

## Research Article

# Bio-Inspired Structure Representation Based Cross-View Discriminative Subspace Learning via Simultaneous Local and Global Alignment

Ao Li <sup>1</sup>, Yu Ding <sup>1</sup>, Xunjiang Zheng <sup>2</sup>, Deyun Chen <sup>1</sup>, Guanglu Sun <sup>1</sup>,  
and Kezheng Lin <sup>1</sup>

<sup>1</sup>School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China

<sup>2</sup>Optical Navigation and Detection Division, Shanghai Aerospace Control Technology Institute, Shanghai 201109, China

Correspondence should be addressed to Xunjiang Zheng; [htb\\_gdb@163.com](mailto:htb_gdb@163.com)

Received 21 August 2020; Revised 13 September 2020; Accepted 16 October 2020; Published 4 November 2020

Academic Editor: Zhile Yang

Copyright © 2020 Ao Li 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.

Recently, cross-view feature learning has been a hot topic in machine learning due to the wide applications of multiview data. Nevertheless, the distribution discrepancy between cross-views leads to the fact that instances of the different views from same class are farther than those within the same view but from different classes. To address this problem, in this paper, we develop a novel cross-view discriminative feature subspace learning method inspired by layered visual perception from human. Firstly, the proposed method utilizes a separable low-rank self-representation model to disentangle the class and view structure layers, respectively. Secondly, a local alignment is constructed with two designed graphs to guide the subspace decomposition in a pairwise way. Finally, the global discriminative constraint on distribution center in each view is designed for further alignment improvement. Extensive cross-view classification experiments on several public datasets prove that our proposed method is more effective than other existing feature learning methods.

## 1. Introduction

Under the modern technique background, there are many artificial intelligence methods inspired by nature, such as machine learning [1–4], reinforcement learning [5], and artificial immune recognition [6]. Among them, machine learning can effectively deal with image recognition problems. However, some researches have indicated that the adaptive ability of traditional machine learning drops sharply, when the learned images have large distribution discrepancy, such as cross-view data [2]. This discrepancy means that data variance in the view space is larger than data variance in the class space. It generates that the different views become the major factor affecting recognition. Therefore, we mainly focus on cross-view subspace learning to deal with the distribution discrepancy problems in this paper.

In recent years, subspace learning (SL) has made great contributions in the field of machine learning and has a wide

application in computer vision, data mining, and so on [7–21]. One of the most typical methods is principal component analysis (PCA) [22], which uses an orthogonal transformation to reduce the dimensionality while preserving unique information (principal component) of the data. However, PCA is an unsupervised dimensionality reduction method, which disregards the discriminative information attached to the semantic component. So, a supervised dimensionality reduction method, linear discriminant analysis (LDA), using semantic component was proposed in [23]. LDA learns a supervised linear combination to adjust the spatial dispersion. However, LDA generates the overfitting phenomenon when processing noisy data. To overcome corruption, rank minimization technique has been spotlighted in recent years. Candès et al. enforced low-rank and sparse constraints to eliminate corrupt information in the data [24]. After that, low-rank representation (LRR) was proposed to restore clean data

through dictionary representation in multiple spaces [7]. In the last decade, LRR model has achieved satisfactory results in various fields [8–16, 25–35], such as domain adaptation [8], clustering [9], transfer learning [25], and low-rank texture structure [26]. However, Liu et al. pointed out that the dictionary of LRR may fail when the data is insufficient. To solve this problem, the latent LRR (LatLRR) was proposed to enhance subspace learning by latent information [10]. However, LatLRR is still unsupervised. Inspired by LRR models and LDA, Li et al. unified linear discriminant constraint and low-rank representation into subspace learning to enhance the learned low-dimensional feature [11]. Afterwards, the low-rank embedding (LRE) proposed in [12] provides a robust embedding subspace learning framework that eliminates reconstruction errors by adopting  $l_{2,1}$ -norm constraint on the projection residual. The latent low-rank and sparse embedding (LLRSE) was developed in [13] based on LRE. LLRSE additionally introduces a reconstructed orthogonal matrix, which makes the projection space contain more unique feature.

Recently, a lot of algorithms dedicated to cross-view feature learning based on above methods have been developed [27–31]. The low-rank common subspace (LRCS) was proposed in [27], which finds a view common subspace by LRR. Nevertheless, LRCS only considers the view label and ignores the discrimination of the class label. From the perspective of multiview data structure, a supervised subspace learning method, namely, multiview discriminant analysis (MVDA), was proposed in [28] by using the discriminative information in the different views. After that, a multiview manifold learning with locality alignment (MVML-LA) framework proposed in [29] provides us with a discriminative low-dimensional latent space. Most recently, the robust cross-view learning (RCVL) was designed to learn a common view-invariance discriminative subspace by adopting a novel rank minimization technique [30]. However, RCVL ignores the global discriminative information.

Human visual perception works through visual circuit, whose function is to understand the visual signal in a layer-wise way. In detail, our brain uses a small feature extractor (each layer in the visual circuit) to obtain some simple features from the real complex signal. Drawing inspiration from layered processing in human visual system, we represent the cross-view data in two different structure layers, class structure layer and view structure layer, respectively, in order to construct the view-consistency feature learning model. Hence, we design two novel discriminative alignment constraints from simultaneous local and global viewpoints, which can not only disentangle the class and view layers but also bridge the gap existent in cross-view data. Our contributions are as follows:

- (1) The dual low-rank representation model is set up to discover the two latent structures existent in cross-view data, which are view and class structures, respectively. These two distinguished structures are conducive to discovering the potential feature for cross-view classification task.

- (2) A local alignment constraint based on two designed local graphs is utilized to transform the neighbouring relationships between each pair of samples in the learned subspace. This constraint can make the view and class structures be separated effectively.
- (3) A global alignment constraint designed in our framework ulteriorly cut down view discrepancy. The projected samples from view and class subspaces are used to compose the discriminative constraint in global alignment by enforcing the mean distance between classes in different views.

Figure 1 illustrates how to learn an aligned subspace in which samples have a large distance between classes and a close distance within class from both of class perspective and view perspective. The structure of this paper is organized as follows. Related works simply review the baseline of our work. Part three presents the proposed model and its solution process. Experiments show the results of the comparison experiments and parameter experiments. At last, conclusions summarize this paper.

## 2. Related Works

Our method has an intimate connection with the two following methods: (1) low-rank representation and (2) linear discriminant analysis.

**2.1. Low-Rank Representation.** LRR is unaffected by errors and can explore the underlying structure of data. Suppose that  $X = [X_1, X_2, \dots, X_k]$  is a matrix of natural data from  $k$  classes. The model of LRR can be expressed by

$$\begin{aligned} \min_{Z, E} \text{rank}(Z) + \lambda \|E\|_1 \\ \text{s.t. } X = XZ + E, \end{aligned} \quad (1)$$

where  $Z$  is the low-rank linear combination coefficient matrix of data  $X$ . Matrix  $E$  with  $l_1$ -norm can fit the corrupted information in real life.  $\lambda > 0$  is used to balance the level of corruption. Therefore, LRR that can simulate the corrupted data with representation framework possesses a feasible skill, which can handle the cross-view data well.

**2.2. Linear Discriminant Analysis.** LDA is a familiar supervised method proposed for dimensional reduction, which is constrained by discriminative semantic information. The main principle of LDA is to find a discriminative subspace with the largest interclass variance and the smallest intraclass variance. Assume that  $n$  samples from  $m$  classes are  $\{X, y\} = \{(x_1, y_1), \dots, (x_n, y_n)\}$ , where  $X$  represents samples and  $y$  represents labels of different samples. In addition,  $\hat{x}$  and  $\hat{x}_i$  denote the center of all samples and samples belonging to the  $i$ th class, respectively. Hence, between-class scatter and within-class scatter are as follows:

