}}, & \text{if } l_i = l_j \\ 0, & \text{if } l_i \neq l_j \end{cases} \tag{12}$$

$$\mathbf{B}_{ij} = \begin{cases} e^{-\frac{||\mathbf{x}_i - \mathbf{x}_j||^2}{t}}, & \text{if } \mathbf{x}_i \in D_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in D_k(\mathbf{x}_i), l_i \neq l_j \\ e^{-\frac{t}{||\mathbf{x}_i - \mathbf{x}_j||^2}}, & \text{if } \mathbf{x}_i \notin D_k(\mathbf{x}_j) \text{ and } \mathbf{x}_j \notin D_k(\mathbf{x}_i), l_i \neq l_j \\ 0, & \text{if } l_i = l_j \end{cases} \tag{13}$$

where $\tilde{\mathbf{x}}_i$ denotes the mean of within-class samples and $D_k(\mathbf{x}_i)$ denotes the k nearest interclass neighbors of $\mathbf{x}_i$. However, unlike GDE, we redefine the weight matrix $\mathbf{W}$ based on the distance between the sample and the mean of within-class samples instead of directly using the distance of samples. Moreover, we also describe the importance of samples through monotonic functions in view of GDE.

According to the constructed two graphs, the intraclass scatter $\mathbf{S}_w$ and the interclass scatter $\mathbf{S}_b$ can be formulated:

$$\mathbf{S}_w = \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 \mathbf{W}_{ij} = \sum_{ij} ||\mathbf{A}^T \mathbf{x}_i - \mathbf{A}^T \mathbf{x}_j||^2 \mathbf{W}_{ij} = 2tr(\mathbf{A}^T \mathbf{XLX}^T \mathbf{A}) \tag{14}$$

$$\mathbf{S}_b = \sum_{ij} ||\mathbf{y}_i - \mathbf{y}_j||^2 \mathbf{B}_{ij} = \sum_{ij} ||\mathbf{A}^T \mathbf{x}_i - \mathbf{A}^T \mathbf{x}_j||^2 \mathbf{B}_{ij} = 2tr(\mathbf{A}^T \mathbf{XL'X}^T \mathbf{A}) \tag{15}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{W}$ and $\mathbf{L'} = \mathbf{D'} - \mathbf{B}$ are the Laplacian matrices. $\mathbf{D}_{ij} = \sum_j \mathbf{W}_{ij}$ and $\mathbf{D'}_{ij} = \sum_j \mathbf{B}_{ij}$ are the diagonal matrices. Therefore, the proposed discriminant criterion function is given:

$$J_{Ours}(\mathbf{A}) = \min_{\mathbf{A}} \frac{tr(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}}\mathbf{A})}{tr(\mathbf{A}^{\mathrm{T}}\mathbf{X}\mathbf{L}'\mathbf{X}^{\mathrm{T}}\mathbf{A})} \tag{16}$$

Consequently, our method can avoid the influence of outliers to the within-class scatter, making the intraclass samples close to each other and interclass samples far away from each other in the projected space.

3.2   Semi-Supervised Extension (SSE)

For unlabeled samples, SDA [6] builds the k-nearest neighbor graph on unlabeled samples to preserve local neighbor relationships. However, the neighbor parameter k in SDA is difficult to define. In the past few years, LRR [7] is proposed, and it uses low-rank and sparse matrix decomposition to express the global structure of the samples, so that the low-rank structure can reflect the intrinsic affinity of samples. Nevertheless, LRR is essentially unconstrained, and it may be affected by noise. In this paper, we use the low rank representation with L_1-norm sparsity constraint to capture their low-rank relationships, and minimize the error function through preserving the low-rank structure of samples. For the given training samples $\mathbf{X}$, we refactor the original samples through the data dictionary $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \ldots, \mathbf{d}_m] \in \mathrm{R}^{d \times m}$. In other words, $\mathbf{X} = \mathbf{D}\mathbf{Z}$, where $\mathbf{Z} = [\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_n] \in \mathrm{R}^{m \times n}$ is the refactor coefficient matrix. Besides, there may exist noise in the distribution of samples, therefore the problem can be modeled as follows [8]:

$$\begin{cases} \min_{z} \|\mathbf{Z}\|_* + \lambda \|\mathbf{E}\|_{2,1} + \gamma \|\mathbf{Z}\|_1 \\ s.t. \quad \mathbf{X} = \mathbf{D}\mathbf{Z} + \mathbf{E} \end{cases} \tag{17}$$

where $\|\mathbf{Z}\|_*$ denotes the kernel norm of $\mathbf{Z}$, that is, the sum of the singular values of the matrix. $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^{n} \sqrt{\sum_{i=1}^{d} (\mathbf{E}_{ij})^2}$ is the $l_{2,1}$-norm of the noise matrix E, which is used to model samples of noise interference. $\|\mathbf{Z}\|_1$ represents the L_1-norm of the matrix, which is used to ensure the sparsity of $\mathbf{Z}$. the parameter λ is used to balance the effects of the noise part; and the parameter γ is to balance the effects of the sparse part on the results. By applying the L_1-norm sparsity constraint and adding the error term to solve noise pollution, we can obtain more robust low-rank relationships.

