IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 25, 2020, accepted March 2, 2020, date of publication March 5, 2020, date of current version March 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2978548

Cross-View Feature Learning via Structures Unlocking Based on Robust Low-Rank Constraint

AO LI[1], (Member, IEEE), YU DING[1], DEYUN CHEN[1], GUANGLU SUN[1], (Member, IEEE), HAILONG JIANG[2], AND QIDI WU[3]

[1]School of Computer Science and Technology, Harbin University of Science and Technology, Harbin 150080, China
[2]Department of Computer Science, Kent State University, Kent, OH 44240, USA
[3]College of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China

Corresponding author: Qidi Wu (373110650@qq.com)

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

ABSTRACT The cross-view multimedia are widely existed and attract many attentions in recent years. Nevertheless, it is noted that the phenomenon, that data in different classes from same view are more similar than that in same class from different views, is usually presented for cross-view multimedia data. The intrinsic imperfection leads disappointing performance for cross-view multimedia recognition or classification. To solve this problem, in this paper, we propose a novel discriminative learning framework with low-rank constraint, which can be applied for view-invariant low-dimensional subspace learning. The advantages of our framework include three aspects. Firstly, to unlock the latent class structure and view structure, a self-expressed model by dual low-rank constraints are presented, which can separate the two manifold structures in the learned subspace. Secondly, two effective discriminative graphs are constructed to guide the affinity relationship of data in the above two low-dimensional projected subspaces respectively. Finally, the joint semantic consensus constraint is designed to be integrated into the learning framework, which can explore the shared and view-specific information for enforcing the view-invariant character in semantic space. Experimental results on several public cross-view multimedia datasets demonstrate that our proposed method outperforms existing excellent subspace learning approaches.

INDEX TERMS Multimedia analysis, multi-view discriminative analysis, cross-view feature learning, low-rank representation.

I. INTRODUCTION

With the rapid development in the age of multimedia applications, data collected from different sensors usually have diverse representations. Even if the data generated from same sources while they can still represent different forms [1]–[7]. Therefore, data processing techniques become particularly important in many areas, such as data transmission [8]–[12], restoration [13],image fusion [14],and feature extraction [2]–[4], [15]–[19] and so on. In the real world, an object can be described under various conditions, including different angles or illumination, multiple modals or descriptions, etc [15], [16], [20]. Hence, cross-view data are frequently seen and attract gradually considerable attentions in multimedia data analysis [17]–[19]. Although multimedia data from different views can provide more completed information for object and promote discrimination ability in terms of theory, it brings some challenges that the data from same view but different classes are much more similar than that within same class but across multiple views due to the large divergence between views. Moreover, for the cross-view data, large discrepancy about the contents of different views also bring difficulties to their applications, e.g. an image obtained

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

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

from certain angle may include some special parts, which are invisible in other views.

Many methods have been proposed to solve the cross-view feature learning in the past decades. One straightforward approach is to re-measure the distance of the cross-view samples by projecting them into an common space. The most typical way is Canonical Correlation Analysis (CCA) [21], of which primary purpose is to respectively learn two transformation matrices of double views to discover a view-invariant space by maximizing the correlation between two projections onto cross-view samples. After that, for enhancing discriminative ability of the common space, Discriminant Canonical Correlation Analysis (DCCA) [22] is put forward by semantic constraint. However, both of these unsupervised and supervised methods are only available to two-views data. To overcome this problem, multi-view CCA [23], [24] and Generalized Multi-view Analysis (GMA) [3] were proposed to adapt to multi-view dataset. To further extend their applications, Multi-view Discrimination Analysis (MvDA) [25] is proposed, which learns the multi-view discriminative subspace by using the consistency constraint of multiple linear transformations.

Similarly, cross-modal data retrieval attracts many attentions and have been widely used in the field of multimedia analysis. Cross-modal data is isomorphic with cross-view data. Each modality has individual distribution character, so how to bridge the gap between different modalities is the most important problem for cross-modal retrieval. There is a way to fix up this difficulty by learning common latent subspace between various modalities, such as Cross-modal Factor analysis [26], generalized multi-view analysis [3] and cross-modal learning matent representation [27]. These methods realize the unsupervised learning in a pair-wise way by projecting samples into the common subspaces. When the semantic labels are available, a set of supervised feature learning methods for cross-modal data have been proposed, such as Supervised coupled-dictionary learning with group structures for multi-modal retrieval [28], local group based consistent feature learning [29] and adaptive hierarchical semantic aggregation [30]. In addition, there are some methods consider semi-supervised cross-modal feature learning, which can solve the problem with part of labeled data, such as generalized semi-supervised structured subspace learning [31], adaptive semi-supervised feature selection [32] and Robust graph-based semi-supervised learning [33].

Recently, low-rank learning [34]–[42] widely catches people's eyes, which can explore the intrinsic latent structures of data. Low-Rank Representation (LRR) [34] adopt low-rank constraint to find a novel data representation. After that, Supervised Regularization-based Robust Subspace (SRRS) [35], [36] is proposed to present the class label as the regularization term to learn a discriminative subspace, which puts the low-rank constraint and dimensionality reduction into a unified learning framework. Most recently, Low-Rank Embedding (LRE) [37] is put forward to learn low-rank representation and embedding structure jointly. In [40], structurally incoherent low-rank 2D locality preserving projection (SILR-2DLPP) is proposed to learn the projection subspace with the unconnected structure of data. However, the above mentioned methods merely project high dimensionality samples into a low dimensionality space, but neglect the intertwined view and class latent structures existed in cross-view data.

By contrast, we broaden the model and assume that there are shared and specific information between different view from the same class, which makes our approach more suitable for describing cross-view multimedia data. Based on mentioned considerations, we design a neoteric cross-view feature learning algorithm that named **J**oint **C**ross-View **S**ubspace **L**earning (JCSL) algorithm via unlocking structures based on dual low-rank constraints to find the discrimination common subspace. In general, the main contributions of our proposed JCSL are as follows:

- We use dual low-rank constraints to disassemble the daedal structures underlying cross-view data into two mutually independent structures that are class structure and view structure respectively. Hence, different from the conventional methods, our proposed method seeks two different low-dimensional manifold spaces for the above two mixed structures, which will make the learned subspace to separate class and view structures effectively.
- We set up an unlocking mechanism to particularly describe class structure and view structure by designing two graphs to guide the neighbor relationships among each pair of samples. The discrimination unlocking makes two potential structures work by minimizing within-class data from between-view and maximizing between-class data from same view.
- We develop a joint discriminative semantic consensus constraint for cross-view data alignment, which can explore their shared and view-specific structures so that our framework can capture the hidden concordant property from different views, and be more adaptive to classification task.

