

Article

Minimum Eigenvector Collaborative Representation Discriminant Projection for Feature Extraction

Haoshuang Hu and Da-Zheng Feng *

National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China; hshu@stu.xidian.edu.cn

* Correspondence: dzfeng@xidian.edu.cn

Received: 27 June 2020; Accepted: 22 August 2020; Published: 24 August 2020



Abstract: High-dimensional signals, such as image signals and audio signals, usually have a sparse or low-dimensional manifold structure, which can be projected into a low-dimensional subspace to improve the efficiency and effectiveness of data processing. In this paper, we propose a linear dimensionality reduction method—minimum eigenvector collaborative representation discriminant projection—to address high-dimensional feature extraction problems. On the one hand, unlike the existing collaborative representation method, we use the eigenvector corresponding to the smallest non-zero eigenvalue of the sample covariance matrix to reduce the error of collaborative representation. On the other hand, we maintain the collaborative representation relationship of samples in the projection subspace to enhance the discriminability of the extracted features. Also, the between-class scatter of the reconstructed samples is used to improve the robustness of the projection space. The experimental results on the COIL-20 image object database, ORL, and FERET face databases, as well as Isolet database demonstrate the effectiveness of the proposed method, especially in low dimensions and small training sample size.

Keywords: collaborative representation; discriminant projection; feature extraction; linear dimensionality reduction; subspace projection

1. Introduction

High-dimensional data widely exists in real applications, such as image recognition, information retrieval, etc. Particularly, in actual data processing problems, one often encounters the so-called high-dimensional small sample size (SSS) problem, in which the number of available samples is smaller than the dimensionality of the sample feature. Besides, high-dimensional data contains a lot of redundant information, and directly processing high-dimensional data will consume a lot of storage and computing resources. Fortunately, some previous research work [1–3] has shown that high-dimensional data is likely lying on or close to a low-dimensional submanifold space, which means that high-dimensional data can be projected into a low-dimensional subspace by some dimensionality reduction (DR) methods without losing important information. In the past few decades, numerous DR theories and methods have been proposed, and part of the work can be found in [4–8].

Principal component analysis (PCA) [4] and linear discriminant analysis (LDA) [5] are the most classic and popular DR methods, and PCA belongs to an unsupervised method while LDA is a supervised method. Despite their simplicity and effectiveness, they still suffer from some limitations in practice. For example, PCA is an unsupervised DR method and fails to provide discrimination information for different classes of data. LDA can find at most $C - 1$ meaningful discriminant projection directions because theoretical analysis shows that the rank of the between-class scatter matrix is at most $C - 1$, where C represents the number of classes. More importantly, they do not use the structural information of the samples, which greatly reduces the discriminativeness of the extracted features, especially when dealing with SSS problems. Most of the DR methods proposed

afterward are based on these two methods or their extended version but consider the structure of the samples. Some DR methods which focus on the local structure of the data have been developed to improve the discrimination performance of the low-dimensional space. A partial list of these methods includes the unsupervised ones such as locality preserving projections (LPP) [9], neighborhood preserving embedding (NPE) [10], and locally linear embedding (LLE) [1], and the supervised ones like marginal Fisher analysis (MFA) [11] and constrained discriminant neighborhood embedding (CDEN) [12]. Several other DR methods that utilize both the local and global structure of the data to improve recognition accuracy have been proposed, such as local Fisher discriminant analysis (LFDA) [13], locally linear discriminant embedding (LLDE) [14], and locality preserving discriminant projections (LPDP) [15].

Recently, some representation-based methods have been used for classification, including sparse representation classification (SRC) [16], sparsity preserving projections (SPP) [17], discriminant sparse neighborhood preserving embedding (DSNPE) [18], and discriminative sparsity preserving projections (DSPP) [19]. However, solving a sparse problem requires an iterative method, which is usually time-consuming. Another type of representation-based DR method, which uses L2 regularization and has a closed solution, has attracted wide attention. It has been proved in [20] that the collaborative representation classification (CRC), with higher efficiency, is competitive to SRC in terms of recognition accuracy. Since then, some collaborative representation based have been proposed, such as collaborative representation based projections (CRP) [21], regularized least square-based discriminative projections (RLSDP) [22], a collaborative representation reconstruction based projections (CRRP) [23], collaborative representation based discriminant neighborhood projections (CRDNP) [24] and collaborative preserving Fisher discriminant analysis (CPFDA) [25].

Except for the linear dimensionality reduction methods introduced above, a class of nonlinear dimensionality reduction methods has been proposed to deal with nonlinear dimensionality reduction problems. Most of the nonlinear DR methods directly use kernel trick to expand linear DR method, likes kernel PCA (KPCA) [26], kernel Fisher discriminant (KFD) [27], kernel direct discriminant analysis (KDDA) [28], and kernel collaborative representation-based projection (KCRP) [29]. Some other recent DR methods can be found in [30–32].

In this article, we study linear DR methods. Although some of the proposed methods make use of both the local structure and the global structure of the sample to extract features, when the reconstruction error of the sample is large, it is difficult for them to maintain the true structural similarity of the sample. In addition, most of the previous dimensionality reduction methods only consider the structural similarity of similar samples when looking for the projection subspace, while ignoring the structural similarity of different types of samples, which could also be used to improve the discriminability of the extracted features.

In order to extract the features with a strong discriminant, a minimum eigenvector collaborative representation discriminant projection (MECRDP) is proposed in this paper. The main contributions of our work are as follows. First, in the collaborative representation of samples, we not only use the information of the sample space but also consider the information of the sample eigenvector space. Specifically, we use the eigenvector corresponding to the smallest non-zero eigenvalue of the sample covariance matrix to reduce the collaborative representation error of each sample. Then, we maintain the collaborative representation relationship of the samples to improve the discriminability of the extracted features. Also, the between-class scatter of the reconstructed samples is used to improve the robustness of the projection subspace. Last, experimental results on four public databases show that MECRDP outperforms other DR methods in terms of recognition accuracy, especially in low dimensions and small training sample size.

The remainder of this paper is organized as follows. In Section 2, we briefly introduce the research work closely related to our method, including LDA and CRP. Section 3, we propose a minimum eigenvector collaborative representation discriminant projection to improve the discriminability of

the projection subspace. The experimental results are presented in Section 4. Finally, the conclusion remarks are given in Section 5.

2. Related Works

For simplicity, suppose a training samples set of C classes is denoted by $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathfrak{X}^{m \times n}$, where $\mathbf{x}_i \in \mathfrak{X}^m$ represents the i th sample, m is the dimension of the sample feature, and n is the number of samples. Besides, suppose the c th class contains n_c samples, and $\sum_{c=1}^C n_c = n$. The method we proposed in this paper has a great relationship with LDA and CRP. In what follows, we briefly review these two methods.

2.1. Linear Discriminant Analysis

The goal of LDA [5] is seeking a projection matrix so that the within-class scatter is minimized, and the between-class scatter is maximized simultaneously. According to the graph embedding [13], the projection matrix of LDA corresponds to the following two optimization problems, respectively,

$$\max_{\mathbf{P}} \sum_{i,j}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j\|_2^2 W_{i,j}^{(b)}, \quad (1)$$

$$\min_{\mathbf{P}} \sum_{i,j}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j\|_2^2 W_{i,j}^{(w)}, \quad (2)$$

where the weights $W_{i,j}^{(b)}$ and $W_{i,j}^{(w)}$ are defined as, respectively,

$$W_{i,j}^{(b)} = \begin{cases} 1/n - 1/n_c, & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the } c\text{th class,} \\ 1/n, & \text{otherwise.} \end{cases} \quad (3)$$

$$W_{i,j}^{(w)} = \begin{cases} 1/n_c, & \text{if } \mathbf{x}_i \text{ and } \mathbf{x}_j \text{ belong to the same class,} \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

Using some algebraic transform, we can rewrite (1) and (2) as

$$\begin{aligned} \max_{\mathbf{P}} \sum_{i,j}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j\|_2^2 W_{i,j}^{(b)} &= \text{tr} \left(\mathbf{P}^T \left(\sum_{i,j=1}^n (\mathbf{x}_i - \mathbf{x}_j) W_{i,j}^{(b)} (\mathbf{x}_i - \mathbf{x}_j)^T \right) \mathbf{P} \right) \\ &= \text{tr} \left(\mathbf{P}^T \left(\sum_{i,j=1}^n \left(\mathbf{x}_i W_{i,j}^{(b)} \mathbf{x}_i^T - \mathbf{x}_i W_{i,j}^{(b)} \mathbf{x}_j^T - \mathbf{x}_j W_{i,j}^{(b)} \mathbf{x}_i^T + \mathbf{x}_j W_{i,j}^{(b)} \mathbf{x}_j^T \right) \right) \mathbf{P} \right) \\ &= 2 \text{tr} \left(\mathbf{P}^T \left(\mathbf{X} \mathbf{D}^{(b)} \mathbf{X}^T - \mathbf{X} \mathbf{W}^{(b)} \mathbf{X}^T \right) \mathbf{P} \right) \\ &= 2 \text{tr} \left(\mathbf{P}^T \left(\mathbf{X} \left(\mathbf{D}^{(b)} - \mathbf{W}^{(b)} \right) \mathbf{X}^T \right) \mathbf{P} \right) \\ &= 2 \text{tr} \left(\mathbf{P}^T \left(\mathbf{X} \mathbf{L}^{(b)} \mathbf{X}^T \right) \mathbf{P} \right) \\ &= \text{tr} \left(\mathbf{P}^T \mathbf{S}_b \mathbf{P} \right) \end{aligned} \quad (5)$$

