

## Research Article

# A Smoothed Matrix Multivariate Elliptical Distribution-Based Projection Method for Feature Extraction

Hong Qiu <sup>1</sup>, Renfang Wang <sup>1</sup>, Dechao Sun <sup>2</sup>, Xinwei Liu <sup>1</sup>, Liang Zhang <sup>1</sup>,  
and Yunpeng Liu <sup>3</sup>

<sup>1</sup>College of Big Data and Software Engineering, Zhejiang Wanli University, Ningbo 315100, China

<sup>2</sup>College of Digital Technology and Engineering, Ningbo University of Finance and Economics, Ningbo 315175, China

<sup>3</sup>College of International Exchange, Ningbo University of Technology, Ningbo 315211, China

Correspondence should be addressed to Renfang Wang; renfang\_wangac@126.com

Received 25 April 2022; Revised 17 July 2022; Accepted 18 August 2022; Published 30 September 2022

Academic Editor: G. Thippa Reddy

Copyright © 2022 Hong Qiu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Big data has the traits such as “the curse of dimensionality,” high storage cost, and heavy computation burden. Self-representation-based feature extraction methods cannot effectively deal with the image-level structural noise in the data, so how to character a better relationship of reconstruction representation is very important. Recently, sparse representation with smoothed matrix multivariate elliptical distribution (SMED) using structural information to handle low-rank error images caused by illumination or occlusion has been proposed. Based on SMED, we present a new method named SMEDP for feature extraction. SMEDP firstly utilizes SMED to automatically construct an adjacency graph and then obtains an optimal projection matrix by maximizing the ratio of the local scatter matrix and the total scatter matrix in the PCA subspace. Experiments on the COIL-20 object database, ORL face database, and CMU PIE face database prove that SMEDP works well and can achieve considerable visual and recognition performance than the relevant methods.

## 1. Introduction

Big data analytics as the premise and the foundation of artificial intelligence can easily cause the problem of “curse of dimensionality” [1], high storage cost, and heavy computation burden [2]. How to deal with such high-dimensional data is always a vital issue in many applications, including machine intelligence, data mining, pattern recognition, and image processing [3–5]. Feature extraction can effectively solve the above problems by reducing the size of a data set while preserving relevant information. Thus, it has become a research hotspot [6–8].

Principle component analysis (PCA) [9] and linear discriminant analysis (LDA) [10] are two commonly used linear feature extraction methods. The former projects the high-dimensional data by maximizing the variance while the latter maximizing the ratio of the between-class scatter matrix and the within-class scatter matrix. Although they are simple and effective for some practical applications, they may fail to

discover the underlying manifold structures of the observed data that are nonlinear.

To discover the intrinsic low-dimensional embedding of the observed data, researchers developed many manifold learning methods. Isometric feature mapping (ISOMAP) [11], Laplacian eigenmaps (LE) [12], and locally linear embedding (LLE) [13] are the most representative methods. They try to preserve the intrinsic geometrical distribution of the high-dimensional data. ISOMAP takes the pairwise geodesic distance between samples as the feature of manifold structure, and LE builds the neighborhood information graph to maintain the neighborhood relationships among the samples, while LLE believes that local relations characterize the manifold structure of data. Although they are excellent in some respects, there are still some problems, such as sensitivity to noise and “out-of-sample” problem. Thus, as improvement methods, neighborhood preserving embedding (NPE) [14], locality preserving projections (LPP) [15], and so on have been proposed. NPE and LPP are the

linearized versions of LE and LLE, and they focus on the local information of the data and linearize the nonlinear structure to handle the “out-of-sample” problem. But apparently, the  $k$ -nearest graph or the  $\epsilon$ -ball graph they use is very sensitive to parameter settings and noise interference [16].

Numerous researchers study it constantly to solve these problems. They found that the methods mentioned above can be unified with a graph embedding framework [17]. It is critical to construct adjacency graphs in this framework, which can effectively reveal the essential structure of the data. More recently, sparse representation (SR)-based methods are adopted to automatically construct adjacency graphs, since it has the advantages of sparsity, robustness to noise, and datum-adaptive neighborhood. For instance, Qiao et al. [18] developed a sparsity preserving projections (SPP) by using a modified SR-based classification (SRC) [19] to automatically build the graph. And then, the projections are obtained such that the sparse reconstructive relationship of the data can be best preserved. But it is a time-consuming process since SPP aims to solve the  $L_1$ -norm optimization problem. Therefore, Yang et al. [20] solved the  $L_2$ -norm optimization problem instead of the  $L_1$ -norm optimization to reduce the running time, and proposed a collaborative representation-based projection (CRP). It benefits from the collaborative representation classifier (CRC) [21] and can obtain a comparable result with SPP. Note that both SPP and CRP choose their neighborhood automatically and contain natural discriminating information, hence can be more practical than some manifold learning methods.

The limitations that still exist though SPP and CRP work well in some applications. First, they are all vector-based methods. That means they need to stretch the image matrix into a vector in advance, which will ignore the spatial structural information of the data. Second, for many error images caused by illumination variations or occlusions, the assumption of independence of error variables may not be appropriate in some practical situations. Many SR works attempt to handle the above problems. Yang et al. [22] and Xu et al. [23] studied the illumination variations and occlusions caused error images and observed that the structure of error variables is low rank or approximately low rank [24]. They pointed out that low-rank constraints can be used to describe the structure noise, whereas it is NP-hard in solving the rank minimization problem [25]. A general way to address such an issue is to substitute the rank function with some tractable surrogate.

To this end, Yang et al. utilized nuclear norm to characterize the illumination variations and occlusions caused error images and developed a nuclear norm-based matrix regression model (NMR) [22]. Based on NMR, Yang et al. [26] explored NMR to construct the graph and gave a NMR-based projection (NMRP). NMRP has shown its promise in feature extraction. Qian et al. [27] integrated error detection and error support into one model and provided a robust nuclear norm regularized regression (RNR). Xie et al. [28] described the structural noise and the mixed noise by the robust nuclear norm and proposed a novel robust matrix regression (RMR). Significantly, these methods all assumed

that each pixel of noise follows Gaussian or Laplacian distribution independently, which is not flexible for practical application. Therefore, some researchers are devoted to the study of low-rank representation (LRR) [29]. Fang et al. [30] improved the latent low-rank representation (LatLRR) [31], which is an excellent LRR-based feature extraction method. They proposed an approximate low-rank projection learning (ALP) and an extended ALPL (EALPL) to learn an optimal projection that can extract more discriminant features than LatLRR. Then, combining with rigid regression, they extended LatLRR to the supervised scenario and proposed a supervised ALPL (SALPL). Zhang et al. [32] introduced orthogonal matrix, row sparsity constraint on the projection matrix, and weighted truncated Schatten  $p$ -norm (WTSN) [32] to improve the robustness of the LRR model. However, these methods cannot handle the “out-of-sample” problem.

Recently, Qiu et al. [33] used a matrix multivariate elliptical distribution to characterize the error image and developed a sparse representation with smoothed matrix multivariate elliptical distribution (SMED). Compared with Gaussian or Laplacian distribution, a matrix multivariate elliptical distribution is more suitable for describing the structure noise. Besides, it introduces regularization terms to smooth the objective function, which makes the model easily to get a globally optimal solution. In short, SMED considers both of the global structure and the local structure based on sparse and low-rank representation. Thus, in this paper, we plan to explore SMED to automatically construct an adjacency graph and then learn an optimal projection matrix in the PCA subspace. The proposed method is named a smoothed matrix multivariate elliptical distribution-based projections (SMEDP). Since the projection matrix can be got directly, SMEDP can solve the “out-of-sample” problem well.

Figure 1 is the flowchart of our method, and the major contributions of this paper include: (1) in order to avoid the adverse effects produced by neglecting structural information, we propose a method named SMEDP. SMEDP can learn an adaptive graph with a modified SMED. (2) A method of iteratively reweighted least squares is designed to obtain the globally optimal weight matrix, which can be used to calculate the final projection matrix. (3) Comprehensive experiments have been carried out to indicate the superiority of SMEDP when compared with several relevant methods.

The rest of the paper is organized as follows. Section 2 gives a brief review of SMED. Section 3 introduces the SMEDP for feature extraction in detail. In Section 4, extensive experiments on three databases are presented to demonstrate the effectiveness of SMEDP. Section 5 draws the conclusions and gives the prospect of the future work.