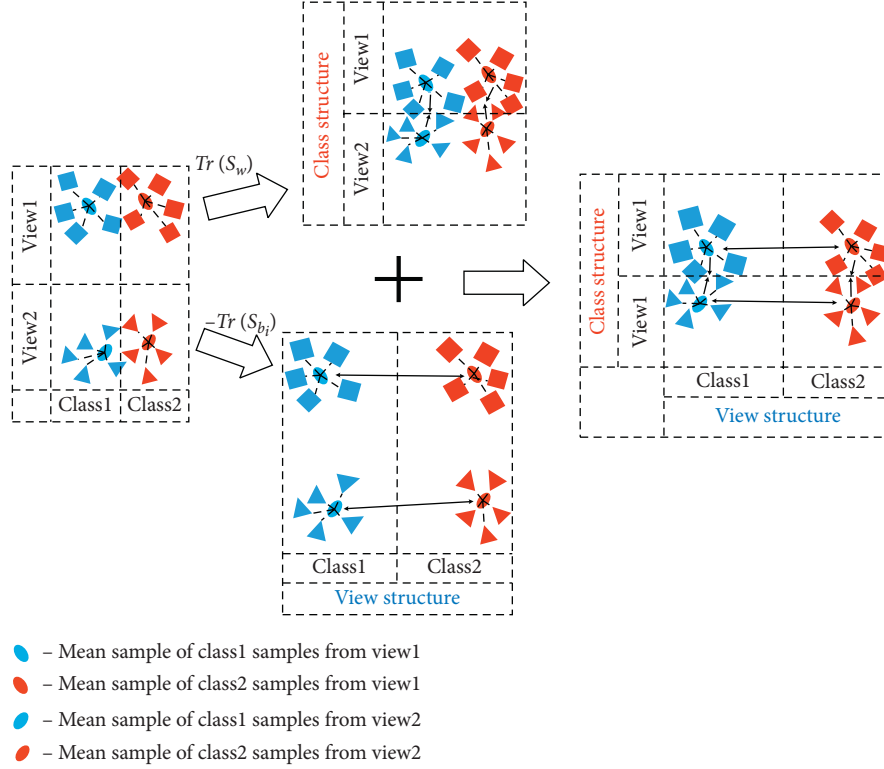


FIGURE 1: Illustration of the discriminative alignment constraint framework.

$$\begin{aligned}
 S_b &= \sum_{i=1}^m n_i (\hat{x}_i - \hat{x})(\hat{x}_i - \hat{x})^T, \\
 S_w &= \sum_{i=1}^m \sum_{x \in X_i} (x - \hat{x}_i)(x - \hat{x}_i)^T,
 \end{aligned} \tag{2}$$

where  $n_i$  is the number of samples from the  $i$ th class. Therefore, the generalized Rayleigh quotient can be described as follows by Fisher discriminant criterion:

$$\max_w \frac{\text{Tr}(w^T S_b w)}{\text{Tr}(w^T S_w w)}, \tag{3}$$

where  $w$  is a projection matrix and  $\text{Tr}(\cdot)$  denotes the trace operator. Furthermore, the solution of equation (3) is relatively complicated, so we transform it into trace-difference problem as follows [36]:

$$\max_w \text{Tr}(w^T S_b w) - \text{Tr}(w^T S_w w). \tag{4}$$

LDA not only can retain as much information for recognition as possible while reducing the dimensionality but also can remove superfluous and dependent variable feature to the disadvantage of classification task. Nevertheless, due to the distribution of cross-view data, the performance of LDA only considered class semantic information is not outstanding.

### 3. Our Proposed Method

This section contains four parts. The first part of this chapter is to specify the symbols in our algorithm. Part two is a detailed introduction and direction about our framework. The third one develops a numerical scheme to obtain the approximate solutions iteratively. The last part discusses the computational complexity of our proposed algorithm.

**3.1. Notations.** Assume that  $X_1$  and  $X_2$  ( $X = [X_1, X_2] \in R^{d \times n}$ ) are two matrices of different views from same  $c$  classes, where  $n$  and  $d$  denote the number and dimensionality of all training samples, respectively. Class structure  $Z_c \in R^{n \times n}$  and view structure  $Z_v \in R^{n \times n}$  are two linear combination matrices, which are included in local graph framework to discover view-invariant structure of cross-view data.  $P \in R^{d \times p}$  is a basis transformation projection matrix, where  $p$  is dimensionality of projected data.  $E \in R^{d \times n}$  is a matrix of error data designed to obtain a robust subspace from noise. In addition,  $V_1, V_2, \hat{V}_1, \hat{V}_2 \in R^{n \times c}$  are four constant coefficient matrices utilized for aligning the global information of cross-view data.

**3.2. Objective Function.** To address the cross-view discriminative analysis, we formulate our subspace learning model with simultaneous local and global alignments as follows:

$$\begin{aligned}
& \min_{Z_c, Z_v, E, P} \overbrace{\|Z_c\|_* + \|Z_v\|_* + \lambda_1 \|E\|_{2,1}}^{D(Z_c, Z_v, E)} + \alpha \overbrace{\left( \text{Tr}\left(P^T X Z_c L_c (P^T X Z_c)^T\right) - \text{Tr}\left(P^T X Z_v L_v (P^T X Z_v)^T\right) \right)}^{\mathcal{L}(P, Z_c, Z_v)} \\
& \quad + \overbrace{\lambda_2 \left( \text{Tr}(S_W(P^T X Z_c)) - \text{Tr}(S_{B1}(P^T X Z_v)) - \text{Tr}(S_{B2}(P^T X Z_v)) \right)}^{G(P, Z_c, Z_v)} \\
& \text{s.t. } X = X(Z_c + Z_v) + E, P^T P = I,
\end{aligned} \tag{5}$$

where  $D(Z_c, Z_v, E)$  is a low-rank framework of two potential manifolds in cross-view data.  $\mathcal{L}(P, Z_c, Z_v)$  enforces the view-specific discriminative local neighbor relationship among instances.  $G(P, Z_c, Z_v)$  performs the discriminative clustering and separation of the global structure in the class manifold and the view structure, respectively. In the following, the above terms are illustrated in detail.

**3.2.1. Dual Low-Rank Representations.** Only one linear combination matrix, which is constrained by the rank minimum, is used by methods based on low-rank model. Nevertheless, unitary low-rank structure gives rise to the failure of linear representation due to differences between distributions of views, which endow cross-view data from a same class with a large divergence. Therefore, the two structure matrices  $Z_c$  and  $Z_v$  are adopted to solve this specific problem, where the between-view samples from the same class are far away and the within-view samples from the different classes are closer. The first term is defined with dual low-rank representations to strip down the class and view structures as follows:

$$\begin{aligned}
D(Z_c, Z_v, E) &= \|Z_c\|_* + \|Z_v\|_* + \lambda_1 \|E\|_{2,1} \\
\text{s.t. } X &= X(Z_c + Z_v) + E,
\end{aligned} \tag{6}$$

where  $\|\cdot\|_*$  is a sign of the nuclear norm, which is close to a representation of the rank minimum problem, and its solution is relatively convenient. Assuming that part of the data from real world contains corruption, we adopt the  $l_{2,1}$ -norm to make matrix  $E$  have the structured sparsity as the noisy data.  $l_{2,1}$ -norm can effectively remove the corruption of specific sample while holding the other clean samples.  $\lambda_1 > 0$  is used to balance the corruption.

**3.2.2. Graph-Based Discriminative Local Alignment.** To introduce the local discriminative constraint, two graph-based constraints are constructed on each pair of synthetic samples with  $Z_c$  and  $Z_v$  from class and view subspaces, respectively, as follows, which can better cluster intraclass samples and decentralize interclass ones.

$$\begin{aligned}
\mathcal{L}_c &= \sum_{i,j} (Y_{c,i} - Y_{c,j})^2 W_{i,j}^c, \\
\mathcal{L}_v &= \sum_{i,j} (Y_{v,i} - Y_{v,j})^2 W_{i,j}^v,
\end{aligned} \tag{7}$$

where  $Y_{c,i}$  and  $Y_{v,i}$  denote the  $i$ th projected samples of cross-view data from the class space  $Y_c = P^T X Z_c$  and view space  $Y_v = P^T X Z_v$ , respectively. Correspondingly,  $Y_{c,j}$  and  $Y_{v,j}$  denote the  $j$ th projected samples.  $W_{i,j}^c$  and  $W_{i,j}^v$  denote graph weight matrices that are defined as follows:

$$\begin{aligned}
W_{i,j}^c &= \begin{cases} 1, & \text{if } x_i \in N_{k_1}^c(x_j), \text{ and } l_i = l_j, \\ 0, & \text{otherwise,} \end{cases} \\
W_{i,j}^v &= \begin{cases} 1, & \text{if } x_i \in N_{k_2}^v(x_j), \text{ but } l_i \neq l_j, \\ 0, & \text{otherwise,} \end{cases}
\end{aligned} \tag{8}$$

where  $l_i$  and  $l_j$  are the labels of samples  $x_i$  and  $x_j$ , respectively.  $x_i \in N_{k_1}^c(x_j)$  denotes that  $x_i$  belongs to the  $k_1$  adjacent datasets of the same sample  $x_j$ .  $x_i \in N_{k_2}^v(x_j)$  means that  $x_i$  belongs to the  $k_2$  adjacent datasets of the same view sample  $x_j$ . Hence,  $\mathcal{L}_c$  can calculate the distance between two similar samples from same class. Similarly,  $\mathcal{L}_v$  denotes the distance between two similar samples from different classes. With the help of Fisher criterion, the pairwise local discriminative constraint  $\mathcal{L}(Z_c, Z_v, P)$  can be rewritten as follows:

$$\mathcal{L}(Z_c, Z_v, P) = \alpha \left( \text{Tr}\left(P^T X Z_c L_c (P^T X Z_c)^T\right) - \text{Tr}\left(P^T X Z_v L_v (P^T X Z_v)^T\right) \right), \tag{9}$$

where  $L_c$  and  $L_v$  denote the Laplacian operators of  $W^c$  and  $W^v$ .  $\alpha$  is a balance parameter. Minimizing the subtraction of trace in

equation (9) can weaken the impact of view information and separate the class structure and the view structure.