The rest of this paper are organized as follows. We briefly introduce some of the techniques involved in our framework in Section II. In Section III, the proposed cross-view feature learning algorithm and its theoretical knowledge and optimization scheme are presented. Experimental setting and results, parameters analysis are provided in Section IV. In the end, the summarization is reached in Section V.

II. RELATED WORKS

Our work involves two kinds of related techniques: 1) low-rank decomposition and 2) linear regression to serve as the basis of our proposed framework. Next, the two related works will be briefly introduced as follows.

A. LOW-RANK REPRESENTATION BASED SUBSPACE LEARNING

LRR seeks a simple and efficient robust subspace to fit noisy and damaged data [43]. Assuming $X = [X_1, X_2, \dots X_k]$ is a matrix including samples from k subspaces, which can be composed of a linear combination of atoms in dictionary and sparse noise matrix. Specially, we use X itself as dictionary, the LRR can be modeled by

$$\min_{Z,E} \|Z\|_* + \lambda \|E\|_1$$
$$s.t. X = XZ + E$$

where Z is the low-rank representation matrix. $\|\cdot\|_*$ denotes the nuclear norm of matrix, which can be seen as a good surrogate for lank minimization and used to constrain the rank of Z. The noisy data E is sparse generally, which can be handled by l_1-norm constraint. λ is the balanced parameter between linear representation and noise. LRR use a rank minimized matrix to build a linear subspace to uncover the latent structures hidden in data, which is feasible for most simple data. However, it may failed for cross-view data in realistic world due to their complicated and changeable structures.

For cross-view data, the samples will present a particular phenomenon that the samples between different view of same class are far away, while ones from different categories but the same view are closer. To address this problem, we hope to establish two subspaces for class and view structures respectively, which can characterize the two potential structures among the cross-view data. In addition, we have designed an unlocking-structure mechanism to make the data within-class more closer by ignoring the view influence, and achieve better results.

B. LINEAR REGRESSION

Assuming that there is a data matrix $X = [X_1, X_2, \dots X_m]$ with m samples, whose column is an n-dimensional feature vector. Also, $Y = [Y_1, Y_2, \dots Y_m]$ denotes its corresponding expected output matrix. Linear regression aims to find a parameter matrix W to map the data to expected output space, and the relationship between data and their outputs can be well established by W. Specially, for the classification task, the label can be used as the expected output to construct the regression loss function as follows.

$$\min_{W} \|Y - WX\|_2^2$$

Although the linear regression is uncomplicated, its extended versions are quite rich. Many extended nonlinear models can also be obtained by introducing hierarchical structures or high-dimensional mappings based on linear models. In many supervised learning, linear regression model is also adopted as empirical risk function and prediction knowledge by semantic constraint frameworks.

Nevertheless, for the cross-view data, the conventional regression model cannot explore the complicated latent structures within semantic space due to their distribution discrepancy. For this reason, we propose an joint learning scenario that enable the view-shared structure and view-special structure within semantic space to be revealed, which can be used to give more reasonable guidance to feature learning for cross-view data.

III. CROSS-VIEW FEATURE LEARNING

In this section, we give a detailed discussion on our neoteric cross-view feature learning framework, which is based on robust low-rank constraint. Then, an feasible solution is developed through iterative scheme.

A. NOTATIONS

Assuming that there is cross-view data $X = [X_1, X_2]$ consisting of two views. The ith view $X_i \in R^{d \times m_i}$ is from L different classes, where d is the dimensionality of training data and m_i is the number of the samples belonging to the ith view ($m = \sum_i m_i$). We construct two graphs to learn two cross-view subspaces that are class structure matrix $Z_c \in R^{m \times m}$ and view structure matrix $Z_v \in R^{m \times m}$. Furthermore, we use the error matrix $E \in R^{p \times m}$ to fit the representation residual. $P \in R^{d \times p}$ is the learned projection feature subspace, where p is dimensionality of common subspace. In addition, we design three regression parameter matrices $W_0, W_1, W_2 \in R^{p \times L}$, where L is the number of classes for different categories of sample data, to explore the shared and specific information from semantic space.

B. OBJECTIVE FUNCTION

We design a cross-view feature learning framework though unlocking class and view subspace structures with low-rank constraint, of which the model can be formulated as follows:

$$\begin{aligned} \min_{\substack{P,Z_c,Z_v,E,\\ W_0,W_1,W_2}} \; & D(Z_c, Z_v, E) + \alpha\Gamma(P, Z_c, Z_v) \\ & + J(W_0, \{W_m\}, P, \{X_m\}, \{Y_m\}) \\ & + R(W_0, \{W_m\}) \\ s.t.\; & P^T X = P^T X(Z_c + Z_v) + E, \\ & P^T P = I, m = 1, 2 \end{aligned} \tag{1}$$

where the first term $D(Z_c, Z_v, E)$ is the dual structure representations for the cross-view data based on low-rank constraints. The second term $\Gamma(P, Z_c, Z_v)$ is the discrimination unlocking term, ensuring that samples from same class close to each other, while the ones from the different classes but within same view far away. The third term $J(W_0, \{W_m\}, P, \{X_m\}, \{Y_m\})$ explores the joint information within the semantic space. The last term $R(W_0, \{W_m\})$ eliminates the trivial solution.

Dual Low-Rank Constraint: Generally, we use a low-rank representation matrix to approximate the internal structure of samples. However, the samples from different views but within same class have large divergence due to these two interlaced structures. On the contrary, the samples from same view but different classes are more similar due to the latent connection from view structure. The relationship of these two significant structures are shown in Figure 1. Hence, we adopt

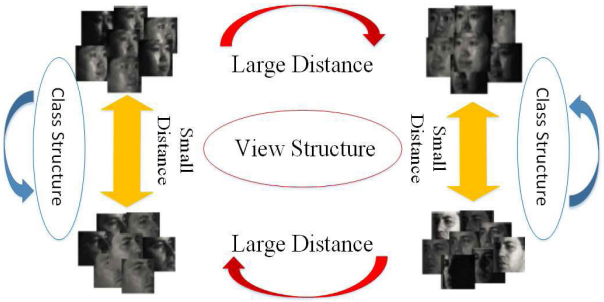


FIGURE 1. Cross-view data with two significant structures.

class structure matrix Z_c and view structure matrix Z_v as two individual subspaces to capture the intrinsic cross-view underlying subspace structures. Thus, we define the first term as follows:

$$D(Z_c, Z_v, E) = \|Z_c\|_* + \|Z_v\|_* + \lambda \|E\|_1 \tag{2}$$

where we use the nuclear norm and l_1-norm to constraint the low-rank representations of two independent structures and the sparsity of the error matrix E, respectively. λ is a positive parameter to balance them.