$$\begin{aligned} \min_{\mathbf{P}} \sum_{i,j}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \mathbf{x}_j\|_2^2 W_{i,j}^{(w)} &= \text{tr} \left(\mathbf{P}^T \left(\sum_{i,j=1}^n (\mathbf{x}_i - \mathbf{x}_j) W_{i,j}^{(w)} (\mathbf{x}_i - \mathbf{x}_j)^T \right) \mathbf{P} \right) \\ &= \text{tr} \left(\mathbf{P}^T \left(\sum_{i,j=1}^n \left(\mathbf{x}_i W_{i,j}^{(w)} \mathbf{x}_i^T - \mathbf{x}_i W_{i,j}^{(w)} \mathbf{x}_j^T - \mathbf{x}_j W_{i,j}^{(w)} \mathbf{x}_i^T + \mathbf{x}_j W_{i,j}^{(w)} \mathbf{x}_j^T \right) \right) \mathbf{P} \right) \\ &= 2 \text{tr} \left(\mathbf{P}^T \left(\mathbf{X} \mathbf{D}^{(w)} \mathbf{X}^T - \mathbf{X} \mathbf{W}^{(w)} \mathbf{X}^T \right) \mathbf{P} \right) \\ &= 2 \text{tr} \left(\mathbf{P}^T \left(\mathbf{X} \left(\mathbf{D}^{(w)} - \mathbf{W}^{(w)} \right) \mathbf{X}^T \right) \mathbf{P} \right) \\ &= 2 \text{tr} \left(\mathbf{P}^T \left(\mathbf{X} \mathbf{L}^{(w)} \mathbf{X}^T \right) \mathbf{P} \right) \\ &= \text{tr} \left(\mathbf{P}^T \mathbf{S}_w \mathbf{P} \right) \end{aligned} \quad (6)$$

where $tr(\cdot)$ denotes matrix trace, $\mathbf{S}_b = 2\mathbf{X}\mathbf{L}^{(b)}\mathbf{X}^T$ and $\mathbf{S}_w = 2\mathbf{X}\mathbf{L}^{(w)}\mathbf{X}^T$ are the between-class scatter matrix, and the within-class scatter matrix, respectively. $\mathbf{L}^{(b)} = \mathbf{D}^{(b)} - \mathbf{W}^{(b)}$ and $\mathbf{L}^{(w)} = \mathbf{D}^{(w)} - \mathbf{W}^{(w)}$ are the Laplacian matrices, in which $\mathbf{D}^{(b)}$ and $\mathbf{D}^{(w)}$ are diagonal matrices with their diagonal entries are $\mathbf{D}_{i,i}^{(b)} = \sum_{j=1}^n W_{i,j}^{(b)}$ and $\mathbf{D}_{i,i}^{(w)} = \sum_{j=1}^n W_{i,j}^{(w)}$, respectively.

Using (5) and (6), the objective function of LDA can be modeled as

$$\max_{\mathbf{P}} \frac{tr(\mathbf{P}^T \mathbf{S}_b \mathbf{P})}{tr(\mathbf{P}^T \mathbf{S}_w \mathbf{P})}, \quad (7)$$

2.2. Collaborative Representation Based Projections

CRP [21] is an unsupervised discriminant projection method based on L2 regularized least squares. The collaborative representation coefficients of the i th sample is gotten by solving the following optimization problem

$$\min_{\mathbf{r}_i} \|\mathbf{x}_i - \mathbf{X}\mathbf{r}_i\|_2^2 + \lambda \|\mathbf{r}_i\|_2^2, \quad s.t. \quad \mathbf{e}_i^T \mathbf{r}_i = 0, \quad (8)$$

where $\mathbf{r}_i \in \mathcal{X}^n$ is the collaborative representation coefficients of the i th sample, $\lambda > 0$ is a regularization parameter, and $\mathbf{e}_i = [0_1, \dots, 0_{i-1}, 1, 0_{i+1}, \dots, 0_n]^T$. The constraint in (1) means the i th sample is represented on all the samples other than itself. The optimal solution of (1) can be easily achieved by using the Lagrangian multiplier method as

$$\mathbf{r}_i = \mathbf{Q} \left(\mathbf{X}\mathbf{x}_i - \frac{\mathbf{e}_i^T \mathbf{Q}\mathbf{X}\mathbf{x}_i}{\mathbf{e}_i^T \mathbf{Q}\mathbf{e}_i} \mathbf{e}_i \right), \quad (9)$$

where $\mathbf{Q} = (\mathbf{X}^T \mathbf{X} + \lambda \mathbf{I})^{-1}$ and $\mathbf{I} \in \mathcal{X}^{n \times n}$ is a identity matrix.

Using the collaborative representation coefficients, the optimal projection matrix $\mathbf{P} \in \mathcal{X}^{m \times d}$ ($d < m$) of CRP is obtained by solving the following two optimization problems, simultaneously

$$\min_{\mathbf{P}} \sum_{i=1}^n \|\mathbf{P}^T \mathbf{x}_i - \sum_{j=1}^n r_{i,j} \mathbf{P}^T \mathbf{x}_j\|_2^2, \quad (10)$$

$$\max_{\mathbf{P}} \sum_{i=1}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \bar{\mathbf{x}}\|_2^2, \quad (11)$$

where $r_{i,j}$ denotes the j th entry of \mathbf{r}_i , and $\bar{\mathbf{x}} = (1/n) \sum_{i=1}^n \mathbf{x}_i$ is the mean of the samples.

With some algebraic formulations, (10) and (11) can be, respectively, rewritten as

$$\begin{aligned} \min_{\mathbf{P}} \sum_{i=1}^n \|\mathbf{P}^T \mathbf{x}_i - \sum_{j=1}^n r_{i,j} \mathbf{P}^T \mathbf{x}_j\|_2^2 &= \sum_{i=1}^n \|\mathbf{P}^T \mathbf{X}\mathbf{e}_i - \mathbf{P}^T \mathbf{X}\mathbf{r}_i\|_2^2 \\ &= tr(\mathbf{P}^T \mathbf{X} (\sum_{i=1}^n (\mathbf{e}_i - \mathbf{r}_i)(\mathbf{e}_i - \mathbf{r}_i)^T) \mathbf{X}^T \mathbf{P}) \\ &= tr(\mathbf{P}^T \mathbf{X} (\mathbf{I} - \mathbf{R}) (\mathbf{I} - \mathbf{R})^T \mathbf{X}^T \mathbf{P}) \\ &= tr(\mathbf{P}^T \mathbf{S}_L \mathbf{P}) \end{aligned} \quad (12)$$

$$\begin{aligned} \max_{\mathbf{P}} \sum_{i=1}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \bar{\mathbf{x}}\|_2^2 &= tr(\mathbf{P}^T (\sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T) \mathbf{P}) \\ &= tr(\mathbf{P}^T \mathbf{S}_T \mathbf{P}) \end{aligned} \quad (13)$$

where $\mathbf{R} = [\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_n]$ is the representation coefficients matrix, $\mathbf{S}_L = (\mathbf{I} - \mathbf{R})(\mathbf{I} - \mathbf{R})^T$ denotes the local scatter matrix and $\mathbf{S}_T = \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T$ is the total scatter matrix.

Then, the optimal projection matrix of the CRP is gotten by solving the following optimization problem

$$\min_{\mathbf{P}} \frac{\text{tr}(\mathbf{P}^T \mathbf{S}_L \mathbf{P})}{\text{tr}(\mathbf{P}^T \mathbf{S}_T \mathbf{P})} \quad (14)$$

3. Minimum Eigenvector Collaborative Representation Discriminant Projection

LDA uses between-class scatter and within-class scatter to improve the discrimination of extracted features, but it ignores the structural relationship of the samples, resulting in a decline in the discrimination of the features in the projection subspace. Although CRP takes into account the structural relationship of the sample, it does not use the class information of the sample, which is not conducive to the improvement of recognition accuracy. In particular, when the number of samples is small, there will be a large representation error, $\|\mathbf{x}_i - \mathbf{X}\mathbf{r}_i\|_2^2$, which fails to maintain the similarity of the sample structure. Here, we propose a new feature extraction method to alleviate the problems mentioned above.

3.1. Method Proposed

The sample is represented on all samples other than itself; therefore, this may cause a large reconstruction error when the number of samples is small. However, any eigenvector corresponding to a non-zero eigenvalue of the sample covariance matrix contains partial information of each sample. In order to improve the reconstruction error of the sample, but only use very little information of the represented sample, we consider constructing an expanded sample matrix $\tilde{\mathbf{X}}$ with the eigenvector corresponding to the smallest non-zero eigenvalue of the sample covariance matrix. Let \mathbf{x}_v be the eigenvector corresponding to the smallest non-zero eigenvalue of $\mathbf{X}\mathbf{X}^T$, then the expanded sample matrix $\tilde{\mathbf{X}}$ is defined as $\tilde{\mathbf{X}} = [\mathbf{X}, \mathbf{x}_v]$. Similar to (8), the collaborative representation coefficients of the i th sample is achieved by

$$\min_{\tilde{\mathbf{r}}_i} \|\mathbf{x}_i - \tilde{\mathbf{X}}\tilde{\mathbf{r}}_i\|_2^2 + \lambda \|\tilde{\mathbf{r}}_i\|_2^2, \quad \text{s.t. } \tilde{\mathbf{e}}_i^T \tilde{\mathbf{r}}_i = 0, \quad (15)$$

where $\tilde{\mathbf{r}}_i \in \mathcal{R}^{n+1}$ is the collaborative representation coefficient and $\tilde{\mathbf{e}}_i = [\mathbf{e}_i^T, 0]^T$. Let $\tilde{\mathbf{R}} = [\tilde{\mathbf{r}}_1, \tilde{\mathbf{r}}_2, \dots, \tilde{\mathbf{r}}_n]$ be the collaborative representation coefficient matrix.