## 2. SMED

The idea of SR-based methods is to derive a series of appropriate representation coefficients. In the following, we give an overview of SMED [33].

Given a dataset of  $n$  training image matrices  $\{\mathbf{X}_1^1, \mathbf{X}_2^1, \dots, \mathbf{X}_{n1}^1, \mathbf{X}_1^2, \mathbf{X}_2^2, \dots, \mathbf{X}_{n2}^2, \dots, \mathbf{X}_1^C, \mathbf{X}_2^C, \dots, \mathbf{X}_{nC}^C\} \in \mathbb{R}^{p \times q}$

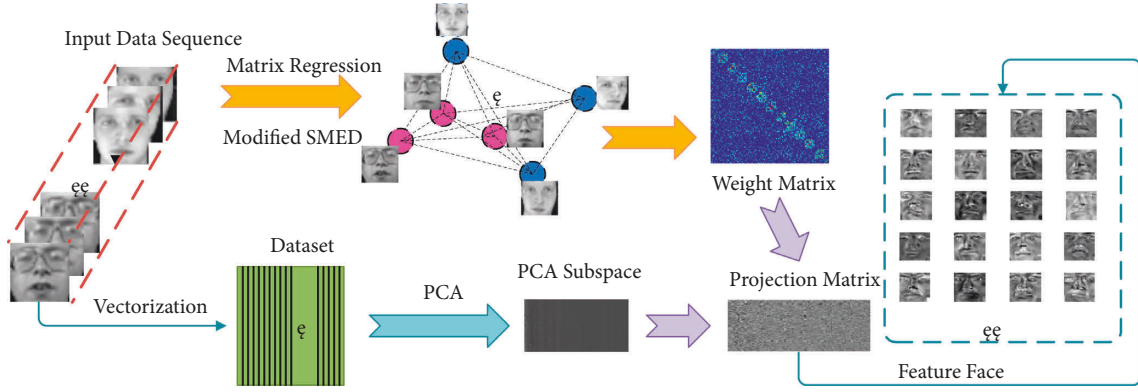


FIGURE 1: The flowchart of SMEDP. Given some sequential data, SMEDP utilizes a modified SMED to automatically build the graph.

and an image matrix  $\mathbf{Y} \in R^{p \times q}$ , we can represent  $\mathbf{Y}$  linearly using training image matrices, that is,

$$\mathbf{Y} = \mathbf{X}_1^1 \alpha_1 + \mathbf{X}_2^1 \alpha_2 + \dots + \mathbf{X}_{n1}^1 \alpha_{n1} + \dots + \mathbf{X}_1^C \alpha_{n-nC} + \dots + \mathbf{X}_{nC}^C \alpha_n + \mathbf{E}, \quad (1)$$

where  $\mathbf{X}_i^c$  represents the  $i$ -th training sample from the  $c$ th class,  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n] \in R^n$  is a vector of representation coefficients, and  $\mathbf{E}$  is the structural noise.

To simplify the formula, linear mapping from  $R^n$  to  $R^{p \times q}$  can be defined as follows:

$$\mathbf{X}(\alpha) = \mathbf{X}_1^1 \alpha_1 + \mathbf{X}_2^1 \alpha_2 + \dots + \mathbf{X}_{n1}^1 \alpha_{n1} + \dots + \mathbf{X}_1^C \alpha_{n-nC} + \dots + \mathbf{X}_{nC}^C \alpha_n. \quad (2)$$

Then, (1) becomes

$$\mathbf{Y} = \mathbf{X}(\alpha) + \mathbf{E}, \quad (3)$$

(3) is a matrix variate-based linear model. Qiu et al. [33] have studied the error image and found that the correlations between pixels in  $\mathbf{E}$  can be effectively alleviated by the proposed matrix multivariate elliptical distribution. Subsequently, a sparse representation with smoothed matrix multivariate elliptical distribution (SMED) model is developed to preserve the spatial information about the structural noise.

SMED assumes that  $\mathbf{E}$  follows the matrix multivariate elliptical distribution and the set of representation coefficients follows the independent Gaussian distribution ( $l=2$ ) or Laplacian distribution ( $l=1$ ). Then, we have

$$\min_{\alpha} \xi(\alpha) = \text{tr}(\mathbf{Y} - \mathbf{X}(\alpha))^T (\mathbf{Y} - \mathbf{X}(\alpha))^2 + \frac{k}{\beta} \sum_{i=1}^n |\alpha_i|^l. \quad (4)$$

Based on the definitions of nuclear norm and  $l_p$ -norm, (4) is equivalent to:

$$\min_{\alpha} \xi(\alpha) = \|\mathbf{Y} - \mathbf{X}(\alpha)\|_*^k + \frac{1}{2} \lambda \|\alpha\|_l^l, \quad (5)$$

where  $\lambda = 2/\beta$ ,  $k > 0$ , and  $0 < l \leq 2$ . Since both two terms in (5) are nonsmooth, SMED introduces regularization terms to smooth them:

$$\min_{\alpha} \xi(\alpha, \mu) = \left\| \begin{bmatrix} \mathbf{Y} - \mathbf{X}(\alpha) \\ \mu \mathbf{I} \end{bmatrix} \right\|_*^k + \frac{1}{2} \lambda \left\| \begin{bmatrix} \alpha \\ \mu \mathbf{1} \end{bmatrix} \right\|_l^l, \quad (6)$$

where  $\mu$  is a penalty parameter,  $\mathbf{I} \in R^{q \times q}$  is an identity matrix, and  $\mathbf{1} \in R^n$  is an all ones vector.

We obtain the optimal solution  $\alpha^*$  by solving the (6), and the corresponding reconstruction error of the  $i$ -th class is defined as follows:

$$e_i(\mathbf{Y}) = \|\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_i\|_* = \|\mathbf{X}(\alpha^*) - \mathbf{X}\delta_i(\alpha^*)\|_*, \quad (7)$$

where  $\hat{\mathbf{Y}}$  represents the reconstructed image of  $\mathbf{Y}$ ,  $\hat{\mathbf{Y}}_i$  means the reconstruction of  $\mathbf{Y}$  in class  $i$ , and  $\delta_i(\alpha^*)$  is a vector with only nonzero entries associated with the  $i$ -th class.

### 3. Algorithm Derivation of SMEDP

It is essential to construct adjacency graphs in graph-based feature extraction methods. Since SMED can better handle with variations of structural noise than several relevant methods, such as SRC and CRC, we explore SMED to automatically construct an adjacency graph and then seek to find a rule to get an optimal projection matrix. Subsequently, a method called smoothed matrix multivariate elliptical distribution-based projection method (SMEDP) is developed.

**3.1. Constructing an Adjacency Graph.** In this section, we will introduce how to construct an adjacency graph. Firstly,  $\mathbf{X} = \{\mathbf{X}_i^c\} \in R^{m \times n}$ , where  $m = p \times q$  is adopted as the vertices of the graph. Secondly, smoothed matrix regression model based on the modified SMED is presented, and the method of iteratively reweighted least squares is used to solve the model to obtain the optimal weight vector  $\alpha_i^*$  for each  $\mathbf{X}_i^c (i=1, 2, \dots, n)$ . The edges of the graph are composed of a weight matrix  $\alpha$  whose elements are  $\alpha_i^*$ . Subsequently, Algorithm 1 describes each step of the procedure in detail.

The smoothed matrix regression model based on the modified SMED is defined as follows:

$$\alpha_i = \text{argmin} \left\{ \left\| \begin{bmatrix} \mathbf{X}_i - \mathbf{X}(\alpha_i) \\ \mu \mathbf{I} \end{bmatrix} \right\|_* + \frac{1}{2} \lambda \left\| \begin{bmatrix} \alpha_i \\ \mu \mathbf{1} \end{bmatrix} \right\|_l \right\}, \quad (8)$$

where the parameters  $k$  and  $l$  in (6) are set to 1,  $\alpha_i = [\alpha_{i1}, \dots, \alpha_{i,i-1}, 0, \alpha_{i,i+1}, \dots, \alpha_{in}]^T$ , and the element  $\alpha_{ij}$  ( $i \neq j$ ) in  $\alpha_i$  represents the contribution of each  $X_i$  to reconstruct  $X_j$ .