**3.2.3. Discriminative Global Alignment.** The discriminative projection of all pairwise samples through  $\mathcal{L}(Z_c, Z_v, P)$  can reduce the impact between views, but the differences between learned features from different classes are not

significant enough. So, to further enhance the separation of two manifolds, we design a global discriminative constraint for cross-view analysis as the third term  $G(P, Z_c, Z_v)$ :

$$G(P, Z_c, Z_v) = \text{Tr}(S_W(P^T X Z_c)) - \text{Tr}(S_{B1}(P^T X Z_v)) - \text{Tr}(S_{B2}(P^T X Z_v)), \quad (10)$$

where  $S_W(P^T X Z_c)$  is within-class and between-view scatter matrix in class manifold and  $S_{Bi}(P^T X Z_v)$  ( $i = 1, 2$ ) is within-view and between-class scatter matrix in view manifold. These scatter matrices are formulated as

$$\begin{aligned} S_W(P^T X Z_c) &= \sum_{j=1}^c (\mu_j^1 - \mu_j^2)(\mu_j^1 - \mu_j^2)^T, \\ S_{Bi}(P^T X Z_v) &= \sum_{j=1}^c (\mu_j^i - \mu^i)(\mu_j^i - \mu^i)^T, \end{aligned} \quad (11)$$

where  $\mu^i$  ( $i = 1, 2$ ) denotes the overall mean feature from the  $i$ th view and  $\mu_j^i$  denotes the mean feature of the  $j$ th class from the  $i$ th view. In this way, the within-class view margin in class structure can be reduced, and the margin of between-class data from same view in view structure can be magnified. The third term can be framed as

$$\begin{aligned} G(P, Z_c, Z_v) &= \lambda_2 (\text{Tr}(S_W(P^T X Z_c)) - \text{Tr}(S_{B1}(P^T X Z_v)) - \text{Tr}(S_{B2}(P^T X Z_v))) \\ &= \lambda_2 (\|P^T X Z_c (V_1 - V_2)\|_F^2 - \|P^T X Z_v (V_1 - \hat{V}_1)\|_F^2 - \|P^T X Z_v (V_2 - \hat{V}_2)\|_F^2), \end{aligned} \quad (12)$$

where  $V_i$  is the coefficient matrix of the within-class mean feature of the  $i$ th view and  $\hat{V}_i$  ( $i = 1, 2$ ) is the coefficient matrix of the overall mean feature of the  $i$ th view. In detail,  $V_i(k, m) = (1/n_i^m)$  only if  $x_k$  belongs to the  $m$ th class from the  $i$ th view, where  $n_i^m$  denotes the number of samples of the  $m$ th class from the  $i$ th view; in other cases,  $V_i(k, m) = 0$ .  $\hat{V}_i(k, m) = (1/n_i)$  only if  $x_k$  belongs to the  $i$ th view, where  $n_i$  denotes the number of samples from the  $i$ th view; in other

cases,  $\hat{V}_i(k, m) = 0$ .  $\lambda_2$  is a trade-off parameter. Equation (12) achieves global alignment by the mean vectors of joint synthetic samples from global representation and further enforces the view-invariant constraint on the same class.

In addition, we add an orthogonal constraint  $P^T P = I$  to neglect trivial solutions. In the end, we rewrite equation (5) with all the terms as

$$\begin{aligned} &\|Z_c\|_* + \|Z_v\|_* + \lambda_1 \|E\|_{2,1} \\ \min_{Z_c, Z_v, E, P} &+ \alpha \left( \text{Tr}(P^T X Z_c L_c (P^T X Z_c)^T) - \text{Tr}(P^T X Z_v L_v (P^T X Z_v)^T) \right) \\ &+ \lambda_2 \left( \|P^T X Z_c (V_1 - V_2)\|_F^2 - \|P^T X Z_v (V_1 - \hat{V}_1)\|_F^2 - \|P^T X Z_v (V_2 - \hat{V}_2)\|_F^2 \right) \\ \text{s.t. } &X = X(Z_c + Z_v) + E, P^T P = I. \end{aligned} \quad (13)$$

**3.3. Optimization Scheme.** To obtain the feasible solution of  $Z_c$  and  $Z_v$ , we adopt two auxiliary variables  $J_c$  and  $J_v$ . Then, equation (13) can be transformed into the following term:

$$\begin{aligned} \min_{Z_c, Z_v, E, P, J_c, J_v} & \|J_c\|_* + \|J_v\|_* + \lambda_1 \|E\|_{2,1} \\ & + \alpha \left( \text{Tr} \left( P^T X Z_c L_c (P^T X Z_c)^T \right) - \text{Tr} \left( P^T X Z_v L_v (P^T X Z_v)^T \right) \right) \\ & + \lambda_2 \left( \|P^T X Z_c (V_1 - V_2)\|_F^2 - \|P^T X Z_v (V_1 - \hat{V}_1)\|_F^2 - \|P^T X Z_v (V_2 - \hat{V}_2)\|_F^2 \right) \\ \text{s.t. } & X = X(Z_c + Z_v) + E, P^T P = I, J_c = Z_c, J_v = Z_v. \end{aligned} \quad (14)$$

For optimization problems with equality constraints, the Augmented Lagrangian method is an effective

solution. The Augmented Lagrangian form of equation (14) is as follows:

$$\begin{aligned} \min_{Z_c, Z_v, E, P, J_c, J_v} & \|J_c\|_* + \|J_v\|_* + \lambda_1 \|E\|_{2,1} \\ & + \alpha \left( \text{Tr} \left( P^T X Z_c L_c (P^T X Z_c)^T \right) - \text{Tr} \left( P^T X Z_v L_v (P^T X Z_v)^T \right) \right) \\ & + \lambda_2 \left( \|P^T X Z_c (V_1 - V_2)\|_F^2 - \|P^T X Z_v (V_1 - \hat{V}_1)\|_F^2 - \|P^T X Z_v (V_2 - \hat{V}_2)\|_F^2 \right) \\ & + \text{Tr} \left( Y_1^T (X - X(Z_c + Z_v) - E) \right) + \text{Tr} \left( Y_2^T (J_c - Z_c) \right) + \text{Tr} \left( Y_3^T (J_v - Z_v) \right) \\ & + \frac{\eta}{2} \left( \|X - X(Z_c + Z_v) - E\|_F^2 + \|J_c - Z_c\|_F^2 + \|J_v - Z_v\|_F^2 \right) \\ \text{s.t. } & P^T P = I, \end{aligned} \quad (15)$$

where  $Y_1, Y_2$ , and  $Y_3$  are the Lagrange multipliers and  $\eta > 0$  is the penalty parameter. We use an alternating solution to optimize iteratively all variables. We define the left bottom of the variable plus  $t$  as the  $t$ -th solution.

First, by ignoring the other variables except  $P$ , equation (15) becomes

$$\begin{aligned} \min_{P_t} & \alpha \left( \text{Tr} \left( P_t^T X Z_{c,t} L_c (P_t^T X Z_{c,t})^T \right) - \text{Tr} \left( P_t^T X Z_{v,t} L_v (P_t^T X Z_{v,t})^T \right) \right) \\ & + \lambda_2 \left( \|P_t^T X Z_{c,t} (V_1 - V_2)\|_F^2 - \|P_t^T X Z_{v,t} (V_1 - \hat{V}_1)\|_F^2 - \|P_t^T X Z_{v,t} (V_2 - \hat{V}_2)\|_F^2 \right) \\ \text{s.t. } & P_t^T P_t = I. \end{aligned} \quad (16)$$

We obtain the projection matrix  $P_t$  one by one, because  $P_t$  is an orthogonal matrix. For the  $i$ th column of  $P_t$ , the objective function is rewritten as

$$\begin{aligned} \min_{P_{i,t}} & \alpha \left( \text{Tr} \left( P_{i,t}^T X Z_{c,t} L_c (P_{i,t}^T X Z_{c,t})^T \right) - \text{Tr} \left( P_{i,t}^T X Z_{v,t} L_v (P_{i,t}^T X Z_{v,t})^T \right) \right) + \varepsilon_{i,t} (P_{i,t}^T P_{i,t} - I) \\ & + \lambda_2 \left( \|P_{i,t}^T X Z_{c,t} (V_1 - V_2)\|_F^2 - \|P_{i,t}^T X Z_{v,t} (V_1 - \hat{V}_1)\|_F^2 - \|P_{i,t}^T X Z_{v,t} (V_2 - \hat{V}_2)\|_F^2 \right). \end{aligned} \quad (17)$$



We enforce the derivative of function (17) to be zero.

$$X \begin{pmatrix} Z_{v,t} \left( (V_1 - \hat{V}_1)(V_1 - \hat{V}_1)^T + (V_2 - \hat{V}_2)(V_2 - \hat{V}_2)^T + \alpha L_v \right) Z_{v,t}^T \\ -Z_{c,t} \left( (V_1 - V_2)(V_1 - V_2)^T + \alpha L_c \right) Z_{c,t}^T \end{pmatrix} X^T P_{i,t} = \varepsilon_{i,t} P_{i,t}. \quad (18)$$

Therefore,  $P_{i,t}$  is the  $i$ th eigenvector of the matrix in equation (18) and  $P$  can be simply solved.