Then, the sample \mathbf{x}_i can be reconstructed as

$$\hat{\mathbf{x}}_i = \tilde{\mathbf{X}}\tilde{\mathbf{r}}_i, \quad (16)$$

and the reconstructed sample matrix is

$$\hat{\mathbf{X}} = [\hat{\mathbf{x}}_1, \hat{\mathbf{x}}_2, \dots, \hat{\mathbf{x}}_n] = \mathbf{X}[\hat{\mathbf{r}}_1, \hat{\mathbf{r}}_2, \dots, \hat{\mathbf{r}}_n] = \tilde{\mathbf{X}}\tilde{\mathbf{R}} \quad (17)$$

In order to maintain the reconstruction similarity of the samples and keep within-class compactness, we modified the optimization problem in (2) as

$$\min_{\mathbf{P}} \sum_{i,j}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \hat{\mathbf{x}}_j\|_2^2 W_{i,j}^{(w)}, \quad (18)$$

Similar to (6), we can simplify (18) as

$$\begin{aligned}
 \min_{\mathbf{P}} \sum_{i,j}^n \|\mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \hat{\mathbf{x}}_j\|_2^2 W_{i,j}^{(w)} &= \text{tr} \left(\mathbf{P}^T \left(\sum_{i,j}^n (\mathbf{x}_i - \hat{\mathbf{x}}_j) W_{i,j}^{(w)} (\mathbf{x}_i - \hat{\mathbf{x}}_j)^T \right) \mathbf{P} \right) \\
 &= \text{tr} \left(\mathbf{P}^T \left(\sum_{i,j}^n (\mathbf{x}_i W_{i,j}^{(w)} \mathbf{x}_i^T - \hat{\mathbf{x}}_j W_{i,j}^{(w)} \mathbf{x}_i^T - \mathbf{x}_i W_{i,j}^{(w)} \hat{\mathbf{x}}_j^T + \hat{\mathbf{x}}_j W_{i,j}^{(w)} \hat{\mathbf{x}}_j^T) \right) \mathbf{P} \right) \\
 &= \text{tr} \left(\mathbf{P}^T (\mathbf{X} \mathbf{D}^{(w)} \mathbf{X}^T - \hat{\mathbf{X}} \mathbf{W}^{(w)} \mathbf{X}^T - \mathbf{X} \mathbf{W}^{(w)} \hat{\mathbf{X}}^T + \hat{\mathbf{X}} \mathbf{D}^{(w)} \hat{\mathbf{X}}^T) \mathbf{P} \right) \\
 &= \text{tr} \left(\mathbf{P}^T \widehat{\mathbf{S}}_w \mathbf{P} \right)
 \end{aligned} \tag{19}$$

where $\widehat{\mathbf{S}}_w = \mathbf{X} \mathbf{D}^{(w)} \mathbf{X}^T - \hat{\mathbf{X}} \mathbf{W}^{(w)} \mathbf{X}^T - \mathbf{X} \mathbf{W}^{(w)} \hat{\mathbf{X}}^T + \hat{\mathbf{X}} \mathbf{D}^{(w)} \hat{\mathbf{X}}^T$ is the collaborative within-class scatter matrix and $\mathbf{W}^{(w)}, \mathbf{D}^{(w)}$ are defined in (4) and (6).

In what follows, we define the collaborative reconstructed between-class scatter matrix as

$$\begin{aligned}
 \hat{\mathbf{S}}_b &= \sum_{i,j}^n (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j) W_{i,j}^{(b)} (\hat{\mathbf{x}}_i - \hat{\mathbf{x}}_j)^T \\
 &= \sum_{i,j}^n (\hat{\mathbf{x}}_i W_{i,j}^{(b)} \hat{\mathbf{x}}_i^T - \hat{\mathbf{x}}_j W_{i,j}^{(b)} \hat{\mathbf{x}}_i^T - \hat{\mathbf{x}}_i W_{i,j}^{(b)} \hat{\mathbf{x}}_j^T + \hat{\mathbf{x}}_j W_{i,j}^{(b)} \hat{\mathbf{x}}_j^T) \\
 &= \hat{\mathbf{X}} \mathbf{D}^{(b)} \hat{\mathbf{X}}^T - \hat{\mathbf{X}} \mathbf{W}^{(b)} \hat{\mathbf{X}}^T - \hat{\mathbf{X}} \mathbf{W}^{(b)} \hat{\mathbf{X}}^T + \hat{\mathbf{X}} \mathbf{D}^{(b)} \hat{\mathbf{X}}^T \\
 &= 2\hat{\mathbf{X}} (\mathbf{D}^{(b)} - \mathbf{W}^{(b)}) \hat{\mathbf{X}}^T \\
 &= 2\hat{\mathbf{X}} \mathbf{L}^{(b)} \hat{\mathbf{X}}^T
 \end{aligned} \tag{20}$$

where $\mathbf{L}^{(b)}$ has been defined (5).

Based on the Fisher criterion, considering both the between-class scatter matrix \mathbf{S}_b and the collaborative reconstructed between-class scatter matrix $\hat{\mathbf{S}}_b$, the objective function of the proposed MECRDP is formulated as

$$\min_{\mathbf{P}} \frac{\text{tr}(\mathbf{P}^T \widehat{\mathbf{S}}_w \mathbf{P})}{\text{tr}(\mathbf{P}^T \widehat{\mathbf{S}}_b \mathbf{P})}, \tag{21}$$

where $\widetilde{\mathbf{S}}_b = \alpha \mathbf{S}_b + (1 - \alpha) \hat{\mathbf{S}}_b$ and $\alpha \in [0, 1]$ is a balance factor between \mathbf{S}_b and $\hat{\mathbf{S}}_b$. According to the definition of \mathbf{S}_b , we know that maximizing $\text{tr}(\mathbf{P}^T \mathbf{S}_b \mathbf{P})$ can improve the discrimination of the projection matrix \mathbf{P} . However, when the dimension of the subspace exceeds a certain size, as the dimension of the projection subspace increases, the discriminant performance is more affected by noise. Consider that the collaborative representation between samples could better describe the similarity of samples in the structure, thereby reducing the impact of the noise, then we can use $\hat{\mathbf{S}}_b$ to improve the robustness of the projection matrix.

The problem (21) is a generalized eigenvalue problem, whose optimal solution could be achieved by the generalized eigenvalue decomposition as follows

$$\widehat{\mathbf{S}}_w \mathbf{p}_i = \lambda_i \widetilde{\mathbf{S}}_b \mathbf{p}_i, \tag{22}$$

where λ_i and \mathbf{p}_i are the eigenvalue and corresponding eigenvector, respectively. Then the projection matrix \mathbf{P} is composed of the eigenvectors corresponding to d —the smallest non-zero eigenvalues—that is $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d]$.

Using \mathbf{x}_v , different from CRP, we not only consider the information in the sample space but also use part of the information in the feature value space, so we can get a smaller representation error, $\|\mathbf{x}_i - \mathbf{X} \mathbf{r}_i\|_2^2$ and keep more collaborative reconstruction information in the feature projection space. In addition, CRP uses the local divergence and global divergence of the sample to obtain

the discriminativeness of the extracted features, and in our method, we use the label information of the sample to keep the sample in the projection space with a small within-class scatter and larger between-class scatter, thereby improving the discriminativeness of the extracted features.

3.2. Algorithmic Procedures

For simplicity, we summarize the algorithmic procedures of the proposed MECRDP as follows:

- Step 1: Project the original high-dimensional sample into an intermediate subspace to remove noise and useless information by PCA, and get the projection matrix \mathbf{P}_V . For simplicity, we still utilize \mathbf{X} to represent the training samples after the projection.
- Step 2: Find the eigenvector \mathbf{x}_v corresponding to the smallest non-zero eigenvalue of $\mathbf{X}\mathbf{X}^T$, then the expanded sample matrix is $\tilde{\mathbf{X}} = [\mathbf{X}, \mathbf{x}_v]$.
- Step 3: Compute the collaborative representation coefficients $\tilde{\mathbf{R}}$ by (15) and then reconstruct the samples matrix $\hat{\mathbf{X}}$ by $\hat{\mathbf{X}} = \tilde{\mathbf{X}}\tilde{\mathbf{R}}$.
- Step 4: Compute the collaborative within-class scatter matrix $\hat{\mathbf{S}}_w$, the between-class scatter matrix \mathbf{S}_b and the collaborative reconstructed between-class scatter matrix $\hat{\mathbf{S}}_b$ by (19), (5) and (20), respectively.
- Step 5: Perform generalized eigenvalue decomposition by $\hat{\mathbf{S}}_w\mathbf{p}_i = \lambda_i\hat{\mathbf{S}}_b\mathbf{p}_i$, and use eigenvectors corresponding to the smallest d eigenvalues to construct the projection matrix $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d]$.
- Step 6: For any input sample $\mathbf{x} \in \mathfrak{R}^n$, its low-dimensional projection is $\mathbf{y} = \mathbf{P}^T\mathbf{P}_V^T\mathbf{x}$.

It is obvious that the proposed method has no iterative steps and its optimal projection matrix can be analytically obtained. Without considering the first step, which is the data preprocessing process that most linear DR methods need to perform, we simply analyze the computational complexity of our proposed method. In step 2, it costs about $O(m^3)$ to find the eigenvector corresponding to the smallest non-zero eigenvalue of $\mathbf{X}\mathbf{X}^T$, step 3 costs about $O(mn^2 + n^3)$, and step 4 needs about $O(mn^2 + m^2n)$, finding the optimal projection matrix in step 5 requires $O(n^3)$, with the total computational complexity being about $O(m^3 + m^2n + mn^2 + n^3)$.

4. Experiments

In this section, some experiments are conducted to evaluate the performance of the proposed MECRDP. We compare the recognition performance of these methods on four public databases, including an image object database COIL20 [33], and two face databases ORL [34], FERET [35], and Isolet [36].

4.1. Preprocessing and Parameter Setting

In our experiments, all the images were converted to grayscale images and were resized. In the following experiments, each sample is stacked into a column vector in column order and is normalized. Besides, to avoid the singular problem caused by small sample size problems, we reduce the dimension by remaining 98% data energy by PCA. Each database is randomly divided into a training set and test set, say randomly selecting s -samples from each class to form a training set, and the remaining samples are used as a test set. For simplicity, without additional explanation, the nearest neighbor method is used to classify the test samples. Considering that most methods compared have adjustable parameters, for the sake of fairness, we look for the satisfactory parameters in a larger range for them. For example, we find the neighbor parameter for LLDE from 1 to $s - 1$, with and empirically set the rescaling coefficient to 1. In MFA, the parameter for the intrinsic graph is empirically set as $s - 1$ and the parameter for the penalty graph is chosen from $\{1C, 3C, 5C, 7C, 9C\}$.