Motivated by SMED, we provide details of using the method of iteratively reweighted least squares to solve Equation (8) [33], and it can be reformulated as follows:

$$\min_{\alpha_i} \text{tr}((X_i - X(\alpha_i))^T (X_i - X(\alpha_i)) + \mu^2 \mathbf{I})^{\frac{1}{2}} + \frac{1}{2} \lambda \left( \sum_{j=1}^n |\alpha_{ij}|^2 + \mu^2 \right)^{\frac{1}{2}}. \quad (9)$$

Let  $\gamma(\alpha_i) = \text{tr}((X_i - X(\alpha_i))^T (X_i - X(\alpha_i)) + \mu^2 \mathbf{I})^{1/2}$  and  $\delta(\alpha_i) = (\sum_{j=1}^n |\alpha_{ij}|^2 + \mu^2)^{1/2}$ . Then,  $\xi(\alpha_i, \mu) = \gamma(\alpha_i) + (\lambda/2)\delta(\alpha_i)$ . The derivative of  $\gamma(\alpha_i)$  is

$$\frac{\partial \gamma(\alpha_i)}{\partial (\alpha_i)} = \frac{(X^T X \alpha_i - X^T X_i)}{((X_i - X(\alpha_i))^T (X_i - X(\alpha_i)) + \mu^2 \mathbf{I})^{-1/2}} \quad (10)$$

$$\triangleq (X^T X \alpha_i - X^T X_i) \mathbf{R},$$

where  $\mathbf{R} = ((X_i - X(\alpha_i))^T (X_i - X(\alpha_i)) + \mu^2 \mathbf{I})^{-1/2}$ . The derivative of  $\delta(\alpha_i)$  is as follows:

$$\frac{\partial \delta(\alpha_i)}{\partial (\alpha_i)} = \frac{\alpha_i}{\left( \sum_{j=1}^n |\alpha_{ij}|^2 + \mu^2 \right)^{-1/2}} \triangleq \alpha_i \mathbf{T}, \quad (11)$$

where  $\mathbf{T} = (\sum_{j=1}^n |\alpha_{ij}|^2 + \mu^2)^{-1/2}$ .

With the derived  $\gamma(\alpha_i)$  and  $\delta(\alpha_i)$ , we get the derivative of  $\xi(\alpha_i, \mu)$ . By setting it to 0, we obtain that:

$$\frac{\partial \xi(\alpha_i)}{\partial (\alpha_i)} = (X^T X \alpha_i - X^T X_i) \mathbf{R} + \frac{1}{2} \lambda \alpha_i \mathbf{T} = 0. \quad (12)$$

Or equivalently,

$$X^T X \alpha_i + \frac{1}{2} \lambda \alpha_i \mathbf{T} \mathbf{R}^{-1} = X^T X_i. \quad (13)$$

As a well-known Sylvester equation, the computational complexity of (13) is  $O(n^3)$  [34–36]. We can use MATLAB command `lyap` to get the weight matrix  $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n] \in R^{n \times n}$ .

Algorithm 1 shows the detailed steps for solving  $\alpha$ .

**3.2. Calculating the Projection Matrix.** In this section, we will introduce how to calculate the projection matrix to improve classification performance. Firstly, we utilize PCA to reduce the dimensions of the raw data. Secondly, we seek to find an optimal projection matrix by maximizing the ratio of the local scatter matrix and the total scatter matrix in the PCA subspace. Subsequently, the final projection matrix is obtained based on the optimal projection matrix and the PCA-transformation matrix.

According to the ideas of CRP [20] and NM RP [26], the local scatter matrix is defined as follows:

$$\sum_{i=1}^n \left\| \mathbf{P}^T \mathbf{x}_i - \sum_{j=1}^n \alpha_{ij} \mathbf{P}^T \mathbf{x}_j \right\|^2 = \mathbf{P}^T \mathbf{X} (\mathbf{I} - \alpha)^T (\mathbf{I} - \alpha) \mathbf{X}^T \mathbf{P}, \quad (14)$$

where  $\mathbf{x}_i = \text{Vec}(\mathbf{X}_i)$  means the vector form of  $\mathbf{X}_i$ .

The total scatter matrix is expressed as follows:

$$\begin{aligned} \sum_{i=1}^n \left\| \mathbf{P}^T \mathbf{x}_i - \mathbf{P}^T \bar{\mathbf{x}} \right\|^2 &= \mathbf{P}^T \left( \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^T \right) \mathbf{P} \\ &= \mathbf{P}^T (\mathbf{X} - \tilde{\mathbf{X}})(\mathbf{X} - \tilde{\mathbf{X}})^T \mathbf{P}, \end{aligned} \quad (15)$$

where  $\bar{\mathbf{x}}$  is the mean vector of the training samples and the matrix  $\tilde{\mathbf{X}}$  is augmented by  $\bar{\mathbf{x}}$ .

So, the objective function of SMEDP is expressed as follows:

$$\min_{\mathbf{P}} \frac{\mathbf{P}^T \mathbf{X} (\mathbf{I} - \alpha)^T (\mathbf{I} - \alpha) \mathbf{X}^T \mathbf{P}}{\mathbf{P}^T (\mathbf{X} - \tilde{\mathbf{X}})(\mathbf{X} - \tilde{\mathbf{X}})^T \mathbf{P}}. \quad (16)$$

Or equivalently,

$$\max_{\mathbf{P}} \frac{\mathbf{P}^T \mathbf{X} \alpha_{\beta} \mathbf{X}^T \mathbf{P}}{\mathbf{P}^T (\mathbf{X} - \tilde{\mathbf{X}})(\mathbf{X} - \tilde{\mathbf{X}})^T \mathbf{P}}, \quad (17)$$

where  $\alpha_{\beta} = \alpha + \alpha^T - \alpha^T \alpha$ . Then, generalized eigenvector decomposition is adopted to solve Equation (17). The optimal projection matrix  $\mathbf{P}$  can be calculated by

$$\mathbf{X} \alpha_{\beta} \mathbf{X}^T \mathbf{P} = \lambda (\mathbf{X} - \tilde{\mathbf{X}})(\mathbf{X} - \tilde{\mathbf{X}})^T \mathbf{P}, \quad (18)$$

where  $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d]$  is the first  $d$  largest eigenvalues of Equation (18).

**3.3. SMEDP.** In this subsection, we summarize the details of SMEDP by Algorithm 2.

## 4. Experiments

We conduct experiments on COIL-20 object database, ORL face databases, and CMU PIE face databases to evaluate the performance of SMEDP. All the experiments are implemented in the same environment (CPU: Intel i7+ Lenovo, RAM: 8 GB, MATLAB: 2020b).

**4.1. Experiments on COIL-20 Object Database.** COIL-20 object database [37] contains 20 objects. For each object, 72 images are taken from different view directions. We use a subset of the COIL-20 for our experiments. This subset includes 50 images of 10 objects (each object has 5 images). All images were in grayscale and resized to  $40 \times 40$  pixels. Figures 2 and 3 show some images.

We test the visual performance of SMEDP and compare it with the classical methods, that is, PCA; some manifold learning methods, that is, LLE, ISOMAP, and LPP; and SR-based methods, that is, SPP, CRP, and NM RP. The number of nearest neighbor point  $k$  in LEE, ISOMAP, and LPP is, respectively, set to 15, 20, and 20. We select Justin Romberg's `llqc_logbarrier` to



**Input:** a dataset of  $n$  training image matrices  $\{\mathbf{X}_i^c\} \in R^{m \times n}$ ,  $m = p \times q$ , and an image matrix  $\mathbf{Y} \in R^{p \times q}$   
**Initialize**  $t = 1$ ,  $\mathbf{R} = \mathbf{I} \in R^{n \times n}$ ,  $\mathbf{T} = \mathbf{e} \in R^n$ ,  $\lambda > 0$ ,  $\mu > 0$  and  $\varepsilon = 1 \times e^{-3}$   
**while** not converged **do**  
 (1) Update  $\alpha_{i(t+1)}$  as follows:  

$$\mathbf{X}^T \mathbf{X} \alpha_i + 1/2 \lambda \alpha_i \mathbf{T} \mathbf{R}^{-1} = \mathbf{X}^T \text{Vec}(\mathbf{X}_i)$$
  
 (2) Update  $\mathbf{R}_{t+1}$  and  $\mathbf{T}_{t+1}$  as follows:  

$$\mathbf{R}_{t+1} = ((\mathbf{X}_i - \text{Mat}(\mathbf{X}(\alpha_{i(t+1)})))^T (\mathbf{X}_i - \text{Mat}(\mathbf{X}(\alpha_{i(t+1)}))) + \mu^2 \mathbf{I})^{-1/2}$$
  

$$\mathbf{T}_{t+1} = (\sum_{j=1}^n |\alpha_{ij(t+1)}|^2 + \mu^2)^{-1/2}$$
  
 (3)  $t = t + 1$ .  
 (4) **If**  $\|\alpha_{t+1} - \alpha_t\|_{\infty} \leq \varepsilon$ , **break**.  
**end while**  
**Output:** the adjacent weight matrix  $\alpha$

ALGORITHM 1: Iterative Algorithm for solving  $\alpha$ .

