Dual Shared-Specific Multiview Subspace Clustering

Tao Zhou, *Member, IEEE*, Changqing Zhang, *Member, IEEE*, Xi Peng, Harish Bhaskar, and Jie Yang

***Abstract*—Multiview subspace clustering has received significant attention as the availability of diverse of multidomain and multiview real-world data has rapidly increased in the recent years. Boosting the performance of multiview clustering algorithms is challenged by two major factors. First, since original features from multiview data are highly redundant, reconstruction based on these attributes inevitably results in inferior performance. Second, since each view of such multiview data may contain unique knowledge as against the others, it remains a challenge to exploit complimentary information across multiple views while simultaneously investigating the uniqueness of each view. In this paper, we present a novel dual shared-specific multiview subspace clustering (DSS-MSC) approach that simultaneously learns the correlations between shared information across multiple views and also utilizes view-specific information to depict specific property for each independent view. Further, we formulate a dual learning framework to capture shared-specific information into the dimensional reduction and self-representation processes, which strengthens the ability of our approach to exploit shared information while preserving view-specific property effectively. The experimental results on several benchmark datasets have demonstrated the effectiveness of the proposed approach against other state-of-the-art techniques.**

***Index Terms*—Complementary information, dual learning, multiview subspace clustering, view-specific property.**

I. Introduction

SUBSPACE clustering is a fundamental technique in real-world applications [1]–[7], in which data points that are drawn from multiple subspace are assumed to correspond to different clusters. In the past decade, a large number of clustering approaches [8]–[11] have been proposed to improve performance incrementally. Among several of these techniques, spectral clustering-based method [12]–[16] has received notable attention. The spectral clustering algorithm, first proposed in [16], aims to learn and cluster a similarity matrix based on the data locality. Several adaptations of subspace clustering approaches based on spectral clustering framework have also been proposed. For example, sparse subspace clustering (SSC) [15] aims to find the sparsest representation for each sample with respect to the entire dataset based on the ℓ_1-norm. Low-rank representation (LRR) [17] clustering attempts to reveal cluster structure by using an LRR. Also, structured SSC (S3C) [18] integrates sparse representation and spectral clustering into a unified framework, which has demonstrated promising results. However, several of the above methods mainly focus on improving the clustering performance from a single view. In practice, the overall performance is limited to single view, and it remains difficult to extend these methods for multiview clustering problems.

Modern datasets are often collected from diverse domains and views in many real-world applications thereby leading to the significantly increased interest in multiview (or multimodal) learning and analysis techniques [19]–[23]. For example, single images and continuous stream videos can both be described by using different visual descriptors, such as SIFT [24], Gabor [25], LBP [26], HOG [27], etc. In general, each type of feature is regarded as an independent view of the original data, and each view of the data may contain some unique information that other views do not. Therefore, the integration of multiple views can provide complementary information to improve clustering accuracy. Multiview clustering approaches have been extensively developed in recent years, and all such approaches can be categorized into three main types. First, graph-based approaches [28]–[32] that explore the correlations among different views using multiple graph fusion strategies. The second type is co-training and co-regularized-based [33], [34] multiview clustering approaches. Finally, the third category of approaches includes subspace learning-based multiview clustering methods [9], [35]–[39], which assumes that there exists an underlying low-dimensional subspace rather than distributed uniformly across the entire feature space.

Despite the progress made within multiview clustering, two major challenges continue to remain. *First*, several existing strategies use the original data directly for reconstruction

Manuscript received December 17, 2017; revised June 20, 2018; accepted May 13, 2019. This work was supported in part by the National Natural Science Foundation of China under Grant 61572315, Grant 61876107, Grant U1803261, Grant 61602337, Grant 61806135, Grant 61625204, and Grant 61836006, in part by 973 Plan under Grant 2015CB856004, and in part by the Fundamental Research Funds for the Central Universities under Grant YJ201949 and Grant 2018SCUH0070. This paper was recommended by Associate Editor S. Ventura. *(Corresponding authors: Changqing Zhang; Jie Yang.)*

T. Zhou is with the Inception Institute of Artificial Intelligence, Abu Dhabi 51133, UAE (e-mail: taozhou.ai@gmail.com).

C. Zhang is with the College of Intelligence and Computing, Tianjin University, Tianjin 300072, China (e-mail: zhangchangqing@tju.edu.cn).

X. Peng is with the College of Computer Science, Sichuan University, Chengdu 610065, China (e-mail: pangsaai@gamil.com).

H. Bhaskar is with Zero One Infinity Consulting Service Ltd., Mississauga, ON L5N6G9, Canada (e-mail: harish@zoiconsulting.ca).

J. Yang is with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: jieyang@sjtu.edu.cn).

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the author.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCYB.2019.2918495

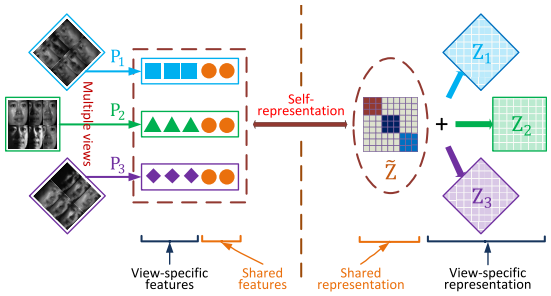


Fig. 1. Basic framework of the proposed DSS-MSC approach.

purposes. However, original (high-dimensional) features contain a high degree of redundancies that could degrade clustering performance. *Second*, as suggested earlier, each independent view offers a unique perspective on the clustering process. Existing techniques are developed either to exploit the correlation across views to boost clustering effectiveness or encapsulate the uniqueness from individual views to strengthen the diversity of the method. However, most approaches cannot simultaneously exploit shared information across multiple views while preserving the uniqueness of particular views during clustering.

To address the limitations above, we propose a dual shared-specific multiview subspace clustering (DSS-MSC) approach. The proposed method simultaneously learns shared information across views to exploit the underlying correlations between them and at the same time, utilizes view-specific information to encapsulate the unique attributes of each independent view. By projecting the original features onto a low-dimensional feature space, we can utilize low-dimensional feature representation to reconstruct data points, thus alleviating the influence of redundancy. Furthermore, we formulate a dual learning framework to capture shared-specific information into the feature projection and self-representation stage, that can strengthen the ability of our method to exploit the shared information and preserve view-specific property effectively. Finally, we build a new similarity matrix by combining shared and view-specific self-representations. The process flow of the proposed framework is depicted in Fig. 1. The main contributions of this paper are listed as follows.

1) We propose a novel multiview subspace clustering approach, that is, *DSS-MSC*, that simultaneously exploits the underlying correlations from multiple views while capturing view-specific information from each independent views.
2) We formulate a *dual learning framework* to capture shared-specific information into feature projection and self-representation stage thus reinforcing the ability of the proposed approach to handle shared-specific information from multiple and independent views.
3) Our strategy utilizes low-dimensional features for data reconstruction, instead of the original high-dimensional data, that enables handling the influence of redundancy.
4) Extensive experiments have been conducted on benchmark datasets, which demonstrate the significant advantages of our proposed approach over other state-of-the-art single view and multiview clustering approaches.

The remainder of this paper is organized as follows. In Section II, we review related works and discuss the differences of our method against others in the literature. Section III presents a detailed description of the proposed multiview subspace clustering approach and provides an optimization solution along with detailed computational complexity analysis and convergence analysis. Further, Section IV illustrates the experimental results on benchmark datasets. Finally, Section V concludes this paper.

II. Related Work

Self-representation-based subspace clustering has demonstrated promising results. In this paper, we adopt a widely accepted assumption that each data point can be expressed as a linear combination of the original data points themselves. Specifically, suppose $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N] \in \mathbb{R}^{L\times N}$ denotes the data matrix, where each column denotes a sample vector; and L and N denote the dimensionality of the features and the number of samples, respectively. A self-representation model can be formulated as

$$\min_{\mathbf{Z}} \mathcal{L}(\mathbf{X}, \mathbf{XZ}) + \lambda\Psi(\mathbf{Z}) \tag{1}$$

where λ is the regularization parameter. $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_N] \in \mathbb{R}^{N\times N}$ is the self-representation matrix and each $\mathbf{z}_i$ is the coding coefficient of the original data point $\mathbf{x}_i$ over the observed data $\mathbf{X}$, and $\mathcal{L}(\cdot)$ and $\Psi(\cdot)$ denote the loss function and regularization term, respectively. Further, to cluster the data into their potential subspaces, a similarity matrix can be computed by using $\mathbf{S} = (|\mathbf{Z}| + |\mathbf{Z}^\top|)/2$. Next, the similarity matrix is used as the input of a spectral clustering algorithm [16] to obtain the final clustering result. A number of subspace clustering approaches have been proposed to explore the relationships among samples using self-representation (e.g., SSC [15], LRR [17], and S3C [18]). However, these approaches only consider the single view features and utilize the original view to reconstruct the data points.

To integrate complementary information from multiple views together, multiview clustering methods [9], [34]–[37] have been developed. The work [9] assumes that multiple views originate from one underlying latent representation, and then performs clustering on the latent representation. Kumar *et al.* [34] proposed a co-regularized multiview spectral clustering method to perform clustering on different views simultaneously using co-regularization constraints. Gao *et al.* [35] unified the representation learning and spectral clustering models into one framework. Cao *et al.* [36] proposed a diversity-induced multiview subspace clustering approach, which utilizes the Hilbert–Schmidt independence criterion [40] to enforce the learned subspace representations to be novel against each other. Xia *et al.* [37] presented a Markov chain method for robust multiview clustering by recovering a shared transition probability matrix via low-rank and sparse decomposition.