Discrimination Unlocking Structure: We have established a framework with dual structure for cross-view data above. Nevertheless, it is difficult to estimate underlying information for cross-view structures. So, we define two supervised unlocking regularizers to cluster within-class samples and decentralize between-class samples as follows:

$$\begin{aligned} \Gamma_c &= \sum\nolimits_{i,j} (Y_{c,i} - Y_{c,j})^2 V^c_{i,j} \\ \Gamma_v &= \sum\nolimits_{i,j} (Y_{v,i} - Y_{v,j})^2 V^v_{i,j} \end{aligned} \tag{3}$$

where $Y_{c,i}$ and $Y_{c,j}$ are the ith and jth columns of the projection matrix for class structure as $Y_c = P^T X Z_c \in R^{p\times m}$. Correspondingly, $Y_{v,i}$ and $Y_{v,j}$ are the ith and jth columns of the projection matrix from view structure as $Y_v = P^T X Z_v \in R^{p\times m}$. $V^c_{i,j}$ and $V^v_{i,j}$ represent two weight matrices of these two supervised regular terms, which are defined as follows:

$$\begin{aligned} V^c_{i,j} &= \begin{cases} 1, & \text{if } x_i \in N^c_{k_1}(x_j), \text{ and } l_i = l_j, \\ 0, & \text{otherwise} \end{cases} \\ V^v_{i,j} &= \begin{cases} 1, & \text{if } x_i \in N^v_{k_2}(x_j), \text{ but } l_i \neq l_j, \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{4}$$

where l_i and l_j are the labels of sample x_i, x_j, respectively. $x_i \in N^c_{k_1}(x_j)$ denotes that x_i is k_1 nearest neighbor of data x_j within the same class. $x_i \in N^v_{k_2}(x_j)$ means that x_i is the k_2 nearest neighbor of the same view data x_j. We use two regularizers by the linear discrimination analysis to implement unlocking between the dual structures, and construct the the following term based on Fisher criterion.

$$\begin{aligned} \Gamma(P, Z_c, Z_v) = &\, tr(P^T X Z_c L_c (P^T X Z_c)^T) \\ &- tr(P^T X Z_v L_v (P^T X Z_v)^T) \end{aligned} \tag{5}$$

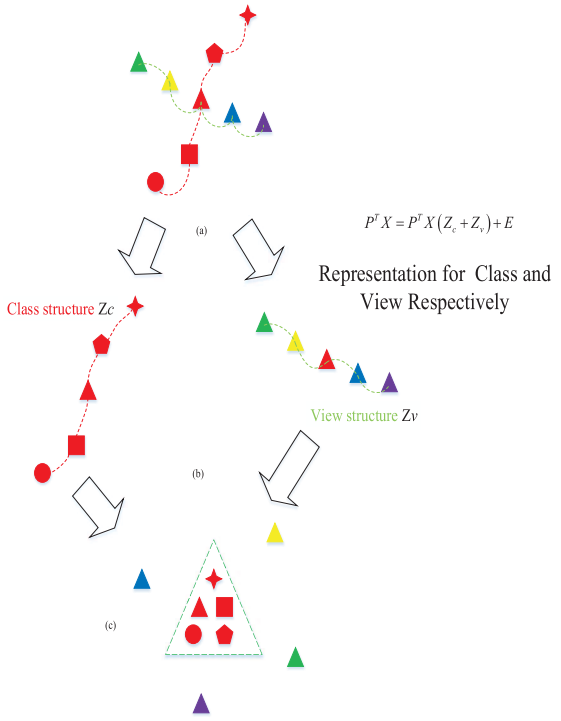


FIGURE 2. Representation for view and class respectively.

where L_c and L_v are the Laplacian matrices of V^c and V^v. In addition, the orthogonal constraint $P^T P = I$ is imposed to eliminate the trivial solution and reduce the redundancy. In this way, two interleaved structures hidden in cross-view data can be unlocked as shown in Figure 2. Hence, by means of discrimination unlocking term, we can get a common subspace in which the within-class data, regardless of their views, are closer while between-class data in the same view are far away from each other.

Joint Learning Constraint: It is noted that the regression within semantic space, which is helpful to earn more discriminant, can be used to guide the feature learning. Nevertheless, different from single view data, the between-view shared and view-specific information are both existed within the semantic space for cross-view data. So, a special regression scheme that explores the shared and view-specific structures simultaneously is expected to give better feedback guidance to cross-view feature learning. To address this problem, we formulate joint learning term in cross-view data in order to reconstitute the feature subspace to capture the view-shared and view-specific structures, with the principles illustrated in Figure 3. Hence, we define the $J(W_0, \{W_m\}, P, \{X_m\}, \{Y_m\})$ term as:

$$\begin{aligned} &J(W_0, \{W_m\}, P, \{X_m\}, \{Y_m\}) \\ &\quad = \left\| (W_0 + W_1)^T P^T X_1 - Y_1 \right\|_F^2 \\ &\qquad + \left\| (W_0 + W_2)^T P^T X_2 - Y_2 \right\|_F^2 \end{aligned} \tag{6}$$

where $Y_i = \left[Y^1, Y^2, \ldots, Y^{m_i}\right]$ is a label matrix for training samples in ith view. If jth sample belong to the kth category,

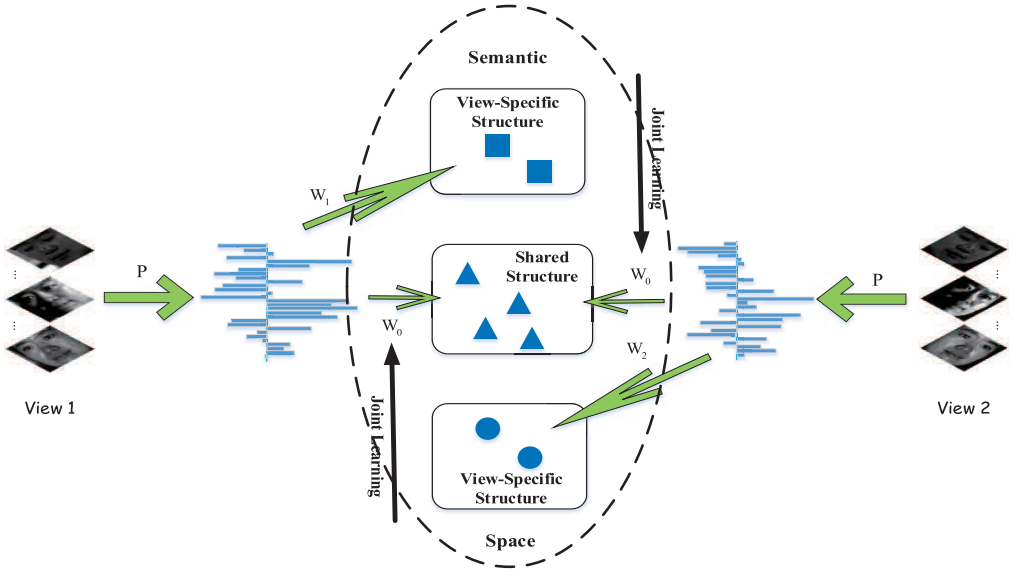


FIGURE 3. Joint shared and view-special structures learning: W_0 and $W_i (i = 1, 2)$ are jointly optimized to explore the shared and view-specific structures within semantic space, which can lead to a more effective and discriminative cross-view feature learning.

the kth element of Y^j is equal to $L-1$ while others to be -1 as $Y^j = [-1, -1, \ldots, L-1, \ldots, -1]^T \in R^L$. We use a weight matrix W_0 to quantify the view-shared structure that is owned jointly by all views. Furthermore, the weight matrices W_1, W_2 are used to elicit the view-specific components. In this way, the view-shared and view-specific structures can adapt to various kinds of cross-view data more effectively.