**Input:** a dataset of  $n$  training image matrices  $\{\mathbf{X}_i^c\} \in R^{m \times n}$ ,  $m = p \times q$ , and an image matrix  $\mathbf{Y} \in R^{p \times q}$   
 (1) Use Algorithm 1 to calculate the weight vector  $\alpha_i$  for each  $\mathbf{X}_i$ .  
 (2) Apply PCA to reduce the dimensions of  $\mathbf{X} = [\text{Vec}(\mathbf{X}_1^1), \text{Vec}(\mathbf{X}_2^1), \dots, \text{Vec}(\mathbf{X}_{n1}^1), \text{Vec}(\mathbf{X}_1^2), \text{Vec}(\mathbf{X}_2^2), \dots, \text{Vec}(\mathbf{X}_{n2}^2), \dots, \text{Vec}(\mathbf{X}_1^C), \text{Vec}(\mathbf{X}_2^C), \dots, \text{Vec}(\mathbf{X}_{nC}^C)]$  and denote  $\mathbf{W}_{\text{PCA}}$  as the PCA-transformation matrix.  
 (3) Compute the local scatter matrix  $\mathbf{X} \alpha_\beta \mathbf{X}^T$  and the total scatter matrix  $(\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T$  in the PCA subspace.  
 (4) Calculate the  $\mathbf{P} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_d]$  of  $\mathbf{X} \alpha_\beta \mathbf{X}^T \mathbf{P} = \lambda (\mathbf{X} - \bar{\mathbf{X}})(\mathbf{X} - \bar{\mathbf{X}})^T \mathbf{P}$  corresponding to the first  $d$  largest eigenvalues.  
**Output:** the optimal projection matrix  $\mathbf{P}$  and the final projection matrix  $\mathbf{W} = \mathbf{P}^T * \mathbf{W}_{\text{PCA}}^T$

ALGORITHM 2: SMEDP.

calculate L1-norm in SPP. Besides, the parameters in SPP, CRP, NMRP, and SMEDP are determined by grid search. In experiments, we start by using PCA to reduce the dimensions of the data to 70, and then adopt the above methods to convert the 70-dimensional representation to a 2-dimensional map. Figure 4 shows the resulting map, where the color of a feature point in the scatterplot represents the class information of data.

From Figure 4, we can find that: (1) PCA confuses the projected data since it has a large variance. (2) LEE and ISOMAP fail to effectively separate data points, which are of different classes. Although they take into account the local structure of the data, the Euclidean distance measure or the geodesic distance measure cannot determine the real local structure. (3) The sparsity properties could enhance the visual performance, so the separation degree of SPP, CRP, and NMRP is better than LEE and ISOMAP. About SPP and CRP, CRP generates a smaller intraclass scatter in our experiment. This result demonstrates that it is the collaborative representation that plays an important role in the effectiveness of SRC. (4) The matrix-based methods usually surpass the vector-based methods (e.g., NMRP and our SMEDP method, which make use of the structural information outperform SPP). Notably, they enhance the performance of intraclass clustering. (5) NMRP can properly embed most classes, but some classes are still mixed. Its performance is inferior to CRP since the nuclear norm used in the matrix regression model is still a loose approximation to the natural rank constraint. (6)



FIGURE 2: Some images from COIL-20.



FIGURE 3: Images of 10 objects from COIL-20.

Due to the introduced regularization term in matrix regression model, our SMEDP method can achieve better results than others. Since SMEDP ignores the class label information, it also exists some overlapping phenomena among classes. We can try to absorb supervised information into the method to overcome the limitation further.

**4.2. Experiments on ORL Face Database.** ORL face database [38] consists 40 people. For each people, 400 images are taken from various illumination and facial expressions. All images were in grayscale and resized to  $32 \times 32$  pixels for

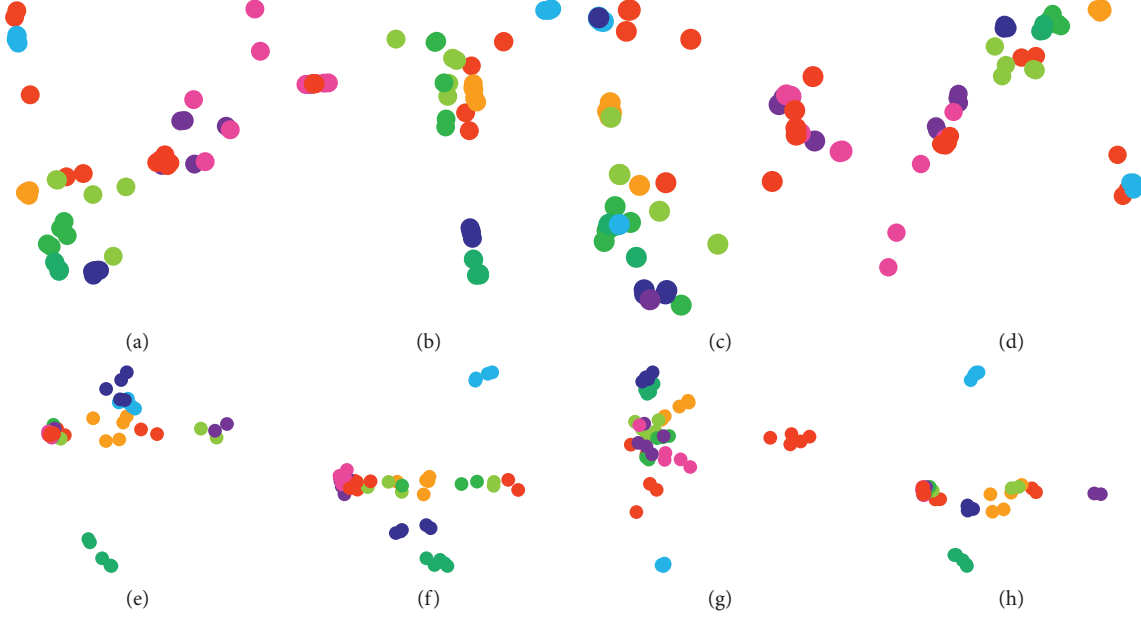


FIGURE 4: Visualizations of 50 images on COIL-20. (a) PCA. (b) LLE. (c) ISOMAP. (d) LPP. (e) SPP. (f) CRP. (g) NMRP. (h) SMEDP.

experiments in this subsection. Figure 5 shows sample images of one person.



FIGURE 5: Sample images of one person from ORL.

**4.2.1. Parameter Selection.** Feature extraction methods need to select insensitive parameters [39]. We need to tune three parameters, that is,  $\lambda$ ,  $\mu$ , and  $\rho$ , for SMEDP. So, an experiment on parameter sensitivity is conducted in this subsection. Since it is difficult to tune the three parameters simultaneously, the strategy of grid search [39] is employed to select the parameter. The core idea of grid search is fixing one parameter and tuning the others. Specifically, 8 images per class are randomly selected in the ORL database for training and the remaining for testing. The number of principal components is set to 290 during the PCA step. Then, we perform SMEDP to convert the 290-dimensional representation to a 110-dimensional map. Finally, we use the NN classifier with cosine distance for classification. The experiment is run 50 times, and the results are reported in Figure 6.