Our research in this paper is closely related to the following studies from [9] and [41]–[43]. The work [9] demonstrates the state-of-the-art performance with the underlying assumption that multiple views originated from one latent representation,

and that the latent representation can effectively exploit correlations among various views. However, this method only considers the shared latent representation, while it ignores specific attributes of individual views. As in this process, all views are cascaded together to share the same self-representation. Canonical correlation analysis (CCA)-based approaches [41], [42] project multiview data into a standard lower-dimensional subspace and then apply any existing clustering algorithm such as the k-means clustering to partition the data. The work [43] proposes a robust subspace clustering approach for multiview data by exploiting correlation consensus. The work [44] performs a supervised classification task by learning the shared components among multiviews. Further, it directly cascades all of the features into a single vector. These studies in [45]–[47] project the original data into a low-dimensional feature space, which is adequate for learning a low-dimensional representation and reducing redundancy. In this case, the observed data matrix is approximated by the production of projection matrix and low-dimensional representation, among which self-representation can approximate the low-dimensional representation. In this paper, we extend the self-representation-based subspace clustering for multiview data and, thus, our method can simultaneously exploit the underlying correlations cross multiple views and capture view-specific properties.

III. Proposed Method

In this section, we introduce the formulation of the proposed multiview subspace clustering approach and then present the optimization steps, computational complexity analysis, and convergence analysis.

A. Formulation

Given a data set $\mathbf{X}_v \in \mathbb{R}^{L_v \times N}$, where $\mathbf{X}_v$ denotes the features matrix of the vth view ($v = 1, 2, \ldots, V$), with L_v and N being the dimension of features from the vth view and the number of samples, respectively. Using linear projection, $\mathbf{X}_v$ can be transformed into low-dimensional representation as

$$\mathbf{H}_v = \mathbf{P}_v \mathbf{X}_v + \mathbf{E}_v^1 \tag{2}$$

where $\mathbf{P}_v \in \mathbb{R}^{D_v \times L_v}$ is the projection matrix and $\mathbf{H}_v \in \mathbb{R}^{D_v \times N}$ is the low-dimensional representation for the vth view. To perform multiview subspace clustering and exploit correlation among various features across multiple views, we extend the above equation into

$$[\mathbf{H}; \mathbf{H}_v] = \mathbf{P}_v \mathbf{X}_v + \mathbf{E}_v^1 \tag{3}$$

where $\mathbf{H} \in \mathbb{R}^{D \times N}$ denotes the shared components across multiple views. As shown in Fig. 1, $\mathbf{H}_v$ denotes the specific features for each view while $\mathbf{H}$ represents their corresponding shared features. We assume that different views share partial features in a low-dimensional feature space, which equalizes the correlation among multiple views. Thus, the objective function is to infer the multiview low-dimensional representation model can be written as

$$\min_{\mathbf{P}_v, \mathbf{H}_v, \mathbf{E}_v^1, \mathbf{H}} \sum_{v}^{V} \left\| [\mathbf{H}; \mathbf{H}_v] - \mathbf{P}_v \mathbf{X}_v - \mathbf{E}_v^1 \right\|_F^2. \tag{4}$$

TABLE I
Main Notations Used in the Proposed Formulation

Notation	Description
$\mathbf{X}_v$	Feature matrix from the v-th view
$\mathbf{P}_v$	Projection matrix for the v-th view
$\mathbf{H}_v$	Low-dimension feature matrix for the v-th view
$\mathbf{H}$	Shared low-dimension features among multi-view data
$\mathbf{Z}_v$	View-specific self-representation matrix for the v-th view
$\widetilde{\mathbf{Z}}$	Shared self-representation matrix among multi-view data
$\mathbf{E}_v$	Error term
L_v	Dimension of original features from the v-th view
D_v	Dimension of view-specific features in low-dimensional space
D	Dimension of shared features in low-dimensional space
V	Number of multiple views
v	View index
λ, β	Regularization parameters

Further, for the corresponding low-dimensional representation $[\mathbf{H}; \mathbf{H}_v]$, the objective function for self-representation-based subspace clustering can be reformulated as

$$\min_{\mathbf{H}_v, \mathbf{Z}_v, \mathbf{E}_v^2, \widetilde{\mathbf{Z}}, \mathbf{H}} \sum_{v}^{V} \left\| [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2 \right\|_F^2 \tag{5}$$

where $\widetilde{\mathbf{Z}}$ denotes shared self-representation coefficient matrix for all views, while $\mathbf{Z}_v$ denotes specific self-representation coefficient matrix for each independent view.

By fusing shared and view-specific information into multiview subspace clustering framework, the objective function of our DSS-MSC turns out to be formulated as

$$\begin{aligned} \min_{\mathbf{P}_v, \mathbf{H}_v, \mathbf{Z}_v, \mathbf{E}_v, \widetilde{\mathbf{Z}}, \mathbf{H}} \quad & \|\widetilde{\mathbf{Z}}\|_* + \lambda \sum\nolimits_v \|\mathbf{E}_v\|_{2,1} + \beta \sum\nolimits_v \|\mathbf{Z}_v\|_1 \\ \text{s.t.} \quad & \mathbf{P}_v \mathbf{X}_v = [\mathbf{H}; \mathbf{H}_v] + \mathbf{E}_v^1 \\ & [\mathbf{H}; \mathbf{H}_v] = [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} + [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v + \mathbf{E}_v^2 \\ & \mathbf{P}_v \mathbf{P}_v^\top = \mathbf{I}, \mathbf{E}_v = \left[\mathbf{E}_v^1; \mathbf{E}_v^2\right], \operatorname{diag}(\mathbf{Z}_v) = 0 \\ & \forall v = 1, \ldots, V \end{aligned} \tag{6}$$

where λ and β are regularization parameters. $\|.\|_*$ is the matrix nuclear norm, which enforces the subspace representation to be low-rank. $||.||_{2,1}$ denotes the $\ell_{2,1}$-norm which encourages the columns of a matrix to be zero [17], [67], that is, $\|\mathbf{E}\|_{2,1} = \sum_{j=1}^{N} \sqrt{\sum_{i=1}^{M} [\mathbf{E}_{ij}]^2}$, where $\mathbf{E} \in \mathbb{R}^{M \times N}$. To provide better clarity to our proposed formulation, we list the main notations in Table I. In addition, there are multiple constraints and regularization terms in the proposed formulation, and details of each term are explained below.

1) The first constraint term (e.g., $\mathbf{P}_v \mathbf{X}_v = [\mathbf{H}; \mathbf{H}_v] + \mathbf{E}_v^1$) is utilized to project original features from each view onto a low-dimensional feature space, and we assume that different views share partial components in the low-dimensional feature space (see left subgraph in Fig. 1).
2) The second constraint term (e.g., $[\mathbf{H}; \mathbf{H}_v] = [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} + [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v + \mathbf{E}_v^2$) is used to reconstruct the data points by using low-dimensional features. In the proposed model, we also assume that different views share self-representations across multiple views while each view still has specific self-representations (see right subgraph in Fig. 1). Note

that the first two constraint terms form a *dual learn framework*, which can effectively exploit the correlation among views by learning shared information while preserving view-specific attributes.

3) The term $\mathbf{P}_v\mathbf{P}_v^\top = \mathbf{I}$ is used to prevent the trivial solution. Without this constraint, $\mathbf{H}$ and $\mathbf{H}_v$ may be pushed arbitrarily.
4) For the constraint $\mathbf{E}_v = [\mathbf{E}_v^1; \mathbf{E}_v^2]$, we vertically concatenate together along the column of errors corresponding to low-dimensional features and self-representations. This mechanism of integration will enforce the columns of $\mathbf{E}_v^1$ and $\mathbf{E}_v^2$ to jointly have consistent magnitude values, and its effectiveness has previously been investigated in [9]. Since $\ell_{2,1}$-norm encourages the columns of $\mathbf{E}_v$ to be zero, the underlying assumption here is that the corruptions are sample-specific, that is, some data vectors are corrupted while the others are clean.
5) The regularization term $\|\mathbf{Z}_v\|_1$ ensures that each specific self-representations for the individual samples from the vth view can be sparsely represented using other samples. Besides, the term $\text{diag}(\mathbf{Z}_v) = 0$ eliminates the trivial solution of considering some points as linear combinations of themselves.

B. Optimization

The problem in (6) can be solved via the alternating direction method of multipliers (ADMMs) [48], which alternatively optimizes one variable at one time with the other variables remaining fixed. To adopt ADMM to our problem, we introduce auxiliary variables $\mathbf{J}$ and $\mathbf{Q}_v$ that make the problem separable and, further, we define the following equivalent problem:

$$\begin{aligned}\min_{\mathbf{P}_v,\mathbf{H}_v,\mathbf{Z}_v,\mathbf{Q}_v,\mathbf{E}_v,\widetilde{\mathbf{Z}},\mathbf{J},\mathbf{H}} \quad & \|\mathbf{J}\|_* + \lambda\sum\nolimits_v \|\mathbf{E}_v\|_{2,1} + \beta\sum\nolimits_v \|\mathbf{Q}_v\|_1 \\ \text{s.t.} \quad & \mathbf{P}_v\mathbf{X}_v = [\mathbf{H}; \mathbf{H}_v] + \mathbf{E}_v^1 \\ & [\mathbf{H}; \mathbf{H}_v] = [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} + [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v + \mathbf{E}_v^2 \\ & \mathbf{E}_v = \left[\mathbf{E}_v^1; \mathbf{E}_v^2\right], \quad \mathbf{P}_v\mathbf{P}_v^\top = \mathbf{I}, \widetilde{\mathbf{Z}} = \mathbf{J} \\ & \mathbf{Z}_v = \mathbf{Q}_v - \text{diag}(\mathbf{Q}_v) \\ & \forall v = 1, \ldots, V. \end{aligned} \tag{7}$$

Thus, the augmented Lagrangian function can be given by

$$\begin{aligned}\mathscr{L}(\mathbf{P}_v, \mathbf{H}_v, \mathbf{Z}_v, \mathbf{Q}_v, \mathbf{E}_v, \widetilde{\mathbf{Z}}, \mathbf{J}, \mathbf{H}) &= \|\mathbf{J}\|_* \\ &+ \lambda\sum\nolimits_v \|\mathbf{E}_v\|_{2,1} + \beta\sum\nolimits_v \|\mathbf{Z}_v\|_1 + \Phi(\mathbf{Y}_z, \widetilde{\mathbf{Z}} - \mathbf{J}) \\ &+ \sum\nolimits_v \Phi\left(\mathbf{Y}_v^1, \mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1\right) \\ &+ \sum\nolimits_v \Phi\left(\mathbf{Y}_v^2, [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\right) \\ &+ \sum\nolimits_v \Phi\left(\mathbf{Y}_v^3, \mathbf{Z}_v - \mathbf{Q}_v + \text{diag}(\mathbf{Q}_v)\right) \\ \text{s.t. } \mathbf{P}_v\mathbf{P}_v^\top = \mathbf{I}, \quad & \mathbf{E}_v = \left[\mathbf{E}_v^1; \mathbf{E}_v^2\right], \forall v = 1, \ldots, V \end{aligned} \tag{8}$$

where $\Phi(\mathbf{Y}, \mathbf{C}) = \langle \mathbf{Y}, \mathbf{C}\rangle + (\mu/2)\|\mathbf{C}\|_F^2$ with $\langle\cdot,\cdot\rangle$ denoting the matrix inner product, and μ is a positive penalty scalar. To find the minimal for $\mathscr{L}$, we update one variable while keeping the other variables fixed. Therefore, we can divide the optimization problem into the following multiple subproblems.