Regularization Constraints: The last term is the regularization term as follows:

$$R(W_0, \{W_m\}) = \varepsilon \|W_1\|_F^2 + \varepsilon \|W_2\|_F^2 + \gamma \|W_0\|_F^2 \quad (7)$$

where ε and γ are the positive parameters to balance the regularization terms. In addition, the Frobenius norm used here aims to prevent the overfitting and enhance the generalization of our model.

In the end, we substitute all the terms into the comprehensive objective function, which can be written as:

$$\begin{aligned} \min_{\substack{P,Z_c,Z_v,E,\\ W_0,W_1,W_2}} & \left\|(W_0+W_1)^T P^T X_1 - Y_1\right\|_F^2 + \varepsilon \|W_1\|_F^2 \\ & + \left\|(W_0+W_2)^T P^T X_2 - Y_2\right\|_F^2 + \varepsilon \|W_2\|_F^2 \\ & + \gamma \|W_0\|_F^2 + \|Z_c\|_* + \|Z_v\|_* + \lambda \|E\|_1 \\ & + \alpha \Gamma(P, Z_c, Z_v) \\ s.t.\ & P^T X = P^T X(Z_c + Z_v) + E, P^T P = I \end{aligned} \quad (8)$$

C. OPTIMIZATION

To solve Eq.(8), we introduce an auxiliary matrix M and rewrite it as the following form:

$$\begin{aligned} \min_{\substack{P,Z_c,Z_v,E,\\ M,W_0,W_1,W_2}} & \left\|(W_0+W_1)^T M^T X_1 - Y_1\right\|_F^2 + \varepsilon \|W_1\|_F^2 \\ & + \left\|(W_0+W_2)^T M^T X_2 - Y_2\right\|_F^2 + \varepsilon \|W_2\|_F^2 \\ & + \gamma \|W_0\|_F^2 + \|Z_c\|_* + \|Z_v\|_* + \lambda \|E\|_1 \\ & + \alpha \Gamma(P, Z_c, Z_v) \\ s.t.\ & P^T X = P^T X(Z_c + Z_v) + E, \\ & P^T P = I, M^T M = I, P = M \end{aligned} \quad (9)$$

The above minimization is an nonconvex problem with the multi-variable, which is difficult to be solved directly. So, we develop an alternating numerical scheme to solve each variable iteratively. Firstly, we transform Eq.(9) to the following augmented Lagrangian function:

$$\begin{aligned} \min_{\substack{P,Z_c,Z_v,E,Q\\ M,W_0,W_1,W_2}} & \left\|(W_0+W_1)^T M^T X_1 - Y_1\right\|_F^2 + \varepsilon \|W_1\|_F^2 \\ & + \left\|(W_0+W_2)^T M^T X_2 - Y_2\right\|_F^2 + \varepsilon \|W_2\|_F^2 \\ & + \gamma \|W_0\|_F^2 + \|Z_c\|_* + \|Z_v\|_* + \lambda \|E\|_1 \\ & + < Q, P^T X - P^T X(Z_c + Z_v) - E > \\ & + \frac{\mu}{2} \left\|P^T X - P^T X(Z_c + Z_v) - E\right\|_F^2 \\ & + \alpha \Gamma(P, Z_c, Z_v) + \beta \|P - M\|_F^2 \end{aligned} \quad (10)$$

where Q is the Lagrange multiplier and μ, β are the penalty parameters. $<, >$ denotes the inner product operator of two matrices. Then, by merging some terms, the function in (10) is rewritten to a new form with a quadratic term as

follows:

$$\begin{aligned}\min_{\substack{P,Z_c,Z_v,E,Q\\M,W_0,W_1,W_2}} & \left\|(W_0+W_1)^T M^T X_1 - Y_1\right\|_F^2 + \varepsilon \|W_1\|_F^2 \\ & + \left\|(W_0+W_2)^T M^T X_2 - Y_2\right\|_F^2 + \varepsilon \|W_2\|_F^2 \\ & + \gamma \|W_0\|_F^2 + \|Z_c\|_* + \|Z_v\|_* + \lambda \|E\|_1 \\ & + \eta(P, Z_c, Z_v, E, Q, \mu) - \frac{1}{\mu}\|Q\|_F^2 \\ & + \beta \|P - M\|_F^2 \end{aligned} \tag{11}$$

where $\eta(P, Z_c, Z_v, E, Q, \mu) = \alpha\Gamma(P, Z_c, Z_v) + \frac{\mu}{2}\left\|P^T X - P^T X(Z_c + Z_v) - E + \frac{Q}{\mu}\right\|_F^2$. Obviously, the variables $P, Z_c, Z_v, M, W_0, W_1, W_2$ and E cannot be addressed simultaneously, but they are solvable individually when fixing other ones. To solve each sub-problem, η is approximated by the first order Taylor expansion. Taking t as the iterative step, then each sub-problem for each variable can be formulated as follows:

Updating Z_c:

$$\min_{Z_c} \frac{1}{\mu\rho}\|Z_c\|_* + \frac{1}{2}\left\|Z_c - Z_{c,t} + \nabla_{Z_c}\eta\right\|_F^2 \tag{12}$$

where $\nabla_{Z_c}\eta = 2\alpha X^T P_t P_t^T X Z_{c,t} L_c - Q_t^T P_t^T X - \mu X^T P_t (P_t^T X - P_t^T X(Z_{c,t} + Z_{v,t}) - E_t)$ and $\rho = \left\|P_t^T X\right\|_2^2$. It is noted that the Eq.(12) is a nuclear-norm minimization problem, which can be addressed by singular value decomposition effectively [44].

Updating Z_v:

$$\min_{Z_v} \frac{1}{\mu\rho}\|Z_v\|_* + \frac{1}{2}\left\|Z_v - Z_{v,t} + \nabla_{Z_v}\eta\right\|_F^2 \tag{13}$$

where $\nabla_{Z_c}\eta = -2\alpha X^T P_t P_t^T X Z_{v,t} L_v - Q_t^T P_t^T X - \mu X^T P_t (P_t^T X - P_t^T X(Z_{c,t+1} + Z_{v,t}) - E_t)$. Similar to Eq.(12), Eq.(13) can be solved in the same way.