According to Eq. (16), we can see that $\mathbf{x}_i = \mathbf{X}\mathbf{z}_i + \mathbf{e}_i$, and $\mathbf{e}_i$ is the column vector of the noise matrix $\mathbf{E}$. It is assumed that the $\mathbf{x}_i$ remains the low-rank structure after linearly projection into the low-dimensional feature space, and the influence of noise on classification can be avoided by minimizing the representation error $\mathbf{e}_i$. Therefore, we can minimize the norm of $\mathbf{e}_i$ as follows:

$$\sum_i \left\|\mathbf{A}^{\mathrm{T}}\mathbf{e}_i\right\|^2 = \sum_i \left\|\mathbf{A}^{\mathrm{T}}(\mathbf{x}_i - \mathbf{X}\mathbf{z}_i)\right\|^2 = tr(\mathbf{A}^{\mathrm{T}}\mathbf{S}_\mathbf{e}\mathbf{A}) \tag{18}$$

where $\mathbf{S}_\mathbf{e}$ can be computed by:

$$\mathbf{S_e} = \sum\nolimits_i (\mathbf{x}_i - \mathbf{Xz}_i)(\mathbf{x}_i - \mathbf{Xz}_i)^{\mathrm{T}} = \mathbf{X}(\mathbf{I} - \mathbf{Z})(\mathbf{I} - \mathbf{Z})^{\mathrm{T}}\mathbf{X}^{\mathrm{T}} \qquad (19)$$

Finally, we integrate the improved graph embedding discriminant criterion and LRR error function for achieving semi-supervised learning. The discriminant criterion function of our extension method is given as following:

$$J_{Ours-SSE}(\mathbf{A}) = \min_{\mathbf{A}} \frac{tr(\mathbf{A}^{\mathrm{T}}\mathbf{XLX}^{\mathrm{T}}\mathbf{A}) + \beta tr(\mathbf{A}^{\mathrm{T}}\mathbf{S_e}\mathbf{A})}{tr(\mathbf{A}^{\mathrm{T}}\mathbf{XL'X}^{\mathrm{T}}\mathbf{A})} \qquad (20)$$

where the regularization parameter β ($\beta \in (0, 1)$) is used to balance the loss of the discriminant criterion function.

3.3   Algorithm

The dimensionality of face images is often very high, thereby we first use the PCA projection $\mathbf{A}_{PCA}$ for dimensionality reduction. For the dataset $\mathbf{X} = [\mathbf{X}_L, \mathbf{X}_U]$ in the PCA subspace, where $\mathbf{X}_L$ is the subset of labeled samples and $\mathbf{X}_U$ is the unlabeled samples. The proposed algorithm is given as follows:

Step1:	For $\mathbf{X}_L$, construct the two adjacency matrices $\mathbf{W}$ and $\mathbf{B}$ based on Eqs. (12) and (13), and compute the intraclass scatter $\mathbf{S}_w$ and the interclass scatter $\mathbf{S}_b$ based on Eqs. (14) and (15).
Step2:	In consideration of unlabeled samples in datasets, compute the reconstruction coefficient matrix $\mathbf{Z}$ based on Eq. (17), then compute the matrix $\mathbf{S_e}$ based on Eq. (19). Let $\mathbf{a}_1$, $\mathbf{a}_2$, …, $\mathbf{a}_p$ be the eigenvectors associated with the p minimal eigenvalues of the general eigenvalue problem $\mathbf{XL'X}^{\mathrm{T}}\mathbf{A} = \lambda(\mathbf{XLX}^{\mathrm{T}} + \beta\mathbf{S_e})\mathbf{A}$, then the optimal projection matrix learned by our extension method can be denoted as $\mathbf{A}_{Ours\text{-}SSE} = [\mathbf{a}_1, \mathbf{a}_2, …, \mathbf{a}_p]$, and the final projection is $\mathbf{A} = \mathbf{A}_{PCA} \cdot \mathbf{A}_{Ours\text{-}SSE}$.
Step3:	For a test samples $\mathbf{x}$, the representation in the low-dimensional space can be represented as $\mathbf{y} = \mathbf{A}^{\mathrm{T}}\mathbf{x}$, and then we can use a classifier to predict its class.

4   Experiments

In this section, we verify the proposed method with multiple classifiers, i.e. NNC (nearest neighbor classifier), SVM (support vector machine), LRC (linear regression classifier) [9] and LLRC (locality-regularized linear regression classifier) [10], on three public face datasets (ORL, AR and FERET), and particularly evaluates its performance under noise and blur conditions. All the experiments are conducted on a computer with Intel Core i7-7700 3.6 GHz Windows 7, RAM-8 GB, and implemented by MATLAB.

