


Received February 17, 2020, accepted March 8, 2020, date of publication March 16, 2020, date of current version March 26, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2980964

Multiset Canonical Correlations Analysis With Global Structure Preservation

HONGJIE ZHANG¹, JINXIN ZHANG², YANWEN LIU¹, AND LING JING¹ 

¹College of Science, China Agricultural University, Beijing 100083, China

²College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, China

Corresponding author: Ling Jing (jingling@cau.edu.cn)

This work was supported by the National Natural Science Foundation of China under Grant 11671032.

ABSTRACT This paper considers unsupervised dimensionality reduction of multi-view data, where locality preserving canonical correlation analysis and a new locality-preserving canonical correlation analysis are two typical effective methods. However, they ignore the global structure while considering the local structure of data, and are sensitive to noises because of the relationship of neighbors based on the Euclidean distance. In this paper, we propose a novel multi-view dimensionality reduction method: multiset canonical correlations analysis based on low-rank representation. Our model introduces the cross-view similarity matrix to consider the correlation of all different points in cross views, which makes it not only preserve the local structure but also the global structure of data. And the cross-view similarity matrix is constructed by using low-rank representation, which can make the model more robust. In addition, a parameter β is introduced to adjust the importance of the correlation of different sample points, enhancing the generalization ability for different datasets. Experiments on four multi-view datasets show our proposed method has better performance than the related methods.

INDEX TERMS Unsupervised learning, multi-view learning, dimensionality reduction, similarity matrix, global structure.

I. INTRODUCTION


In the fields of machine learning and pattern recognition, due to the advancement of data acquisition technology, a large number of data are presented in various forms, which forms multi-view data [1]–[4]. For example, an object can be described in various forms such as its shape, color or texture. Multi-view data has more feature information than single view data, and the feature information in each view tends to be complementary [5]. This makes the study of multi-view data more meaningful than single view data [6].

Because multi-view data is often represented in high-dimensional feature space, directly learning multi-view data not only cause a huge waste of time and cost, but also leads to “dimension disaster”, so multi-view data learning is also facing more severe challenges [7]–[10].

Dimensionality reduction of multi-view data is one of the approaches to overcome this difficulty and has attracted widespread attention from researchers [11]–[14]. Dimensionality reduction can be divided into supervised, semi-supervised and unsupervised [15]. This paper is

concerned with the last one, which is formulated as follows: Suppose that m views for N patterns are given as $\{X^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_N^{(i)}) \in R^{D_i \times N}\}_{i=1}^m$, where D_i denotes the dimension of the i -th view. We hope to find m suitable projection matrix $\{P^{(i)} \in R^{D_i \times d}\}_{i=1}^m$ ($d < D_i$) to reduce the dimensions of multi-view data into a lower common dimension by $\{P^{(i)T} X^{(i)}\}_{i=1}^m$.

Among recent methods [16]–[20] the most representative is to apply the canonical correlation analysis (CCA) [16], [21], [22] to two views of data. The idea of CCA is to extract the canonical variables $P^{(1)T} X^{(1)}$ and $P^{(2)T} X^{(2)}$ from the mean-normalized two-view data $X^{(1)} = [x_1^{(1)}, \dots, x_N^{(1)}]$ and $X^{(2)} = [x_1^{(2)}, \dots, x_N^{(2)}]$ by maximizing $(P^{(1)T} X^{(1)}, P^{(2)T} X^{(2)})$, where $(P^{(1)T} X^{(1)}, P^{(2)T} X^{(2)})$ represents the correlation between $P^{(1)T} X^{(1)}$ and $P^{(2)T} X^{(2)}$. Or equivalently by maximizing $\sum_{i=1}^N (P^{(1)T} x_i^{(1)}, P^{(2)T} x_i^{(2)})$. Clearly the quantity $\sum_{i=1}^N (P^{(1)T} x_i^{(1)}, P^{(2)T} x_i^{(2)})$ can be considered as the overall correlation and $(P^{(1)T} x_i^{(1)}, P^{(2)T} x_i^{(2)})$ as the correlation of the same sample point in cross views. So CCA is to seek the maximum sum of the correlation of the same sample point in cross views after dimension reduction. Later the above CCA for two-view data is extended to deal

The associate editor coordinating the review of this manuscript and approving it for publication was Tossapon Boongoen .

with multi-view data, e. g. MCCA [23]. Note that both CCA and MCCA are only suitable for dimensionality reduction of samples with linear distribution [24], [25]. In order to break the limitation, kernel function is introduced. Firstly, the sample points are mapped nonlinearly into the high-dimensional kernel function space, so that the sample points with non-linear distribution are transformed into the ones with linear distribution. Then, the dimensionality reduction of the sample points is carried out through the traditional CCA in the kernel function space. This is the well-known kernel canonical correlation analysis (KCCA) [26]. However, the KCCA also has obvious shortcomings: for all sample points in each view, the uniform non-linear mapping makes KCCA difficult to find a suitable and effective kernel function. Even if such a kernel function exists, the generalization ability of the model will be greatly weakened [27]. Thus in order to protect local manifold structure, Sun and Chen proposed the locality preserving canonical correlation analysis (LPCCA) [27] for two-view data, which is based on locally linear embedding (LLE) [28] and locality preserving projections (LPP) [29] for single view data. LPCCA embeds near-neighbor information by deleting non-neighbor points when calculating correlation matrix with CCA, protecting the local manifold structure of sample points. It divides data into multiple small neighborhoods where the data in each small neighborhood can be considered as linearly distributed [28]. However, when calculating the overall correlation it only considers the correlation of the same sample point in cross views, but does not require the correlation of different sample points in cross views. As a further improvement, a new locality-preserving canonical correlation analysis (ALPCCA) [30] was proposed by Wang and Zhang, when calculating overall correlation, not only the correlation of the same point in cross views, but also the correlation of the neighbors in cross views is taken into account. This method adds local neighbor information of the data to the model so that the local manifold structure of the data can be maintained. Compared with LPCCA, ALPCCA improves the discriminating ability of the model to some extent [30]. And because ALPCCA chooses “add strategy” instead of “delete strategy”, it is more effective than LPCCA to solve “small sample problem”. However, both LPCCA and ALPCCA only pay attention to the protection of the local structure of the sample points and neglect the global structure of the data. And their embedding of local neighbors in the model depends on the Euclidean distance between neighbors, which is sensitive to noisy data [31]. In addition, because the fixed parameters are used to generate the similarity matrix, these methods cannot generate adaptive neighborhoods and greatly reduce the generalization ability of the model. Moreover, both LPCCA and ALPCCA are suitable for dimensionality reduction of two views, but not for the case of more than two views.