Update  $J_c$ :

$$\begin{aligned} J_{c,t+1} &= \min_{J_{c,t}} \|J_{c,t}\|_* + \text{Tr}(Y_2^T (J_{c,t} - Z_{c,t})) + \frac{\eta_t}{2} \|J_{c,t} - Z_{c,t}\|_F^2 \\ &= \frac{1}{\eta_t} \|J_{c,t}\|_* + \frac{1}{2} \left\| J_{c,t} - \left( Z_{c,t} + \left( \frac{Y_{2,t}}{\eta_t} \right) \right) \right\|_F^2. \end{aligned} \quad (19)$$

The singular value thresholding is an approximate method to solve the above kernel norm minimization equations [37].

Update  $J_v$ :

$$\begin{aligned} J_{v,t+1} &= \min_{J_{v,t}} \|J_{v,t}\|_* + \text{Tr}(Y_3^T (J_{v,t} - Z_{v,t})) + \frac{\eta_t}{2} \|J_{v,t} - Z_{v,t}\|_F^2 \\ &= \frac{1}{\eta_t} \|J_{v,t}\|_* + \frac{1}{2} \left\| J_{v,t} - \left( Z_{v,t} + \left( \frac{Y_{3,t}}{\eta_t} \right) \right) \right\|_F^2. \end{aligned} \quad (20)$$

Equation (20) can be addressed in the same way as equation (19).

We enforce the derivative of equation (15) with respect to  $Z_c$  to be zero.

$$\begin{aligned} &(X^T P_t P_t^T X)^{-1} (I + X^T X) Z_{c,t+1} + Z_{c,t+1} \left( \frac{(\lambda_2 (V_1 - V_2)(V_1 - V_2)^T + \alpha L_c)}{\eta_t} \right) \\ &= (X^T P_t P_t^T X)^{-1} \left( X^T (X - XZ_{v,t} - E_t) + J_{c,t} + \frac{(X^T Y_{1,t} - Y_{2,t})}{\eta_t} \right). \end{aligned} \quad (21)$$

It is obvious that equation (21) is a Sylvester equation. We can easily solve it by [38]. Similarly, we enforce the derivative of equation (15) with respect to  $Z_v$  to be zero.

$$\begin{aligned} &(X^T P_t P_t^T X)^{-1} (I + X^T X) Z_{v,t+1} \\ &+ Z_{v,t+1} \left( \frac{(\lambda_2 (V_1 - \hat{V}_1)(V_1 - \hat{V}_1)^T + \lambda_2 (V_2 - \hat{V}_2)(V_2 - \hat{V}_2)^T + \alpha L_v)}{\eta_t} \right) \\ &= (X^T P_t P_t^T X)^{-1} \left( X^T (X - XZ_{c,t} - E_t) + J_{v,t} + \frac{(X^T Y_{1,t} - Y_{3,t})}{\eta_t} \right). \end{aligned} \quad (22)$$

Equation (22) can be addressed in the same way as equation (21).

Update  $E$ :

$$E_{t+1} = \min_{E_t} \frac{\lambda_1}{\eta_t} \|E_t\|_{2,1} + \frac{1}{2} \|E_t - \left( X - X(Z_{c,t} + Z_{v,t}) + \frac{Y_1}{\eta_t} \right)\|_F^2. \quad (23)$$

The above equation is a  $L_{2,1}$ -norm minimization problem whose solution is shown in [39].

The entire numerical iterative scheme for equation (14) is shown in Algorithm 1, where the parameters  $\rho, \theta, t_{\max}, \eta$ , and  $\eta_{\max}$  are set empirically. Moreover, the matrices  $Z_c, Z_v, E, Y_1, Y_2$ , and  $Y_3$  are initialized as 0 and the parameters  $\alpha, \lambda_1$ , and  $\lambda_2$  are tuned by the experiments.

**3.4. Complexity Analysis.** According to the above computational process and Algorithm 1, we discuss the computational complexity of the proposed algorithm in detail. In

**Input:** data matrix  $X$ , parameters  $\alpha, \lambda_1$  and  $\lambda_2$   
**Initialize:**  $\rho = 1.3$ ,  $\theta = 10^{-9}$ ,  $t = 0$ ,  $t_{\max} = 200$ ,  $\eta_0 = 0.1$  and  $\eta_{\max} = 10^{10}$ ;  
**while not converged or  $t \leq t_{\max}$  do**  
 (1) Solving  $P_{t+1}$  by equation (18);  
 (2) Solving  $J_{c,t+1}$  by equation (19);  
 (3) Solving  $J_{v,t+1}$  by equation (20);  
 (4) Solving  $Z_{c,t+1}$  by equation (21);  
 (5) Solving  $Z_{v,t+1}$  by equation (22);  
 (6) Solving  $E_t$  by equation (23);  
 (7) Updating  $Y_{1,t+1}$  by  $Y_{1,t+1} = Y_{1,t} + \eta_t (X - X(Z_{c,t+1} + Z_{v,t+1}) - E)$ ;  
 (8) Updating  $Y_{2,t+1}$  by  $Y_{2,t+1} = Y_{2,t} + \eta_t (J_{c,t+1} - Z_{c,t+1})$ ;  
 (9) Updating  $Y_{3,t+1}$  by  $Y_{3,t+1} = Y_{3,t} + \eta_t (J_{v,t+1} - Z_{v,t+1})$ ;  
 (10) Updating the parameter  $\eta_{t+1}$  by  $\eta_{t+1} = \min(\eta_{\max}, \rho\eta_t)$ ;  
 (11) Checking convergence by  $\max(\|X - X(Z_{c,t+1} + Z_{v,t+1}) - E_{t+1}\|_{\infty}, \|J_{c,t+1} - Z_{c,t+1}\|_{\infty}, \|J_{v,t+1} - Z_{v,t+1}\|_{\infty}) < \theta$ ;  
 (12)  $t = t + 1$ .  
**end while**  
**Output:**  $J_c, J_v, Z_c, Z_v, E, P$

ALGORITHM 1: Algorithm for our proposed framework.

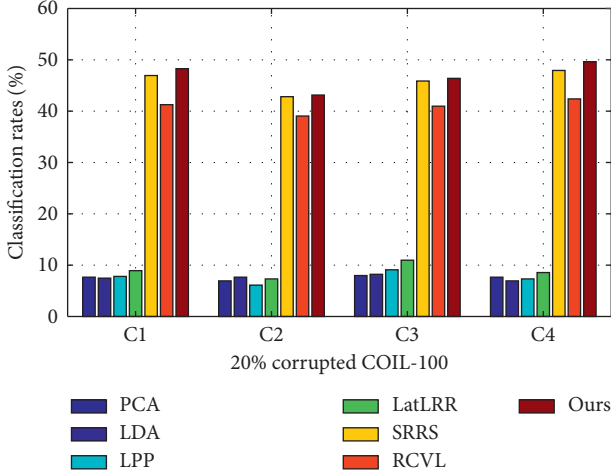


FIGURE 2: Classification results of different experiments on 20% corrupted COIL-100.

Algorithm 1, the main factor of algorithm complexity depends on Steps 1–5. Equation (18) is a typical characteristic equation, which costs  $O(n^3)$ , where  $n$  is the number of the training samples. The SVD decomposition in Steps 2 and 3 takes about  $O(n^3)$ , but it can be reduced  $O(n^2)$  due to the low-rank matrices  $J_c$  and  $J_v$ , where  $n$  is the rank of low-rank matrix. Equations (21) and (22) are two Sylvester equations whose computational complexity is  $O(n^3)$ . In summary, the computational complexity of our proposed method is  $O(n^3)$ .