4.1   Experiments on ORL Dataset

The ORL face dataset contains 400 images of 40 people. All images are in grayscale and manually cropped to 92×112 pixels. On ORL face dataset, we select 5 images of each person for training samples and the remaining images for testing. Furthermore, the
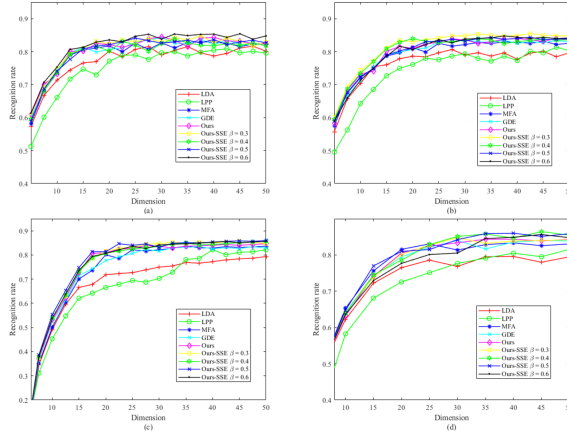
first r (r = 2, 3) images of each person are labeled. The NNC, SVM, LRC and LLRC are picked for classification. Firstly, the dimension of PCA subspace is set to 50. In our method, the value of the interclass neighbor parameter k is 4, and the tuning parameter of adjacency matrices t is set to 2^3. The neighbor parameter k in LPP is 4. For MFA and GDE, the intraclass neighbor parameter k_1 is picked as 1, and the interclass neighbor parameter k_2 is 40.

Table 1 reports the maximal recognition rates of different methods on ORL dataset. It can be seen that the recognition performance of our method is better than other methods in various classifiers, especially GDE. When unlabeled samples are existing in dataset, the L_1-norm sparse constraint is introduced, which makes our extension method is more effective in classifying these samples. Furthermore, the regularization parameter β which is applied for the loss of the discriminant criterion function is critical to the recognition performance, and too large value of β will make the recognition rate degraded. In addition, Fig. 1 shows the recognition rate curves of all methods under different classifiers as the projection axes varies. It can be easily found that the recognition performance of all method outperforms under LLRC, and our extension method achieves the best performance, that further proves the efficiency of our extension method.

Table 1. Maximal recognition rates (%) and corresponding dimensions (shown in the parentheses) of different methods under four classifiers on ORL dataset.

Methods	$r = 2$				$r = 3$			
	NNC	SVM	LRC	LLRC	NNC	SVM	LRC	LLRC
LDA	77.48	77.87	75.12	76.28	80.15	79.56	76.87	78.50
	(25)	(32)	(30)	(28)	(20)	(30)	(35)	(35)
LPP	76.32	77.40	78.54	78.89	80.87	81.32	82.15	82.50
	(50)	(50)	(45)	(45)	(45)	(45)	(40)	(40)
MFA	74.26	77.86	79.20	79.65	81.75	82.55	82.95	83.05
	(25)	(25)	(30)	(30)	(20)	(25)	(25)	(25)
GDE	75.12	78.54	80.25	80.56	82.45	82.87	83.12	84.00
	(30)	(30)	(30)	(35)	(30)	(35)	(30)	(30)
Ours	**76.78**	**79.20**	**80.75**	**81.35**	**82.33**	**83.56**	**83.20**	**84.33**
	(35)	**(30)**	**(35)**	**(35)**	**(35)**	**(30)**	**(35)**	**(35)**
Ours-SSE $\beta = 0.3$	**78.36**	**79.47**	**80.97**	**82.87**	**82.93**	**84.77**	**84.65**	**84.75**
	(35)	**(38)**	**(30)**	**(30)**	**(25)**	**(30)**	**(30)**	**(30)**
Ours-SSE $\beta = 0.4$	**80.15**	**80.96**	**81.37**	**82.48**	**81.74**	**83.87**	**84.48**	**85.46**
	(35)	**(35)**	**(35)**	**(35)**	**(40)**	**(34)**	**(38)**	**(33)**
Ours-SSE $\beta = 0.5$	**80.75**	**81.25**	**82.75**	**82.84**	**82.67**	**83.95**	**84.84**	**85.87**
	(38)	**(34)**	**(35)**	**(35)**	**(38)**	**(37)**	**(37)**	**(35)**
Ours-SSE $\beta = 0.6$	**79.68**	**80.34**	**81.25**	**81.87**	**84.78**	**84.17**	**84.13**	**84.76**
	(35)	**(37)**	**(35)**	**(36)**	**(38)**	**(35)**	**(34)**	**(38)**

Fig. 1. Recognition rate curves of all methods under different classifiers when the labeled images of each class $r = 3$. (a) NNC; (b) SVM; (c) LRC; (d) LLRC.
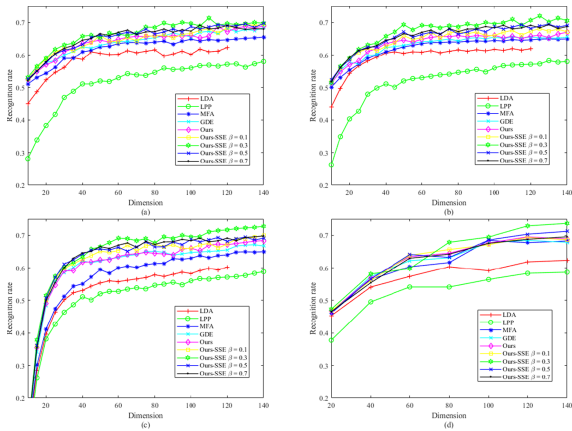
4.2   Experiments on AR Dataset