Updating E:

$$\min_E \frac{1}{2}\left\|E - (P_t^T(X - X(Z_{c,t+1} + Z_{v,t+1})) + \frac{Q_t}{\mu})\right\|_F^2 + \frac{\lambda}{\mu}\|E\|_1 \tag{14}$$

The above equation is a standard element-wise l_1-norm minimization problem, which can be solved by thresholding shrinkage algorithm directly [45].

Enforcing the derived function to be zero, P can be updated by

$$P_{t+1} = (2\alpha X Z_n X^T + \mu X_n X_n^T)^{-1}(\beta M + X_n(E - Q/\mu)^T) \tag{15}$$

where we define $Z_n = Z_{c,t+1} L_c Z_{c,t+1}^T - Z_{v,t+1} L_v Z_{v,t+1}^T$ and $X_n = X - X(Z_{c,t+1} + Z_{v,t+1})$ for simplicity.

Algorithm 1

Input: data matrices X_1,X_2,Y_1,Y_2,parameters λ,α,ε,γ,L_c,L_v
Initialize: $E_0 = 0, \theta = 10^{-6}, \beta = 1, \psi = 1.1, \mu = 0.1$, $\mu_{max} = 10^6, t_{max} = 10^2, t = 0$;
while not converged or $t \le t_{max}$ **do**
1. Optimize $Z_{c,t+1}$ in Eq.(12);
2. Optimize $Z_{v,t+1}$ in Eq.(13);
3. Optimize E_{t+1} in Eq.(14);
4. Optimize P_{t+1} in Eq.(15);
5. Optimize M_{t+1} in Eq.(16);
6. Optimize $W_{0,t+1}$ in Eq.(17);
7. Optimize $W_{1,t+1}$ in Eq.(18);
8. Optimize $W_{2,t+1}$ in Eq.(19);
9. $P_{t+1} \leftarrow$ orthogonal(P_{t+1});
10. Optimize the multiplier Q_{t+1} by $Q_{t+1} = Q_t + \mu(P_{t+1}^T(X - X(Z_{c,t+1} + Z_{v,t+1})) - E_{t+1})$
11. Update the parameter μ by $\mu = \min(\mu_{max}, \psi\mu)$;
12. Check convergence by $\left\|P_{t+1}^T(X - X(Z_{c,t+1} + Z_{v,t+1})) - E_{t+1}\right\|_\infty < \theta$;
13. $t = t + 1$.
end while
Output: $Z_c, Z_v, E, P, M, W_0, W_1, W_2$

Updating M:

$$\begin{aligned}\min_M & \left\|(W_0+W_1)^T M^T X_1 - Y_1\right\|_F^2 \\ & + \left\|(W_0+W_2)^T M^T X_2 - Y_2\right\|_F^2 + \beta\|P - M\|_F^2 \\ s.t.\ & M^T M = I \end{aligned} \tag{16}$$

It is difficult to solve the problem(16) with non-convex constraints directly on Eucildean space. We use a gradient-based approach to optimize the problem on the Stiefel manifold [46].

Updating W_0:

$$\begin{aligned} W_{0,t+1} = & (M^T X_1 X_1^T M + M^T X_2 X_2^T M + \varepsilon I)^{-1} \\ & \times (M^T X_1 (Y_1^T - X_1^T M W_1) \\ & + M^T X_2 (Y_2^T - X_2^T M W_2)) \end{aligned} \tag{17}$$

Updating W_1:

$$W_{1,t+1} = (M^T X_1 X_1^T M + \varepsilon I)^{-1} M^T X_1 (Y_1^T - X_1^T M W_0) \tag{18}$$

Updating W_2:

$$W_{2,t+1} = (M^T X_2 X_2^T M + \varepsilon I)^{-1} M^T X_2 (Y_2^T - X_2^T M W_0) \tag{19}$$

In the end, we summarize the detailed optimization for problem (11) in **Algorithm 1**, in which we set those parameters $\mu, \mu_{max}, \theta, \psi, t_{max}$ empirically and tune them through the experiments.

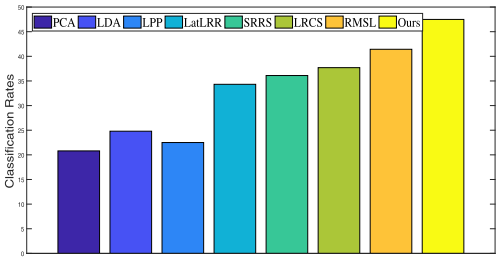


FIGURE 4. Average recognition rates of comparison methods on the Wikipedia dataset.

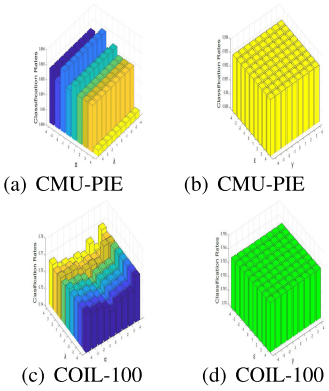


FIGURE 5. Parameters analysis of our proposed method on CMU-PIE faces Case8(a,b) and COIL-100 Case1(c,d), where the value from -4 to 4 denotes $\left[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}, 10^{3}, 10^{4},\right]$.

D. COMPLEXITY ANALYSIS

In this section, we would analyze the computational complexity of our proposed algorithm. In Algorithm 1, the most complicated iteration are steps 1,2 and 4-8. The computational complexity of the SVD decomposition applied in steps 1 and 2 is $O(n^3)$. Since the matrices Z_c and Z_v are two low-rank matrices, the cost is reduced to $O(rn^2)$, in which r is the rank of the low-rank matrix and n is the number of samples. The computational complexity of an inverse operation for an mth order matrix is $O(m^3)$. Therefore, step 4 and 5 will cost $O(d^3)$, where d is the dimensionality of each sample.Steps 6-8 are just the matrix multiplication with $O(p^3)$, where p is dimensionality of learned feature subspace. In fact, r and p are far less than d. Hence, the comprehensive computational complexity of the proposed algorithm is $O(d^3)$.

IV. EXPERIMENTS

In this section, we firstly introduce the four multimedia datasets used in our experiments. Then, we figure the experiment settings. We compared our proposed method to several existing excellent feature learning methods. The classification experiments are tested on the learned features of comparison methods. We select fifty percentage of the dataset as the training set, and the testing set includes the rest of data. All the unknown matrices in Algorithm 1 are initialized randomly. The parameters are empirically set to obtain the best performance and their analysis can be found in Figure 5. Furthermore, each experiment are conducted 5 times and the average performance evaluations are reported to show the advantages of our proposed method.