In this paper we propose a novel multi-view data dimension reduction method: multiset canonical correlations analysis based on low-rank representation (LRMCCA), which protects both the local structure and the global structure

of sample points. First, we perform low-rank representation [32]–[37] of the data, spontaneously learn the low-rank representation matrix of the sample points in each view. Then, to maintain the global structure [35] of the data, we construct the cross-view similarity matrix by using the low-rank representation matrix, so that we can get the similarity between the sample points, including all sample points in cross views. Moreover, our LRMCCA is robust to noisy data due to using low-rank representation instead of using Euclidean distance to learn similarity matrix.

Our LRMCCA has the following three characteristics:

- The cross-view similarity matrix is introduced to reflect the structure property of both the local and the global of data.
- The cross-view similarity matrix is constructed through low-rank representation instead of Euclidean distance. This makes our model not sensitive to the outliers and no need to select neighbor points.
- A parameter β is introduced to adjust the importance of the correlation of different sample points in cross views to enhance the generalization ability for different datasets.

The rest of the paper is organized as follows: Multiset canonical correlation analysis and low-rank representation are reviewed in Section II in brief. In Section III, our LRMCCA are presented, as well as the solution and complexity analysis of our method. A series of comparative experiments are shown to verify the advantages of our LRMCCA in Section IV. Section V gives the summarization, including the possibility to extend our LRMCCA to supervised and semi-supervised learning.

II. USEFUL WORKS

Four works, MCCA, LPCCA, ALPCCA and LRR are useful for our study later.

A. MCCA

Consider the multi-view dimensionality reduction problem proposed in Section I, MCCA seeks m projection matrix $\{P^{(i)} \in R^{D_i \times d}\}_{i=1}^m$ ($d < D_i$), to maximize the sum of the pairwise correlation between the transformed variables $\{P^{(i)T} X^{(i)}\}_{i=1}^m$ under some constraints. More precisely, it constructs the optimization problem:

$$\begin{aligned} & \max_{P^{(1)}, P^{(2)}, \dots, P^{(m)}} \sum_{i=1}^m \sum_{j=1, j \neq i}^m P^{(i)T} C^{(ij)} P^{(j)} \\ & s.t. \sum_{i=1}^m P^{(i)} C^{(ii)} P^{(i)} = I \quad (i = 1, \dots, m) \end{aligned} \quad (1)$$

where $I \in R^{d \times d}$ is identity matrix, $C^{(ii)} = \sum_{k=1}^N x_k^{(i)} x_k^{(i)T} = X^{(i)} X^{(i)T}$ is the within-set scatter matrix in the i -th views, $C^{(ij)} = \sum_{k=1}^N x_k^{(i)} x_k^{(j)T} = X^{(i)} X^{(j)T}$ ($i \neq j$) is the between-set scatter matrix between the i -th and j -th view. The solution can be obtained by solving a generalized eigenvalue problem [23].

B. LPCCA

LPCCA [27] is applied to solve two-view feature extraction problem. It aims at exploring locality correlations between different views. LPCCA is defined by the following optimization problem:

$$\begin{aligned} \max_{P^{(1)}, P^{(2)}} \quad & \text{tr} \left(P^{(1)T} \sum_{i,j=1}^N S_{ij}^{(1)} (x_i^{(1)} - x_j^{(1)}) S_{ij}^{(2)} (x_i^{(2)} - x_j^{(2)})^T P^{(2)} \right) \\ \text{s.t.} \quad & P^{(1)T} \sum_{i,j=1}^N S_{ij}^{(1)} (x_i^{(1)} - x_j^{(1)}) S_{ij}^{(1)} (x_i^{(1)} - x_j^{(1)})^T P^{(2)} = I \\ & P^{(2)T} \sum_{i,j=1}^N S_{ij}^{(2)} (x_i^{(2)} - x_j^{(2)}) S_{ij}^{(2)} (x_i^{(2)} - x_j^{(2)})^T P^{(2)} = I \quad (2) \end{aligned}$$

where the weight values $S_{ij}^{(1)}$ and $S_{ij}^{(2)}$ are calculated by:

$$S_{ij}^{(1)} = \begin{cases} \exp \left(- \left\| x_i^{(1)} - x_j^{(1)} \right\|_2^2 / t \right) & \text{if } x_i^{(1)} \text{ and } x_j^{(1)} \\ & \text{are the neighbors} \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

$$S_{ij}^{(2)} = \begin{cases} \exp \left(- \left\| x_i^{(2)} - x_j^{(2)} \right\|_2^2 / t \right) & \text{if } x_i^{(2)} \text{ and } x_j^{(2)} \\ & \text{are the neighbors} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where $t > 0$ is a given parameter.

C. ALPCCA

ALPCCA [30] not only considers the correlation of the same points in cross views, but also the correlation of the neighbors in cross views. ALPCCA is defined by the following optimization problem:

$$\begin{aligned} \max_{P^{(1)T}, P^{(2)T}} \quad & \text{tr} \left(P^{(1)T} \bar{C}_{12} P^{(2)T} \right) \\ \text{s.t.} \quad & P^{(1)T} C_{11} P^{(2)T} = I \\ & P^{(1)T} C_{22} P^{(2)T} = I \quad (5) \end{aligned}$$

where $\bar{C}_{12} = \sum_{i=1}^N x_i^{(1)} (x_i^{(2)})^T + \sum_{i=1}^N \sum_{j=1}^N S_{ij}^{(1)} x_i^{(1)} (x_j^{(2)})^T + \sum_{i=1}^N \sum_{j=1}^N S_{ij}^{(2)} x_i^{(2)} (x_j^{(1)})^T$ and $C_{11} = X^{(1)} X^{(1)T}$ and $C_{22} = X^{(2)} X^{(2)T}$.