The AR face dataset contains over 3000 face images of 120 people. All images are manually cropped to 50×40 pixels. On AR face dataset, we only use the images without occlusion, and the first 7 images of each person are taken for training and 14–20 images are applied for testing. Furthermore, the first r ($r = 3, 4, 5$) images of each person are labeled. The NNC, SVM, LRC and LLRC are employed for classification. Before the experiments, the dimension of PCA subspace is set to 150. The parameters in the experiments are set as follows: In our method, the value of the interclass neighbor parameter k is 6, and the tuning parameter of adjacency matrices t is chosen as 2^4. The neighbor parameter k in LPP is set to 6. For MFA and GDE, the intraclass neighbor parameter k_1 is picked as 1, and the interclass parameter k_2 is set to 120.

Table 2 reports the maximal recognition rates of different methods on AR dataset. The results indicate that compared with GDE, our method has a small improvement (<1%) under four classifiers, but the recognition performance of our extension method has obvious advantages when unlabeled samples are existing in datasets. Meanwhile, the more labeled samples, the better performance our extension method achieves. In addition, the recognition rate curves of different methods under various classifiers as the projection axes varies are shown in Fig. 2. We can see that LPP performs the worst, but when the regularization parameter β is set to 0.3, the curves in Fig. 2 demonstrate that our extension method performs better than other methods, which proves that the L_1-norm sparse constraint we introduced can eliminate the influence of unlabeled samples on the experimental results.

Table 2. Maximal recognition rates (%) and corresponding dimensions of different methods under four classifiers on AR dataset.

Method	$r = 3$				$r = 4$				$r = 5$			
	NNC	SVM	LRC	LLRC	NNC	SVM	LRC	LLRC	NNC	SVM	LRC	LLRC
LDA	56.18	57.18	56.15	57.33	58.73	59.38	68.56	59.33	61.25	61.55	60.25	61.87
	(125)	(125)	(130)	(110)	(110)	(110)	(110)	(110)	(115)	(115)	(120)	(125)
LPP	49.35	50.12	49.75	51.28	53.65	54.20	55.75	56.15	57.14	58.24	57.35	58.33
	(120)	(120)	(125)	(125)	(115)	(120)	(125)	(120)	(120)	(130)	(125)	(125)
MFA	60.25	61.32	62.55	64.18	63.54	62.87	63.59	65.83	65.55	64.95	64.88	67.75
	(135)	(125)	(130)	(130)	(130)	(125)	(125)	(125)	(140)	(130)	(125)	(125)
GDE	60.88	62.45	63.37	65.28	65.24	63.14	64.87	66.67	67.87	65.25	66.72	68.45
	(130)	(130)	(130)	(130)	(135)	(135)	(130)	(130)	(130)	(130)	(125)	(125)
Ours	**61.75**	**62.45**	**64.78**	**66.54**	**65.76**	**63.84**	**65.87**	**67.75**	**68.74**	**66.48**	**67.50**	**68.87**
	(125)	**(125)**	**(130)**	**(130)**	**(130)**	**(130)**	**(130)**	**(130)**	**(130)**	**(135)**	**(135)**	**(135)**
Ours-SSE $\beta = 0.1$	**63.54**	**65.47**	**66.34**	**67.25**	**66.48**	**66.78**	**67.85**	**69.45**	**68.88**	**67.96**	**68.85**	**69.34**
	(120)	**(130)**	**(125)**	**(130)**	**(125)**	**(130)**	**(135)**	**(130)**	**(125)**	**(135)**	**(125)**	**(125)**
Ours-SSE $\beta = 0.3$	**67.76**	**69.72**	**68.68**	**69.25**	**68.37**	**70.14**	**70.36**	**71.48**	**69.34**	**71.33**	**72.05**	**72.87**
	(110)	**(135)**	**(135)**	**(130)**	**(120)**	**(134)**	**(130)**	**(135)**	**(125)**	**(135)**	**(125)**	**(125)**
Ours-SSE $\beta = 0.5$	**66.87**	**67.26**	**66.28**	**67.39**	**67.46**	**68.24**	**67.52**	**68.25**	**68.90**	**69.26**	**68.98**	**70.34**
	(120)	**(135)**	**(135)**	**(135)**	**(120)**	**(137)**	**(136)**	**(135)**	**(120)**	**(130)**	**(125)**	**(125)**
Ours-SSE $\beta = 0.7$	**65.33**	**66.12**	**65.76**	**66.15**	**66.55**	**67.32**	**67.33**	**68.12**	**67.87**	**68.95**	**68.37**	**68.77**
	(130)	**(130)**	**(130)**	**(135)**	**(125)**	**(125)**	**(125)**	**(130)**	**(130)**	**(130)**	**(125)**	**(125)**

Fig. 2. Recognition rate curves of all methods under different classifiers when the labeled images of each class $r = 5$. (a) NNC; (b) SVM; (c) LRC; (d) LLRC.