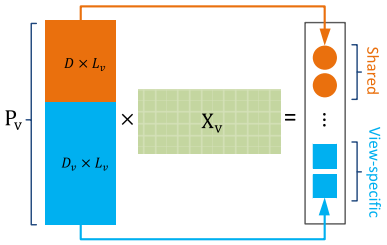


Fig. 2. Illustration of the processing flow to obtain shared features and view-specific features in a low-dimensional feature space by using a projection operation.

P_v-Subproblem: While keeping the other variables fixed except $\mathbf{P}_v$, the objective function in (8) is equivalent to minimizing

$$\min_{\mathbf{P}_v} \Phi\left(\mathbf{Y}_v^1, \mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1\right), \text{ s.t. } \mathbf{P}_v\mathbf{P}_v^\top = \mathbf{I}. \tag{9}$$

Equation (9) contains a matrix orthogonality constraint which has been used in [9] and [49]. We use the same method from [9] and [49] to obtain an optimal $\mathbf{P}_v$.

H_v-Subproblem: The associated optimization problem with respect to $\mathbf{H}_v$ can be drafted as

$$\begin{aligned}\min_{\mathbf{H}_v} \quad & \Phi\left(\mathbf{Y}_v^1, \mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1\right) \\ & + \Phi\left(\mathbf{Y}_v^2, [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\right). \end{aligned} \tag{10}$$

By ignoring some of the irrelevant terms of the equation, we have the following optimization problem:

$$\begin{aligned}\min_{\mathbf{H}_v} \quad & \Phi\left(\mathbf{Y}_v^{1,c}, \mathbf{P}_v^c\mathbf{X}_v - \mathbf{H}_v - \mathbf{E}_v^{1,c}\right) \\ & + \Phi\left(\mathbf{Y}_v^{2,c}, \mathbf{H}_v - \mathbf{H}_v\widetilde{\mathbf{Z}} - \mathbf{H}_v\mathbf{Z}_v - \mathbf{E}_v^{2,c}\right) \end{aligned} \tag{11}$$

where $\mathbf{P}_v = [\mathbf{P}_v^s; \mathbf{P}_v^c]$, $\mathbf{P}_v^s \in \mathbb{R}^{D\times L_v}$ denotes the components to project each view onto shared ones, and $\mathbf{P}_v^c \in \mathbb{R}^{D_v\times L_v}$ denotes the components to project each view onto view-specific ones (as shown in Fig. 2). Similarly, $\mathbf{E}_v^1 = [\mathbf{E}_v^{1,s}; \mathbf{E}_v^{1,c}]$ and $\mathbf{E}_v^2 = [\mathbf{E}_v^{2,s}; \mathbf{E}_v^{2,c}]$.

By taking the derivative of the above objective with respect to $\mathbf{H}_v$ and setting it to zero, we obtain the following closed-form solution:

$$\begin{aligned}\mathbf{H}_v = & \left(\mathbf{P}_v^c\mathbf{X}_v - \mathbf{E}_v^{1,c} + \frac{\mathbf{Y}_v^{1,c}}{\mu} + \mathbf{E}_v^{2,c}\mathbf{Z}_v'^\top - \frac{\mathbf{Y}_v^{2,c}\mathbf{Z}_v'^\top}{\mu}\right) \\ & \left(\mathbf{I} + \mathbf{Z}_v'\mathbf{Z}_v'^\top\right)^{-1} \end{aligned} \tag{12}$$

where $\mathbf{Z}_v' = \mathbf{I} - \widetilde{\mathbf{Z}} - \mathbf{Z}_v$, and $\mathbf{I}$ is an identity matrix.

Z_v-Subproblem: By fixing the other variables to constant, we update $\mathbf{Z}_v$ by solving

$$\begin{aligned}\min_{\mathbf{Z}_v} \quad & \Phi\left(\mathbf{Y}_v^3, \mathbf{Z}_v - \mathbf{Q}_v + \text{diag}(\mathbf{Q}_v)\right) \\ & + \Phi\left(\mathbf{Y}_v^2, [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\right). \end{aligned} \tag{13}$$

By reducing $\mathbf{B}_v = [\mathbf{H}; \mathbf{H}_v]$ and $\mathbf{A}_v = [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - \mathbf{E}_v^2$, the above problem is equivalent to minimizing

$$\min_{\mathbf{Z}_v} \ \Phi\Big(\mathbf{Y}_v^3, \mathbf{Z}_v - \mathbf{Q}_v + \operatorname{diag}(\mathbf{Q}_v)\Big) + \Phi\Big(\mathbf{Y}_v^2, \mathbf{A}_v - \mathbf{B}_v\mathbf{Z}_v\Big). \tag{14}$$

Similar to above, by taking the derivative of the above objective function with respect to $\mathbf{Z}_v$ and setting it to zero, we have the following closed-form solution for $\mathbf{Z}_v$ as:

$$\mathbf{Z}_v = \Big(\mathbf{B}_v^\top\mathbf{B}_v + \mathbf{I}\Big)^{-1} \Big(\mathbf{Q}_v - \operatorname{diag}(\mathbf{Q}_v) + \mathbf{B}_v^\top\mathbf{A}_v + \frac{\mathbf{B}_v^\top\mathbf{Y}_v^2 - \mathbf{Y}_v^3}{\mu}\Big). \tag{15}$$

$\mathbf{Q}_v$-Subproblem: Further, we update $\mathbf{Q}_v$ by solving the following problem:

$$\min_{\mathbf{Q}_v} \ \beta\|\mathbf{Q}_v\|_1 + \Phi\Big(\mathbf{Y}_v^3, \mathbf{Z}_v - \mathbf{Q}_v + \operatorname{diag}(\mathbf{Q}_v)\Big). \tag{16}$$

This is equivalent to optimizing

$$\min_{\mathbf{Q}_v} \ \frac{\beta}{\mu}\|\mathbf{Q}_v\|_1 + \frac{1}{2}\Big\|\mathbf{Q}_v - \operatorname{diag}(\mathbf{Q}_v) - \Big(\mathbf{Z}_v + \mathbf{Y}_v^3/\mu\Big)\Big\|_F^2. \tag{17}$$

This subproblem has a closed-form solution given by

$$\mathbf{Q}_v = \mathbf{Q}_v' - \operatorname{diag}\big(\mathbf{Q}_v'\big) \tag{18}$$

where $\mathbf{Q}_v' = \Im_{(\beta/\mu)}(\mathbf{Z}_v + \mathbf{Y}_v^3/\mu)$, and $\Im$ is an element-wise shrinkage thresholding operator [50].

$\mathbf{E}_v$-Subproblem: The error term $\mathbf{E}_v$ can be updated by solving

$$\begin{aligned}\mathbf{E}_v &= \operatorname*{argmin}_{\mathbf{E}_v} \ \lambda\|\mathbf{E}_v\|_{2,1} + \Phi\Big(\mathbf{Y}_v^1, \mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1\Big) \\ &\quad + \Phi\Big(\mathbf{Y}_v^2, [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\Big) \\ &= \operatorname*{argmin}_{\mathbf{E}_v} \ \frac{\lambda}{\mu}\|\mathbf{E}_v\|_{2,1} + \frac{1}{2}\|\mathbf{E}_v - \mathbf{G}_v\|_F^2\end{aligned} \tag{19}$$

where $\mathbf{G}_v$ is formed by vertically concatenating the matrices $\mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] + \mathbf{Y}_v^1/\mu$ and $[\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v + \mathbf{Y}_v^2/\mu$. This subproblem can be efficiently solved by [17, Lemma 3.2].

***J*-Subproblem:** When all the other variables remain fixed except $\mathbf{J}$, the optimization problem is equivalent to minimizing

$$\min_{\mathbf{J}} \|\mathbf{J}\|_* + \Phi\big(\mathbf{Y}_z, \widetilde{\mathbf{Z}} - \mathbf{J}\big). \tag{20}$$

This can be further reduced to

$$\min_{\mathbf{J}} \frac{1}{\mu}\|\mathbf{J}\|_* + \frac{1}{2}\big\|\mathbf{J} - \big(\widetilde{\mathbf{Z}} + \mathbf{Y}_z/\mu\big)\big\|_F^2. \tag{21}$$

The above problem can be solved by using a singular value thresholding operator [51].