Particularly, each experiment was independently repeated 20 times to avoid the bias caused by random selection, and the average results are recorded and reported. All the experiments are

implemented in MATLAB R2011b installed on a personal computer (Intel Core i5-4590, 3.30 GHz, 8 Gb RAM).

4.2. Experimental Results and Analysis

4.2.1. Experimental on COIL20 Database

The COIL20 [33] is an object recognition database which has 1440 images from 20 objects with 72 images in each object. We resize each image into 32×32 and randomly select 4, 6, 8, 10 images from each class to form the training sample. The parameters λ and α in MECRDP are set to 0.5 and 0.1, respectively.

Table 1 records the maximum average recognition accuracy, the standard derivations, the corresponding dimension, and the average running time of each method on COIL20. Figure 1 shows the average recognition accuracy versus the subspace dimensions on COIL20. From the results in Table 1 and Figure 1, we observe that the proposed MECRDP achieves better recognition accuracy than other methods on COIL20 database. Compared with other methods, the maximum average recognition accuracy of MECRDP is improved about 3%. Table 1 shows the running time of our method is almost the same as other methods, which implies the efficiency of our method. In addition, the optimal recognition result of MECRDP is usually obtained in a lower projection dimension than other methods. The same result could be verified more intuitively in Figure 1. The result in Figure 1 shows that when the dimension of the projection subspace is low, MECRDP is significantly better than other methods. This result means that the features extracted by our method have obvious discriminativeness even in the low-dimensional space. Table 2 shows the maximum average recognition accuracy with different dimension by using nearest neighbor classifier. To evaluate the effectiveness of MECRDP further, we compare it with other dimension reduction methods under another classifier two-layer neural network; the results are shown in Table 3. Tables 2 and 3 show that the recognition accuracy achieved by the neural network is not as good as that of the nearest neighbor method, but the proposed method always achieves the best results. The result of COIL20 verifies the effectiveness of our method for feature extraction.

Table 1. The maximum average recognition accuracy (%) \pm the standard derivations (%), the corresponding dimension, and the average running time (seconds) in parentheses of each method in the COIL20 database.

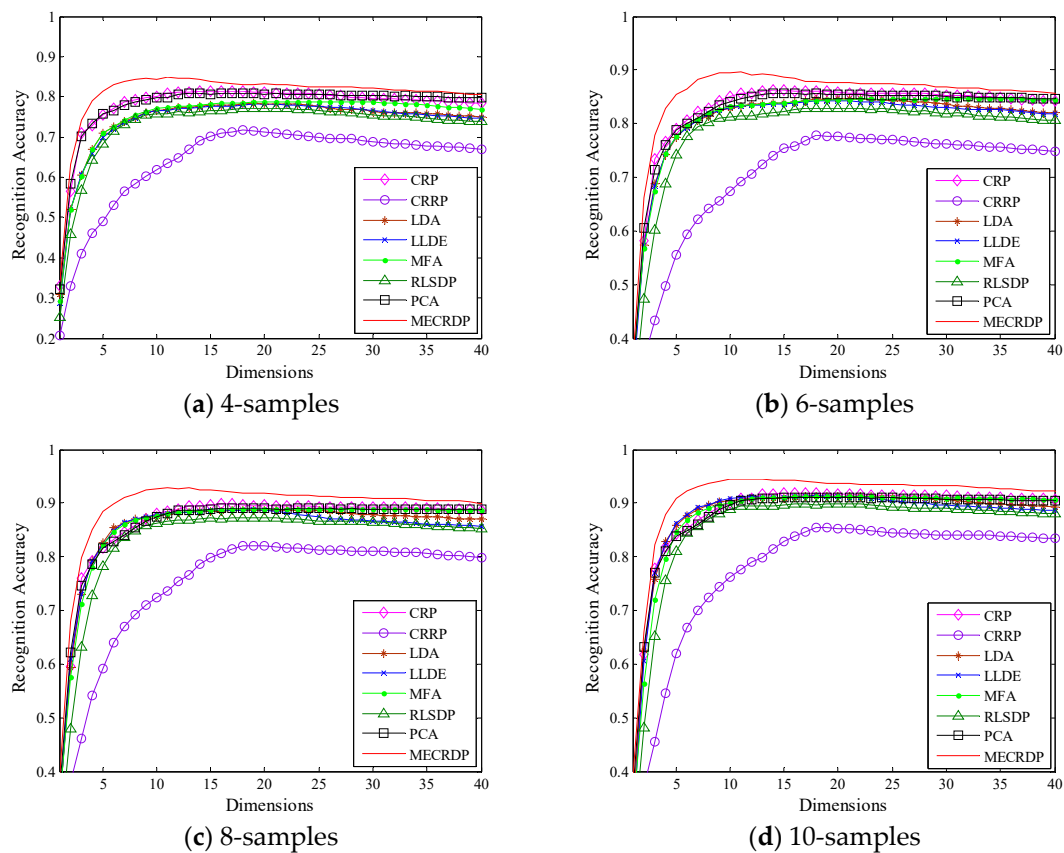
Methods	4-Samples	6-Samples	8-Samples	10-Samples
CRP	81.51 \pm 2.08 (16, 0.008)	85.35 \pm 2.12 (14, 0.011)	89.90 \pm 1.07 (16, 0.017)	91.84 \pm 1.39 (17, 0.025)
CRRP	71.70 \pm 2.90 (18, 0.007)	77.75 \pm 2.59 (18, 0.016)	82.12 \pm 2.04 (19, 0.026)	85.47 \pm 1.54 (19, 0.042)
LDA	78.36 \pm 2.93 (19, 0.008)	84.70 \pm 2.28 (22, 0.008)	88.73 \pm 1.74 (22, 0.016)	91.30 \pm 1.55 (19, 0.017)
LLDE	78.22 \pm 3.16 (20, 0.009)	84.27 \pm 2.39 (19, 0.015)	88.78 \pm 1.73 (19, 0.019)	91.38 \pm 1.59 (18, 0.029)
MFA	78.84 \pm 2.86 (26, 0.005)	84.65 \pm 2.13 (32, 0.010)	88.95 \pm 1.80 (21, 0.017)	91.33 \pm 1.35 (18, 0.023)
RLSDP	77.25 \pm 3.32 (19, 0.005)	82.91 \pm 2.50 (18, 0.010)	87.30 \pm 1.85 (20, 0.013)	89.83 \pm 1.73 (16, 0.020)
PCA	80.84 \pm 1.75 (18, 0.004)	85.67 \pm 1.68 (14, 0.006)	89.10 \pm 1.38 (23, 0.012)	91.14 \pm 1.25 (18, 0.016)
MECRDP	84.89 \pm 2.26 (11, 0.007)	89.60 \pm 2.00 (11, 0.010)	92.79 \pm 1.36 (13, 0.018)	94.50 \pm 1.07 (13, 0.026)

Table 2. The maximum average recognition accuracy (%) in the COIL20 database with different dimension by using nearest neighbor classifier.

Dimension	5	10	15	20	25	30	35	40
CRP	84.31	90.13	91.82	91.75	91.52	91.38	91.05	90.72
CRRP	62.02	76.25	82.75	85.29	84.48	84.15	83.83	83.41
LDA	85.83	90.47	90.92	91.27	91.07	90.59	89.80	89.22
LLDE	86.27	90.80	91.14	91.19	90.45	89.65	89.07	88.61
MFA	84.73	90.10	91.13	91.28	91.06	90.96	90.88	90.73
RLSDP	81.14	88.89	89.76	89.77	89.23	89.00	88.38	87.97
PCA	83.90	89.66	90.98	91.02	90.99	90.76	90.61	90.46
MECRDP	90.80	94.40	94.27	93.71	93.42	93.20	93.69	92.20

Table 3. The maximum average recognition accuracy (%) on the COIL20 database with different dimension by using neural network classifier.

Dimension	5	10	15	20	25	30	35	40
CRP	77.21	87.16	89.63	89.12	88.15	87.48	86.37	86.05
CRRP	60.76	74.60	79.71	80.67	80.57	79.98	79.27	78.78
LDA	84.46	87.19	85.19	83.10	82.73	82.65	81.71	81.84
LLDE	85.48	87.75	85.13	83.51	82.62	82.66	82.57	80.67
MFA	83.62	87.52	86.19	83.40	82.98	82.30	82.00	81.44
RLSDP	79.93	85.79	84.63	81.58	81.85	81.88	81.57	81.02
PCA	77.71	86.70	89.85	89.38	88.25	87.66	86.58	85.83
MECRDP	88.77	91.67	91.36	90.85	89.78	89.14	88.03	87.49

**Figure 1.** The average recognition accuracy vs. the subspace dimensions of each method in the COIL20 database.

4.2.2. Experiment on the ORL Database

The ORL [34] is the face database and consists of 400 images from 40 individuals with 10 images for each individual. These images are varying in lighting, facial expressions, and details. In our experiments, we cropped and resized each image to 32×32 , and randomly selected 3, 4, 5, 6 images from each class to form the training set. We set λ and α in MECRDP to 0.05 and 0.9, respectively.