4.3   Experiments on FERET Dataset

The FERET face dataset contains over 1000 face images of 200 people. All images are in grayscale and manually cropped to 80×80 pixels. On FERET face dataset, the first 4 images of each person are employed for training and the remaining for testing. Similar to the experiment on ORL dataset, the first r ($r = 2, 3$) images are labeled. The NNC, SVM, LRC and LLRC are chosen for classification. First of all, the dimension of PCA subspace is set to 150. The parameters of this experiment are set as follows: In our method, the interclass neighbor parameter k is picked as 3, and the tuning parameter of adjacency matrices t is 2^3. For LPP, the neighbor parameter k is set to 3, and for MFA and GDE, the intraclass neighbor parameter k_1 is set to 1 and the interclass neighbor parameter k_2 is 200.
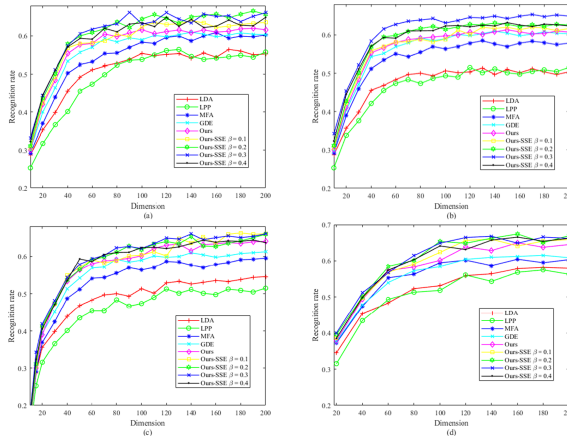
Fig. 3. Recognition rate curves of different methods under different classifiers when the labeled samples of each class $r = 3$. (a) NNC; (b) SVM; (c) LRC; (d) LLRC.
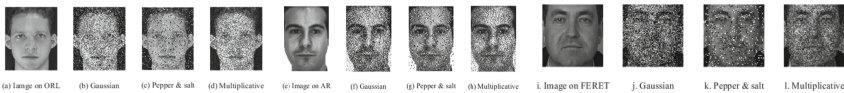
Table 3 reports the maximal recognition rates of different methods on FERET dataset. We can know that the overall recognition rate of all methods in this experiment has decreased, but compared with other methods, our method still has certain advantages. Furthermore, Fig. 3 represents the recognition rate curves of all methods under four classifiers as the projection axes varies. The results show that compared with the other three classifiers, all methods on LLRC will achieve better results, and when the value of β is set to 0.3, in most cases the curve of our extension method reaches the best performance, which also validates the efficiency of our extension method.

Table 3. Maximal recognition rates (%) and corresponding dimensions of different methods under four classifiers on FERET dataset.

Method	$r = 2$				$r = 3$			
	NNC	SVM	LRC	LLRC	NNC	SVM	LRC	LLRC
LDA	48.85	47.36	49.32	50.87	55.20	50.36	53.27	55.96
	(120)	(130)	(125)	(125)	(125)	(120)	(125)	(125)
LPP	53.00	49.63	51.76	54.16	56.33	51.48	53.25	56.14
	(125)	(130)	(130)	(120)	(125)	(125)	(120)	(120)
MFA	57.25	54.37	56.84	58.21	59.78	57.87	58.65	60.20
	(130)	(125)	(130)	(120)	(125)	(125)	(120)	(125)
GDE	58.13	57.33	58.67	59.72	59.89	60.32	59.96	60.74
	(125)	(130)	(125)	(125)	(130)	(130)	(125)	(130)
Ours	**58.87**	**58.67**	**59.17**	**60.78**	**61.63**	**60.75**	**63.28**	**63.86**
	(130)	**(125)**	**(130)**	**(125)**	**(130)**	**(125)**	**(130)**	**(125)**
Ours-SSE $\beta = 0.1$	**60.25**	**59.74**	**61.02**	**63.34**	**63.55**	**61.67**	**64.85**	**65.88**
	(135)	**(130)**	**(130)**	**(130)**	**(135)**	**(140)**	**(135)**	**(130)**
Ours-SSE $\beta = 0.2$	**63.11**	**61.24**	**62.95**	**64.72**	**65.67**	**62.57**	**65.28**	**66.25**
	(135)	**(135)**	**(130)**	**(130)**	**(135)**	**(135)**	**(140)**	**(135)**
Ours-SSE $\beta = 0.3$	**63.24**	**65.72**	**64.15**	**65.28**	**66.14**	**64.92**	**65.98**	**66.80**
	(135)	**(140)**	**(140)**	**(135)**	**(140)**	**(140)**	**(140)**	**(140)**
Ours-SSE $\beta = 0.4$	**62.17**	**60.36**	**63.24**	**63.62**	**64.90**	**62.34**	**64.05**	**65.73**
	(140)	**(140)**	**(140)**	**(135)**	**(140)**	**(140)**	**(145)**	**(145)**

4.4 Experiments Under Noise Condition

In view of the effect of noise, in this section, Gaussian noise (mean $\mu = 0$, variance $\sigma = 0.05$), pepper & salt noise (noise density $d = 0.1$) and multiplicative noise (mean $\mu = 0$, variance $\sigma = 0.1$) are applied to face samples on ORL, AR and FERET datasets. Face samples under three different noises on ORL, AR and FERET datasets are given in Fig. 4. The comparative results of the recognition rates of different methods on three datasets are shown in Tables 4, 5, and 6.