A. DATASETS AND EXPERIMENTAL SETTING

Extended YaleB face database consists of 16128 images, including 64 images in different illumination conditions and 9 postures. We crop the face images to 32×32. We divide each person's images into four poses P1,P2,P3,P4 in experiments, which are approximately positive, so that the samples can be more relevant to our experiments. We further partition each two poses into one group, where V1[P1,P2] and V2[P3,P4].

CMU-PIE face database contains 68 different persons, and each subject has 9 different poses with 21 different illumination variations. We adopt samples from 5 poses, including P05, P09, P14, P27, P29. We select different classes of poses to constitute different pairs of training and testing set, and crop the face images to 64×64.

COIL-100 object database includes 100 objects, a total of 7200 images. Each object obtained 72 images that are caught with 5 degree rotation. The object images are cropped to 32×32 and divided into two subsets as 'C1' and 'C2'. In addition, C1 contains the images in two point of view V1[0°, 85°] and V2[185°, 265°]. Similarly, C2 obtains the images in V3[90°, 175°] and V4[270°, 355°].

Wikipedia dataset consists of 10 semantic categories with 2866 image-text pairs. Each pair of samples contains an image feature with a dimension of 4096 and a 100-dimensional text feature. Due to the inconsistency of image features and text features in the database, we adopt PCA to reduce the dimensionality of the image feature to match text feature.

B. EXPERIMENTAL RESULTS AND ANALYSIS

In comparison experiments, we primarily select some existing excellent feature learning algorithms, which are PCA [47], LDA [48], LPP [49], LatLRR [50], SRRS [36], LRCS [51], and Robust multi-view subspace learning (RMSL) [52]. Among them, PCA, LPP, LatLRR and LRCS belong to the unsupervised learning; while LDA, SRRS, RMSL, and the proposed method are supervised learning. The k-Nearest Neighbor Classifier is a simple and classical classification algorithm. We use the kNN classifier for performance evaluation of the extracted feature by comparison methods. For CMU-PIE faces with 5 poses, we divide each two poses into same group, where Case1:{P05,P09}, Case2:{P05,P14}, Case3:{P05,P27}, Case4:{P05,P29}, Case5:{P09,P14}, Case6:{P09,P27}, Case7:{P09,P29}, Case8:{P14,P27}, Case9:{P14,P29}, Case10:{P27,P29}. The results of comparison methods on CMU-PIE are shown in Tables 1 and 2. For extended YaleB faces, we select one from V1 and one from V2 as training set from each set of perspectives, and another perspective as a test set. There are four experimental groups to evaluate the performance of all experimental methods, as shown in Table 3. For COIL-100 objects, our experimental setup is similar to

TABLE 1. Average recognition rates (%) of comparison methods on the CMU-PIE face dataset (Case1-5).

Methods	Case1	Case2	Case3	Case4	Case5
PCA	48.81±0.73	50.89±1.00	50.50±1.14	49.07±1.04	48.36±1.33
LDA	62.48±0.78	66.31±1.33	66.76±1.67	62.16±1.44	61.27±1.35
LPP	62.40±1.08	59.25±0.02	60.17±0.03	61.97±0.14	65.72±0.11
LatLRR	65.07±1.00	65.36±1.75	66.61±1.23	62.47±1.8	63.65±3.11
SRRS	85.31±1.02	82.04±2.12	82.33±1.03	85.22±1.08	83.37±1.12
LRCS	95.68±1.01	91.83±0.64	92.30±0.59	95.48±0.43	89.60±1.08
RMSL	97.14±0.02	92.97±0.01	93.70±0.07	97.26±0.17	91.85±0.04
Ours	**98.39±0.03**	**93.82±0.04**	**94.50±0.05**	**98.19±0.11**	**92.47±0.12**

TABLE 2. Average recognition rates (%) of comparison methods on the CMU-PIE face dataset (Case6-10).

Methods	Case6	Case7	Case8	Case9	Case10
PCA	48.43±0.73	45.51±1.62	55.28±1.37	48.64±1.76	49.68±0.86
LDA	61.50±1.49	56.54±1.05	66.96±1.32	61.10±1.56	61.83±1.48
LPP	66.13±0.65	63.34±0.05	59.29±0.14	58.10±0.01	63.72±0.13
LatLRR	63.09±0.97	61.04±1.39	66.1±1.84	60.78±2.01	60.42±1.26
SRRS	86.17±0.44	82.89±0.32	77.45±0.64	81.64±0.78	82.18±0.83
LRCS	89.23±0.62	95.57±1.02	87.42±0.54	90.88±0.92	90.64±0.71
RMSL	92.99±0.11	97.55±0.08	88.47±0.01	92.02±0.12	92.36±0.06
Ours	**93.95±0.08**	**98.35±0.04**	**89.48±0.07**	**92.19±0.01**	**93.89±0.15**

TABLE 3. Average recognition rates (%) of comparison methods on the extended YaleB face dataset.

Methods	Case1	Case2	Case3	Case4
PCA	68.73±1.13	59.17±0.91	52.15±1.07	42.87±2.19
LDA	75.92±1.47	69.74±1.61	61.33±2.01	56.93±1.39
LPP	56.20±0.24	70.29±0.03	71.19±0.15	55.89±0.09
LatLRR	72.70±1.51	67.96±2.1	59.83±2.0	51.85±0.84
SRRS	72.32±1.12	73.12±0.63	82.48±0.49	84.68±0.90
LRCS	71.73±1.08	73.25±0.86	75.91±0.75	72.95±1.11
RMSL	74.04±0.02	75.33±0.05	83.31±0.12	85.36±0.17
Ours	**82.72±0.15**	**83.63±0.03**	**87.15±0.07**	**95.31±0.20**

TABLE 4. Average recognition rates (%) of comparison methods on the COIL-100 object dataset.

Methods	Case1	Case2	Case3	Case4
PCA	40.03±0.86	40.57±0.78	41.13±0.92	39.37±1.20
LDA	39.84±1.05	39.63±1.22	39.48±1.20	40.51±1.58
LPP	45.32±0.36	47.57±0.60	46.98±0.78	46.07±0.82
LatLRR	61.09±1.48	63.24±1.03	65.43±1.80	60.98±1.93
SRRS	65.44±1.04	68.23±0.89	69.25±0.33	62.97±1.50
LRCS	62.89±0.20	66.06±1.06	67.19±1.31	61.75±0.88
RMSL	67.83±0.09	61.72±0.17	60.08±0.12	66.33±0.08
Ours	**76.36±0.06**	**70.89±0.11**	**71.78±0.08**	**76.17±0.12**

extended YaleB faces. The evaluation results of the comparison methods are shown in Table 4. Then, we take the image feature as the first view and text feature as the second view.