Two cases are used to evaluate the effects of different parameter settings: (1) Fix  $\rho = 1.1$ , tune  $\lambda$  and  $\mu$  from  $\{0.0001, 0.001, 0.01, 0.1, 1\}$ . Figure 6(a) shows the results, and it is obvious that the parameter  $\lambda$  and  $\mu$  do affect the performance of SMEDP. When  $\lambda = 0.01, 0.1, 1$  and  $\mu = 0.01, 0.001, 0.0001$ , the performance of SMEDP changes a little and the recognition rate is more than 90%. When  $\lambda = 0.1$  and  $\mu = 0.01$ , SMEDP achieves the best effect; (2) fix  $\lambda = 0.1$  and  $\mu = 0.01$ , vary  $\rho$  from  $\{0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1.0, 1.05, 1.1, 1.15, 1.2, 1.3, 1.35, 1.4, 1.45, 1.5, 1.55\}$ . Figure 6(b) gives the average recognition rate curves of SMEDP with different values of  $\rho$ . From Figure 6(b), we can find that SMEDP is robust with different  $\rho$ , which means parameter  $\rho$  has a certain degree of influence on the recognition rate, but the influence is insignificance. To facilitate the analysis of data,

the part of Figure 6(b), where the recognition rate ranges from 87% to 93%, is shown in Figure 6(c). Besides, from the results, we can find that  $\rho$  with a value of 1.1 produces better performance. Overall, the above analysis provides a guidance of parameters selection for SMEDP in follow-up experiments.

**4.2.2. Classification Results with Different Expressions and Illuminations.** We construct classification experiments to evaluate the robustness of SMEDP with different expressions and illuminations. We also compare SMEDP with PCA, LPP, SPP, CRP, and NMRP. The parameters of each method are set as follows: (1) nearest neighbor point  $k$  in LPP is set to 20; (2) Justin Romberg's `l1qc_logbarrier` is used to calculate L1-norm and  $\lambda$  is set to  $1e-3$  in SPP; (3) in CRP and NMRP,  $\lambda$  is set to 0.1 and 1, respectively; (4) for SMEDP, the parameters  $\lambda$ ,  $\mu$ , and  $\rho$  are set as 0.1, 0.01, and 1.1, respectively. We run the experiments 50 times. In each run,  $l$  ( $= 5, 6, 7$ , and 8) images are randomly selected from each subject as a training set and the rest  $10-l$  images are treated as a test set. Besides, SPP, CRP, NMRP, and SMEDP all apply a PCA preprocess step to reduce the dimensions of the raw data. For  $l$  ( $= 5, 6, 7$ , and 8), the numbers of principal components are set to 170, 210, 250, and 290, respectively. Finally, the numbers of subspace dimensions by all the methods except SPP are varied from 50 to the numbers of principal components at intervals of 30, 40, 50, and 60. Since SPP is sensitive to large variations in pose, the numbers of subspace dimensions for ORL database are varied from 30 to 50 to maintain a relatively high level of recognition rate. Figure 7

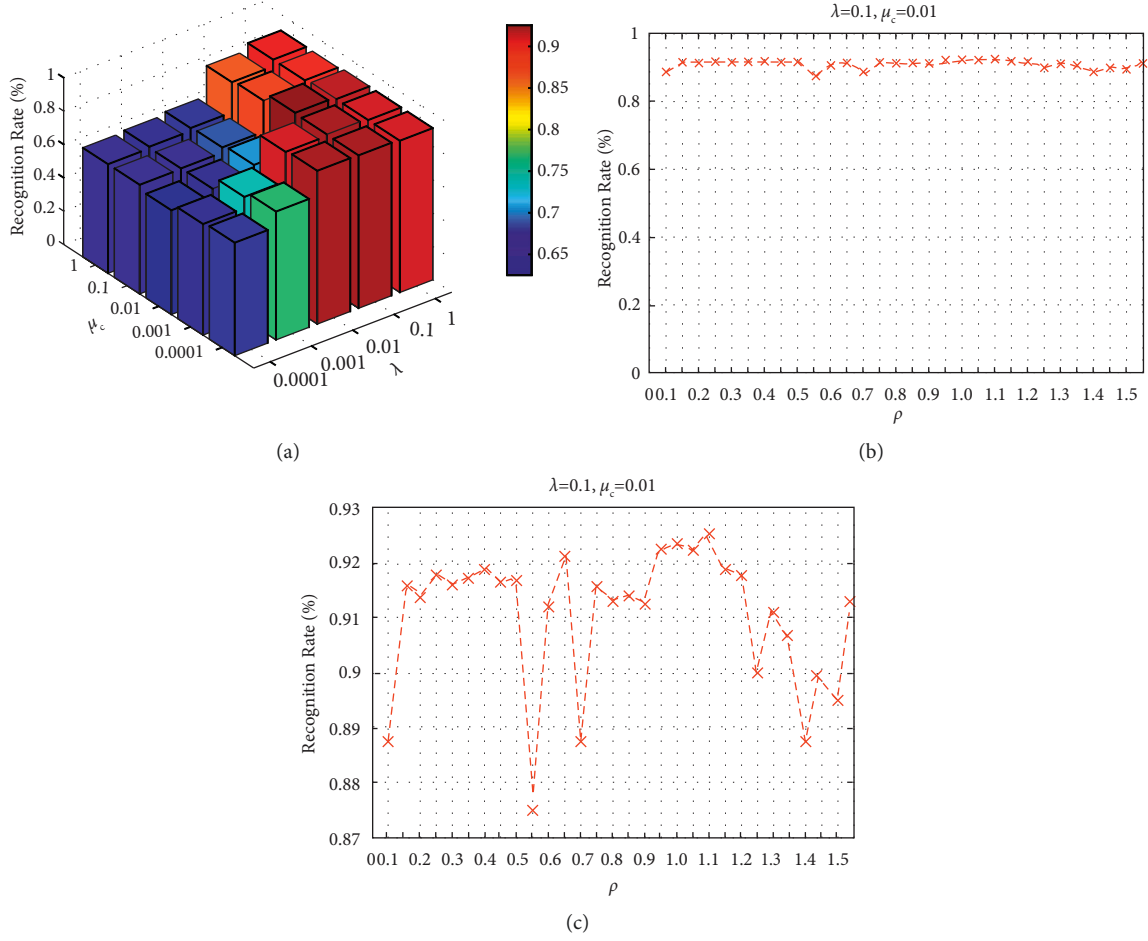


FIGURE 6: The classification performance of SMEDP for varying values of  $\lambda$ ,  $\mu_c$ , and  $\rho$  on ORL. (a) The average recognition rate of SMEDP with different values of  $\lambda$  and  $\mu_c$ . (b) The average recognition rate curves of SMEDP with varying values of  $\rho$ . (c) The part of Figure 6(b) where the recognition rate ranges from 87% to 93%.

shows the average recognition results of different methods, and Table 1 (the best performing method has been highlighted) gives the average maximal recognition rate of each method.

From the results, we can see that SMEDP performs the best, which verifies that it can preserve the primary discriminant information for classification. Specifically, SMEDP can outperform the others by about 2% improvement on  $l (= 5, 6)$  and about 1% improvement on  $l (= 7, 8)$ . Since NMRP and SMEDP get an adjacent weight matrix by matrix regression, their results are closer. Moreover, CRP, which is based on vector regression, achieves a performance comparable to NMRP in most cases. This is mainly because collaborative representation can improve the discriminating capability to some extent. Since the limitations of PCA, LPP, and SPP mentioned above, the classification performance of them is poor.

**4.3. Experiments on CMU PIE Face Database.** CMU PIE face database [40] consists 41,368 images of 68 people. All images are taken from different poses, illumination, and facial expressions for each individual. We choose C09 (one of five

subsets of near frontal poses, which contains 1632 images) for the test. All images were in grayscale and resized to  $64 \times 64$  pixels. Figure 8 shows sample images of one person.

**4.3.1. Classification Results with Different Illuminations.** For evaluating the robustness of SMEDP to the data with different illuminations, we conduct experiments and systematically compare it with the methods mentioned in Section 4.2.2. Specifically, we choose the first  $l (= 7, 9, 11, 13)$  images of one person for training and the rest  $24-l$  images for testing. It is worth noting that LPP, SPP, CRP, NMRP, and SMEDP all involve the selection of parameters, and they are adjusted based on grid search. In this subsection, they are set as follows: (1) nearest neighbor point  $k$  in LPP is set to 20; (2) in SPP, Justin Romberg's `l1qc_logbarrier` is used to calculate L1-norm and  $\lambda$  is set to  $1e-4$ ; (3) in CRP and NMRP,  $\lambda$  is set to 0.01 and 0.1, respectively; (4) for SMEDP,  $\lambda$ ,  $\mu_c$ , and  $\rho$  are set as 0.01, 0.01, and 1.1, respectively. PCA is first used to preprocess the images. We set the dimensions of the PCA subspace to 300, 250, 200, and 150, which correspond to  $l (= 7, 9, 11, 13)$ , respectively. Figure 9, Tables 2 and 3 show the results. Note that numbers in bold mean the best results in Tables 2 and 3.