(a) Iamge on ORL   (b) Gaussian   (c) Pepper & salt   (d) Multiplicative   (e) Image on AR   (f) Gaussian   (g) Pepper & salt   (h) Multiplicative   i. Image on FERET   j. Gaussian   k. Pepper & salt   l. Multiplicative

Fig. 4. Face samples under three different noises on ORL, AR and FERET datasets.

It can be seen from Tables 4, 5, and, 6 that under three noise conditions, our method performs better than LDA, LPP, MFA and GDE, due to the improvement in the definition of the similarities of intraclass samples. Especially, since the low rank

representation of L_1-norm sparsity constraint is introduced, the recognition rate of our extension method has obvious advantages under three noise conditions, which further proves the anti-noise ability of our extension method. Furthermore, as the number of samples increases and the faces become more complex, our method still has a better and robust performance under three noise conditions.

Table 4. Maximal recognition rates (%) of different methods on ORL dataset under different noise conditions ($r = 3$).

Method	No noise			Gaussian noise			Salt & pepper noise			Multiplicative noise		
	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC
LDA	79.56	76.87	78.50	77.65	74.35	73.28	74.34	74.17	73.56	76.54	75.13	72.84
	(30)	(35)	(35)	(30)	(25)	(27)	(15)	(20)	(35)	(20)	(18)	(26)
LPP	81.32	82.15	82.50	78.56	76.72	70.85	72.25	73.24	74.18	78.32	72.19	73.27
	(45)	(40)	(40)	(35)	(28)	(32)	(30)	(25)	(27)	(25)	(28)	(20)
MFA	82.55	82.95	83.05	76.25	76.56	74.33	78.33	75.83	76.25	77.56	73.44	72.18
	(25)	(25)	(25)	(25)	(30)	(26)	(25)	(28)	(28)	(30)	(30)	(30)
GDE	82.87	83.12	84.00	75.33	77.84	73.28	77.45	74.36	73.57	76.33	75.17	76.34
	(35)	(30)	(30)	(25)	(32)	(25)	(30)	(31)	(32)	(35)	(23)	(25)
Ours	**83.56**	**83.20**	**84.33**	**80.67**	**78.68**	**75.03**	**78.87**	**76.33**	**77.18**	**79.67**	**76.25**	**76.84**
	(30)	**(35)**	**(35)**	**(38)**	**(30)**	**(26)**	**(35)**	**(25)**	**(19)**	**(30)**	**(31)**	**(27)**
Ours-SSE	84.77	84.65	84.75	84.25	79.34	76.56	82.67	77.54	78.65	81.25	76.87	77.25
$\beta = 0.3$	(30)	(30)	(30)	(35)	(25)	(30)	(30)	(20)	(30)	(25)	(30)	(35)

Table 5. Maximal recognition rates (%) of different methods on AR dataset under different noise conditions ($r = 5$).
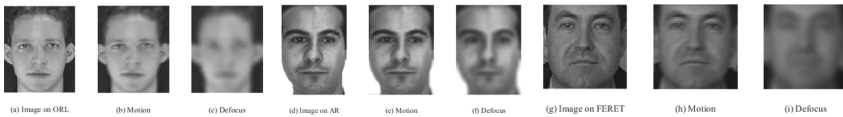
Method	No noise			Gaussian noise			Salt & pepper noise			Multiplicative noise		
	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC
LDA	61.55	60.25	61.87	57.50	54.25	53.75	54.72	53.86	55.47	57.36	57.43	55.71
	(115)	(120)	(125)	(120)	(125)	(115)	(125)	(115)	(130)	(120)	(105)	(125)
LPP	58.24	57.35	58.33	54.20	55.18	53.87	56.74	55.74	55.98	55.25	56.33	54.87
	(130)	(125)	(125)	(125)	(100)	(125)	(130)	(125)	(110)	(125)	(105)	(120)
MFA	64.95	64.88	67.75	59.34	57.65	56.73	58.37	56.35	57.76	57.87	57.48	58.67
	(130)	(125)	(125)	(125)	(130)	(125)	(125)	(115)	(130)	(130)	(105)	(120)
GDE	65.25	66.72	68.45	58.78	56.98	57.47	59.25	56.44	53.85	56.74	56.48	57.65
	(130)	(125)	(125)	(125)	(115)	(120)	(130)	(125)	(130)	(135)	(120)	(105)
Ours	**66.48**	**67.50**	**68.87**	**60.67**	**58.36**	**58.45**	**61.33**	**57.68**	**56.38**	**58.56**	**58.37**	**58.87**
	(135)	**(135)**	**(135)**	**(125)**	**(130)**	**(125)**	**(135)**	**(125)**	**(110)**	**(130)**	**(120)**	**(105)**
Ours-SSE	71.33	72.05	72.87	61.76	59.65	58.76	63.18	58.36	57.75	59.75	59.45	59.79
$\beta = 0.3$	(135)	(125)	(125)	(135)	(125)	(130)	(130)	(120)	(130)	(125)	(130)	(135)

Table 6. Maximal recognition rates (%) of different methods on FERET dataset under different noise conditions ($r = 3$).