$\widetilde{\mathbf{Z}}$-Subproblem: While considering the variable $\widetilde{\mathbf{Z}}$ assuming that the others remain fixed, the objective function in (8) is equivalent to minimizing

$$\begin{aligned}\min_{\widetilde{\mathbf{Z}}} \ &\Phi\big(\mathbf{Y}_z, \widetilde{\mathbf{Z}} - \mathbf{J}\big) \\ &+ \sum\nolimits_v \Phi\Big(\mathbf{Y}_v^2, [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\Big).\end{aligned} \tag{22}$$

For convenience, two auxiliary variables $\mathbf{C}_v$ and $\mathbf{D}_v$ can be introduced as

$$\min_{\widetilde{\mathbf{Z}}} \ \Phi\big(\mathbf{Y}_z, \widetilde{\mathbf{Z}} - \mathbf{J}\big) + \sum\nolimits_v \Phi\Big(\mathbf{Y}_v^2, \mathbf{C}_v - \mathbf{D}_v\widetilde{\mathbf{Z}}\Big) \tag{23}$$

where $\mathbf{C}_v = [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2$ and $\mathbf{D}_v = [\mathbf{H}; \mathbf{H}_v]$.

Similar to the other variables, by taking the derivative of the above objective function with respect to $\widetilde{\mathbf{Z}}$ and setting it to zero, we have the following closed-form solution:

$$\widetilde{\mathbf{Z}} = \Big(\mathbf{I} + \sum\nolimits_v \mathbf{D}_v^\top\mathbf{D}_v\Big)^{-1}\Big(\mathbf{J} - \frac{\mathbf{Y}_z}{\mu} + \sum\nolimits_v \mathbf{D}_v^\top\Big(\mathbf{C}_v + \frac{\mathbf{Y}_v^2}{\mu}\Big)\Big) \tag{24}$$

where $\mathbf{I}$ is an identity matrix.

***H*-Subproblem:** Finally, by dropping some of the unrelated terms, we optimize $\mathbf{H}$ as

$$\begin{aligned}\min_{\mathbf{H}} \ &\sum\nolimits_v \Phi\Big(\mathbf{Y}_v^{1,s}, \mathbf{P}_v^s\mathbf{X}_v - \mathbf{H} - \mathbf{E}_v^{1,s}\Big) \\ &+ \sum\nolimits_v \Phi\Big(\mathbf{Y}_v^{2,s}, \mathbf{H} - \mathbf{H}\widetilde{\mathbf{Z}} - \mathbf{H}\mathbf{Z}_v - \mathbf{E}_v^{2,s}\Big)\end{aligned} \tag{25}$$

whose closed-form solution can be calculated using

$$\begin{aligned}\mathbf{H} = &\sum\nolimits_v \Big(\mathbf{P}_v^s\mathbf{X}_v - \mathbf{E}_v^{1,s} + \frac{\mathbf{Y}_v^{1,s}}{\mu} + \mathbf{E}_v^{2,s}\mathbf{Z}_v'^\top - \frac{\mathbf{Y}_v^{2,s}\mathbf{Z}_v'^\top}{\mu}\Big) \\ &\Big(\sum\nolimits_v \Big(\mathbf{I} + \mathbf{Z}_v'\mathbf{Z}_v'^\top\Big)\Big)^{-1}\end{aligned} \tag{26}$$

where $\mathbf{Z}_v' = \mathbf{I} - \widetilde{\mathbf{Z}} - \mathbf{Z}_v$, and $\mathbf{I}$ is an identity matrix.

Multiplier: Besides, all multipliers $\mathbf{Y}_z$, $\mathbf{Y}_v^1$, $\mathbf{Y}_v^2$, and $\mathbf{Y}_v^3$ ($v = 1, \ldots, V$) can be updated by

$$\begin{cases}\mathbf{Y}_z \coloneqq \mathbf{Y}_z + \mu\big(\widetilde{\mathbf{Z}} - \mathbf{J}\big) \\ \mathbf{Y}_v^1 \coloneqq \mathbf{Y}_v^1 + \mu\big(\mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1\big) \\ \mathbf{Y}_v^2 \coloneqq \mathbf{Y}_v^2 + \mu\big([\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\big) \\ \mathbf{Y}_v^3 \coloneqq \mathbf{Y}_v^3 + \mu(\mathbf{Z}_v - \mathbf{Q}_v + \operatorname{diag}(\mathbf{Q}_v)).\end{cases} \tag{27}$$

After updating all the variables as mentioned above, we repeat the updates iteratively until convergence. The details of solving (6) using ADMM algorithm is summarized in Algorithm 1. By obtaining optimal $\widetilde{\mathbf{Z}}$ and $\mathbf{Z}_v (v = 1, \ldots, V)$, we can deduce the final clustering results by implementing the spectral clustering algorithm [16] on the similarity matrix $\mathbf{S}$, that is, $\mathbf{S} = (1/V)\sum_v(|\widetilde{\mathbf{Z}} + \mathbf{Z}_v| + |(\widetilde{\mathbf{Z}} + \mathbf{Z}_v)^\top|)$. It is important to note that the similarity matrix $\mathbf{S}$ simultaneously utilizes the shared and specific self-representations, which effectively exploits the correlations across multiple views and incorporates view-specific properties of individual views that benefit clustering performance.

C. Computational Complexity Analysis

In this section, we provide a detailed computational complexity analysis of our algorithm. A major share of the computational demand of Algorithm 1 can be attributed to: 1) matrix inversion and multiplication operations that are used for updating $\mathbf{H}_v$, $\mathbf{Z}_v$, $\mathbf{E}_v$, $\widetilde{\mathbf{Z}}$, and $\mathbf{H}$ and 2) SVD operation for

Algorithm 1: Solving Problem (8) via ADMM Algorithm

Input: Multiview data: $X_1, \ldots, X_V$, parameters λ, β, D and D_v.
Initialize: $\mathbf{Y}_z = 0$, $\mathbf{Y}_1^1, \ldots, \mathbf{Y}_V^1 = 0$, $\mathbf{Y}_1^2, \ldots, \mathbf{Y}_V^2 = 0$, $\mathbf{Y}_1^3, \ldots, \mathbf{Y}_V^3 = 0$, $\varepsilon = 10^{-6}$, $\rho = 1.2$, $\mu = 10^{-6}$, $max_\mu = 10^6$, initialize $\mathbf{H}$ and $\mathbf{H}_v$ with random values.
Output: $\widetilde{\mathbf{Z}}$, $\mathbf{Z}_{1,\ldots,V}$.
while *not converged* **do**
  for *v=1,...,V* **do**
    Update $\mathbf{P}_v$, $\mathbf{H}_v$, $\mathbf{Z}_v$, $\mathbf{Q}_v$ and $\mathbf{E}_v$ via Eq. (9), Eq. (12), Eq. (15), Eq. (18) and Eq. (19), respectively;
  end
  Update $\mathbf{J}$, $\widetilde{\mathbf{Z}}$ and $\mathbf{H}$ via Eq. (21), Eq. (24) and Eq. (26), respectively;
  Update $\mathbf{Y}_z$, $\mathbf{Y}_v^1$, $\mathbf{Y}_v^2$ and $\mathbf{Y}_v^3$ ($v = 1, \ldots, V$) via Eq. (27);
  Update the parameter μ via $\mu := \min(\rho\mu, max_\mu)$;
  Check the convergence conditions:
  $\|\widetilde{\mathbf{Z}} - \mathbf{J}\|_\infty < \varepsilon$ and $\|\mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1\|_\infty < \varepsilon$ and $\|[\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\|_\infty < \varepsilon$ and $\mathbf{Z}_v - \mathbf{Q}_v + \mathrm{diag}(\mathbf{Q}_v))\|_\infty < \varepsilon$.
end

updating $\mathbf{P}_v$ and $\mathbf{J}$. To account for the complexity in implementing the former, that is, to update $\mathbf{H}_v$ and $\mathbf{Z}_v$ would cost about $O(N^3)$, while the complexity of optimizing $\mathbf{E}_v$ is $O(N^2)$ and updating $\widetilde{\mathbf{Z}}$ and $\mathbf{Z}_v$ would costs about $O(N^3)$. On the other hand, for each view, the complexity of updating $\mathbf{P}_v$ is $O(L_v^2 N)$, and the complexity of updating $\mathbf{J}$ is $O(N^3)$ due to the SVD operation. Thus, the total complexity is determined as $O(T(VL_v^2 N + VN^2 + N^3))$, where T is the number of iterations and V is the number of multiple views. Furthermore, considering $N \gg V$ for multiview data settings, the final complexity of our proposed algorithm can be reported as $O(T(VL_v^2 N + N^3))$.

D. Convergence Analysis

We analyze the convergence property of the proposed optimization algorithm in this section. Although it is difficult to guarantee its convergence to a local minimum, empirical evidence on real-world data have demonstrated a stable convergence behavior of Algorithm 1 as shown in Fig. 10. Following [52] and [53], we also present a proof of weak convergence of Algorithm 1 by showing that under mild conditions, any limit point of the iteration sequence generated by Algorithm 1 is a stationary point that satisfies the Karush−Kuhn−Tucker (KKT) conditions [54]. Note that it is worth providing that any convergence point must be a point that satisfies the KKT conditions, since they are the necessary condition to obtain a local optimal solution. This result indicates an assurance about the convergence behavior of the proposed optimization algorithms.