Table 4 reports the maximum average recognition accuracy, the standard derivations, the corresponding dimension, and the average running time of each method on the ORL database. Tables 5 and 6 list the average recognition rates obtained by the nearest neighbor classifier and neural network classifier for each method, respectively. Figure 2 plots the average recognition accuracy versus the subspace dimensions on ORL. The results show that our method is superior to other methods, especially when the number of training samples is small; the performance improvement of our method is more obvious. For example, when the training samples of each class are 3 and 4, the maximum

average recognition accuracy of MECRDP is improved by about 4.1% and 3.1%, respectively. Besides, we note that, for each method, the recognition accuracy is increased as the training sample increases, while that of MECRDP always outperforms other methods. Figure 2 shows that as the projection dimension increases, the performance of some methods, such as LDA, LLDE, and RLSDP, will decrease because they can extract some noise information when the projection dimension is large. However, our method has shown a certain degree of robustness. Comparing Tables 5 and 6, we find that the recognition accuracy obtained by the nearest neighbor classifier is better than that of the neural network classifier. What's more, the proposed MECRDP can achieve the best results in almost all the selected feature dimensions. The experimental results on the ORL database verify the advantages of the proposed algorithm.

Table 4. The maximum average recognition accuracy (%) \pm the standard derivations (%), the corresponding dimension, and the average running time (seconds) in parentheses of each method in the ORL database.

Methods	3-Samples	4-Samples	5-Samples	6-Samples
CRP	79.55 \pm 3.12 (42, 0.019)	84.48 \pm 2.32 (52, 0.024)	88.08 \pm 2.17 (62, 0.044)	90.56 \pm 3.06 (72, 0.058)
CRRP	81.02 \pm 2.79 (40, 0.023)	87.46 \pm 2.25 (40, 0.034)	90.58 \pm 1.85 (40, 0.074)	92.75 \pm 1.71 (38, 0.085)
LDA	84.20 \pm 2.87 (38, 0.013)	88.75 \pm 4.18 (42, 0.018)	92.95 \pm 1.85 (42, 0.030)	95.25 \pm 1.68 (44, 0.035)
LLDE	78.66 \pm 3.26 (34, 0.016)	89.23 \pm 2.00 (40, 0.025)	92.80 \pm 1.57 (40, 0.045)	94.97 \pm 1.62 (44, 0.054)
MFA	84.07 \pm 3.19 (58, 0.018)	89.50 \pm 1.43 (48, 0.027)	92.90 \pm 1.55 (42, 0.034)	95.03 \pm 1.17 (48, 0.045)
RLSDP	82.86 \pm 2.86 (40, 0.012)	88.94 \pm 2.34 (40, 0.018)	92.78 \pm 1.57 (40, 0.031)	94.62 \pm 1.45 (38, 0.038)
PCA	75.07 \pm 2.95 (80, 0.008)	81.44 \pm 2.63 (74, 0.012)	85.00 \pm 2.81 (68, 0.027)	88.22 \pm 2.54 (78, 0.024)
MECRDP	87.92 \pm 2.26 (38, 0.014)	93.43 \pm 1.62 (40, 0.023)	95.28 \pm 1.70 (40, 0.037)	96.66 \pm 1.42 (38, 0.054)

Table 5. The maximum average recognition accuracy (%) in the ORL database with different dimensions by using the nearest neighbor classifier.

Dimension	10	20	30	40	50	60	70	80
CRP	81.78	87.53	88.25	88.87	89.19	90.16	90.28	90.25
CRRP	72.16	86.03	90.56	92.69	92.28	92.00	91.97	91.84
LDA	88.41	93.12	93.75	95.03	95.00	94.50	93.84	92.81
LLDE	88.28	93.63	93.84	94.94	94.34	93.97	93.75	93.25
MFA	89.00	93.50	94.06	94.75	94.97	94.81	94.81	94.91
RLSDP	88.34	92.56	93.50	94.41	94.12	94.09	92.81	92.31
PCA	80.91	86.19	87.22	87.69	87.87	88.00	88.19	88.19
MECRDP	94.16	96.25	96.25	96.62	96.59	96.06	96.00	95.91

Table 6. The maximum average recognition accuracy (%) in the ORL database with different dimensions by using the neural network.

Dimension	10	20	30	40	50	60	70	80
CRP	81.47	88.09	86.72	84.88	83.13	80.56	76.69	75.94
CRRP	71.00	81.28	80.87	78.13	77.53	76.19	72.47	73.56
LDA	85.72	87.53	86.06	78.28	77.69	76.38	75.28	74.69
LLDE	85.97	88.78	84.31	78.09	78.13	77.44	74.94	73.44
MFA	86.97	88.06	84.66	78.97	76.12	73.19	70.88	72.34
RLSDP	85.62	87.38	83.53	79.03	78.31	76.09	74.97	74.31
PCA	80.59	88.44	86.94	84.38	82.12	81.09	79.75	78.22
MECRDP	92.03	90.50	88.39	86.15	84.11	82.28	79.63	78.47

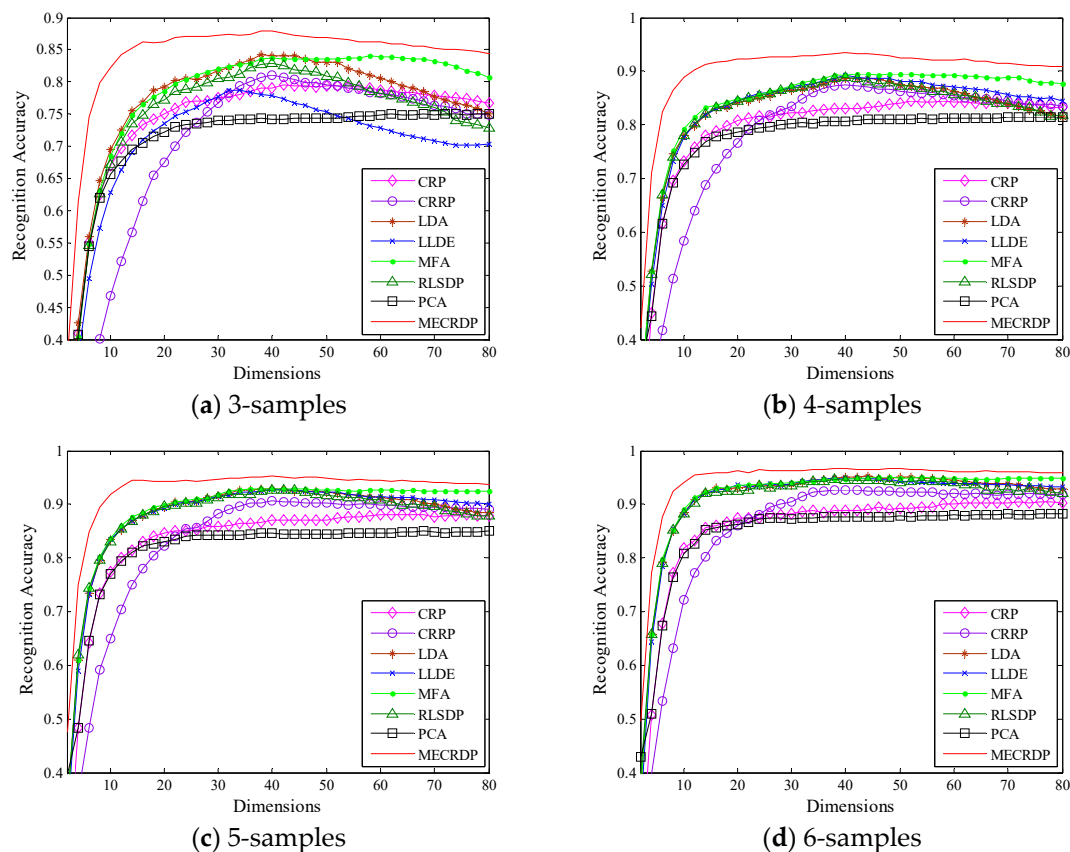


Figure 2. The average recognition accuracy vs. the subspace dimensions of each method in the ORL database.

4.2.3. Experiment on the FERET Database

The FERET [35] is the face database and consists of 13,539 images from 1565 individuals, where we select a subset containing 1400 images from 200 individuals, with seven images for each individual. These images are varying in facial expression, illumination, and pose. We cropped and resized each image to 40×40 , and randomly selected 3, 4, 5 images from each class to form the training set. We set λ and α in MECRDP to 0.1 and 0.1, respectively.

Table 7 shows the maximum average recognition accuracy, the standard derivations, the corresponding dimensions, and the average running time of each method on the FERET database. The average recognition rates obtained by the nearest neighbor classifier and neural network classifier for each method are shown in Tables 8 and 9, respectively. Figure 3 illustrates the average recognition accuracy versus the subspace dimensions on FERET. The results in Table 3 and Figure 3 show that the proposed MECRDP outperforms other methods. It is worth noting that FERET has 200 classes but only seven samples for each class. On the FERET database, the average recognition accuracy is greatly improved by MECRDP. For example, the training samples of each class are 3 and 4, the maximum average recognition accuracy of MECRDP is improved by about 26.2% and 14.1%, respectively. Besides, Figure 3 shows that when the projection dimension increases, the recognition accuracy of MECRDP will decrease but it is still higher than other methods. The results in Tables 8 and 9 show that the nearest neighbor classifier could achieve higher recognition accuracy than the neural network classifier. Table 8 shows that MECRDP has achieved the highest recognition accuracy in each selected feature dimension. Table 8 shows that MECRDP has achieved the highest recognition accuracy in each selected feature dimension, while Table 9 shows that MECRDP performs better than other feature extraction methods in most cases when the neural network classifiers are used. In general, the experimental results show that the features extracted by MECRDP are more discriminative in most cases.

Table 7. The maximum average recognition accuracy (%) \pm the standard derivations (%), the corresponding dimension, and the average running time (seconds) in parentheses of each method in the FERET database.