From the recognition rates (Tables 1,2,3,4 and Figure 4), we can conclude that the learned feature subspace by our proposed method is more effective than that of other comparison ones for classification task. Moreover, performance based on low-rank algorithm is generally better than others due to exploring latent subspace structures.

C. PROPERTY EVALUATION

In this part, we will test the influences of the parameters and feature dimensionality, selected in the experiments, on the performance.

Our framework has four parameters$\lambda, \alpha, \varepsilon, \gamma$. We evaluate them respectively on CMU-PIE faces Case8 and COIL-100 Case1. Generally, we upgrade one parameter while fixing other parameters. However, considering different values of four parameters is too complicated. To simplify, we only upgrade two parameters ε and γ (λ and α) while fixing the other two parameters λ and α (ε and γ). The reason we set the parameter group in this way is that parameters ε and γ both belong to the joint learning structures that can learn the shared information and view-specific information of cross-view data, while parameters λ and α belong to the dual low-rank discriminative unlocking structures. The experimental results are shown in Figure 5. From Figure 5, we can see that the classification results are hardly sensitive to $\lambda, \varepsilon, \gamma$. For the parameter α, the classification rate presents small fluctuations

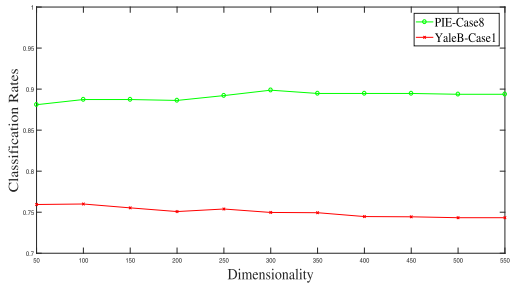


FIGURE 6. Classification rates of different dimensionality on CMU-PIE faces Case8 and Extend YaleB faces Case1.

in a narrow range. We can obtain almost consistent classification results in a wide range. The results point out that the proposed method is nearly stable with different parameters selection.

Afterwards, we evaluate the dimensionality influence of our method in CMU-PIE faces Case8 and Extend YaleB faces Case1, and the Figure 6 shows the evaluation results. The results show that the performance is not sensitive to the dimensions. For the CMU-PIE faces Case8, classification performance increases slightly when the dimensionality goes up. Performance reaches the highest around 300. In the Extend YaleB faces Case1, classification performance decrease slightly with dimensionality increasing.

V. CONCLUSION

To solve the cross-view multimedia data recognition and classification problems, this paper provides a cross-view discriminative subspace learning method based on low-rank constraint for feature learning of multimedia data. The proposed framework learns projection subspace by seeking two manifold fields with dual low-rank representations, which can be used to unlock the class and view structures. To further align the cross-view data, a discriminative joint regression constraint is designed in semantic space. Our method seamlessly put the representation and semantic alignment into a unified framework. Meanwhile, the numerical scheme is developed to guarantee the convergence. The parameters presented in the proposed framework are also analyzed in detail. We test our method on four kinds of multimedia datasets, and find that it outperforms other conventional comparison methods and achieves competitive performance.

REFERENCES

[1] Y. Tian, S. Chen, M. Shyu, T. Huang, P. Sheu, and A. D. Bimbo, "Multimedia big data," *IEEE MultimediaMag.*, vol. 22, no. 3, pp. 93–95, 2015.

[2] G. Andrew, R. Arora, J. Bilmes, and K. Livescu, "Deep canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, 2013, pp. 1247–1255.

[3] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2160–2167.

[4] 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.

[5] J. Liu, C. Wang, H. Su, B. Du, and D. Tao, "Multistage GAN for fabric defect detection," *IEEE Trans. Image Process.*, vol. 29, pp. 3388–3400, 2020.

[6] F. Luo, B. Du, L. Zhang, L. Zhang, and D. Tao, "Feature learning using spatial-spectral hypergraph discriminant analysis for hyperspectral image," *IEEE Trans. Cybern.*, vol. 49, no. 7, pp. 2406–2419, Jul. 2019.

[7] B. Du, Q. Wei, and R. Liu, "An improved quantum-behaved particle swarm optimization for endmember extraction," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 6003–6017, Aug. 2019.

[8] Q. Wu, Y. Li, and Y. Lin, "The application of nonlocal total variation in image denoising for mobile transmission," *Multimedia Tools Appl.*, vol. 76, no. 16, pp. 17179–17191, Jul. 2016.

[9] S. Liu, Z. Pan, and X. Cheng, "A novel fast fractal image compression method based on distance clustering in high dimensional sphere surface," *Fractals*, vol. 25, no. 04, Jul. 2017, Art. no. 1740004.

[10] S. Liu, W. Bai, G. Liu, W. Li, and H. M. Srivastava, "Parallel fractal compression method for big video data," *Complexity*, vol. 2018, pp. 1–16, Oct. 2018.

[11] Y. Lin, H. Tao, Y. Tu, and T. Liu, "A node self-localization algorithm with a mobile anchor node in underwater acoustic sensor networks," *IEEE Access*, vol. 7, pp. 43773–43780, 2019.

[12] Z. Dou, G. Si, Y. Lin, and M. Wang, "An adaptive resource allocation model with anti-jamming in IoT network," *IEEE Access*, vol. 7, pp. 93250–93258, 2019.

[13] W. Qidi, L. Yibing, L. Yun, and Y. Xiaodong, "The nonlocal sparse reconstruction algorithm by similarity measurement with shearlet feature vector," *Math. Problems Eng.*, vol. 2014, pp. 1–8, 2014.

[14] J. Sun, W. Wang, K. Zhang, L. Zhang, Y. Lin, Q. Han, Q. Da, and L. Kou, "A multi-focus image fusion algorithm in 5G communications," *Multimedia Tools Appl.*, vol. 78, no. 20, pp. 28537–28556, 2019.

[15] J.-F. Hu, W.-S. Zheng, J. Lai, and J. Zhang, "Jointly learning heterogeneous features for RGB-D activity recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2186–2200, Nov. 2017.

[16] Y. Yan, F. Nie, W. Li, C. Gao, Y. Yang, and D. Xu, "Image classification by cross-media active learning with privileged information," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2494–2502, Dec. 2016.

[17] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4333–4341.

[18] J. Pang, F. Tao, C. Zhang, W. Zhang, Q. Huang, and B. Yin, "Robust latent Poisson deconvolution from multiple features for Web topic detection," *IEEE Trans. Multimedia*, vol. 18, no. 12, pp. 2482–2493, Dec. 2016.

[19] Z. Tao, H. Liu, S. Li, Z. Ding, and Y. Fu, "From ensemble clustering to multi-view clustering," in *Proc. 26th Int. Joint Conf. Artif. Intell.*, Aug. 2017, pp. 2843–2849.