## 4. Experiments

In this section, we evaluate the performance of our proposed method with classification task. Firstly, we introduce four cross-view datasets: face database, object databases, image-text database, and experimental setup. Secondly, we adopt several excellent subspace learning algorithms for comparison with ours. The initializations of all unknown parameters are tuned to get the best experimental results. The

analysis of parameters is shown in Figure 2. In addition, each experiment is repeated 10 times and the average classification results are shown.

**4.1. Experimental Datasets.** **CMU-PIE Face dataset** contains face images in 9 postures and 21 illumination conditions of 68 people. In our experiment, we chose 4 kinds of postures, which are P05, P09, P27, and P29. The original dimension of the cropped face images is  $64 \times 64$ . To enhance efficiency of our algorithm, the 300-dimensional principal feature extracted by PCA was adopted to our experiment.

**Wikipedia dataset**, which is an image-text bimodal dataset, consists of 2866 pairwise samples from 10 classes. The dimensions of the image and text are 4096 and 100, respectively. Therefore, due to inconsistency of dimensionality of the two features, we use PCA to adjust the image dimension.

**COIL-20 object dataset** is composed of 20 objects from a level 360-degree view. There is  $5^\circ$  between every two adjacent images, so each category has 72 samples. We divide the 72 images into two groups, G1 and G2. In addition, G1 is composed of samples from V1 [ $0^\circ, 85^\circ$ ] and V2 [ $185^\circ, 265^\circ$ ]. Similarly, G2 is composed of samples from V3 [ $90^\circ, 175^\circ$ ] and V4 [ $270^\circ, 355^\circ$ ].

**COIL-100 object dataset** is an extension of the COIL-20. The only difference is that the COIL-100 is composed of 100 objects from a level 360-degree view. Therefore, the set of the COIL-100 dataset is similar to that of the COIL-20 dataset.

**4.2. Experimental Results and Analysis.** In experiments, we need not use any information about the test set, including class and view information. We select several subspace learning methods as comparison methods, that is, PCA, LDA, locality preserving projections (LPP) [40], LatLRR, SRRS, and RCVL. After extracting feature with the comparison methods, we uniformly choose KNN as classifier to evaluate their performance. In addition, we also add 10



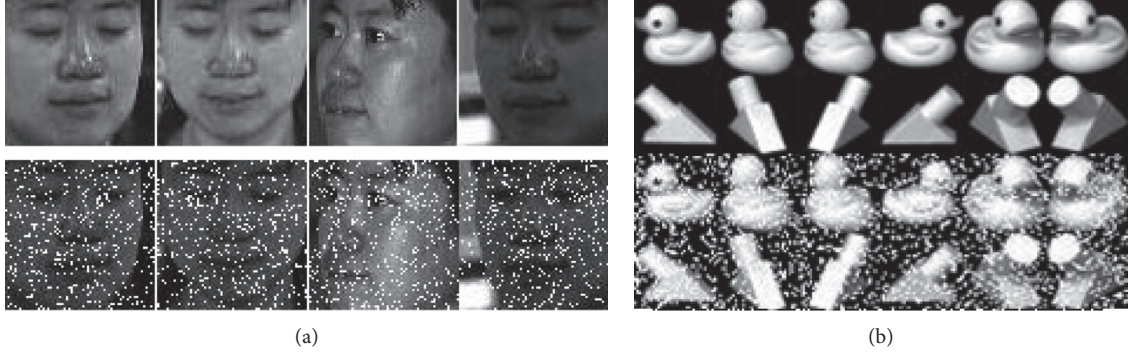


FIGURE 3: The instances of original data and corrupted data. (a) CMU-PIE face dataset. (b) COIL-20 object dataset.

TABLE 1: Classification results (%) of all methods on original CMU-PIE dataset.

| Methods | C1                      | C2                      | C3                      | C4                      | C5                      | C6                      |
|---------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| LDA     | 62.96 $\pm$ 0.66        | 66.76 $\pm$ 0.58        | 62.16 $\pm$ 1.20        | 61.50 $\pm$ 1.14        | 56.54 $\pm$ 0.55        | 61.83 $\pm$ 0.88        |
| PCA     | 48.28 $\pm$ 0.59        | 50.50 $\pm$ 0.72        | 49.07 $\pm$ 0.95        | 48.43 $\pm$ 1.11        | 45.51 $\pm$ 0.84        | 49.68 $\pm$ 0.43        |
| LatLRR  | 65.10 $\pm$ 0.97        | 66.61 $\pm$ 1.57        | 62.47 $\pm$ 1.34        | 63.09 $\pm$ 2.08        | 61.04 $\pm$ 1.87        | 60.42 $\pm$ 0.73        |
| LPP     | 62.40 $\pm$ 0.80        | 60.17 $\pm$ 0.27        | 61.97 $\pm$ 0.33        | 62.13 $\pm$ 0.45        | 58.34 $\pm$ 0.10        | 60.72 $\pm$ 0.28        |
| SRRS    | 95.35 $\pm$ 1.05        | 91.66 $\pm$ 1.84        | 95.82 $\pm$ 1.35        | 90.22 $\pm$ 0.26        | 96.04 $\pm$ 1.07        | 87.16 $\pm$ 0.55        |
| RCVL    | 97.14 $\pm$ 0.09        | <b>93.70</b> $\pm$ 0.59 | 97.26 $\pm$ 0.04        | <b>92.99</b> $\pm$ 0.12 | 97.55 $\pm$ 0.06        | 88.60 $\pm$ 0.03        |
| Ours    | <b>98.27</b> $\pm$ 0.11 | 93.47 $\pm$ 0.06        | <b>98.15</b> $\pm$ 0.08 | 91.69 $\pm$ 0.08        | <b>98.53</b> $\pm$ 0.02 | <b>90.27</b> $\pm$ 0.14 |

TABLE 2: Classification results (%) of all methods on 10% corrupted CMU-PIE dataset.

| Methods | C1                      | C2                      | C3                      | C4                      | C5                      | C6                      |
|---------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| LDA     | 20.45 $\pm$ 0.04        | 21.01 $\pm$ 0.06        | 19.16 $\pm$ 0.17        | 16.42 $\pm$ 0.12        | 13.64 $\pm$ 0.15        | 14.96 $\pm$ 0.05        |
| PCA     | 48.09 $\pm$ 0.74        | 46.90 $\pm$ 0.25        | 47.74 $\pm$ 0.31        | 47.99 $\pm$ 0.08        | 46.90 $\pm$ 0.10        | 46.63 $\pm$ 0.06        |
| LatLRR  | 59.14 $\pm$ 1.82        | 60.38 $\pm$ 1.10        | 57.94 $\pm$ 0.81        | 58.10 $\pm$ 1.24        | 57.03 $\pm$ 0.56        | 56.98 $\pm$ 0.96        |
| LPP     | 38.47 $\pm$ 0.27        | 34.97 $\pm$ 0.80        | 37.85 $\pm$ 0.47        | 38.06 $\pm$ 0.24        | 33.16 $\pm$ 0.63        | 35.77 $\pm$ 0.57        |
| SRRS    | 69.54 $\pm$ 0.41        | 66.91 $\pm$ 0.96        | 70.02 $\pm$ 0.96        | 68.89 $\pm$ 0.31        | 70.28 $\pm$ 0.97        | 65.64 $\pm$ 0.51        |
| RCVL    | 87.14 $\pm$ 0.04        | 78.10 $\pm$ 0.07        | 85.52 $\pm$ 0.12        | 77.75 $\pm$ 0.07        | 86.87 $\pm$ 0.10        | 76.09 $\pm$ 0.05        |
| Ours    | <b>88.68</b> $\pm$ 0.05 | <b>79.22</b> $\pm$ 0.02 | <b>87.31</b> $\pm$ 0.04 | <b>79.75</b> $\pm$ 0.11 | <b>90.69</b> $\pm$ 0.09 | <b>78.39</b> $\pm$ 0.04 |

TABLE 3: Classification results (%) of all methods on 20% corrupted CMU-PIE dataset.