Assume that the proposed algorithm reaches a stationary point, the KKT conditions for (7) can be derived as follows (since the procedure of solving $\mathbf{P}_v$ do not involve in the Lagrange multiplier, thus we ignore to proof the KKT conditions for it):

$$\begin{cases} \mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1 = 0, \mathbf{Z}_v - \mathbf{Q}_v + \mathrm{diag}(\mathbf{Q}_v) = 0 \\ [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2 = 0, \widetilde{\mathbf{Z}} - \mathbf{J} = 0 \\ \frac{\partial \mathcal{F}}{\partial \mathbf{H}_v} = \mathbf{Y}_v^{2,c}(\mathbf{I} - \mathbf{Z}_v^\top - \widetilde{\mathbf{Z}}^\top) - \mathbf{Y}_v^{1,c} = 0 \\ \frac{\partial \mathcal{F}}{\partial \mathbf{Z}_v} = \mathbf{Y}_v^3 - [\mathbf{H}; \mathbf{H}_v]^\top \mathbf{Y}_v^2 = 0 \\ \frac{\partial \mathcal{F}}{\partial \widetilde{\mathbf{Z}}} = \mathbf{Y}_z - \sum_v [\mathbf{H}; \mathbf{H}_v]^\top \mathbf{Y}_v^2 = 0 \\ \frac{\partial \mathcal{F}}{\partial \mathbf{H}} = \sum_v \mathbf{Y}_v^{1,s} - \sum_v \mathbf{Y}_v^{2,s}(\mathbf{I} - \widetilde{\mathbf{Z}} - \mathbf{Z}) = 0 \\ \left[\mathbf{Y}_v^1; \mathbf{Y}_v^2\right] \in \lambda\partial\|\mathbf{E}_v\|_{2,1}, \mathbf{Y}_v^3 \in \lambda\partial\|\mathbf{Q}_v\|_1, \mathbf{Y}_z \in \partial\|\mathbf{J}\|_*. \end{cases} \tag{28}$$

Then, we can obtain the following equation as:

$$\mathbf{E}_v = \Im_{\frac{\lambda}{\mu}}(o) \tag{29}$$

where $\Im_\eta(t) \triangleq \mathcal{P}(\eta, t)$ is the $\ell_{2,1}$ minimization operation [17], and $o = [\mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] + \mathbf{Y}_v^1/\mu; \mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v + \mathbf{Y}_v^2/\mu]$. We also obtain the following equation from the last one relationship as:

$$\mathbf{J} = \Omega_{\frac{1}{\mu}}\left(\widetilde{\mathbf{Z}} + \frac{\mathbf{Y}_z}{\mu}\right) \tag{30}$$

where $\Omega_\eta(\mathbf{Z})(\mathbf{Z} = \widetilde{\mathbf{Z}} + (\mathbf{Y}_z/\mu))$ denotes the singular value thresholding operation, which is computed by using

$$\Omega(\mathbf{Z}, \eta) = \mathbf{U}\mathbf{S}_\eta(\Sigma)\mathbf{V}^T \tag{31}$$

where $\mathbf{S}_\eta(\Sigma_{ii}) = \mathrm{sgn}(\Sigma_{ii})\max(0, |\Sigma_{ii} - \eta|)$ is the soft-thresholding operator and $\mathbf{Z} = \mathbf{U}\Sigma\mathbf{V}^T$ is the SVD of $\mathbf{Z}$ [51].

Besides, we can obtain the following relation as:

$$\mathbf{Q}_v' = \mathcal{S}\left(\mathbf{Z}_v + \frac{\mathbf{Y}_v^3}{\mu}, \frac{\beta}{\mu}\right) \tag{32}$$

where $\mathcal{S}(x, \tau) = \mathrm{sgn}(x)\max(|x| - \tau)$. Therefore, we have the following KKT conditions as:

$$\begin{cases} \mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1 = 0, \mathbf{Z}_v - \mathbf{Q}_v + \mathrm{diag}(\mathbf{Q}_v) = 0 \\ [\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2 = 0, \widetilde{\mathbf{Z}} - \mathbf{J} = 0 \\ \mathbf{Y}_v^{2,c}\left(\mathbf{I} - \mathbf{Z}_v^\top - \widetilde{\mathbf{Z}}^\top\right) - \mathbf{Y}_v^{1,c} = 0, \mathbf{Y}_v^3 - [\mathbf{H}; \mathbf{H}_v]^\top \mathbf{Y}_v^2 = 0 \\ \mathbf{Y}_z - \sum_v [\mathbf{H}; \mathbf{H}_v]^\top \mathbf{Y}_v^2 = 0 \\ \sum_v \mathbf{Y}_v^{1,s} - \sum_v \mathbf{Y}_v^{2,s}\left(\mathbf{I} - \widetilde{\mathbf{Z}} - \mathbf{Z}\right) = 0 \\ \mathbf{E}_v = \mathcal{P}\left(o, \frac{\lambda}{\mu}\right), \mathbf{J} = \Omega\left(\widetilde{\mathbf{Z}} + \frac{\mathbf{Y}_z}{\mu}, \frac{1}{\mu}\right) \\ \mathbf{Q}_v' = \mathcal{S}\left(\mathbf{Z}_v + \frac{\mathbf{Y}_v^3}{\mu}, \frac{\beta}{\mu}\right). \end{cases} \tag{33}$$

Based on the above analysis, in the following section, we prove the proposed algorithm converges to a point that satisfies the above KKT conditions.

Theorem 1: $\theta \triangleq (\mathbf{P}_v, \mathbf{H}_v, \mathbf{Z}_v, \mathbf{Q}_v, \mathbf{E}_v, \widetilde{\mathbf{Z}}, \mathbf{J}, \mathbf{H})$ and $\{\theta\}_{j=1}^\infty$ can be generated by Algorithm 1. Assume $\{\theta\}_{j=1}^\infty$ is bounded and $\lim_{j\to\infty}\{\theta^{j+1} - \theta^j\} = 0$. Then, any point of $\{\theta\}_{j=1}^\infty$ satisfies the KKT conditions. In particular, whenever $\{\theta\}_{j=1}^\infty$ converges, it converges to a KKT point.

The detailed proofs of Theorem 1 can be found in the supplementary material.

TABLE II
RESULTS (MEAN ± STANDARD DEVIATION) ON YALE DATASET

Feature	Method	NMI	ACC	AR	F-score	Precision	Recall
Single	$\text{Single}_{\text{best}}$	0.654±0.012	0.616±0.030	0.440±0.011	0.475±0.011	0.457±0.011	0.495±0.010
	LRR_{best}	0.709±0.011	0.697±0.001	0.515±0.004	0.547±0.007	0.529±0.003	0.567±0.004
	S3C_{best}	0.593±0.008	0.554±0.013	0.362±0.010	0.403±0.010	0.380±0.011	0.428±0.013
Multiple	FeatConcate	0.648±0.030	0.607±0.043	0.434±0.042	0.471±0.039	0.447±0.043	0.497±0.032
	ConcatePCA	0.665±0.037	0.578±0.038	0.396±0.011	0.434±0.011	0.419±0.012	0.450±0.009
	Co-Reg SPC	0.648±0.002	0.564±0.001	0.436±0.002	0.466±0.001	0.455±0.004	0.491±0.003
	ConReg SPC	0.673±0.023	0.611±0.035	0.466±0.032	0.501±0.030	0.476±0.032	0.532±0.029
	Min-Dis	0.645±0.005	0.615±0.043	0.433±0.006	0.470±0.006	0.446±0.005	0.496±0.006
	MVSC	0.650±0.015	0.652±0.023	0.425±0.009	0.501±0.010	0.451±0.010	0.532±0.011
	LMSC	0.735±0.021	0.752±0.026	0.551±0.011	0.564±0.019	0.543±0.015	0.571±0.013
	DSS-MSC	**0.779±0.012**	**0.782±0.013**	**0.601±0.009**	**0.613±0.012**	**0.592±0.010**	**0.622±0.015**

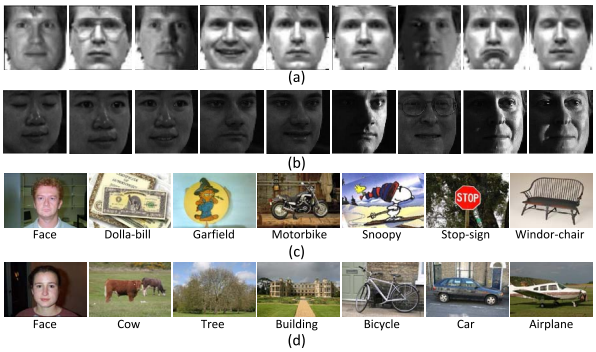


Fig. 3. Sampling images from (a) Yale, (b) CMU-PIE, (c) Caltech101, and (d) MSRCV1 datasets.

IV. EXPERIMENTAL RESULTS

In this section, we first give the details about five real-word multiview datasets, state-of-the-art methods, and basic experimental settings that have been used in validating our proposed framework. Further, we present experimental results and evaluate some properties of the proposed model.

A. Experimental Setup

We evaluate the effectiveness of the proposed approach by using the following five benchmark datasets.

1) *Yale*[1]: This dataset contains 165 gray-scale images of 15 individuals. Each has 11 images, with different facial expressions and other configurations, including center-light, with glasses, happy, left-light, without glasses, normal, right-light, sad, sleepy, surprised, and wink [Fig. 3(a)]. Following [11], three types of features, that is, intensity (view1), LBP (view2), and Gabor (view3), are used.
2) *MSRCV1*[2]: This dataset consists of 240 images and 8 object classes. Similar to [9] and [55], we select seven classes, that is, Cow, Tree, Building, Airplane, Face, Car, and Bicycle [Fig. 3(d)]. Following [9], we extract six types features on in the Caltech101-7 dataset, that is, CENTRIST (view1), CMT (view2), GIST (view3), HOG (view4), LBP (view5), and SIFT (view6).
3) *Caltech101-7*[3]: This is a widely used subset of the image dataset Caltech101 which contains 101 categories. In this subset, there are 441 total images selected from seven categories, including Dolla-bill, Face, Garfield, Motorbike, Snoopy, Stop-sign, and Windsor-chair [Fig. 3(c)]. Specifically, six types of features can be extracted, including CENTRIST (view1), CMT (view2), GIST (view3), HOG (view4), LBP (view5), and SIFT (view6) [56].
4) *BBCSport*[4]: The dataset includes the documents from the BBC Sport website corresponding to sports news under five popular areas, which is associated with two views [9], [37].
5) *CMU-PIE*[5]: This multiview face dataset consists of 68 subjects in total, with large variances within the same subject but in different poses [Fig. 3(b)]. We randomly select 15 samples from each subject to construct 1020 facial images in the evaluation subset, and all face images are cropped to 64×64 size. Besides, three types of features: intensity (view1), LBP (view2), and HOG (view3) are used in the dataset.

We compare the proposed subspace clustering approach with some recent state-of-the-art methods.