In (5), $S_{ij}^{(1)}$ and $S_{ij}^{(2)}$ are defined by (3) and (4), respectively.

D. LRR

Suppose that there is a dictionary $A = [a_1, a_2, \dots, a_N] \in R^{D \times N}$. Low-rank representation finds a matrix $Z \in R^{N \times N}$ with the lowest rank such that the data set $X = \{x_1, x_2, \dots, x_N\} \in R^{D \times N}$ can be represented as a linear combination AZ . The corresponding optimization problem can be written as

$$\min_Z \|Z\|_*, \quad \text{s.t. } X = AZ \quad (6)$$

where $\|\cdot\|_*$ represents the nuclear norm of a matrix. In real applications, we often choose sample matrix X as dictionary matrix A . The above problem become

$$\min_Z \|Z\|_*, \quad \text{s.t. } X = XZ \quad (7)$$

or approximately

$$\min_{Z, E} \|Z\|_* + \alpha \|E\|_\ell, \quad \text{s.t. } X = XZ + E \quad (8)$$

where α is a positive parameter, E is a sparse additive error matrix. For noisy data sets, low-rank representation has strong denoising ability. For different distribution of noise in datasets, $\|E\|_\ell$ can be selected as $\|E\|_F^2$, $\|E\|_1$ and $\|E\|_{2,1}$ etc. [32].

III. MULTISSET CANONICAL CORRELATIONS ANALYSIS BASED ON LOW-RANK REPRESENTATION

Now we are in a position to show our LRMCCA. Its model and the working procedure are given in Section A and FIGURE 1 respectively, and the corresponding solution and the complexity analysis are shown in Sections B and C.

A. LRMCCA MODEL

Consider the multi-view dimensionality reduction problem proposed in Section I. Previous studies have shown that LRR mentioned in Section II can spontaneously learn the low-rank representation coefficient matrix [38], [39]. And based on the coefficient matrix, the cross-view similarity matrix of sample points can be constructed, which does not need to select the global parameters as K-nearest neighbor algorithm [40] does, but can automatically obtain the adaptive neighborhood of sample points. Furthermore, it captures the local structure and the global structure of data.

More precisely, the optimization problem of low-rank representation of sample points $X^{(i)}$ in the i -th view is as follows:

$$\begin{aligned} \min_{Z^{(i)}, E^{(i)}} \quad & \|Z^{(i)}\|_* + \alpha \|E^{(i)}\|_\ell \\ \text{s.t.} \quad & X^{(i)} = X^{(i)} Z^{(i)} + E^{(i)} \quad (i = 1, \dots, m) \quad (9) \end{aligned}$$

The solution $Z^{(i)}$ can be used to get the lower dimension sample points

$$\tilde{X}^{(i)} = X^{(i)} Z^{(i)}, \quad i = 1, \dots, m \quad (10)$$

with denoising property. It can also be used to give the similarity [35] between the k_1 -th sample point and the k_2 -th sample point in the i -th view

$$s_{k_1 k_2}^{(i)} = \frac{1}{2} \left(\left| z_{k_1 k_2}^{(i)} \right| + \left| z_{k_2 k_1}^{(i)} \right| \right) \quad (11)$$

Thus, we get the symmetric similarity matrix $S^{(i)}$, $i = 1, \dots, m$ of sample points under all views.

Furthermore, the similarity between the k_1 -th sample point in the i -th view and the k_2 -th sample point in the j -th view should be

$$w_{k_1 k_2}^{(ij)} = \frac{1}{2} (s_{k_1 k_2}^{(i)} + s_{k_1 k_2}^{(j)}) \quad (12)$$

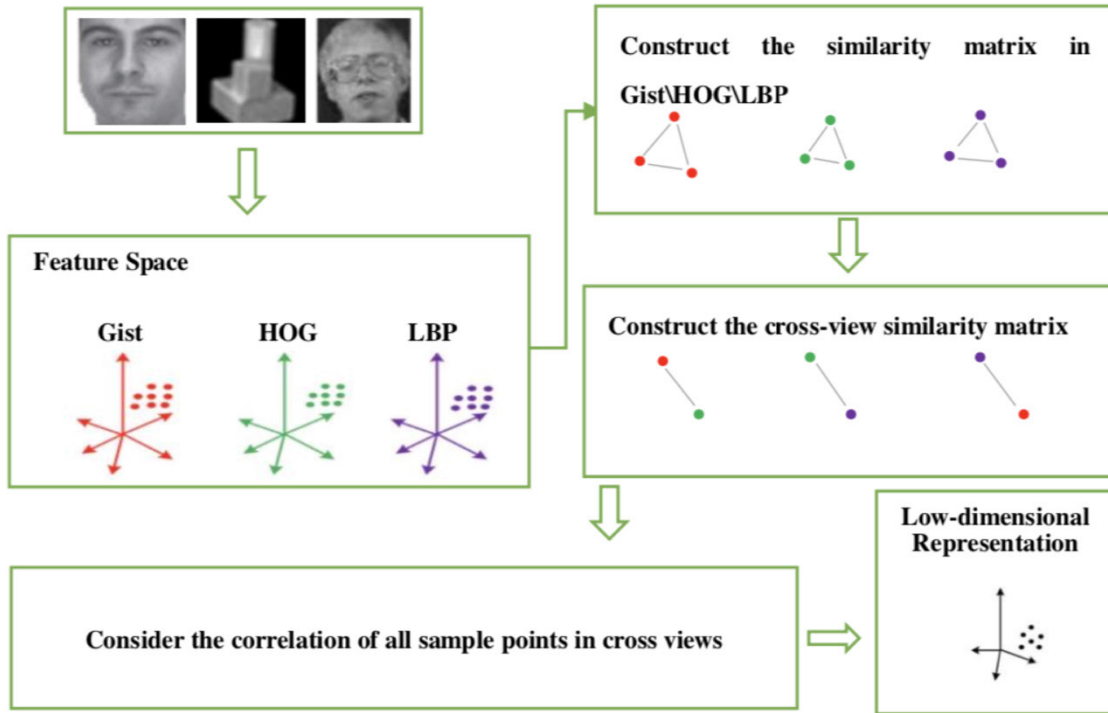


FIGURE 1. The working procedure of LRMCCA. Firstly, the low-rank representation coefficient matrix is learned in each view. Secondly, the cross-view similarity matrix is constructed by using the coefficient matrix. Finally, the cross-view similarity matrix is used to find the projection matrices to maintain the local and global structure in low-dimensional space.