[20] S. Liu, G. Liu, and H. Zhou, "A robust parallel object tracking method for illumination variations," *Mobile Netw. Appl.*, vol. 24, no. 1, pp. 5–17, Oct. 2018.

[21] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, Dec. 1936.

[22] T. Kim, J. Kittler, and R. Cipolla, "Learning discriminative canonical correlations for object recognition with image sets," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 251–262.

[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] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Proc. Conf. Data Mining Data Warehouses*, 2010, pp. 1–4.

[25] 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.

[26] D. Li, N. Dimitrova, M. Li, and I. K. Sethi, "Multimedia content processing through cross-modal association," in *Proc. 11th ACM Int. Conf. Multimedia*, 2003, pp. 604–611.

[27] F. Wu, X. Jiang, X. Li, S. Tang, W. Lu, Z. Zhang, and Y. Zhuang, "Cross-modal learning to rank via latent joint representation," *IEEE Trans. Image Process.*, vol. 24, no. 5, pp. 1497–1509, May 2015.

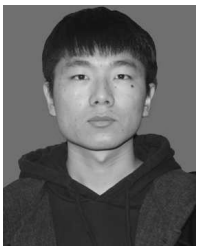
[28] Y. Zhuang, Y. Wang, F. Wu, Y. Zhang, and W. Lu, "Supervised coupled dictionary learning with group structures for multi-modal retrieval," in *Proc. 27th AAAI Conf. Artif. Intell.*, 2013, pp. 1070–1076.

[29] C. Kang, S. Xiang, S. Liao, C. Xu, and C. Pan, "Learning consistent feature representation for cross-modal multimedia retrieval," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 370–381, Mar. 2015.

[30] Y. Hua, S. Wang, S. Liu, A. Cai, and Q. Huang, "Cross-modal correlation learning by adaptive hierarchical semantic aggregation," *IEEE Trans. Multimedia*, vol. 18, no. 6, pp. 1201–1216, Jun. 2016.

[31] L. Zhang, B. Ma, G. Li, Q. Huang, and Q. Tian, "Generalized semi-supervised and structured subspace learning for cross-modal retrieval," *IEEE Trans. Multimedia*, vol. 20, no. 1, pp. 128–141, Jan. 2018.
[32] E. Yu, J. Sun, J. Li, X. Chang, X.-H. Han, and A. G. Hauptmann, "Adaptive semi-supervised feature selection for cross-modal retrieval," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1276–1288, May 2019.
[33] B. Du, T. Xinyao, Z. Wang, L. Zhang, and D. Tao, "Robust graph-based semisupervised learning for noisy labeled data via maximum correntropy criterion," *IEEE Trans. Cybern.*, vol. 49, no. 4, pp. 1440–1453, Apr. 2019.
[34] 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.
[35] S. Li and Y. Fu, "Robust subspace discovery through supervised low-rank constraints," in *Proc. SIAM Int. Conf. Data Mining*, Apr. 2014, pp. 163–171.
[36] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2160–2173, Nov. 2016.
[37] 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.
[38] A. Li, D. Chen, Z. Wu, G. Sun, and K. Lin, "Self-supervised sparse coding scheme for image classification based on low rank representation," *PLoS ONE*, vol. 13, no. 6, Jun. 2018, Art. no. e0199141.
[39] T. Zhou, F. Liu, H. Bhaskar, and J. Yang, "Robust visual tracking via online discriminative and low-rank dictionary learning," *IEEE Trans. Cybern.*, vol. 48, no. 9, pp. 2643–2655, Sep. 2018.
[40] Y. Lu, C. Yuan, X. Li, Z. Lai, D. Zhang, and L. Shen, "Structurally incoherent low-rank 2DLPP for image classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 29, no. 6, pp. 1701–1714, Jun. 2019.
[41] Q. Yao, J. T. Kwok, T. Wang, and T.-Y. Liu, "Large-scale low-rank matrix learning with nonconvex regularizers," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2628–2643, Nov. 2019.
[42] A. Li, X. Liu, Y. Wang, D. Chen, K. Lin, G. Sun, and H. Jiang, "Subspace structural constraint-based discriminative feature learning via nonnegative low rank representation," *PLoS ONE*, vol. 14, no. 5, May 2019, Art. no. e0215450.
[43] G. Liu, Z. Lin, and Y. Yong, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learn.*, 2010, pp. 663–670.
[44] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, Jan. 2010.
[45] Z. Lin, M. Chen, and Y. Ma, "The augmented Lagrange multiplier method for exact recovery of corrupted low-rank matrices," 2010, *arXiv:1009.5055*. [Online]. Available: http://arxiv.org/abs/1009.5055
[46] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Program.*, vol. 142, nos. 1–2, pp. 397–434, Aug. 2012.
[47] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
[48] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[49] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst.*, 2004, pp. 153–160.
[50] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. Int. Conf. Comput. Vis.*, Nov. 2011, pp. 1615–1622.
[51] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 110–119.
[52] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Proc. 13th AAAI Conf. Artif. Intell.*, 2016, pp. 1181–1187.

AO LI (Member, IEEE) received the B.S. and Ph.D. degrees from Harbin Engineering University, Harbin, China, in 2009 and 2014, respectively. From 2017 to 2018, he was a Research Assistant with Wright State University. He currently works with the Harbin University of Science and Technology, Harbin. His current research interests include sparse representation, pattern recognition, machine learning, and their applications to computer vision.



YU DING received the B.S. degree from Northeast Agricultural University, Harbin, China, in 2018. He is currently pursuing the master's degree with the School of Computer Science and Technology, Harbin University of Science and Technology, Harbin. His research interests are in computer vision, pattern recognition, and machine learning.

DEYUN CHEN received the Ph.D. degree from the Harbin University of Science and Technology, Harbin, China, in 2006. He is currently the Dean of the School of Computer Science and Technology and a Full Professor with the Harbin University of Science and Technology. He has published over 100 scientific articles. His current research interests include electrical capacitance tomography, image processing, and pattern recognition.

GUANGLU SUN (Member, IEEE) received the B.S., M.S., and Ph.D. degrees from the Harbin Institute of Technology. From 2014 to 2015, he was a Visiting Scholar with Northwestern University, USA. He is currently a Professor and the Director of the Center of Information Security and Intelligent Technology, Harbin University of Science and Technology, Harbin, China. His current research interests include computer networks and security, machine learning, and intelligent information processing.

HAILONG JIANG received the bachelor's degree in electronic science and technology from Xidian University, in 2014, and the master's degree in integrated circuits from the University of Chinese Academy of Sciences, in 2017. He is currently pursuing the Ph.D. degree with the Department of Computer Science, Kent State University, Kent, OH, USA. His research interests are in high performance computing and cloud computing, involving visualization, program analysis, and machine learning.

QIDI WU is currently pursuing the Ph.D. degree with Harbin Engineering University. Her current research interests include sparse representation and image processing.