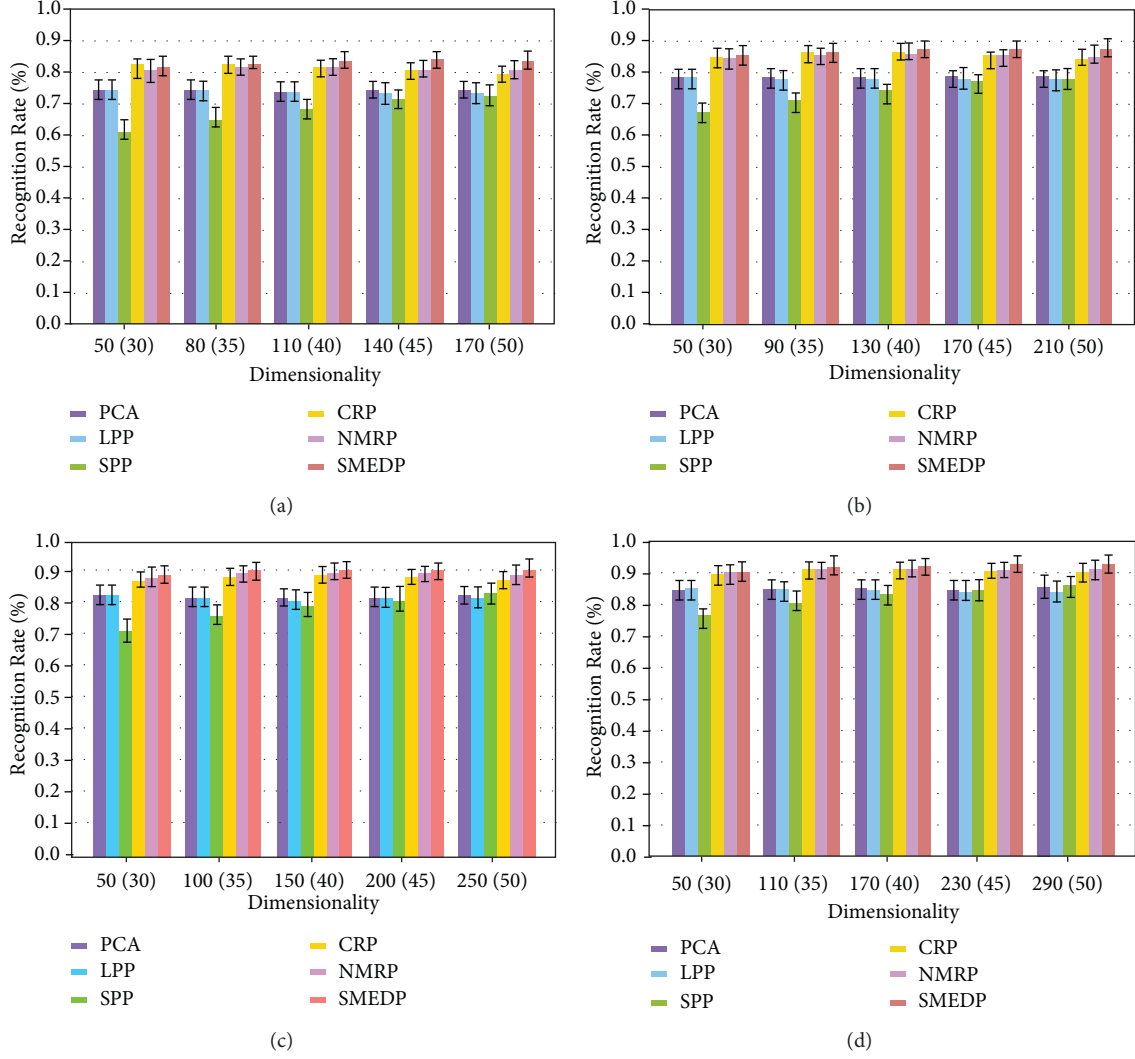


FIGURE 7: Recognition rate (%) versus subspace dimension on ORL. (a)  $l=5$ . (b)  $l=6$ . (c)  $l=7$ . (d)  $l=8$ .

TABLE 1: Average recognition rate (%) and standard deviation of different methods on ORL.

Methods	$l=5$	$l=6$	$l=7$	$l=8$
PCA	$0.7466 \pm 0.03254$	$0.7886 \pm 0.02666$	$0.8245 \pm 0.02511$	$0.8593 \pm 0.03828$
LPP	$0.7455 \pm 0.02937$	$0.7876 \pm 0.03026$	$0.8233 \pm 0.02761$	$0.8563 \pm 0.03456$
SPP	$0.7265 \pm 0.03429$	$0.7806 \pm 0.03386$	$0.8315 \pm 0.03525$	$0.8643 \pm 0.03805$
CRP	$0.8230 \pm 0.02229$	$0.8658 \pm 0.02426$	$0.8950 \pm 0.02630$	$0.9140 \pm 0.02797$
NMRP	$0.8185 \pm 0.02423$	$0.8595 \pm 0.02735$	$0.8975 \pm 0.02860$	$0.9170 \pm 0.02597$
SMEDP	<b><math>0.8427 \pm 0.02548</math></b>	<b><math>0.8776 \pm 0.02273</math></b>	<b><math>0.9073 \pm 0.02671</math></b>	<b><math>0.9253 \pm 0.02521</math></b>

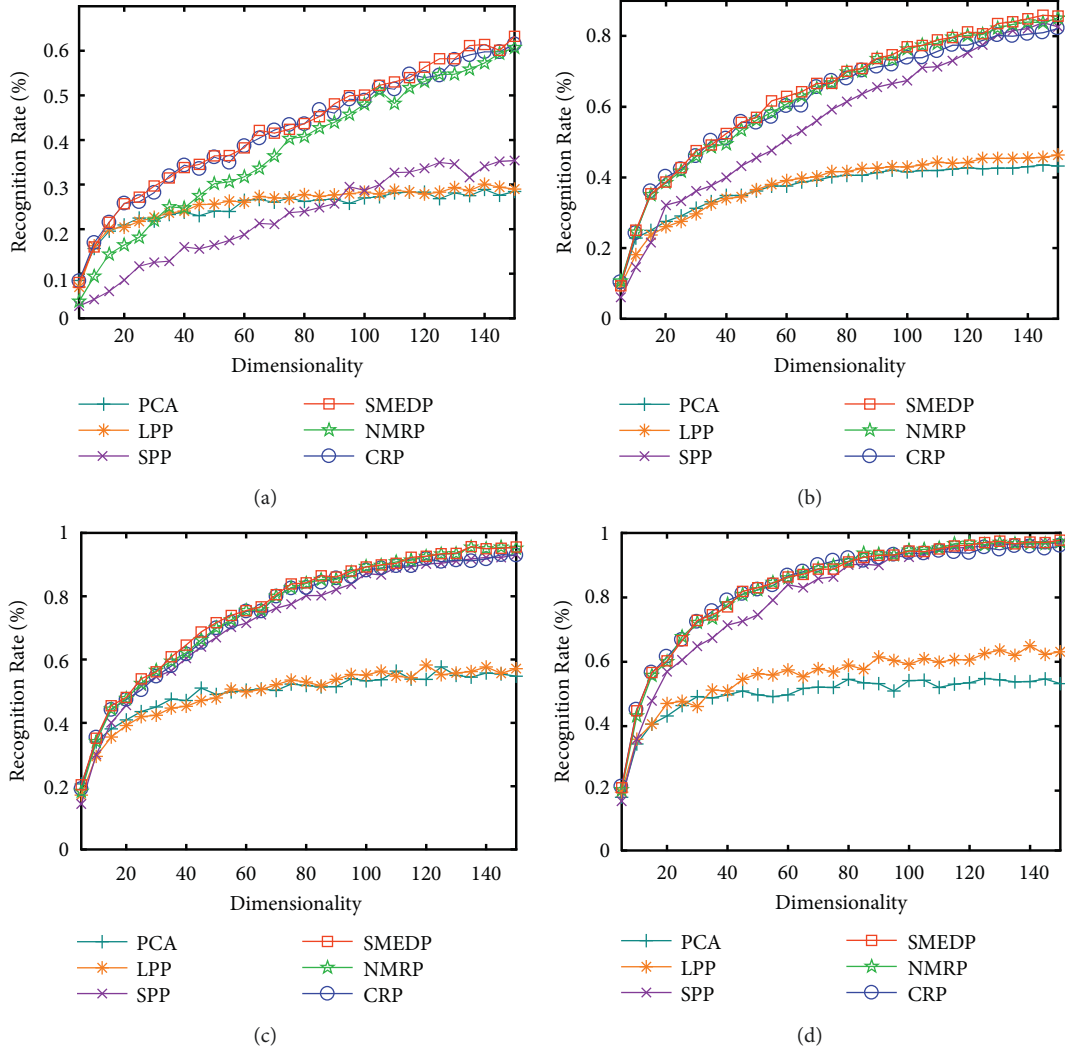


FIGURE 8: Sample images of one person from CMU PIE.