| Methods | C1                      | C2                      | C3                      | C4                      | C5                      | C6                      |
|---------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|-------------------------|
| LDA     | 9.23 $\pm$ 0.10         | 8.66 $\pm$ 0.07         | 8.57 $\pm$ 0.13         | 6.27 $\pm$ 0.11         | 6.82 $\pm$ 0.12         | 6.59 $\pm$ 0.10         |
| PCA     | 25.98 $\pm$ 0.06        | 25.99 $\pm$ 0.15        | 26.58 $\pm$ 0.14        | 21.97 $\pm$ 0.10        | 22.01 $\pm$ 0.19        | 20.15 $\pm$ 0.17        |
| LatLRR  | 38.26 $\pm$ 1.04        | 34.67 $\pm$ 1.18        | 35.00 $\pm$ 0.84        | 36.11 $\pm$ 1.12        | 34.97 $\pm$ 0.74        | 35.76 $\pm$ 0.78        |
| LPP     | 34.25 $\pm$ 0.22        | 30.08 $\pm$ 0.67        | 29.89 $\pm$ 0.93        | 33.27 $\pm$ 0.25        | 30.95 $\pm$ 0.51        | 31.07 $\pm$ 0.59        |
| SRRS    | 74.30 $\pm$ 0.14        | 63.02 $\pm$ 0.19        | 72.79 $\pm$ 0.15        | 59.21 $\pm$ 0.36        | 68.73 $\pm$ 0.17        | 54.98 $\pm$ 0.27        |
| RCVL    | 70.63 $\pm$ 0.08        | 61.39 $\pm$ 0.16        | 71.44 $\pm$ 0.10        | 57.34 $\pm$ 0.16        | 65.02 $\pm$ 0.09        | 53.05 $\pm$ 0.07        |
| Ours    | <b>74.58</b> $\pm$ 0.03 | <b>64.74</b> $\pm$ 0.12 | <b>73.17</b> $\pm$ 0.07 | <b>60.86</b> $\pm$ 0.08 | <b>69.06</b> $\pm$ 0.06 | <b>56.56</b> $\pm$ 0.10 |

TABLE 4: Classification results (%) of all methods on original COIL-20 dataset.

| Methods | C1                      | C2                      | C3                      | C4                      |
|---------|-------------------------|-------------------------|-------------------------|-------------------------|
| LDA     | 63.85 $\pm$ 1.89        | 61.23 $\pm$ 1.02        | 64.12 $\pm$ 2.07        | 62.77 $\pm$ 1.73        |
| PCA     | 65.3 $\pm$ 0.76         | 63.42 $\pm$ 1.88        | 66.83 $\pm$ 1.45        | 70.31 $\pm$ 1.60        |
| LatLRR  | 80.65 $\pm$ 1.57        | 81.7 $\pm$ 2.13         | 80.39 $\pm$ 1.05        | 86.8 $\pm$ 1.29         |
| LPP     | 76.06 $\pm$ 1.17        | 74.6 $\pm$ 1.04         | 77.53 $\pm$ 1.45        | 78.99 $\pm$ 1.86        |
| SRRS    | 85.56 $\pm$ 1.05        | 83.89 $\pm$ 0.63        | 83.47 $\pm$ 0.84        | 94.58 $\pm$ 0.77        |
| RCVL    | 83.22 $\pm$ 0.97        | 79.04 $\pm$ 1.25        | 81.44 $\pm$ 0.75        | 90.23 $\pm$ 1.00        |
| Ours    | <b>86.39</b> $\pm$ 0.55 | <b>84.61</b> $\pm$ 0.99 | <b>85.67</b> $\pm$ 0.26 | <b>95.06</b> $\pm$ 0.57 |

TABLE 5: Classification results (%) of all methods on 10% corrupted COIL-20 dataset.

| Methods | C1                                 | C2                                 | C3                                 | C4                                 |
|---------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| LDA     | 34.13 $\pm$ 1.11                   | 35.12 $\pm$ 0.87                   | 35.27 $\pm$ 1.24                   | 38.55 $\pm$ 1.47                   |
| PCA     | 49.44 $\pm$ 0.97                   | 51.69 $\pm$ 0.94                   | 48.31 $\pm$ 1.49                   | 50.81 $\pm$ 0.61                   |
| LatLRR  | 58.71 $\pm$ 1.58                   | 57.74 $\pm$ 1.49                   | 57.66 $\pm$ 1.27                   | 60.00 $\pm$ 1.51                   |
| LPP     | 57.63 $\pm$ 1.28                   | 54.84 $\pm$ 1.42                   | 58.25 $\pm$ 1.31                   | 58.62 $\pm$ 1.01                   |
| SRRS    | 80.72 $\pm$ 0.70                   | 74.36 $\pm$ 0.79                   | 75.50 $\pm$ 0.41                   | 89.06 $\pm$ 0.57                   |
| RCVL    | 78.20 $\pm$ 1.07                   | 71.92 $\pm$ 1.02                   | 73.67 $\pm$ 0.80                   | 87.13 $\pm$ 0.55                   |
| Ours    | <b>82.22 <math>\pm</math> 0.90</b> | <b>75.14 <math>\pm</math> 0.65</b> | <b>77.08 <math>\pm</math> 0.31</b> | <b>90.28 <math>\pm</math> 0.93</b> |

TABLE 6: Classification results (%) of all methods on 20% corrupted COIL-20 dataset.

| Methods | C1                                 | C2                                 | C3                                 | C4                                 |
|---------|------------------------------------|------------------------------------|------------------------------------|------------------------------------|
| LDA     | 21.45 $\pm$ 1.22                   | 20.76 $\pm$ 1.57                   | 21.15 $\pm$ 1.06                   | 20.77 $\pm$ 1.39                   |
| PCA     | 43.49 $\pm$ 1.46                   | 44.16 $\pm$ 0.64                   | 43.98 $\pm$ 1.10                   | 45.03 $\pm$ 1.42                   |
| LatLRR  | 33.71 $\pm$ 1.13                   | 34.92 $\pm$ 0.88                   | 30.73 $\pm$ 1.28                   | 33.55 $\pm$ 1.36                   |
| LPP     | 35.13 $\pm$ 1.04                   | 31.26 $\pm$ 0.97                   | 32.72 $\pm$ 1.60                   | 34.16 $\pm$ 1.31                   |
| SRRS    | 75.64 $\pm$ 1.02                   | 66.72 $\pm$ 0.84                   | 70.56 $\pm$ 1.17                   | 81.14 $\pm$ 0.45                   |
| RCVL    | 73.20 $\pm$ 0.94                   | 64.96 $\pm$ 0.48                   | 66.03 $\pm$ 0.56                   | 79.02 $\pm$ 1.26                   |
| Ours    | <b>76.25 <math>\pm</math> 0.57</b> | <b>68.36 <math>\pm</math> 0.24</b> | <b>71.69 <math>\pm</math> 1.10</b> | <b>82.64 <math>\pm</math> 1.16</b> |

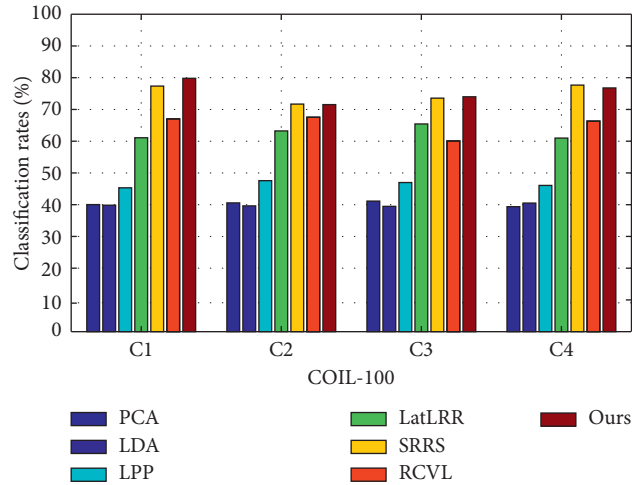


FIGURE 4: Classification results of different experiments on COIL-100.

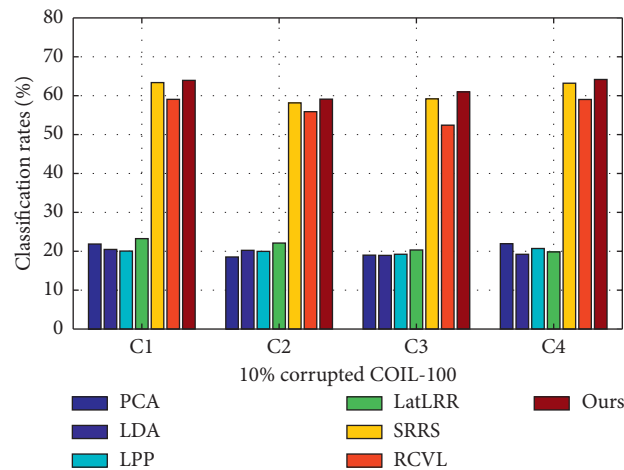


FIGURE 5: Classification results of different experiments on 10% corrupted COIL-100.