Method	No noise			Gaussian noise			Salt & pepper noise			Multiplicative noise		
	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC
LDA	50.36	53.27	55.96	54.25	53.75	55.12	48.63	50.13	48.76	50.76	51.25	51.78
	(120)	(125)	(125)	(100)	(125)	(90)	(85)	(95)	(80)	(75)	(115)	(135)
LPP	51.48	53.25	56.14	54.73	54.33	55.36	45.76	48.49	47.36	49.38	52.48	53.76
	(125)	(120)	(120)	(95)	(125)	(130)	(100)	(110)	(120)	(95)	(90)	(120)
MFA	57.87	58.65	60.20	57.65	56.75	54.87	48.56	49.76	50.17	51.45	53.56	52.85
	(125)	(120)	(125)	(90)	(80)	(95)	(95)	(85)	(110)	(80)	(130)	(115)
GDE	60.32	59.96	60.74	58.37	57.67	56.84	46.93	49.35	47.85	52.37	53.64	52.87
	(130)	(125)	(130)	(85)	(120)	(125)	(90)	(120)	(110)	(80)	(120)	(125)
Ours	**60.75**	**63.28**	**63.86**	**59.76**	**58.87**	**57.43**	**49.67**	**51.28**	**50.96**	**54.85**	**56.37**	**54.84**
	(125)	**(130)**	**(125)**	**(110)**	**(115)**	**(110)**	**(80)**	**(125)**	**(100)**	**(75)**	**(120)**	**(100)**
Ours-SSE	**64.92**	**65.98**	**66.80**	**62.43**	**59.17**	**58.25**	**51.27**	**51.45**	**51.73**	**56.84**	**57.74**	**55.28**
$\beta = 0.3$	**(140)**	**(140)**	**(140)**	**(105)**	**(125)**	**(130)**	**(85)**	**(120)**	**(115)**	**(90)**	**(110)**	**(105)**

4.5   Experiments Under Blur Condition

In consideration of the blurred images in dataset, in this section, motion blur (the motion angle is 10° and the distance of motion is 5) and defocus blur (the radius is 5) are employed to the face images on ORL, AR and FERET datasets. Figure 5 shows face samples under different blur conditions on ORL, AR and FERET datasets. The comparative results of the recognition rate of all methods on three datasets are shown in Tables 7, 8, and 9.

(a) Image on ORL   (b) Motion   (c) Defocus   (d) Image on AR   (e) Motion   (f) Defocus   (g) Image on FERET   (h) Motion   (i) Defocus

Fig. 5. Face samples under two blurs on ORL, AR and FERET datasets.

According to the results in Tables 7, 8, and 9, it can be found that our method has an advantage over other methods in recognition rate under two blur conditions. Moreover, our extension method uses sparse representation to optimize the similarity of intraclass samples, which makes it perform better than GDE. However, as the number of samples increases and the faces become more complex, the recognition performance will be degraded. It can be seen from Table 9 that when the defocused blurred images are existing, the maximal recognition rate of various methods is low, not exceeding 40%, but the performance of our extension method still has a little advantage, which indicates the anti-interference ability of our extension method.

Table 7. Maximal recognition rates (%) of different methods on ORL dataset under different blur conditions.

Method	Original			Defocus blurred			Motion blurred		
	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC
LDA	79.56 (30)	76.87 (35)	78.50 (35)	52.18 (20)	51.37 (25)	53.24 (20)	53.61 (20)	51.48 (25)	53.74 (25)
LPP	81.32 (45)	82.15 (40)	82.50 (40)	50.37 (30)	51.48 (30)	52.65 (25)	43.56 (19)	47.96 (18)	46.47 (25)
MFA	82.55 (25)	82.95 (25)	83.05 (25)	53.65 (35)	52.87 (30)	53.76 (30)	55.78 (20)	54.76 (28)	56.83 (20)
GDE	82.87 (35)	83.12 (30)	84.00 (30)	53.87 (40)	54.73 (35)	53.76 (20)	46.52 (17)	48.75 (20)	50.34 (30)
Ours	**83.56** (30)	**83.20** (35)	**84.33** (35)	**57.34** (35)	**55.15** (35)	**54.87** (25)	**52.96** (25)	**55.46** (28)	**57.75** (30)
Ours-SSE $\beta = 0.3$	**84.77** (30)	**84.65** (30)	**84.75** (30)	**58.36** (25)	**55.84** (20)	**55.25** (30)	**57.87** (22)	**56.17** (25)	**58.25** (25)

Table 8. Maximal recognition rates (%) of different methods on AR dataset under different blur conditions.