From the experimental results, it can be seen that the recognition rate of these methods generally increases with the growth of the feature dimensions. Our proposed

SMEDP always obtains the best results, which demonstrates its robustness to illumination variations. One can also note that SMEDP can outperform the others by about 2% improvements on 150 dimensions when  $l=7$  and 11. Besides, NMRP and CRP are comparable to SMEDP in some cases since the former takes advantage of matrix structure and the latter uses collaborative representation to preserve the primary discriminant information. SPP is superior to PCA and LPP, but inferior to SMEDP and NMRP. The superiority is distinctive because the sparsity properties are considered in the construction of graph.



FIGURE 9: Recognition rate (%) versus subspace dimension on CMU PIE. (a)  $l=7$ . (b)  $l=9$ . (c)  $l=11$ . (d)  $l=13$ .TABLE 2: Recognition rates (%) and versus subspace dimension of different methods on CMU PIE with  $l=7, 9$ .

Methods	$l=7$					$l=9$				
	$d=30$	60	90	120	150	$d=30$	60	90	120	150
PCA	0.2180	0.2647	0.2673	0.2829	0.2846	0.3127	0.3755	0.4137	0.4275	0.4324
LPP	0.2266	0.2604	0.2768	0.2785	0.2898	0.2971	0.3912	0.4265	0.4422	0.4637
SPP	0.1254	0.1877	0.2569	0.3365	0.3538	0.3618	0.5078	0.6559	0.7539	0.8265
CRP	0.2820	<b>0.3867</b>	0.4602	0.5398	0.6151	0.4598	0.6013	0.7127	0.7745	0.8216
NMRP	0.2197	0.3175	0.4394	0.5311	0.6073	0.4598	0.6078	0.7363	0.8020	0.8510
SMEDP	<b>0.2967</b>	0.3824	<b>0.4810</b>	<b>0.5631</b>	<b>0.6324</b>	<b>0.4759</b>	<b>0.6294</b>	<b>0.7369</b>	<b>0.8122</b>	<b>0.8571</b>

PCA and LPP perform worse than others. The reason is that they only take global structure or local structure of data into account.

**4.3.2. Classification Results with Block Occlusions.** In this subsection, we verify the robustness of SMEDP. We use the similar experiment setting as in Section 4.3.1. First 13 samples are selected as a training set and the rest samples with random occlusions of black blocks or white blocks are

treated as a test set. The block size determines the occlusion level of an image. We use PCA to reduce the dimension of data to 150, and then, classification tests were carried out on the test samples with varying degrees of random occlusions. Figure 10 shows the results.

From Figure 10, it is evident that the recognition rates of PCA and LPP are low whether the samples under the occlusions of black blocks or white blocks. It means that they cannot handle the structural noise very well. The proposed SMEDP is superior to some SR-based feature extract

TABLE 3: Recognition rate (%) and versus subspace dimension of different methods on CMU PIE with  $l=11, 13$ .

Methods	$l=11$					$l=13$				
	$d=30$	60	90	120	150	$d=30$	60	90	120	150
PCA	0.3127	0.3755	0.4138	0.4275	0.4326	0.4920	0.4973	0.5321	0.5348	0.5321
LPP	0.2971	0.3912	0.4265	0.4422	0.4637	0.4612	0.5749	0.6150	0.6056	0.6310
SPP	0.3618	0.5078	0.6559	0.7529	0.8265	0.6484	0.8396	0.8997	0.9559	0.9679
CRP	<b>0.5098</b>	<b>0.6333</b>	0.7127	0.7745	0.8216	0.7233	0.8670	0.9225	0.9372	0.9599
NMRP	0.4598	0.6078	<b>0.7363</b>	0.8020	0.8510	0.7219	0.8610	<b>0.9345</b>	<b>0.9679</b>	0.9733
SMEDP	0.4059	0.5794	0.7069	<b>0.8022</b>	<b>0.8571</b>	<b>0.7286</b>	<b>0.8636</b>	0.9305	0.9626	<b>0.9779</b>

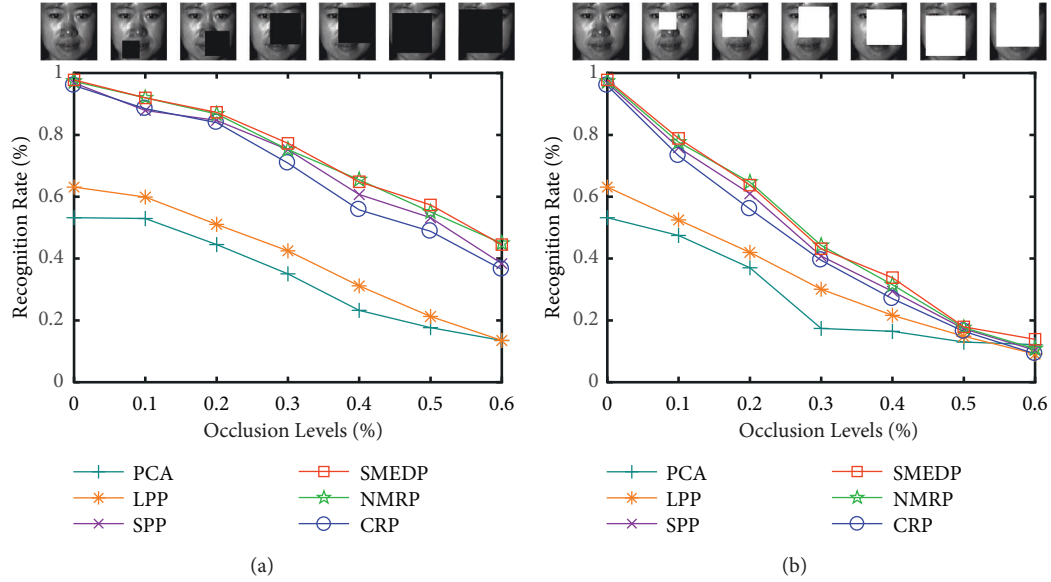


FIGURE 10: Recognition rate (%) under the different occlusion levels on CMU PIE. (a) The occlusion of black block. (b) The occlusion of white block.

methods (SPP and CRP). But the performance of NMRP is very close to SMEDP when the occlusion level is no more than 30%, that is, because they are all based on matrix regression, which can get more discriminative information from corrupted data. As the level of occlusion rises, NMRP no longer has an advantage, and the smoothed matrix multivariate elliptical distribution plays a role in SMEDP. Overall, this experiment confirms SMEDP is more robust than the others with block occlusions.

## 5. Conclusions

A method called smoothed matrix multivariate elliptical distribution-based projections (SMEDP) is proposed in this paper. Specifically, we use a modified SMED to automatically construct an adjacency graph and then learn an optimal projection matrix in the PCA subspace. Since SMED considers both the global structure and the local structure based on sparse and low-rank representation, SMEDP can get more discriminative information from corrupted data. Experimental results on three databases show that SMEDP achieves considerable visual

and recognition performance. Specifically, its robustness to illumination variations and block occlusions.