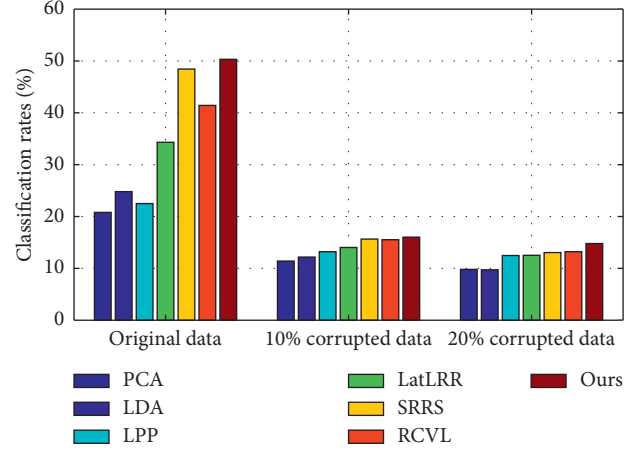
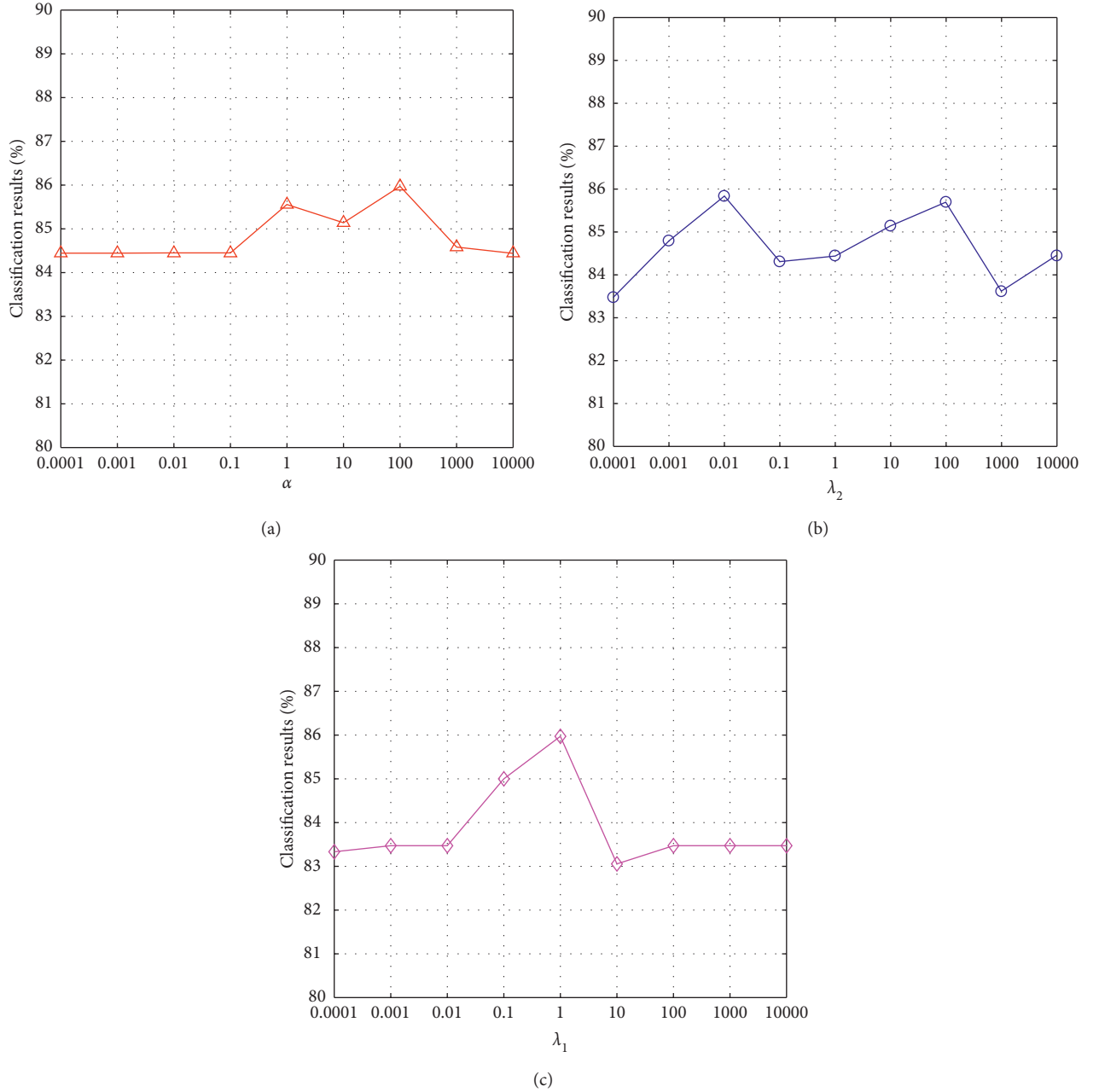


FIGURE 6: Classification results of comparison experiments on original and corrupted Wikipedia.

FIGURE 7: Classification results with different parameters (a)  $\alpha$ . (b)  $\lambda_2$ . (c)  $\lambda_1$ .

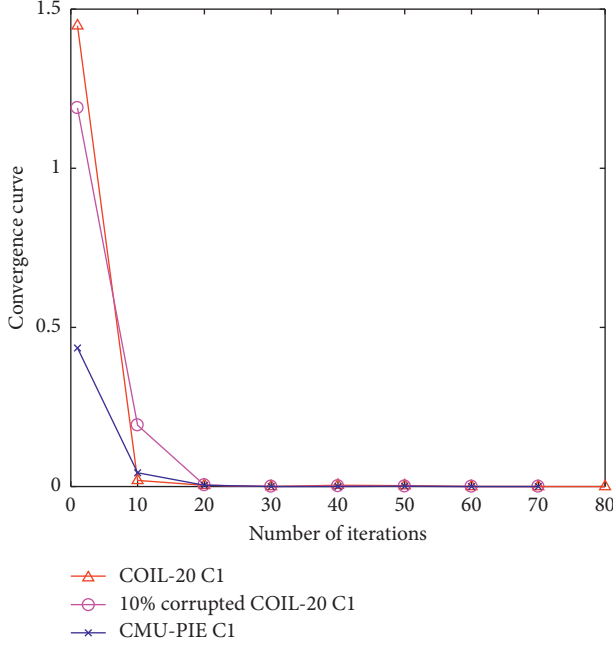


FIGURE 8: Convergence curves on different datasets.

percent and 20 percent of random noise to part of the datasets to demonstrate the adaptability of our subspace learning algorithm to different levels of corrupted data and some instances are shown in Figure 3.

For CMU-PIE, we randomly perform cross-view subspace learning on two poses, with a total of 6 experimental groups, which are C1{P05,P09}, C2{P05,P27}, C3{P05,P29}, C4{P09,P27}, C5{P09,P29}, and C6{P27,P29}. Tables 1–3 respectively show the classification results of all experimental algorithms on the original data, 10% noisy data, and 20% noisy data. For COIL-20 and COIL-100 object datasets, we select two sets of samples from G1 and G2 as a cross-view training set, respectively, and the others as a test set. So we get 4 experimental groups from COIL-20 and COIL-100 datasets, including C1{V1,V3}, C2{V1,V4}, C3{V2,V4}, and C4{V2,V3}. Tables 4–6 display the classification results of all experimental algorithms on the original data, 10% noisy data, and 20% noisy data from COIL-20. Figures 2, 4, and 5 show the experimental results of four groups from original COIL-100 dataset, 10% corrupted COIL-100 dataset, and 20% corrupted COIL-100 dataset. For Wikipedia, we use the reduced dimensionality image feature and text feature as two views and Figure 6 displays the results of comparison experiments on the original data, 10% noisy data, and 20% noisy data.

The results of experiments prove that our method achieves the persistent higher classification results compared to other methods. For noisy data, the classification results of most methods based on LRR are more robust than those of other methods. It is due to the fact that low-rank representation framework can restore raw information from corrupted data by learning latent structure. Besides, another result also can be found that the classification results of methods used for cross-view data are better than those of

other comparisons. Our proposed method projects data into the discriminative view-invariant subspace via dual low-rank representations framework, so that the method can better learn from cross-view data.

**4.3. Performance Evaluations.** In this part, we test what parameters should we assign to ensure that the performance of our method can get a best grade. Then, we show the convergence of our algorithm.

There are three tunable parameters  $\alpha$ ,  $\lambda_1$ , and  $\lambda_2$  in our framework. We evaluate the effect of parameters on COIL-20 C1.  $\alpha$ ,  $\lambda_2$  are two parameters to adjust discriminative local alignment and discriminative global alignment. From Figures 7(a) and 7(b), it can be seen that our method gets the best result, when  $\alpha = 100$  and  $\lambda_2 = 0.01$ . Furthermore,  $\lambda_1$  is a parameter to constrain the corrupted data and the classification result is optimal around 1.

In the end, we show the convergence analysis of our method on different datasets: the original, the 10% corrupted COIL-20 C1, and the CMU-PIE C1. The maximum value of  $\|X - X(Z_{c,t+1} + Z_{v,t+1}) - E_{t+1}\|_{\infty}$ ,  $\|J_{c,t+1} - Z_{c,t+1}\|_{\infty}$ , and  $\|J_{v,t+1} - Z_{v,t+1}\|_{\infty}$  is used as convergence criterion in each iteration. The variation of the maximum value with the increase of the number of iterations is shown in Figure 8. The curves point out that the proposed algorithm converges steadily and efficiently after 20 iterations.