In this way, we can get the cross-view similarity matrix

$$W^{(ij)} = \begin{bmatrix} w_{11}^{(ij)} & \cdots & w_{1N}^{(ij)} \\ \vdots & \ddots & \vdots \\ w_{N1}^{(ij)} & \cdots & w_{NN}^{(ij)} \end{bmatrix}, \quad i, j = 1, \dots, m \quad (13)$$

between the sample points of any cross views.

It should be noted that the element values representing the similarity between sample points does not have a specific physical meaning as distance does, it is only a relative value. So the similarity matrix $S^{(i)}$ should be normalized: divide each element in $S^{(i)}$ by its maximum element $A^{(i)} = \max S^{(i)}, i = 1, \dots, m$. This yields

$$W^{(ij)} = \frac{1}{2} \left(\frac{S^{(i)}}{A^{(i)}} + \frac{S^{(j)}}{A^{(j)}} \right), \quad i, j = 1, \dots, m. \quad (14)$$

Now let us turn to our ultimate goal to find m projection matrices $\{P^{(i)} \in R^{D_i \times d}\}_{i=1}^m$. We emphasize to protect the global structure of data by considering the correlation of all sample points in cross views, yielding our model—the optimization problem with the variables $\{P^{(i)} \in R^{D_i \times d}\}_{i=1}^m$

$$\begin{aligned} & \max_{P^{(1)}, P^{(2)}, \dots, P^{(m)}} \sum_{i=1}^m \sum_{j=1, j \neq i}^m P^{(i)T} \tilde{C}^{(ij)} P^{(j)} \\ & s.t. \sum_{i=1}^m P^{(i)T} \tilde{C}^{(ii)} P^{(i)} = 1 \quad (i = 1, \dots, m) \end{aligned} \quad (15)$$

$$\begin{aligned} \tilde{C}^{(ij)} &= \sum_{k=1}^N \tilde{x}_k^{(i)} \tilde{x}_k^{(j)} + \beta w_{k_1 k_2}^{(ij)} \sum_{k_1=1}^N \sum_{k_2=1}^N \tilde{x}_{k_1}^{(i)} \tilde{x}_{k_2}^{(j)} \\ &= \tilde{X}^{(i)} \tilde{X}^{(j)T} + \beta \tilde{X}^{(i)} W^{(ij)} \tilde{X}^{(j)T} \\ &= \tilde{X}^{(i)} \tilde{X}^{(j)T} + \beta \tilde{X}^{(i)} H^{(ij)} \tilde{X}^{(j)T} \\ &= \tilde{X}^{(i)} (I + \beta H^{(ij)}) \tilde{X}^{(j)T} \end{aligned} \quad (16)$$

where $\tilde{X}^{(i)} = [x_1^{(i)}, \dots, x_N^{(i)}]$ and $W^{(ij)}$ are given by (10) (11) and (14) respectively, $\tilde{C}^{(ii)}$ represents the within-set scatter matrix in the i -th view, $H^{(ij)}$ is the weight matrix obtained by zeroing the diagonal elements of matrix $W^{(ij)}$ such that there is no duplication, β is a positive parameter to adjust the importance of the correlation of different sample points in cross views. Note that the correlation of different sample points in cross views after dimension reduction cannot be greater than that of the same point, so β should lie between 0 and 1. And the model degenerates into MCCA when $\beta=0$.

B. LRMCCA SOLUTION

The solutions of optimization problem (9) and problem (15) are considered here. In problem (9), let $\ell=2,1$, this means we adopt $L_{2,1}$ -norm for the matrix $E^{(i)}$, introduce variable J to separate the objective function and get:

$$\begin{aligned} & \min_{Z^{(i)}, E^{(i)}, J^{(i)}} \|Z^{(i)}\|_* + \alpha \|E^{(i)}\|_{2,1} \\ & s.t. X^{(i)} = X^{(i)}Z^{(i)} + E^{(i)} \\ & \quad Z^{(i)} = J^{(i)} \quad (i = 1, \dots, m) \end{aligned} \quad (17)$$

TABLE 1. Description of the benchmark datasets.

Datasets	Number of sample points	Number of training points	Number of test points	Reduced dimensions
AR	2600	1000	1600	150
COIL-20	1440	300	1140	150
ORL	400	240	160	150
Yale	165	105	60	50

TABLE 2. Highest classification accuracy (mean%) of different methods on four datasets.

Method	AR	COIL-20	ORL	Yale
MCCA	60.51	90.20	88.04	77.67
LRR+MCCA	60.52	91.88	90.42	77.67
MVPLS	57.77	92.63	90.37	80.78
coLPP	63.72	91.85	90.17	83.61
LPCCA	64.19	90.67	88.68	79.11
ALPCCA	62.71	91.69	91.87	83.67
LRMCCA	65.10	93.05	92.04	85.22



FIGURE 2. Sample images of one identity in datasets.

Using augmented Lagrange function leads to

$$\begin{aligned}
 &L^{(i)}(Z^{(i)}, E^{(i)}, J^{(i)}, M_1, M_2, \mu) \\
 &= \|J^{(i)}\|_* + \alpha \|E^{(i)}\|_{2,1} \\
 &\quad + \text{tr}(M_1^T (X^{(i)} - X^{(i)}Z^{(i)} - E^{(i)})) + \text{tr}(M_2^T (Z^{(i)} - J^{(i)})) \\
 &\quad + \frac{\mu}{2} (\|X^{(i)} - X^{(i)}Z^{(i)} - E^{(i)}\|_F^2 + \|Z^{(i)} - J^{(i)}\|_F^2) \quad (18)
 \end{aligned}$$

where $M_1^{(i)}, M_2^{(i)}$ is a Lagrange multiplier, $\mu_i > 0$ is a penalty parameter and $i = 1, \dots, m$ is a view label. Problem (18) is an unconstrained optimization problem and can be solved by iteration alternatively. In fact, fixing the variable $Z^{(i)}, E^{(i)}$, consider the objective function $L^{(i)}$ as a function of $J^{(i)}$ and get the new $J^{(i)}$. Then fix other variables in the same way to get the new $Z^{(i)}, E^{(i)}$ in turn [32], [41].