1) *Single_{best}:* The method performs the standard spectral clustering algorithm [16] by selectively using the most informative view.
2) *LRR_{best}:* The method performs LRR [17] by selectively using the most informative view with.
3) *S3C_{best} [18]:* The method carries out clustering on every single view and reports the view that demonstrated the best performance.
4) *FeatConcate:* The method first concatenates features from all views and then applies the standard spectral clustering.
5) *ConcatePCA:* The method first concatenates features from all views and then applies PCA [57] to obtain a low-dimensional subspace representation. Further, it applies standard spectral clustering on the low-dimensional representation.

[1]http://cvc.yale.edu/projects/yalefaces/yalefaces.html
[2]http://research.microsoft.com/en-us/projects/objectclassrecognition/
[3]http://www.vision.caltech.edu/Image_Datasets/Caltech101/
[4]http://mlg.ucd.ie/datasets/
[5]http://vasc.ri.cmu.edu/idb/html/face/

TABLE III
RESULTS (MEAN ± STANDARD DEVIATION) ON MRSCV1 DATASET

Feature	Method	NMI	ACC	AR	F-score	Precision	Recall
Single	$\text{Single}_{\text{best}}$	0.574±0.032	0.668±0.051	0.536±0.010	0.535±0.043	0.571±0.009	0.612±0.009
	LRR_{best}	0.569±0.008	0.676±0.009	0.502±0.010	0.524±0.009	0.543±0.009	0.587±0.007
	S3C_{best}	0.612±0.005	0.688±0.009	0.514±0.006	0.583±0.006	0.572±0.009	0.594±0.010
Multiple	FeatConcate	0.613±0.042	0.672±0.031	0.505±0.032	0.575±0.024	0.566±0.021	0.586±0.027
	ConcatePCA	0.621±0.022	0.702±0.015	0.541±0.009	0.607±0.014	0.595±0.011	0.617±0.015
	Co-Reg SPC	0.569±0.013	0.653±0.017	0.512±0.010	0.587±0.018	0.543±0.010	0.583±0.011
	ConReg SPC	0.670±0.013	0.651±0.011	0.577±0.009	0.637±0.014	0.601±0.021	0.656±0.010
	Min-Dis	0.657±0.017	0.745±0.044	0.567±0.008	0.628±0.007	0.615±0.015	0.643±0.010
	MVSC	0.615±0.012	0.695±0.008	0.506±0.014	0.573±0.015	0.525±0.013	0.616±0.012
	LMSC	0.653±0.011	0.806±0.013	0.599±0.017	0.652±0.017	0.612±0.012	0.663±0.011
	DSS-MSC	**0.743±0.015**	**0.846±0.011**	**0.681±0.014**	**0.726±0.021**	**0.711±0.011**	**0.743±0.013**

TABLE IV
RESULTS (MEAN ± STANDARD DEVIATION) ON CALTECH101 DATASET

Feature	Method	NMI	ACC	AR	F-score	Precision	Recall
Single	$\text{Single}_{\text{best}}$	0.589±0.009	0.629±0.007	0.523±0.012	0.576±0.009	0.586±0.014	0.566±0.003
	LRR_{best}	0.639±0.002	0.646±0.003	0.580±0.001	0.649±0.002	0.631±0.001	0.623±0.003
	S3C_{best}	0.578±0.000	0.611±0.007	0.504±0.009	0.559±0.007	0.568±0.010	0.551±0.006
Multiple	FeatConcate	0.603±0.017	0.641±0.020	0.526±0.034	0.601±0.023	0.624±0.021	0.579±0.024
	ConcatePCA	0.651±0.012	0.672±0.017	0.558±0.009	0.634±0.013	0.639±0.014	0.621±0.010
	Co-Reg SPC	0.623±0.003	0.590±0.005	0.549±0.005	0.620±0.004	0.645±0.005	0.598±0.003
	ConReg SPC	0.607±0.002	0.695±0.003	0.570±0.004	0.637±0.004	0.666±0.003	0.610±0.003
	Min-Dis	0.624±0.004	0.701±0.023	0.552±0.007	0.623±0.006	0.645±0.006	0.603±0.007
	MVSC	0.683±0.002	0.712±0.003	0.596±0.004	0.675±0.004	0.566±0.003	0.667±0.003
	LMSC	0.652±0.013	0.710±0.012	0.593±0.002	0.664±0.009	0.656±0.014	0.661±0.012
	DSS-MSC	**0.691±0.002**	**0.737±0.001**	**0.635±0.002**	**0.703±0.006**	**0.698±0.009**	**0.710±0.003**

6) *Co-Reg SPC [34]:* The pairwise multiview spectral clustering method co-regularizes the clustering hypotheses to enforce corresponding data points in each view to have the same cluster membership.
7) *ConReg SPC [58]:* A common representation of all views is first learned. Then, a standard spectral clustering is applied to the similarity matrix.
8) *Min-Dis [59]:* This method creates a bipartite graph and tries to minimize the disagreement. The final result is obtained through spectral clustering.
9) *MVSC [35]:* This method performs subspace clustering on individual views and fuses them to obtain the final result of clustering.
10) *LMSC [9]:* This method assumes that each view is originated from one underlying latent representation.

In order to evaluate the clustering performance of the proposed algorithm and compare it against competing baselines, six popular metrics, including normalized mutual information (NMI), accuracy (ACC), adjusted Rand index (AR), *F*-score, precision, and recall are utilized in this paper. Each metric penalizes or favors different properties in the clustering task and, hence, we report results on these different measures to employ a comprehensive evaluation of our method against all state-of-the-art methods. Note that a higher value of each metric indicates better clustering quality.

For a fair comparison in all experiments, we obtain the best performance for all comparison methods by using the source codes from the authors with the default or suggestion parameter settings, or directly cite the best experimental results from their original papers. Specifically, for LRR, we tune the parameter λ in the range of $\{0.001, 0.01, 0.02, 0.05, 0.1, 0.2, 1, 2, 5\}$ according the authors' suggestion. For S3C, the parameter spaces are $\lambda \in \{10^{-5}, 10^{-4}, \ldots, 10^{5}\}$, $\alpha \in [0.03, 0.3]$. For ConcatePCA, the optimal dimensionality is searched in a range $\{100, 200, \ldots, 500\}$. For Co-Reg SPC, its parameter λ is searched in a range of $\{0, 0.02, \ldots, 0.1\}$ according to authors' suggestion. For MVSC, its parameters λ_1 and λ_2 are search in a range of $\{10^{-4}, 10^{-3}, \ldots, 10^{3}\}$. For LMSC, its parameter λ is searched from $\{10^{-4}, 10^{-3}, \ldots, 10^{3}\}$ according to authors' suggestion.

For our method, we tune the dimension of the view-specific features in the range of $D_v \in \{10, 20, \ldots, 100\}$, and tune the dimension of the shared features in the range of $D \in \{5, 10, \ldots, 40\}$. We also tune the parameters λ and β in the range of $\{10^{-5}, \ldots, 10^{2}\}$. Finally, we report the mean values and standard deviations for all methods over 30 independent trials.

B. Performance Comparison

We report the detailed clustering results in Tables II–VI. In each table, the values in bold indicate the best performance.

Table II shows the clustering results on the Yale dataset. From Table II, it can be observed that our method obtains much better clustering performance than the competing single-view and multiview methods. The main reason is that our approach combines shared and specific self-representations

TABLE V
RESULTS (MEAN ± STANDARD DEVIATION) ON BBCSPORT DATASET

Feature	Method	NMI	ACC	AR	F-score	Precision	Recall
Single	$\text{Single}_{\text{best}}$	0.781±0.007	0.849±0.018	0.784±0.019	0.834±0.015	0.845±0.008	0.827±0.022
	LRR_{best}	0.690±0.002	0.787±0.003	0.745±0.004	0.770±0.002	0.812±0.003	0.767±0.004
	S3C_{best}	0.781±0.008	0.851±0.016	0.787±0.019	0.837±0.015	0.844±0.007	0.831±0.023
Multiple	FeatConcate	0.795±0.021	0.830±0.033	0.789±0.035	0.834±0.042	0.843±0.029	0.836±0.041
	ConcatePCA	0.809±0.027	0.847±0.048	0.806±0.024	0.852±0.031	0.851±0.022	0.854±0.036
	Co-Reg SPC	0.792±0.005	0.764±0.022	0.805±0.012	0.851±0.010	0.858±0.005	0.844±0.014
	ConReg SPC	0.814±0.000	0.898±0.001	0.841±0.000	0.879±0.000	0.872±0.002	0.887±0.001
	Min-Dis	0.793±0.002	0.885±0.048	0.792±0.008	0.840±0.006	0.848±0.003	0.835±0.011
	MVSC	0.785±0.018	0.879±0.049	0.781±0.045	0.834±0.030	0.830±0.064	0.842±0.020
	LMSC	0.826±0.007	0.912±0.006	0.842±0.011	0.887±0.008	0.873±0.007	0.877±0.012
	DSS-MSC	**0.884±0.012**	**0.966±0.004**	**0.898±0.009**	**0.923±0.007**	**0.927±0.007**	**0.918±0.008**

TABLE VI
RESULTS (MEAN ± STANDARD DEVIATION) ON CMU-PIE DATASET

Feature	Method	NMI	ACC	AR	F-score	Precision	Recall
Single	$\text{Single}_{\text{best}}$	0.763±0.004	0.623±0.011	0.489±0.009	0.487±0.009	0.428±0.009	0.502±0.009
	LRR_{best}	0.792±0.005	0.667±0.015	0.514±0.017	0.501±0.017	0.452±0.021	0.562±0.011
	S3C_{best}	0.747±0.004	0.555±0.008	0.505±0.008	0.473±0.008	0.404±0.009	0.497±0.008
Multiple	FeatConcate	0.719±0.012	0.579±0.020	0.512±0.024	0.453±0.021	0.459±0.023	0.481±0.024
	ConcatePCA	0.772±0.009	0.619±0.023	0.508±0.024	0.504±0.024	0.478±0.029	0.557±0.021
	Co-Reg SPC	0.728±0.003	0.601±0.003	0.514±0.004	0.477±0.014	0.483±0.013	0.493±0.015
	ConReg SPC	0.739±0.003	0.607±0.021	0.517±0.004	0.481±0.004	0.493±0.003	0.514±0.005
	Min-Dis	0.723±0.010	0.613±0.021	0.528±0.017	0.477±0.017	0.497±0.015	0.523±0.020
	MVSC	0.782±0.013	0.704±0.013	0.568±0.014	0.528±0.009	0.517±0.013	0.619±0.015
	LMSC	0.858±0.009	0.754±0.023	0.599±0.017	0.605±0.013	0.562±0.015	0.685±0.021
	DSS-MSC	**0.878±0.008**	**0.810±0.014**	**0.651±0.011**	**0.653±0.011**	**0.611±0.014**	**0.723±0.010**

to form a new ideal similarity matrix, which can simultaneously exploit the correlations among multiple views and learn particular representations to preserve view-specific property for improving clustering performance. It is also important to note that our method utilizes the low-dimensional features to reconstruct the data points, which eliminates the influence of redundancy to benefit the improved clustering results. Besides, merely concatenating features from multiview is not a promising proposition since it does not effectively address the correlations among multiple views, thereby resulting in poor performance.