## 5. Conclusions

We proposed a subspace learning algorithm with discriminative constraint via low-rank representation to solve the cross-view recognition task. Our method can learn a distribution-invariant subspace from cross-view data by designing two substantial structures with dual low-rank constraints. We also integrate the local alignment and the global alignment into our framework to eliminate the interference caused by the view manifold in the subspace. Meanwhile, we also design a feasible iterative scheme to ensure that the model converges and obtains the optimal solution. Extensive experiments on several public datasets prove that our proposed method has strong robustness and stability for cross-view classification tasks.

## Data Availability

The datasets used in this paper can be downloaded through the following links: (1) CMU-PIE face dataset: <http://vasc.ricmu.edu/idb/html/face/>. (2) Wikipedia dataset: <http://www.svcl.ucsd.edu/projects/crossmodal/>. (3) COIL-20 dataset: <https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>. (4) COIL-100 dataset: <https://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php>.

## Conflicts of Interest

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

## Acknowledgments

This work was supported in part by the National Natural Science Foundation of China (Grants 61501147 and 62071157), University Nursing Program for Young Scholars with Creative Talents in Heilongjiang Province (Grant UNPYSCT-2018203), Natural Science Foundation of Heilongjiang Province (Grant YQ2019F011), Fundamental Research Foundation for University of Heilongjiang Province (Grant LGYC2018JQ013), and Postdoctoral Foundation of Heilongjiang Province (Grant LBH-Q19112).

## References

- [1] G. Guo, G. Mu, Y. Fu et al., "Human age estimation using bio-inspired features," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 27, pp. 112–119, 2009.
- [2] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" *Neural Information Processing Systems*, vol. 27, pp. 3320–3328, 2014.
- [3] H. Kuang, X. Zhang, Y.-J. Li, L. L. H. Chan, and H. Yan, "Nighttime vehicle detection based on bio-inspired image enhancement and weighted score-level feature fusion," *IEEE Transactions on Intelligent Transportation Systems*, vol. 18, no. 4, pp. 927–936, 2017.
- [4] P. Hu, D. Peng, Y. Sang, and Y. Xiang, "Multi-view linear discriminant analysis network," *IEEE Transactions on Image Processing*, vol. 28, no. 11, pp. 5352–5365, 2019.
- [5] D. Liu, X. Yang, D. Wang, and Q. Wei, "Reinforcement-learning-based robust controller design for continuous-time uncertain nonlinear systems subject to input constraints," *IEEE Transactions on Cybernetics*, vol. 45, no. 7, pp. 1372–1385, 2015.
- [6] Y. Zhong, L. Zhang, B. Huang et al., "An unsupervised artificial immune classifier for multi/hyperspectral remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 2, pp. 420–431, 2006.
- [7] 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.
- [8] C. Peng, Z. Kang, and Q. Cheng, "Subspace clustering via variance regularized ridge regression," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 35, pp. 2931–2940, 2017.
- [9] Z. Xue, J. Du, D. Du, and S. Lyu, "Deep low-rank subspace ensemble for multi-view clustering," *Information Sciences*, vol. 482, pp. 210–227, 2019.
- [10] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," *IEEE International Conference on Computer Vision*, vol. 35, pp. 1615–1622, 2011.
- [11] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 11, pp. 2160–2173, 2016.
- [12] W. K. Wong, Z. Lai, J. Wen, X. Fang, and Y. Lu, "Low-rank embedding for robust image feature extraction," *IEEE Transactions on Image Processing*, vol. 26, no. 6, pp. 2905–2917, 2017.
- [13] Z. Ren, Q. Sun, B. Wu, X. Zhang, and W. Yan, "Learning latent low-rank and sparse embedding for robust image feature extraction," *IEEE Transactions on Image Processing*, vol. 29, pp. 2094–2107, 2020.
- [14] A. Li, D. Chen, Z. Wu et al., "Self-supervised sparse coding scheme for image classification based on low-rank representation," *PloS One*, vol. 13, no. 6, pp. 1–15, 2018.
- [15] A. Li, X. Liu, Y. Wang et al., "Subspace structural constraint-based discriminative feature learning via nonnegative low-rank representation," *PloS One*, vol. 14, no. 5, pp. 1–19, 2019.
- [16] A. Li, Y. Ding, D. Chen, G. Sun, H. Jiang, and Q. Wu, "Cross-view feature learning via structures unlocking based on robust low-rank constraint," *IEEE Access*, vol. 8, pp. 46851–46860, 2020.
- [17] A. Li, Y. Ding, D. Chen, G. Sun, and H. Jiang, "Discriminative subspace learning for cross-view classification with simultaneous local and global alignment," *Neural Computing for Advanced Applications*, vol. 1265, pp. 168–179, 2020.
- [18] S. Yi, Z. He, Y.-M. Cheung, and W.-S. Chen, "Unified sparse subspace learning via self-contained regression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 28, no. 10, pp. 2537–2550, 2018.
- [19] L. Zhang, B. Ma, G. Li, Q. Huang, and Q. Tian, "Generalized semi-supervised and structured subspace learning for cross-modal retrieval," *IEEE Transactions on Multimedia*, vol. 20, no. 1, pp. 128–141, 2018.
- [20] Z. Lai, D. Mo, J. Wen, L. Shen, and W. K. Wong, "Generalized robust regression for jointly sparse subspace learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 29, no. 3, pp. 756–772, 2019.
- [21] C. Peng, Y. Chen, Z. Kang, C. Chen, and Q. Cheng, "Robust principal component analysis: a factorization-based approach with linear complexity," *Information Sciences*, vol. 513, pp. 581–599, 2020.
- [22] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [23] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.
- [24] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, 2011.
- [25] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *International Journal of Computer Vision*, vol. 109, no. 1–2, pp. 74–93, 2014.
- [26] H. Mobahi, Z. Zhou, A.-Y. Yang et al., "Holistic 3d reconstruction of urban structures from low-rank textures," *IEEE International Conference on Computer Vision Workshops*, vol. 35, pp. 593–600, 2011.
- [27] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," *IEEE International Conference on Data Mining*, vol. 35, pp. 110–119, 2014.
- [28] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 188–194, 2016.
- [29] Y. Zhao, X. You, S. Yu et al., "Multi-view manifold learning with locality alignment," *Pattern Recognition*, vol. 78, pp. 154–166, 2018.
- [30] Z. Ding and Y. Fu, "Dual low-rank decompositions for robust cross-view learning," *IEEE Transactions on Image Processing*, vol. 28, no. 1, pp. 194–204, 2019.
- [31] A. Li, Z. Wu, H. Lu, D. Chen, and G. Sun, "Collaborative self-regression method with nonlinear feature based on multi-task

- learning for image classification,” *IEEE Access*, vol. 6, pp. 43513–43525, 2018.
- [32] C. Peng, C. Chen, Z. Kang et al., “RES-PCA: a scalable approach to recovering low-rank matrices,” *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, vol. 35, pp. 7309–7317, 2019.
  - [33] Z. Kang, H. Pan, S. C. H. Hoi, and Z. Xu, “Robust graph learning from noisy data,” *IEEE Transactions on Cybernetics*, vol. 50, no. 5, pp. 1833–1843, 2020.
  - [34] A. Li, R. An, D. Chen et al., “Semi-supervised subspace learning for pattern classification via robust low rank constraint,” *Mobile Networks & Applications*, vol. 35, 2020.
  - [35] Z. Kang, X. Lu, Y. Lu, C. Peng, W. Chen, and Z. Xu, “Structure learning with similarity preserving,” *Neural Networks*, vol. 129, no. 3, pp. 138–148, 2020.
  - [36] Y.-F. Guo, S.-J. Li, J.-Y. Yang et al., “A generalized Foley–Sammon transform based on generalized Fisher discriminant criterion and its application to face recognition,” *Pattern Recognition Letters*, vol. 24, no. 1–3, pp. 147–158, 2003.
  - [37] J. F. Cai, E. J. Candès, and A. Shen, “A singular value thresholding algorithm for matrix completion,” *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
  - [38] R. H. Bartels and G. W. Stewart, “Solution of the matrix equation  $AX + XB = C$  [F4],” *Communications of the ACM*, vol. 15, no. 9, pp. 820–826, 1972.
  - [39] J. Yang, W. Yin, Y. Zhang, and Y. Wang, “A fast algorithm for edge-preserving variational multichannel image restoration,” *SIAM Journal on Imaging Sciences*, vol. 2, no. 2, pp. 569–592, 2009.
  - [40] X. He and P. Niyogi, “Locality preserving projections,” *Advances in Neural Information Processing Systems*, vol. 2, pp. 153–160, 2004.