Now turn to the solution of problem (15). Similarly to that of MCCA, it can also be transformed into a generalized eigenvalue decomposition problem

$$\begin{bmatrix} 0 & \tilde{C}^{(12)} & \dots & \tilde{C}^{(1m)} \\ \tilde{C}^{(21)} & 0 & \dots & \tilde{C}^{(2m)} \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{C}^{(m1)} & \tilde{C}^{(m2)} & \dots & 0 \end{bmatrix} \begin{bmatrix} P^{(1)} \\ P^{(2)} \\ \vdots \\ P^{(m)} \end{bmatrix}$$

$$= \lambda \begin{bmatrix} \tilde{C}^{(11)} & & & \\ & \tilde{C}^{(22)} & & \\ & & \ddots & \\ & & & \tilde{C}^{(mm)} \end{bmatrix} \begin{bmatrix} P^{(1)} \\ P^{(2)} \\ \vdots \\ P^{(m)} \end{bmatrix} \quad (19)$$

The above discussion leads to the following algorithm:

C. COMPLEXITY ANALYSIS

We now briefly analyze the major computational complexity of our proposed LRMCCA, which is composed of two parts. The first part is mainly to minimize the nuclear norm in Step 1, which costs at most $O(D^3)$ (Suppose $D = \max\{D_1, D_2, \dots, D_m\}$ and $D > N$) in each iteration through SVD operation. So, the computational complexity is at most $O(mtD^3)$, where t denotes the loops in the inner iteration of Step 1. Fortunately, according to [32], [42], the SVD could be speeded up to $O(mtd^2D)$ where d is usually a small one. In the second part, the computational complexity of covariance matrices is $O(m^2)$ in Step 5. Therefore, the major computational complexity of LRMCCA is $O(mtd^2D + m^2)$.

IV. EXPERIMENTS AND RESULTS

In this part, in order to show the performance of our LRMCCA, we compare it with a series of state-of-art

Algorithm 1 LRMCCA

Input: data matrices $X^{(1)}, X^{(2)}, \dots, X^{(m)}$ parameters α, β .

Output: projection matrices $P^{(1)}, P^{(2)}, \dots, P^{(m)}$

Initialize:

$$Z^{(i)} = J^{(i)} = 0, \quad E^{(i)} = 0, \quad M_1 = 0, \quad M_2 = 0, \\ \mu = 10^{-6}, \quad \mu_{\max} = 10^6, \quad \rho = 1.1, \quad \varepsilon = 10^{-8}.$$

1. while not converged do (1) Update $J^{(i)}$ by

$$J = \arg \min \frac{1}{\mu} \|J^{(i)}\|_* + \frac{1}{2} \|J^{(i)} - (Z^{(i)} + M_2/\mu)\|_F^2,$$

which is solved via the Singular Value Thresholding (SVT) operator.

(2) Update $Z^{(i)}$ by

$$Z^{(i)} = (I + X^{(i)T}X^{(i)})^{-1}(X^{(i)T}(X^{(i)} - E^{(i)}) + J^{(i)} + (X^{(i)T}M_1 - M_2)/\mu).$$

(3) Update $E^{(i)}$ by

$$E^{(i)} = \arg \min \frac{\alpha_i}{\mu} \|E^{(i)}\|_{2,1} + \frac{1}{2} \|E^{(i)} - (X^{(i)} - X^{(i)}Z^{(i)} + M_1/\mu)\|_F^2.$$

(4) Update the multipliers M_1, M_2 by

$$M_1 = M_1 + \mu(X^{(i)} - X^{(i)}Z^{(i)} - E^{(i)}), \\ M_2 = M_2 + \mu(Z^{(i)} - J^{(i)}).$$

(5) Update the parameter μ by $\mu = \min(\rho\mu, \mu_{\max})$.

(6) Check the convergence conditions:

$$\|X^{(i)} - X^{(i)}Z^{(i)} - E^{(i)}\|_{\infty} < \varepsilon \text{ and } \|Z^{(i)} - J^{(i)}\|_{\infty} < \varepsilon.$$

end while

2. Get the denoised sample point by (10).
3. Construct and normalize the matrix $S^{(i)}$ by (11).
4. Calculate the cross-view similarity matrix $W^{(ij)}$ by (14). Set the weight matrix $H^{(ij)}$ to be the matrix that returns the diagonal elements of $W^{(ij)}$ to zero.
5. Calculate two scatter matrices $\tilde{C}^{(ij)}$ and $\tilde{C}^{(ii)}$ by (16).
6. Calculate the projection matrices $P^{(1)}, P^{(2)}, \dots, P^{(m)}$ by (19).

multi-view data reduction methods, including MCCA, LRR + MCCA, MVPLS, coLPP, LPCCA, ALPCCA (LRR + MCCA is used to denoise first, and then MCCA is used to reduce dimension) in four groups of real multi-view data. The nearest neighborhood classifier is used to obtain the classification accuracy, and the classification accuracy is used to evaluate the experimental results. Finally, the influence of parameters α, β on the experimental results is analyzed.

A. DATASETS

We perform experiments on four datasets. The sample images of one identity in each dataset and description of the datasets are given in FIGURE 2 and TABLE 1 respectively.

1) AR DATASET

The AR dataset was created by Aleix Martinez and Robert Benavente in the Computer Vision Center (CVC) of UAB. It is the most widely used standard database. It contains over 2600 face images of 100 individual (50 men and 50 women), including frontal facial images with different facial expressions, lighting conditions and occlusions. We randomly choose ten images of each person for training and the rest images for testing, so that here are 1000 training images and 1600 testing images in total.