However, each method has its own strengths and weaknesses. SMEDP separates graph embedding and projection learning into two independent processes, which makes it difficult to learn the optimal projection matrix once the graph embedding is invalid. Furthermore, although SMEDP contains the natural discriminating information, it ignores the label information of the data, which will still affect the discriminative ability of the model. In the future, we will consider to simultaneously perform graph embedding and projection learning in a unified objective function and introduce the class label information in graph construction to extend SMEDP to the supervised scenario. We believe that the improvement of SMEDP will have better performance for pattern recognition task.

## Data Availability

The dataset can be accessed upon request (<https://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html>).

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## Acknowledgments

This work was supported by the National Natural Science Foundation of China (61906170 and 62106228), Project of the Science and Technology Plan for Zhejiang Province (LGF21F020023, LQ22F020003, and LGF21F020022), Plan Project of Ningbo Municipal Science and Technology (2021Z050 and 2022S002), and Ningbo Natural Science Foundation (202003N4324 and 2021J175).

## References

- [1] Z. Liu, J. Wang, G. Liu, and L. Zhang, "Discriminative low-rank preserving projection for dimensionality reduction," *Applied Soft Computing*, vol. 85, pp. 105768–105811, 2019.
- [2] A. Abbasi, A. R. Javed, C. Chakraborty, J. Nebhen, W. Zehra, and Z. Jalil, "ElStream: an ensemble learning approach for concept drift detection in dynamic social big data stream learning," *IEEE Access*, vol. 9, pp. 66408–66419, 2021.
- [3] H. Cai, V. W. Zheng, and K. C. C. Chang, "A comprehensive survey of graph embedding: problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 9, pp. 1616–1637, 2018.
- [4] W. Zhou, J. Yong, and F. He, "A collaborative feature learning method for accurate and robust tracking," in *Proceedings of the 2019 IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pp. 221–226, Porto, Portugal, May 2019.
- [5] Z. Liu, Xi. Wang, J. Pu, L. Wang, and L. Zhang, "Nonnegative low-rank representation based manifold embedding for semi-supervised learning," *Knowledge-Based Systems*, vol. 136, pp. 121–129, 2017.
- [6] H. Guo, H. Zou, and J. Tan, "Semi-supervised dimensionality reduction via sparse locality preserving projection," *Applied Intelligence*, vol. 50, no. 4, pp. 1222–1232, 2020.
- [7] Y. Liu, R. Zhang, F. Nie, X. Li, and C. Ding, "Supervised dimensionality reduction methods via recursive regression," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 9, pp. 3269–3279, 2020.
- [8] L. Zhu, Q. Hu, J. Yang, J. Zhang, P. Xu, and N. Ying, "EEG signal classification using manifold learning and matrix-variate Gaussian model," *Computational Intelligence and Neuroscience*, vol. 2021, pp. 1–12, 2021.
- [9] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [10] Q. L. Ye, C. X. Zhao, H. F. Zhang, and X. B. Chen, "Recursive concave-convex Fisher linear discriminant with applications to face, handwritten digit and terrain recognition," *Pattern Recognition*, vol. 45, no. 1, pp. 54–65, 2012.
- [11] J. B. Tenenbaum, V. D. Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [12] M. Belkin and P. Niyogi, "Laplacian Eigenmaps for dimensionality reduction and data representation," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [13] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [14] X. He, D. Cai, S. Yan, and H. Zhang, "Neighborhood preserving embedding," *Neural Computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [15] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang, "Face recognition using Laplacianfaces" 2005," *International Conference on Computer Vision (ICCV)*, vol. 2, no. 23, pp. 1208–1213, 2005.
- [16] W. Zhang, P. Kang, X. Fang, L. Teng, and N. Han, "Joint sparse representation and locality preserving projection for feature extraction," *International journal of machine learning and cybernetics*, vol. 10, no. 7, pp. 1731–1745, 2019.
- [17] S. Yan, D. Xu, B. Zhang, H. j. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 40–51, 2007.
- [18] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, no. 1, pp. 331–341, 2010.
- [19] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.
- [20] W. Yang, Z. Wang, and C. Sun, "A collaborative representation based projections method for feature extraction," *Pattern Recognition*, vol. 48, no. 1, pp. 20–27, 2015.
- [21] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?" in *Proceedings of the 2011 International Conference on Computer Vision*, pp. 471–478, Barcelona, Spain, November 2011.
- [22] J. Yang, L. Luo, J. J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear norm based matrix regression with applications to face recognition with occlusion and illumination changes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 1, pp. 156–171, 2017.
- [23] Y. Xu, X. Fang, J. Wu, X. Li, and D. Zhang, "Discriminative transfer subspace learning via low-rank and sparse representation," *IEEE Transactions on Image Processing*, vol. 25, no. 2, pp. 850–863, 2016.
- [24] J. W. Liu, C. L. Peng, and L. X. Lin, "Structured sparse models," *Chinese Journal of Computers*, vol. 40, no. 6, pp. 1309–1337, 2017.
- [25] Z. Zha, X. Yuan, B. Wen, J. Zhou, J. Zhang, and C. Zhu, "A benchmark for sparse coding: when group sparsity meets rank minimization," *IEEE Transactions on Image Processing*, vol. 29, pp. 5094–5109, 2020.
- [26] W. Yang, J. Li, H. Zheng, and R. Y. D. Xu, "A nuclear norm based matrix regression based projections method for feature extraction," *IEEE Access*, vol. 6, pp. 7445–7451, 2018.
- [27] J. J. Qian, L. Luo, J. Yang, F. L. Zhang, and Z. C. Lin, "Robust nuclear norm regularized regression for face recognition with occlusion," *Pattern Recognition*, vol. 48, no. 10, pp. 3145–3159, 2015.
- [28] J. Xie, J. Yang, J. J. Qian, Y. Tai, and H. M. Zhang, "Robust nuclear norm-based matrix regression with applications to robust face recognition," *IEEE Transactions on Image Processing*, vol. 26, no. 5, pp. 2286–2295, 2017.
- [29] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation,"

- IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.
- [30] X. Fang, N. Han, J. Wu et al., “Approximate low-rank projection learning for feature extraction,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 29, no. 11, pp. 5228–5241, 2018.
  - [31] G. Liu and S. C. Yan, “Latent low-rank representation for subspace segmentation and feature extraction,” in *Proceedings of the 2011 International Conference on Computer Vision*, pp. 1615–1622, Barcelona, Spain, November 2011.
  - [32] X. Zhang, Z. Tan, H. Sun, Z. Wang, and M. Qin, “Orthogonal low-rank projection learning for robust image feature extraction,” *IEEE Transactions on Multimedia*, pp. 1–13, 2021.
  - [33] H. Qiu, W. Wang, and J. Zheng, “Sparse representation with smoothed matrix multivariate elliptical distribution,” *Acta Automatica Sinica*, vol. 45, no. 8, pp. 1548–1563, 2019.
  - [34] P. Benner, R. C. Li, and N. Truhar, “On the ADI method for Sylvester equations,” *Journal of Computational and Applied Mathematics*, vol. 233, no. 4, pp. 1035–1045, 2009.
  - [35] H. Chen, L. Wang, and T. Li, “A note on the solvability for generalized Sylvester equations,” *Revista de la Real Academia de Ciencias Exactas, Físicas y Naturales. Serie A, Matemáticas*, vol. 115, no. 2, pp. 64–67, 2021.
  - [36] B. Li, C. Y. Lu, C. C. Leng, and L. B. Jin, “Robust low rank subspace clustering based on local graph laplace constraint,” *Acta Automatica Sinica*, vol. 41, no. 11, pp. 1971–1980, 2015.
  - [37] A. Hjouji, J. El-Mekkaoui, M. Jourhmane, and B. Bouikhalene, “New set of non-separable orthogonal invariant moments for image recognition,” *Journal of Mathematical Imaging and Vision*, vol. 62, no. 4, pp. 606–624, 2020.
  - [38] S. A. Angadi and V. C. Kagawade, “A robust face recognition approach through symbolic modeling of polar fft features,” *Pattern Recognition*, vol. 71, pp. 235–248, 2017.
  - [39] H. Qu, L. Li, Z. Li, and J. Zheng, “Supervised discriminant isomap with maximum margin graph regularization for dimensionality reduction,” *Expert Systems with Applications*, vol. 180, pp. 115055–115114, 2021.
  - [40] T. N. Le, D. B. Giap, J. W. Wang, and C. C. Wang, “Tensor-compensated color face recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 3339–3354, 2021.