Method	Original			Defocus blurred			Motion blurred		
	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC
LDA	61.55 (115)	60.25 (120)	61.87 (125)	49.75 (85)	53.84 (95)	46.72 (100)	50.88 (80)	49.76 (105)	51.74 (110)
LPP	58.24 (130)	57.35 (125)	58.33 (125)	48.26 (100)	50.25 (85)	48.83 (95)	49.20 (90)	48.75 (80)	50.37 (100)
MFA	64.95 (130)	64.88 (125)	67.75 (125)	50.33 (95)	47.26 (90)	51.87 (110)	52.83 (95)	50.78 (85)	51.34 (90)
GDE	65.25 (130)	66.72 (125)	68.45 (125)	49.76 (105)	48.20 (85)	49.25 (100)	51.56 (75)	49.95 (100)	50.83 (95)
Ours	**66.48** (135)	**67.50** (135)	**68.87** (135)	**50.68** (110)	**51.78** (90)	**52.50** (105)	**55.47** (95)	**52.67** (90)	**52.86** (90)
Ours-SSE $\beta = 0.3$	**71.33** (135)	**72.05** (125)	**72.87** (125)	**52.47** (115)	**52.15** (95)	**52.73** (105)	**57.83** (90)	**53.25** (75)	**54.28** (105)

Table 9. Maximal recognition rates (%) of different methods on FERET dataset under different blur conditions.

Method	Original			Defocus blurred			Motion blurred		
	SVM	LRC	LLRC	SVM	LRC	LLRC	SVM	LRC	LLRC
LDA	50.36	53.27	55.96	34.63	33.84	35.25	35.16	34.87	36.73
	(120)	(125)	(125)	(105)	(85)	(90)	(95)	(90)	(85)
LPP	51.48	53.25	56.14	33.75	35.25	35.78	35.78	38.26	36.48
	(125)	(120)	(120)	(110)	(75)	(85)	(105)	(90)	(95)
MFA	57.87	58.65	60.20	35.74	37.26	38.67	36.36	36.18	37.64
	(125)	(120)	(125)	(100)	(76)	(95)	(95)	(90)	(105)
GDE	60.32	59.96	60.74	34.95	38.20	36.47	35.47	36.76	38.78
	(130)	(125)	(130)	(96)	(85)	(105)	(105)	(85)	(100)
Ours	**60.75**	**63.28**	**63.86**	**36.53**	**38.78**	**39.16**	**36.50**	**39.65**	**39.74**
	(125)	**(130)**	**(125)**	**(90)**	**(90)**	**(95)**	**(100)**	**(105)**	**(90)**
Ours-SSE	**64.92**	**65.98**	**66.80**	**36.87**	**39.15**	**39.77**	**36.75**	**40.25**	**41.38**
$\beta = 0.3$	**(140)**	**(140)**	**(140)**	**(110)**	**(95)**	**(100)**	**(95)**	**(110)**	**(105)**

5   Conclusion

In this paper, we propose a novel method to eliminate the influence of outliers in face recognition. It builds the intrinsic graph in the light of the intraclass samples and their with-class mean, and constructs the penalty graph from the perspective of GDE. In addition, a low-rank representation with sparse constraints is introduced to explore the global low rank relationship of unlabeled samples. The experimental results on three datasets under different noise and blur conditions prove the effectiveness and robust of the proposed method.

References

1. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586–591. IEEE Computer Society Press, Los Alamitos (1991)
2. Belhumeur, P.N., Hepanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997)
3. He, X.F., Niyogi, P.: Locality preserving projections. In: Proceedings of the Neural Information Processing System Conference, pp. 100–115. MIT Press, Cambridge (2003)

4. Yan, S., Xu, D., Zhang, B., Zhang, H.J., Yang, Q., Lin, S.: Graph embedding and extensions: a general framework for dimensionality reduction. IEEE Trans. Pattern Anal. Mach. Intell. **29**(40), 40–50 (2007)
5. Huang, P., Li, T., Gao, G., Yang, G.: Feature extraction based on graph discriminant embedding and its applications to face recognition. Softw. Comput. **23**(16), 7015–7028 (2018). https://doi.org/10.1007/s00500-018-3340-5
6. Cai, D., He, X., Han, J.: Semi-supervised discriminant analysis. In: 11th IEEE International Conference on Computer Vision, Rio de Janeiro, Brazil, pp. 1–7. IEEE Press (2007)
7. Liu, G., Lin, Z., Yan, S., Sun, J., Yu, Y., Ma, Y.: Robust recovery of subspace structures by low-rank representation. IEEE Trans. Pattern Anal. Mach. Intell. **35**(1), 171–184 (2013)
8. Hai-Hong, G., Jie, M., Jian-Long, L.: Feature extraction based on low rank description. Sci. Technol. Eng. **16**(10), 200–204 (2016)
9. Naseem, I., Togneri, R., Bennamoun, M.: Linear regression for face recognition. IEEE Trans. Pattern Anal. Mach. Intell. **32**(11), 2106–2112 (2010)
10. Brown, D., Li, H.X., Gao, Y.S.: Locality-regularized linear regression for face recognition. In: Proceedings of the 21st International Conference on Pattern Recognition, Tsukuba, pp. 1586–1589 (2012)