Tables IV and III show the comparison of clustering results on the MRSCV1 and Caltech101 datasets, respectively. From Tables III and IV, it can be noticed that our proposed method obtains the best results in all of the six evaluation criteria. Relatively, MVSC and LMSC also obtain comparable performances on the Caltech101 dataset. Despite reporting clustering performance by combining two of the best views, the Min-Dis method still performs poorly compared to LMSC and MVSC. One possible reason is the feature redundancy when Min-Dis method performs clustering on the original features.

Similar trends can also be inferred from Tables II and III. A majority of the methods from the literature reports lower performance compared to the proposed method as shown in Tables V and VI. Against other baselines, LMSC and MVSC again obtain more promising performance. Co-Reg SPC also obtains relatively better performance, particularly, on the BBCSport dataset. It could be concluded that the co-regularized-based clustering algorithm is suitable for constructing one similarity matrix on this dataset. Overall, our method demonstrates much better clustering performance than other state-of-the-art methods on six evaluation metrics, which verifies the effectiveness of the proposed subspace clustering approach.

C. Model Property Evaluation

1) Critical Components Verification: Experimental results so far have clearly demonstrated the effectiveness of the proposed framework. In addition, there are three critical components in our proposed method: 1) dual learning framework; 2) different views share partial features in a low-dimensional feature space; and 3) our model uses low-dimensional features to reconstruct the data points. In the following texts, we will verify the effectiveness of these critical components in our proposed multiview clustering framework.

First, if we do not consider formulating a dual learning framework for exploiting shared-specific information into the dimensional reduction and self-representation processes, our method will degrade to

$$\begin{aligned}\min_{\mathbf{P}_v,\mathbf{H}_v,\mathbf{E}_v,\widetilde{\mathbf{Z}}} \quad & \|\widetilde{\mathbf{Z}}\|_* + \lambda \sum\nolimits_v \|\mathbf{E}_v\|_{2,1} \\ \text{s.t.} \quad & \mathbf{P}_v\mathbf{X}_v = \mathbf{H}_v + \mathbf{E}_v^1, \mathbf{H}_v = \mathbf{H}_v\widetilde{\mathbf{Z}} + \mathbf{E}_v^2 \\ & \mathbf{P}_v\mathbf{P}_v^\top = \mathbf{I}, \mathbf{E}_v = \left[\mathbf{E}_v^1; \mathbf{E}_v^2\right], \forall v = 1, \ldots, V. \end{aligned} \quad (34)$$

Remarks: We first project the original features from each view into a low-dimensional space to obtain the view-specific

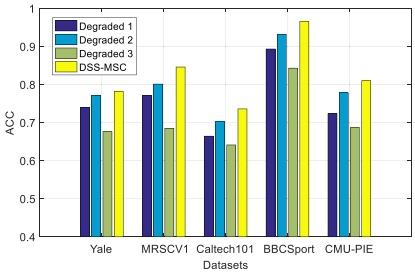


Fig. 4. Performance comparison between the proposed multiview subspace clustering approach (DSS-MSC) with the three degraded ones.

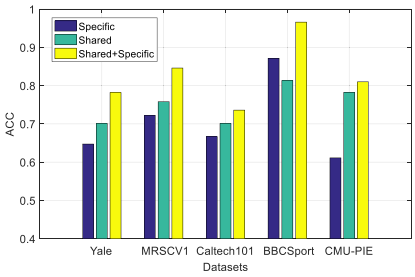


Fig. 5. Performance variation of our algorithm respects to different critical components (e.g., specific, shared, and combining shared with specific self-representations).

representations. And then, we assume that different views share self-representations across multiple views. Thus, the degraded formulation only employs single shared-specific learning process, and it is denoted as "Degraded 1."

Second, if we do not consider the shared feature across multiple views in the low-dimensional feature space, our method will degrade to

$$\begin{aligned}\min_{\mathbf{P}_v,\mathbf{H}_v,\mathbf{Z}_v,\mathbf{E}_v,\widetilde{\mathbf{Z}}} \quad & \|\widetilde{\mathbf{Z}}\|_* + \lambda \sum\nolimits_v \|\mathbf{E}_v\|_{2,1} + \beta \sum\nolimits_v \|\mathbf{Z}_v\|_1 \\ \text{s.t.} \quad & \mathbf{P}_v\mathbf{X}_v = \mathbf{H}_v + \mathbf{E}_v^1 \\ & \mathbf{H}_v = \mathbf{H}_v\widetilde{\mathbf{Z}} + \mathbf{H}_v\mathbf{Z}_v + \mathbf{E}_v^2 \\ & \mathbf{P}_v\mathbf{P}_v^\top = \mathbf{I}, \mathbf{E}_v = \left[\mathbf{E}_v^1; \mathbf{E}_v^2\right], \text{diag}(\mathbf{Z}_v) = 0 \\ & \forall v = 1, \ldots, V. \end{aligned} \tag{35}$$

Remarks: In this formulation, we only exploit the correlation across multiple views in self-representation process by using the low-dimensional feature representations. We denote it as "Degraded 2." Note that the degraded formulation can also be regarded as one-time shared-specific learning framework, that is, we exploit the correlation across multiple views and preserve view-specific property for each view based on the low-dimensional features.

Third, if we directly use the original features to reconstruct the data points without projecting them into a low-dimensional feature space, our method will degrade to

$$\begin{aligned}\min_{\mathbf{Z}_v,\mathbf{E}_v,\widetilde{\mathbf{Z}}} \quad & \left\|\widetilde{\mathbf{Z}}\right\|_* + \lambda \sum\nolimits_v \|\mathbf{E}_v\|_{2,1} + \beta \sum\nolimits_v \|\mathbf{Z}_v\|_1 \\ \text{s.t.} \quad & \mathbf{X}_v = \mathbf{X}_v\widetilde{\mathbf{Z}} + \mathbf{X}_v\mathbf{Z}_v + \mathbf{E}_v, \text{diag}(\mathbf{Z}_v) = 0 \\ & \forall v = 1, \ldots, V. \end{aligned} \tag{36}$$

Remarks: In this formulation, we use the original features to reconstruct the data points without projecting them into a low-dimensional feature space, and we denote it as "Degraded 3."

Fig. 4 shows the performance comparison (in terms of ACC) of our proposed framework and the three degraded counterparts. From Fig. 4, we can have the following observations.

1) It can be clearly seen that our proposed approach performs better than its three degraded formulations.
2) The dual learning framework performs better than that using one-time shared-specific learning framework, which aims to strengthen the ability of the proposed approach for exploiting the shared information and preserve view-specific property effectively.
3) The comparison results also verify the clear advantages of the low-dimensional features induced multiview subspace clustering over the degraded formulation by using the original features.

In addition, as described in the previous sections, our approach effectively learns shared representations to exploit the correlations among multiple views and simultaneously captures view-specific representations for each view to depict its unique property. In this manner, our method combined shared and specific self-representations to improve subspace clustering performance. In order to demonstrate the effectiveness of this fusion process, combining shared and specific self-representations, against using either specific [i.e., $\mathbf{S} = (1/V)\sum_v(|\mathbf{Z}_v| + |\mathbf{Z}_v^T|)$] or shared self-representations (i.e., $\mathbf{S} = |\widetilde{\mathbf{Z}}| + |\widetilde{\mathbf{Z}}^T|$) individually, we show the comparison results in Fig. 5. It can be seen that our method combining view-specific and shared self-representations obtain the best performance in comparison to either specific self-representations or shared self-representations independently.

Further, in order to intuitively investigate the effectiveness of our method, we visualize different views and different self-representations by using t-distributed stochastic neighbor embedding (t-SNE) algorithm [60] on the MSRCV1 dataset. It is observed that the learned self-representations [as shown in Fig. 6(g)–(i)] characterize the underlying cluster structure much better than all original feature representations [as shown in Fig. 6(a)–(f)]. Especially, Fig. 6(i) (combining specific and shared self-representations) shows the advantage of the proposed subspace clustering method, which also characterizes the underlying cluster structure much better than specific or shared self-representations.