2) COIL-20 DATASET

The COIL-100 Dataset was created by Columbia University in 1996. It contains 100 object images, each containing 72 color information. We select 20 object images and convert the color information of these 20 images into three view features in three different ways. We randomly choose fifteen images of each object for training and the rest images for testing, so that here are 300 training images and 1140 testing images in total.

3) ORL DATASET

The ORL face dataset contains 400 images of 40 different people. It was created by Olivetti Research Laboratory in Cambridge from April 1992 to April 1994. All images are stored in PGM format and are all sized 112×92 pixels with 256-level gray scale. Each image of the same person is captured in different time, light, facial expressions (open eyes/close eyes, smile/no smile) and facial details (glasses/no glasses). All images were taken in a darker uniform background, with the front face (some with a slight sideslip). We randomly choose six images of each person for training and the rest images for testing, so that here are 240 training images and 160 testing images in total.

4) YALE DATASET

It was created by Yale University Computer Vision and Control Center, it contains 165 GIF-formatted gray images of 15 volunteers. Each volunteer had 11 pictures, including changes in light, expression and posture. In experiments, we randomly choose seven pictures in each group for the training, and the rest four pictures for the testing, so that there are 105 training pictures and 60 testing pictures in total.

B. SELECTION OF PARAMETERS α AND β

The parameter $\alpha > 0$ is used to balance the importance of low rank and noise when the data is represented by a low-rank matrix. In general, the choice of this parameter depends on a priori knowledge of the error level of the data. The parameter $\beta \geq 0$ is used to balance the importance of the correlation of different sample points in cross views when calculating overall correlation in dimensional reduction. Because the correlation of different sample points in cross views cannot be greater than the correlation of the same points, we should limit β to between 0 and 1.

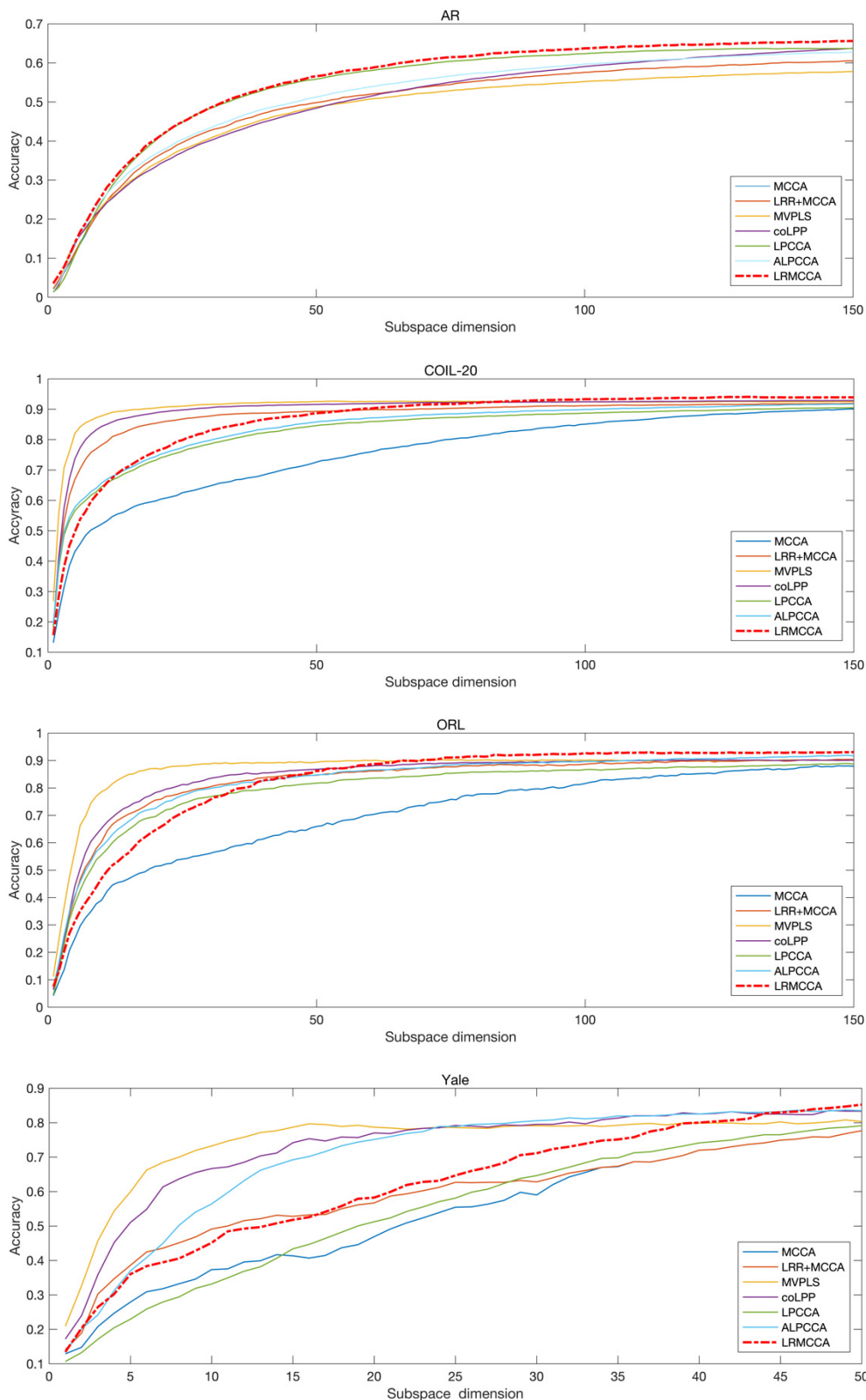


FIGURE 3. Classification accuracy on four datasets with different dimensions.

As we all know, it is very challenging to choose an appropriate parameter to achieve the highest classification accuracy. In the experimental part of this paper, we searched

within $\alpha = [1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10, 100]$ and $\beta = [1e-4, 1e-3, 1e-2, 1e-1, 1]$ by the method of five-fold cross-validation in order to obtain optimal parameters α and β .