Methods	3-Samples	4-Samples	5-Samples
CRP	31.45 \pm 1.40 (80, 0.561)	34.99 \pm 1.63 (80, 1.189)	37.85 \pm 2.67 (80, 2.005)
CRRP	48.08 \pm 1.61 (34, 1.362)	64.32 \pm 1.69 (32, 2.479)	72.91 \pm 1.31 (24, 3.924)
LDA	52.03 \pm 1.82 (34, 0.443)	68.64 \pm 1.92 (32, 0.809)	77.30 \pm 1.82 (26, 1.314)
LLDE	52.32 \pm 1.86 (34, 0.551)	68.89 \pm 1.61 (26, 1.029)	77.84 \pm 1.89 (24, 1.630)
MFA	52.28 \pm 2.04 (36, 0.454)	68.95 \pm 1.99 (22, 0.864)	78.19 \pm 1.57 (24, 1.387)
RLSDP	48.57 \pm 1.71 (36, 0.480)	66.35 \pm 1.54 (26, 0.873)	76.01 \pm 1.27 (24, 1.374)
PCA	26.52 \pm 0.99 (80, 0.245)	30.15 \pm 1.11 (80, 0.460)	34.45 \pm 2.03 (80, 0.806)
MECRDP	78.13 \pm 1.71 (18, 0.563)	83.66 \pm 1.12 (20, 1.038)	87.00 \pm 1.26 (24, 1.727)

Table 8. The maximum average recognition accuracy (%) in the FERET database with different dimensions by using the nearest neighbor classifier.

Dimension	10	20	30	40	50	60	70	80
CRP	23.18	28.36	30.01	32.56	34.40	35.70	36.99	37.85
CRRP	66.75	74.49	72.58	71.66	70.66	69.55	68.65	67.42
LDA	70.65	76.89	76.99	76.05	75.10	73.93	72.85	71.42
LLDE	71.09	77.57	77.80	76.56	75.36	74.69	73.50	72.55
MFA	71.10	77.71	77.74	77.05	75.44	74.13	73.14	71.73
RLSDP	69.68	75.43	75.53	74.51	73.30	72.14	71.39	70.28
PCA	25.35	29.21	30.16	32.56	33.39	33.76	34.21	34.45
MECRDP	80.55	86.89	86.11	84.55	82.42	79.81	77.56	75.65

Table 9. The maximum average recognition accuracy (%) in the FERET database with different dimensions by using the neural network.

Dimension	10	20	30	40	50	60	70	80
CRP	17.06	35.78	42.63	46.09	47.26	45.73	43.16	41.14
CRRP	51.83	58.00	52.88	47.17	42.04	37.91	34.28	31.39
LDA	55.03	61.86	56.45	50.16	45.00	40.05	36.50	33.14
LLDE	54.88	61.86	56.11	50.15	45.22	40.37	35.74	32.75
MFA	56.83	63.69	57.98	51.37	45.81	40.21	35.33	32.76
RLSDP	53.85	59.84	54.48	48.88	43.14	39.68	35.45	32.90
PCA	17.65	35.99	44.18	47.09	47.51	46.80	45.05	43.19
MECRDP	68.91	78.88	72.30	65.06	57.85	50.26	44.90	42.92

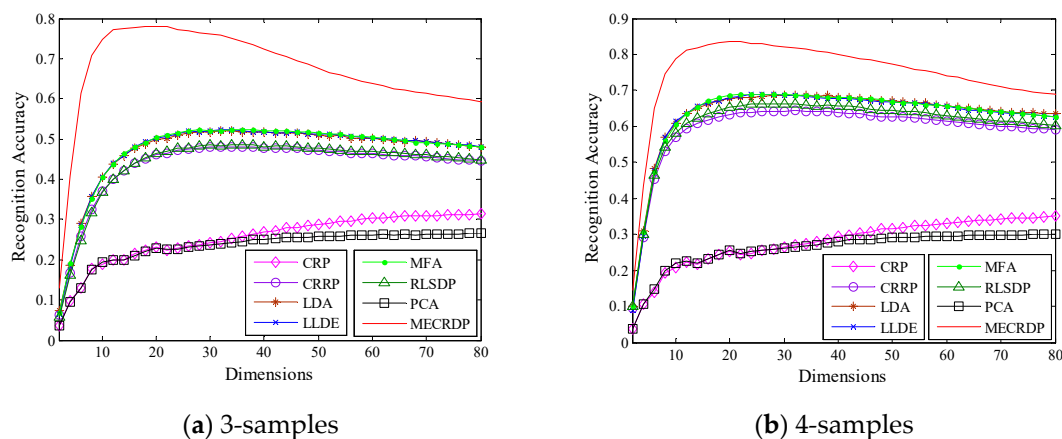
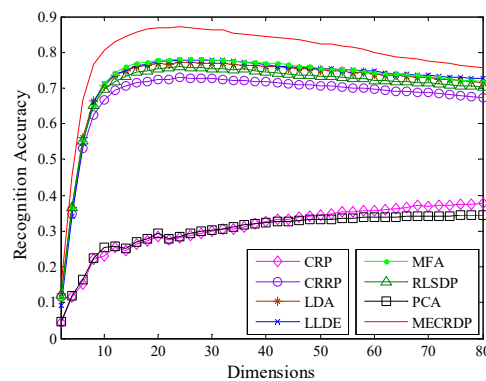


Figure 3. Cont.



(c) 5-samples

Figure 3. The average recognition accuracy vs. the subspace dimensions of each method in the FERET database.

4.2.4. Experiment on the Isolet Database

The Isolet database [36] was generated as follows. One hundred fifty subjects spoke the name of each letter of the alphabet twice. Hence, there were 52 training examples from each speaker and the size of each sample was 617×1 . The speakers were grouped into sets of 30 speakers each, and were referred to as isolet1, isolet2, isolet3, isolet4, and isolet5. In our experiment, we compared the performance of each feature extraction method in Isolet1. We set λ and α in MECRDP to 1 and 0.1, respectively.

Table 10 shows the maximum average recognition accuracy, the standard derivations, the corresponding dimensions, and the average running time of each method on the Isolet1 database. Tables 11 and 12 report the average recognition rates obtained by the nearest neighbor classifier and neural network classifier for each method, respectively. Figure 4 illustrates the average recognition accuracy versus the subspace dimensions in Isolet1. The results in Table 10 show that under different numbers of training samples, the best results are always obtained by MECRDP. Figure 4 shows that as the number of training samples increases, the recognition accuracy of all the feature extraction algorithms is improved. Although the advantages of the proposed MECRDP over other methods will gradually decrease, it can maintain a high recognition accuracy. For example, when the number of training samples for each class is five, the recognition accuracy of MECRDP is improved by at least 3.4% compared to other methods. When the number of training samples for each class is increased to 20, the performance of our method is improved by only 1.2%, and its best recognition accuracy is 94.05%. Tables 11 and 12 show that MECRDP performs better than other feature extraction methods in most cases. Based on the above experimental results, we believe that the proposed MECRDP is an effective feature extraction method.

Table 10. The maximum average recognition accuracy (%) \pm the standard derivations (%), the corresponding dimensions, and the average running time (seconds) in parentheses of each method in the Isolet1 database.

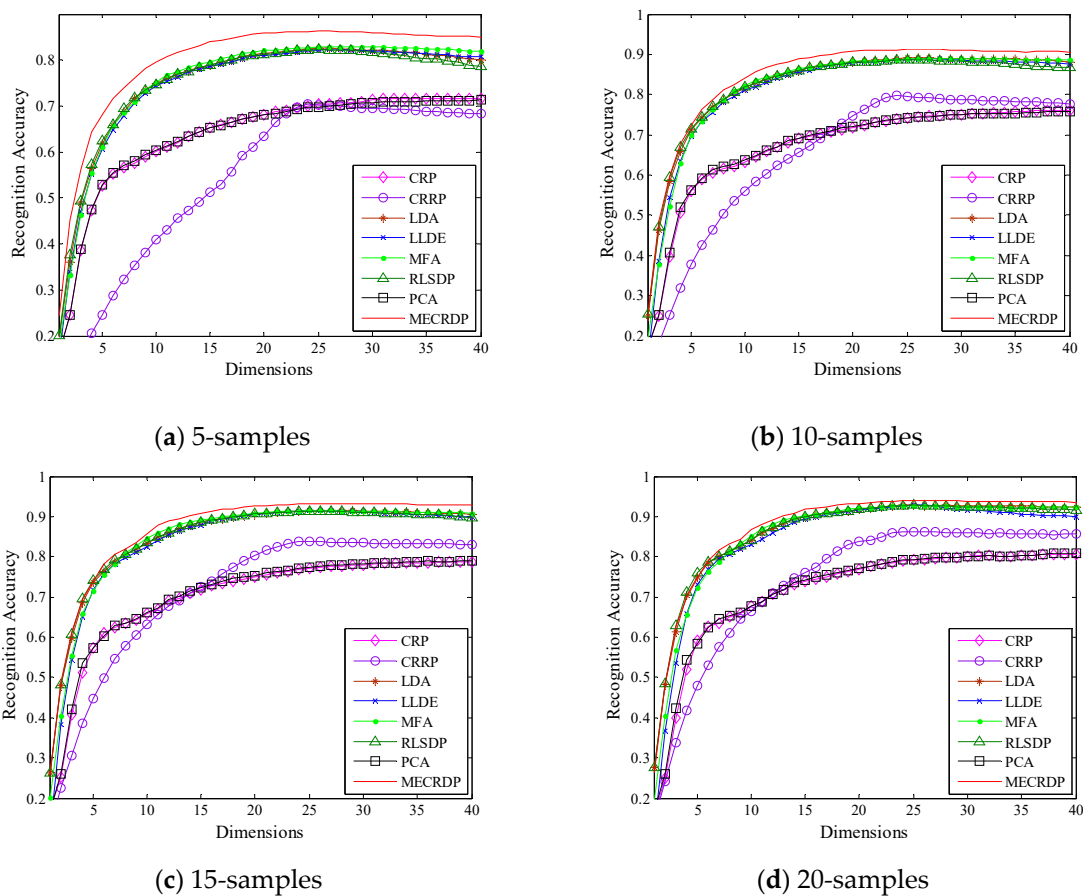
Methods	5-Samples	10-Samples	15-Samples	20-Samples
CRP	71.59 \pm 1.24 (35, 0.015)	76.11 \pm 1.39 (39, 0.054)	78.80 \pm 0.78 (40, 0.125)	80.72 \pm 1.01 (40, 0.209)
CRRP	70.47 \pm 2.25 (24, 0.025)	79.59 \pm 1.21 (24, 0.073)	83.79 \pm 1.17 (25, 0.223)	86.25 \pm 1.10 (24, 0.412)
LDA	82.59 \pm 1.87 (25, 0.012)	89.12 \pm 1.08 (25, 0.030)	91.55 \pm 0.76 (25, 0.064)	92.80 \pm 0.68 (24, 0.118)
LLDE	82.48 \pm 1.89 (26, 0.016)	88.89 \pm 0.94 (26, 0.055)	91.65 \pm 0.75 (26, 0.118)	92.86 \pm 0.59 (26, 0.217)
MFA	82.94 \pm 1.72 (29, 0.013)	88.93 \pm 0.98 (25, 0.043)	91.59 \pm 0.87 (25, 0.099)	92.71 \pm 0.82 (24, 0.169)
RLSDP	82.31 \pm 1.72 (25, 0.011)	88.71 \pm 0.97 (26, 0.034)	91.42 \pm 0.53 (25, 0.077)	92.80 \pm 0.67 (25, 0.144)
PCA	71.29 \pm 1.57 (40, 0.007)	75.86 \pm 1.45 (40, 0.022)	79.09 \pm 0.87 (39, 0.049)	80.91 \pm 1.13 (40, 0.089)
MECRDP	86.40 \pm 1.61 (25, 0.013)	91.36 \pm 1.05 (26, 0.046)	93.23 \pm 0.77 (25, 0.103)	94.05 \pm 0.68 (25, 0.199)

Table 11. The maximum average recognition accuracy (%) in the Isolet1 database with different dimensions by using the nearest neighbor classifier.

Dimension	5	10	15	20	25	30	35	40
CRP	59.25	67.74	73.92	76.97	79.15	80.00	80.34	80.72
CRRP	48.04	66.51	75.96	83.81	86.23	86.02	85.75	85.59
LDA	75.03	84.04	89.88	91.83	92.79	92.73	92.64	92.36
LLDE	73.01	82.88	89.40	91.58	92.83	91.69	90.58	89.92
MFA	72.20	85.12	90.28	92.03	92.68	92.37	92.24	92.33
RLSDP	75.94	84.21	89.91	91.75	92.80	92.37	92.14	91.63
PCA	58.51	67.83	74.08	77.25	79.32	80.04	80.23	80.91
MECRDP	74.72	86.62	91.82	93.26	94.05	93.82	93.73	93.57

Table 12. The maximum average recognition accuracy (%) in the Isolet1 database with different dimensions by using the neural network classifier.

Dimension	5	10	15	20	25	30	35	40
CRP	63.29	74.57	83.81	85.74	87.16	87.41	87.55	87.13
CRRP	50.40	66.66	75.35	80.59	81.92	81.40	80.44	79.51
LDA	73.79	81.86	85.65	86.40	85.50	84.89	84.07	83.75
LLDE	70.66	80.75	85.86	86.70	86.14	85.39	85.08	84.30
MFA	71.56	82.97	86.58	87.63	86.38	84.95	84.04	83.00
RLSDP	74.24	81.78	85.86	86.29	85.90	85.10	84.17	83.53
PCA	63.65	75.17	84.34	86.29	87.08	87.81	87.75	87.83
MECRDP	73.49	85.08	89.17	88.94	89.54	89.19	89.23	89.06

**Figure 4.** The average recognition accuracy vs. the subspace dimensions of each method in the Isolet1 database.

4.2.5. Parameter Sensitivity Analysis

Here we analyze the influence of parameters on the proposed algorithm through experiments on the four databases.

Figures 5 and 6 plot the maximum average recognition accuracy of MECRDP versus the parameter λ and α for the four samples of each database, respectively. Figures 4 and 5 show that the parameter selection of MECRDP has a great relationship with the sample database. Within a certain range, such as $\lambda \in [0, 1]$, there is a little change in the recognition accuracy of the four databases. While α has a greater impact on the recognition accuracy for COIL20, ORL, and FERET, particularly for Isolet1, the recognition accuracy of MECRDP was robust for parameter α .

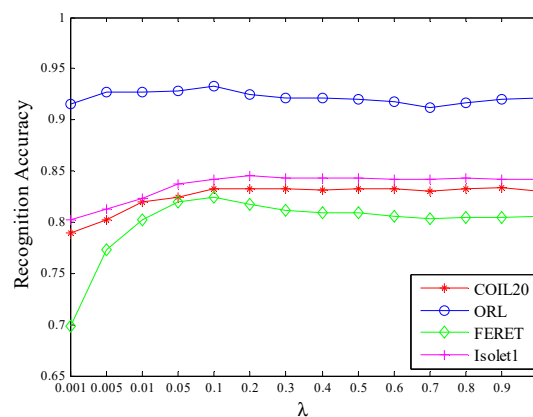


Figure 5. The maximal average recognition accuracy of MECRDP vs. the regularization parameter λ in the four databases.

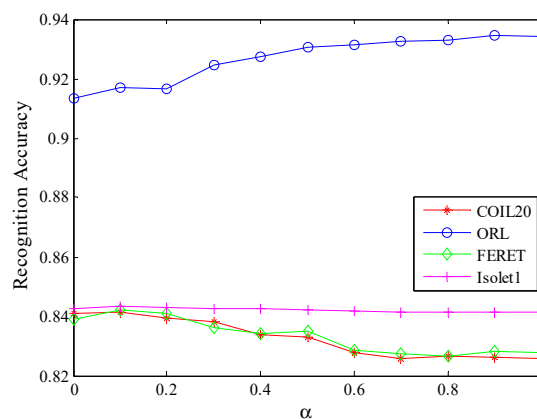


Figure 6. The maximal average recognition accuracy of MECRDP vs. the balance factor α in the four databases.

4.2.6. Visualization

In order to compare the distribution of extracted features in low-dimensional space more intuitively, we randomly selected six classes of samples from the ORL database and then projected them into two-dimensional space. Figure 7 shows the projection results of all methods in two dimensions. The results show that most methods have good clustering results except CRP and PCA, which are unsupervised feature extraction methods. This shows that the label information helps to improve the discriminative feature extraction. However, in some classes (class 1 and class 6), some methods, such as CRRP, LDA, MFA, and RLSDP, have not achieved a large between-class distance. Figure 7 shows that the features extracted by MECRDP not only have a small within-class compactness, but also have a larger between-class distance.

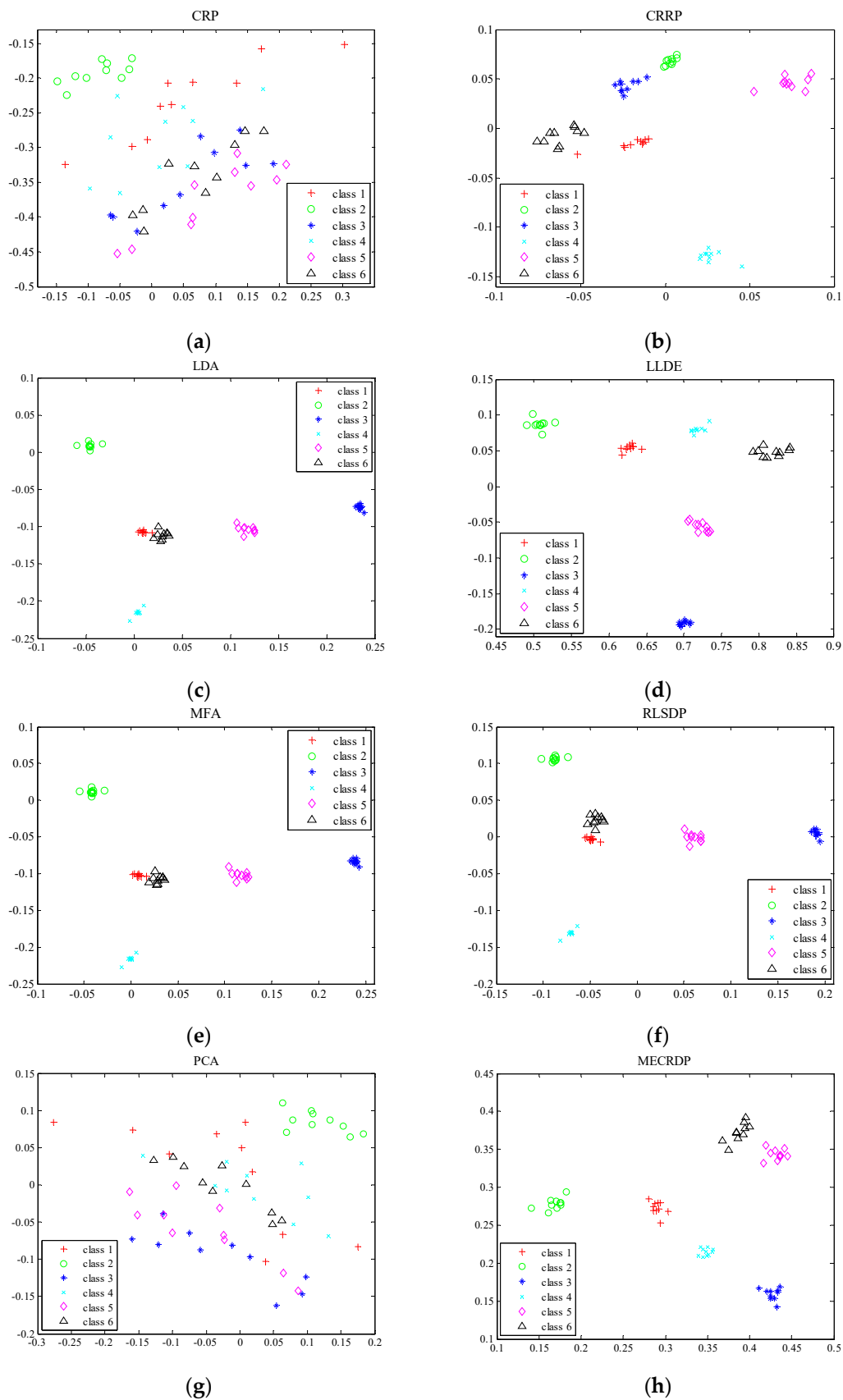


Figure 7. The visualization of six individuals in the ORL database. (a) CRP, (b) CRRP, (c) LDA, (d) LLDE, (e) MFA, (f) RLSDP, (g) PCA, (h) MECRDP.