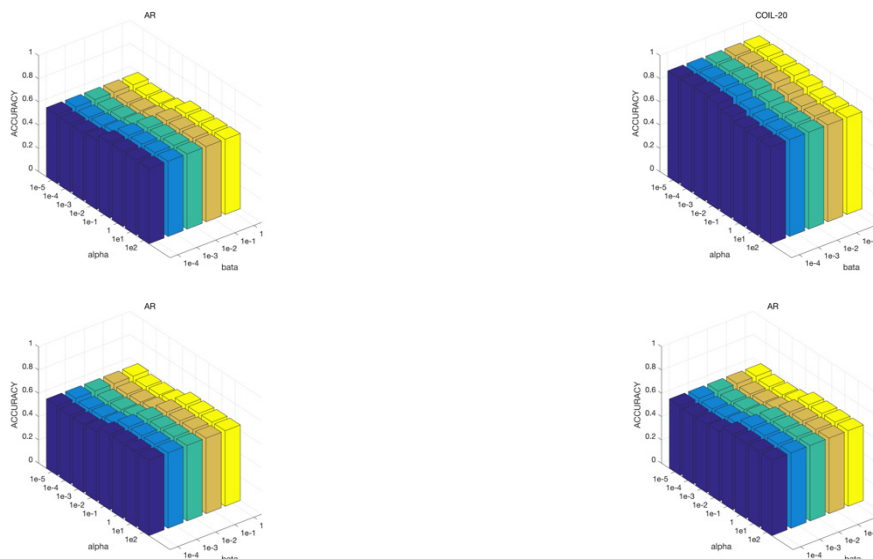


FIGURE 4. Classification accuracy changes with different parameters on four datasets.

The highest classification accuracy of different methods is given in TABLE 2, where the bold value is the maximum value among all methods in each dataset. And the classification accuracy with different dimensions is given in FIGURE 3. From these results, we can see that our method can lead to better classification effect on four datasets after dimension reduction.

C. EFFECTS OF PARAMETERS α AND β

In LRMCCA, the choice of α and β will affect the dimension reduction effect of the model. In order to make the reader to intuitively understand the influence of parameter changes on the dimensionality reduction effect of the model, we respectively give three-dimensional graphics of the classification accuracy corresponding to each group of parameters in FIGURE 4.

V. CONCLUSION

This paper presents a novel dimension reduction model for multi-view data: LRMCCA. It first uses the low-rank representation to learn the cross-view similarity matrix, then transforms the cross-view similarity matrix into the weight matrix, and finally uses the weight matrix for calculating the overall correlation. Thus the correlation between any two sample points in cross views is reasonably considered. For the data, this model protects not only its local structure, but also its global structure. Moreover, the low-rank representation is exploited, which generates adaptive neighborhoods. This makes the model have better generalization and noise resistance. Our numerical experiments validated the advantage of our approach.

It should be interesting to extend our approach from unsupervised learning to both supervised learning and semi-supervised learning. For supervised learning, what we need to

do is to add label information to LRMCCA; the sample points belonging to the same class should correspond to larger correlation. For semi-supervised learning, the similarity matrix obtained from the low-rank representation may be exploited to deduce the possibility of belonging to each class. This extra label information should be helpful in each view.

REFERENCES

- [1] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, Dec. 2013.
- [2] G. Li, K. Chang, and S. C. H. Hoi, "Multiview semi-supervised learning with consensus," *IEEE Trans. Knowl. Data Eng.*, vol. 24, no. 11, pp. 2040–2051, Nov. 2012.
- [3] C. Xu, D. Tao, and C. Xu, "Multi-view learning with incomplete views," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5812–5825, Dec. 2015.
- [4] Q. Yin, S. Wu, and L. Wang, "Unified subspace learning for incomplete and unlabeled multi-view data," *Pattern Recognit.*, vol. 67, pp. 313–327, Jul. 2017.
- [5] Y. Yang, W. Zhang, and Y. Xie, "Image automatic annotation via multi-view deep representation," *J. Vis. Commun. Image Represent.*, vol. 33, pp. 368–377, Nov. 2015.
- [6] F. Zhuang, G. Karypis, X. Ning, Q. He, and Z. Shi, "Multi-view learning via probabilistic latent semantic analysis," *Inf. Sci.*, vol. 199, pp. 20–30, Sep. 2012.
- [7] J. Zhao, X. Xie, X. Xu, and S. Sun, "Multi-view learning overview: Recent progress and new challenges," *Inf. Fusion*, vol. 38, pp. 43–54, Nov. 2017.
- [8] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.
- [9] Y. Li, M. Yang, and Z. Zhang, "A survey of multi-view representation learning," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 10, pp. 1863–1883, Oct. 2019.
- [10] L. Zhang, Q. Zhang, L. Zhang, D. Tao, X. Huang, and B. Du, "Ensemble manifold regularized sparse low-rank approximation for multiview feature embedding," *Pattern Recognit.*, vol. 48, no. 10, pp. 3102–3112, Oct. 2015.
- [11] H. Pan, J. He, Y. Ling, L. Ju, and G. He, "Graph regularized multiview marginal discriminant projection," *J. Vis. Commun. Image Represent.*, vol. 57, pp. 12–22, Nov. 2018.
- [12] J. Yu, D. Tao, Y. Rui, and J. Cheng, "Pairwise constraints based multiview features fusion for scene classification," *Pattern Recognit.*, vol. 46, no. 2, pp. 483–496, Feb. 2013.

- [13] Y. Han, F. Wu, D. Tao, J. Shao, Y. Zhuang, and J. Jiang, "Sparse unsupervised dimensionality reduction for multiple view data," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 10, pp. 1485–1496, Oct. 2012.
- [14] X. Chen, S. Chen, H. Xue, and X. Zhou, "A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data," *Pattern Recognit.*, vol. 45, no. 5, pp. 2005–2018, May 2012.
- [15] J. C. Ang, A. Mirzal, H. Haron, and H. N. A. Hamed, "Supervised, unsupervised, and semi-supervised feature selection: A review on gene selection," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, vol. 13, no. 5, pp. 971–989, Sep. 2016.
- [16] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, Dec. 2004.
- [17] X. Xu, Y. Yang, C. Deng, and F. Nie, "Adaptive graph weighting for multi-view dimensionality reduction," *Signal Process.*, vol. 165, pp. 186–196, Dec. 2019.
- [18] T. Apasiba Abeo, X.-J. Shen, B.-K. Bao, Z.-J. Zha, and J. Fan, "A generalized multi-dictionary least squares framework regularized with multi-graph embeddings," *Pattern Recognit.*, vol. 90, pp. 1–11, Jun. 2019.
- [19] X.-J. Shen, X.-Z. Luo, T. A. Abeo, Y. Yang, X. Shao, and S.-Y. Li, "Laplacian regularized kernel canonical correlation ensemble for remote sensing image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 7, pp. 1150–1154, Jul. 2019.
- [20] X.-J. Shen, S.-X. Liu, B.-K. Bao, C.-H. Pan, Z.-J. Zha, and J. Fan, "A generalized least-squares approach regularized with graph embedding for dimensionality reduction," *Pattern Recognit.*, vol. 98, Feb. 2020, Art. no. 107023, doi: [10.1016/j.patcog.2019.107023](https://doi.org/10.1016/j.patcog.2019.107023).
- [21] J. Wan, H. Wang, and M. Yang, "Cost sensitive semi-supervised canonical correlation analysis for multi-view dimensionality reduction," *Neural Process. Lett.*, vol. 45, no. 2, pp. 411–430, Apr. 2017.
- [22] L. Han, X.-Y. Jing, and F. Wu, "Multi-view local discrimination and canonical correlation analysis for image classification," *Neurocomputing*, vol. 275, pp. 1087–1098, Jan. 2018.
- [23] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 293–305, Mar. 2002.
- [24] Y.-Y. Lin, T.-L. Liu, and C.-S. Fuh, "Multiple Kernel learning for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1147–1160, Jun. 2011.
- [25] K. Yoshida, J. Yoshimoto, and K. Doya, "Sparse kernel canonical correlation analysis for discovery of nonlinear interactions in high-dimensional data," *BMC Bioinf.*, vol. 18, no. 1, p. 108, Dec. 2017.
- [26] T. Melzer, M. Reiter, and H. Bischof, "Appearance models based on Kernel canonical correlation analysis," *Pattern Recognit.*, vol. 36, no. 9, pp. 1961–1971, Sep. 2003.
- [27] T. Sun and S. Chen, "Locality preserving CCA with applications to data visualization and pose estimation," *Image Vis. Comput.*, vol. 25, no. 5, pp. 531–543, May 2007.
- [28] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.
- [29] W. Yu, X. Teng, and C. Liu, "Face recognition using discriminant locality preserving projections," *Image Vis. Comput.*, vol. 24, no. 3, pp. 239–248, Mar. 2006.
- [30] F. Wang and D. Zhang, "A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction," *Neural Process. Lett.*, vol. 37, no. 2, pp. 135–146, Apr. 2013.
- [31] J. Zhang, J. Wang, and X. Cai, "Sparse locality preserving discriminative projections for face recognition," *Neurocomputing*, vol. 260, pp. 321–330, Oct. 2017.
- [32] J. Chen and J. Yang, "Robust subspace segmentation via low-rank representation," *IEEE Trans. Cybern.*, vol. 44, no. 8, pp. 1432–1445, Aug. 2014.
- [33] W. K. Wong, Z. Lai, J. Wen, X. Fang, and Y. Lu, "Low-rank embedding for robust image feature extraction," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2905–2917, Jun. 2017.
- [34] L. Wei, X. Wang, A. Wu, R. Zhou, and C. Zhu, "Robust subspace segmentation by self-representation constrained low-rank representation," *Neural Process. Lett.*, vol. 48, no. 3, pp. 1671–1691, Dec. 2018.
- [35] Y. Zheng, X. Zhang, S. Yang, and L. Jiao, "Low-rank representation with local constraint for graph construction," *Neurocomputing*, vol. 122, pp. 398–405, Dec. 2013.
- [36] S. Zhan, J. Wu, N. Han, J. Wen, and X. Fang, "Unsupervised feature extraction by low-rank and sparsity preserving embedding," *Neural Netw.*, vol. 109, pp. 56–66, Jan. 2019.
- [37] L. Xie, M. Yin, X. Yin, Y. Liu, and G. Yin, "Low-rank sparse preserving projections for dimensionality reduction," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5261–5274, Nov. 2018.
- [38] C.-H. Zheng, Y.-F. Hou, and J. Zhang, "Improved sparse representation with low-rank representation for robust face recognition," *Neurocomputing*, vol. 198, pp. 114–124, Jul. 2016.
- [39] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
- [40] Y. Song, J. Liang, J. Lu, and X. Zhao, "An efficient instance selection algorithm for K nearest neighbor regression," *Neurocomputing*, vol. 251, pp. 26–34, Aug. 2017.
- [41] Y. Zhang, M. Xiang, and B. Yang, "Low-rank preserving embedding," *Pattern Recognit.*, vol. 70, pp. 112–125, Oct. 2017.
- [42] Z. Ding and Y. Fu, "Robust multiview data analysis through collective low-rank subspace," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1986–1997, May 2018.



HONGJIE ZHANG received the B.S. degree from Hebei Normal University, Hebei, China. He is currently pursuing the M.S. degree with the College of Science, China Agricultural University. His current research interests include dimensionality reduction and image classification.



JINXIN ZHANG received the B.S. and M.S. degrees from China Agricultural University (CAU), where he is currently pursuing the Ph.D. degree with the College of Information and Electrical Engineering. His current research interests include dimensionality reduction, image classification, and semi-supervised learning.



YANWEN LIU received the B.S. degree from Inner Mongolia University, Hohhot, China, in 2017. She is currently pursuing the M.S. degree with the College of Science, China Agricultural University, Beijing, China. Her current research interest includes dimensionality reduction and clustering.



LING JING received the Ph.D. degree from the Beijing University of Aeronautics and Astronautics (BUAA). She is currently a Professor with the College of Science of China Agricultural University (CAU). Her main research interests include machine learning, algorithm design, bioinformatics, and optimization. She is a Council Member of the Beijing Operations Research Society (BORS) and the Operations Research Society of China (ORSC).