In a word, these experimental results imply that using the minimum eigenvector to reduce the sample reconstruction error, and maintaining the collaborative representation relationship between the samples helps to improve the discriminability of the extracted features, especially when the number of training samples is small.

5. Conclusions

A new linear dimensionality reduction method based on minimum eigenvector collaborative representation discriminant projection has been proposed for feature extraction in this paper, which can be viewed as an extended collaborative representation projection method. This method employs the eigenvector corresponding to the smallest non-zero eigenvalue of the sample covariance matrix to reduce the collaborative representation error. Meanwhile, the collaborative representation relationship of the samples is maintained in the projection subspace. In addition, the between-class scatter of the reconstructed samples is used to improve the robustness of the projection subspace. The experimental results on four public databases demonstrated the superiority of the proposed method in terms of recognition accuracy as compared with other commonly used linear DR methods. What's more, the experiments show that the proposed method is especially suitable for dealing with small sample size problems, and it can also work well when the number of training samples is large. Thus, we believe that MECRDP is a general algorithm for feature extraction.

Note that the proposed MECRDP is a parameterized method, and its performance will inevitably be affected by the choice of the parameters, which also happens to other parameterized methods such as CRP, MFA, and RLSDP. In addition, sometimes, as the feature dimension increases, the performance of the algorithm will decrease. How to utilize other information of the sample, such as spatial distribution information, local structure information, and high-order statistical information to further improve the performance and robustness of the algorithm, is also an interesting direction for future study.

Author Contributions: Conceptualization, Methodology, and Software, H.H.; Data and Formal Analysis, H.H.; Writing—Original Draft Preparation, H.H.; Writing—Review & Editing, H.H. and D.-Z.F.; Supervision, H.H. and D.-Z.F.; Project Administration, D.-Z.F.; Funding Acquisition, D.-Z.F. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China, grant number 61271293.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Roweis, S.T.; Saul, L.K. Nonlinear dimensionality reduction by locally linear embedding. *Science* **2000**, *290*, 2323–2326. [[CrossRef](#)] [[PubMed](#)]
2. Tenenbaum, J.B.; De Silva, V.; Langford, J.C. A global geometric framework for nonlinear dimensionality reduction. *Science* **2000**, *290*, 2319–2323. [[CrossRef](#)] [[PubMed](#)]
3. Belkin, M.; Niyogi, P. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Comput.* **2003**, *15*, 1373–1396. [[CrossRef](#)]
4. Turk, M.; Pentland, A. Eigenfaces for recognition. *J. Cogn. Neurosci.* **1991**, *3*, 71–86. [[CrossRef](#)] [[PubMed](#)]
5. Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D.J. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **1997**, *19*, 711–720. [[CrossRef](#)]
6. Hyvärinen, A.; Oja, E. Independent component analysis: Algorithms and applications. *Neural Netw.* **2000**, *13*, 411–430. [[CrossRef](#)]
7. Li, X.; Shu, L. Kernel Based Nonlinear Dimensionality Reduction and Classification for Genomic Microarray. *Sensors* **2008**, *8*, 4186–4200. [[CrossRef](#)]
8. Rosdi, B.A.; Shing, C.W.; Suandi, S.A. Finger vein recognition using local line binary pattern. *Sensors* **2011**, *11*, 11357–11371. [[CrossRef](#)]
9. He, X.; Yan, S.; Hu, Y.; Niyogi, P.; Zhang, H.-J. Face recognition using Laplacianfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* **2005**, *27*, 328–340.

10. He, X.; Cai, D.; Yan, S.; Zhang, H.-J. Neighborhood preserving embedding. In Proceedings of the 10th IEEE International Conference on Computer Vision (ICCV 2005), Beijing, China, 17–20 October 2005; pp. 1208–1213.
11. 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.* **2007**, *29*, 40–51. [[CrossRef](#)]
12. Li, B.; Lei, L.; Zhang, X.-P. Constrained discriminant neighborhood embedding for high dimensional data feature extraction. *Neurocomputing* **2016**, *173*, 137–144. [[CrossRef](#)]
13. Sugiyama, M. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *J. Mach. Learn. Res.* **2007**, *8*, 1027–1061.
14. Li, B.; Zheng, C.-H.; Huang, D.-S. Locally linear discriminant embedding: An efficient method for face recognition. *Pattern Recognit.* **2008**, *41*, 3813–3821. [[CrossRef](#)]
15. Gui, J.; Jia, W.; Zhu, L.; Wang, S.-L.; Huang, D.-S. Locality preserving discriminant projections for face and palmprint recognition. *Neurocomputing* **2010**, *73*, 2696–2707. [[CrossRef](#)]
16. Zhang, L.; Chen, S.; Qiao, L. Graph optimization for dimensionality reduction with sparsity constraints. *Pattern Recognit.* **2012**, *45*, 1205–1210. [[CrossRef](#)]
17. Qiao, L.; Chen, S.; Tan, X. Sparsity preserving projections with applications to face recognition. *Pattern Recognit.* **2010**, *43*, 331–341. [[CrossRef](#)]
18. Gui, J.; Sun, Z.; Jia, W.; Hu, R.; Lei, Y.; Ji, S. Discriminant sparse neighborhood preserving embedding for face recognition. *Pattern Recognit.* **2012**, *45*, 2884–2893. [[CrossRef](#)]
19. Gao, Q.; Huang, Y.; Zhang, H.; Hong, X.K.; Wang, L.Y. Discriminative sparsity preserving projections for image recognition. *Pattern Recognit.* **2015**, *48*, 2543–2553. [[CrossRef](#)]
20. Zhang, L.; Yang, M.; Feng, X. Sparse representation or collaborative representation: Which helps face recognition? In Proceedings of the IEEE International Conference on Computer Vision Workshops, ICCV 2011 Workshops, Barcelona, Spain, 6–13 November 2011; pp. 471–478.
21. Yang, W.; Wang, Z.; Sun, C. A collaborative representation based projections method for feature extraction. *Pattern Recognit.* **2015**, *48*, 20–27. [[CrossRef](#)]
22. Yang, W.; Sun, C.; Zheng, W. A regularized least square based discriminative projections for feature extraction. *Neurocomputing* **2016**, *175*, 198–205. [[CrossRef](#)]
23. Yin, J.; Wei, L.; Song, M.; Zeng, W. Optimized projection for collaborative representation based classification and its applications to face recognition. *Pattern Recognit. Lett.* **2016**, *73*, 83–90. [[CrossRef](#)]
24. Wang, G.; Shi, N. Collaborative representation-based discriminant neighborhood projections for face recognition. *Neural Comput. Appl.* **2020**, *32*, 5815–5832. [[CrossRef](#)]
25. Yuan, M.D.; Feng, D.Z.; Shi, Y.; Liu, W.J. Dimensionality reduction by collaborative preserving Fisher discriminant analysis. *Neurocomputing* **2019**, *356*, 228–243. [[CrossRef](#)]
26. Kim, K.I.; Jung, K.; Kim, H.J. Face recognition using kernel principal component analysis. *IEEE Signal. Process. Lett.* **2002**, *9*, 40–42.
27. Mika, S.; Rätsch, G.; Weston, J.; Schölkopf, B.; Müller, K.-R. Fisher discriminant analysis with kernels. In Proceedings of the IEEE Signal Processing Society Workshop on Neural Network for Signal Processing, Madison, WI, USA, 23–25 August 1998; pp. 41–48.
28. Lu, J.; Plataniotis, K.N.; Venetsanopoulos, A.N. Face recognition using kernel direct discriminant analysis algorithms. *IEEE Trans. Neural Netw.* **2003**, *14*, 117–126.
29. Jun, Y.; Lai, Z.; Yan, H. Dimension Reduction using Kernel Collaborative Representation based Projection. *AEU Int. J. Electron. Commun.* **2017**, *81*, 23–30.
30. Hoppenstedt, B.; Reichert, M.; Kammerer, K.; Probst, T.; Schlee, W.; Spiliopoulou, M.; Pryss, R. Dimensionality Reduction and Subspace Clustering in Mixed Reality for Condition Monitoring of High-Dimensional Production Data. *Sensors* **2019**, *19*, 3903. [[CrossRef](#)]
31. Yu, W.; Miao, Z.; Yi, S. Learning a local manifold representation based on improved neighborhood rough set and LLE for hyperspectral dimensionality reduction. *Signal. Process.* **2019**, *164*, 20–29. [[CrossRef](#)]
32. Xu, X.; Yang, Y.; Deng, C.; Nie, F. Adaptive graph weighting for multi-view dimensionality reduction. *Signal. Process.* **2019**, *165*, 186–196. [[CrossRef](#)]
33. Nene, S.A.; Nayar, S.K.; Murase, H. *Technical Report CUCS-005-96, Columbia Object Image Library (COIL-20)*; Springer: New York, NY, USA, 1996.

34. Samaria, F.S.; Harter, A.C. Parameterisation of a stochastic model for human face identification. In Proceedings of the Second IEEE Workshop on Applications of Computer Vision, Sarasota, FL, USA, 1 January 1994; pp. 138–142.
35. Phillips, P.J.; Wechsler, H.; Huang, J.; Rauss, P.J. The FERET database and evaluation procedure for face-recognition algorithms. *Image Vis. Comput.* **1998**, *16*, 295–306. [[CrossRef](#)]
36. Dua, D.; Graff, C. UCI Machine Learning Repository. Irvine, CA: University of California. School of Information and Computer Science. 2019. Available online: <http://archive.ics.uci.edu/ml> (accessed on 16 August 2020).



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).