2) Parameters Sensitivity: In our proposed approach, there are two regularization parameters, that is, λ and β in (6), which are used to balance the two constraint terms. Fig. 7 shows the performance of the proposed approach with respect to ACC measure for different values of λ and β values on five benchmark datasets. From Fig. 7, it can be seen that the clustering performance is good when the values of λ are not small, and the value of β is set to smaller than the value of λ on these datasets (please refer to see the details of parameters' settings in Section IV-A). Specifically, our proposed multiview clustering approach can obtain better performance when the values of λ and β fall in $\{0.1, 1, 10\}$ and $\{0.001, 0.01, 0.1\}$, respectively, which indicates that suitable values of λ and β can guarantee much better clustering performance. In addition, Figs. 8 and 9

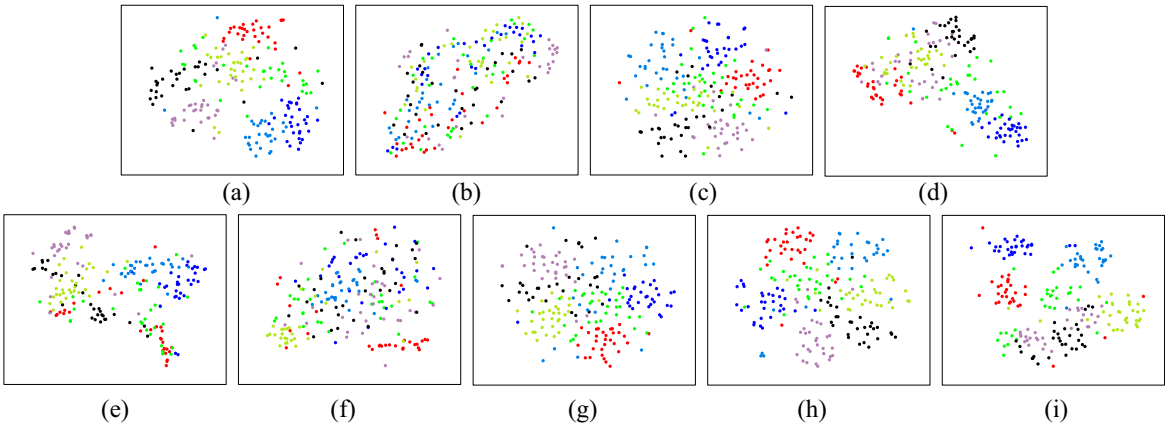


Fig. 6. Visualization of different views and different self-representations with t-SNE [60]. (a) View1. (b) View2. (c) View3. (d) View4. (e) View5. (f) View6. (g) Specific. (h) Shared. (i) Shared + Specific.

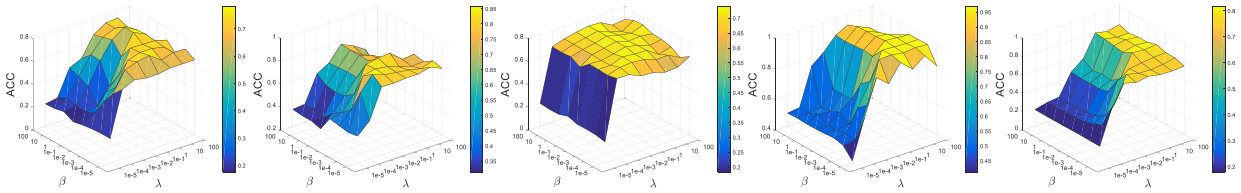


Fig. 7. Performance of the proposed approach with respect to ACC measure when varying parameters λ and β on five benchmark datasets (from left to right: Yale, MRSCV1, Caltech101, BBCSport, and CMU-PIE).

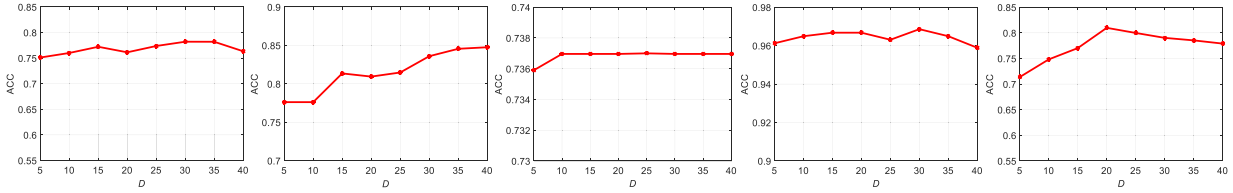


Fig. 8. Performance of the proposed approach with respect to ACC measure when varying the parameter D on five benchmark datasets (from left to right: Yale, MRSCV1, Caltech101, BBCSport, and CMU-PIE).

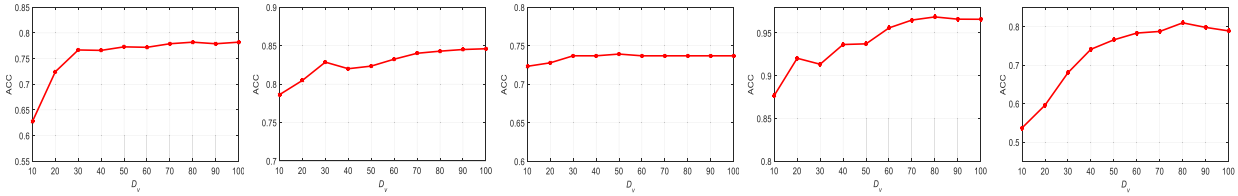


Fig. 9. Performance of the proposed approach with respect to ACC measure when varying the parameter D_v on five benchmark datasets (from left to right: Yale, MRSCV1, Caltech101, BBCSport, and CMU-PIE).

show the performance of the proposed approach with respect to ACC measure for different values of D and D_v. From Fig. 8, it is can be seen that our approach obtain relatively better clustering performance when $D \geq 20$, especially, the performance of our approach is not sensitive to the value of D on the Caltech101 and BBCSport datasets. Besides, from Fig. 9, it can be seen that our approach obtain relatively better clustering performance when $D_v \geq 60$, especially, the performance of our approach is not sensitive to the value of D_v on the Caltech101 datasets. Some previous works have indicated how to identify the optimal values of different parameters are data dependent and still an open problem [53], [61]. Overall, our method can obtain much better performance when the parameters λ and β are tuned in a small range, that is, $\{0.1, 1, 10\}$ and $\{0.001, 0.01, 0.1\}$, respectively, and with setting $D \geq 20$ and $D_v \geq 60$, respectively.

3) Convergence Analysis: We compute the relative errors, that is, $\|\mathbf{P}_v\mathbf{X}_v - [\mathbf{H}; \mathbf{H}_v] - \mathbf{E}_v^1\|_\infty$ as "error1," $\|[\mathbf{H}; \mathbf{H}_v] - [\mathbf{H}; \mathbf{H}_v]\widetilde{\mathbf{Z}} - [\mathbf{H}; \mathbf{H}_v]\mathbf{Z}_v - \mathbf{E}_v^2\|_\infty$ as "error2," $\|\mathbf{Z}_v - \mathbf{Q}_v + \text{diag}(\mathbf{Q}_v)\|_\infty$ as "error3," and $\|\widetilde{\mathbf{Z}} - \mathbf{J}\|_\infty$ as "error4" to better characterize the convergence of our approach. For ease of illustration, the errors are normalized and summed across views. Fig. 10 demonstrates the behavior of convergence conditions for the presented optimization algorithm. It can be seen that on all five real-world datasets, the optimization algorithm converges within 40–60 iterations.

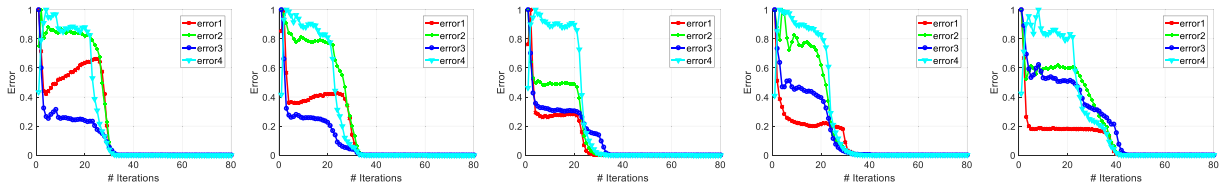


Fig. 10. Convergence curves using the sum of normalized errors across views for the presented optimization algorithm on five benchmark datasets (from left to right: Yale, MRSCV1, Caltech101, BBCSport, and CMU-PIE).

V. Conclusion

In this paper, we have proposed a DSS-MSC approach, which learns the shared information to exploit the underlying correlations cross multiple views and simultaneously capture view-specific details to depict specific property for each independent view. Moreover, a dual learning framework is presented to strengthen the ability of our method in exploiting the correlation and preserving view-specific properties without being influenced by redundancies or high-dimensionality of data. Extensive experiments on five benchmark datasets have demonstrated the effectiveness of the proposed approach against state-of-the-art methods, including single-view as well as multiview clustering approaches. Our validation of critical components have also verified that the enhanced shared and view-specific information could be regarded as valuable complements to benefit multiview subspace clustering performance. In the future, to capture more complex correlations, some nonlinear methods [62] and deep networks [63] will be introduced in our model. Moreover, our model can be extended to unsupervised features learning [64] and other classification tasks [65], [66].

References

[1] X. Peng, J. Lu, Z. Yi, and R. Yan, "Automatic subspace learning via principal coefficients embedding," *IEEE Trans. Cybern.*, vol. 47, no. 11, pp. 3583–3596, Nov. 2017.

[2] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Proc. Int. Joint Conf. Artif. Intell.*, Phoenix, AZ, USA, 2016, pp. 1181–1187.

[3] H. F. Bassani and A. F. R. Araujo, "Dimension selective self-organizing maps with time-varying structure for subspace and projected clustering," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 3, pp. 458–471, Mar. 2015.

[4] X. Peng, Z. Yu, Z. Yi, and H. Tang, "Constructing the l2-graph for robust subspace learning and subspace clustering," *IEEE Trans. Cybern.*, vol. 47, no. 4, pp. 1053–1066, Apr. 2017.

[5] C.-G. Li, C. You, and R. Vidal, "Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework," *IEEE Trans. Image Process.*, vol. 26, no. 6, pp. 2988–3001, Jun. 2017.

[6] J. Wang *et al.*, "Distance metric learning for soft subspace clustering in composite kernel space," *Pattern Recognit.*, vol. 52, pp. 113–134, Apr. 2016.

[7] L. Zhai, J. Zhu, Q. Zheng, S. Pang, Z. Li, and J. Wang, "Multi-view spectral clustering via partial sum minimisation of singular values," *Electron. Lett.*, vol. 55, no. 6, pp. 314–316, Mar. 2019.

[8] M. Rahmani and G. K. Atia, "Innovation pursuit: A new approach to subspace clustering," *IEEE Trans. Signal Process.*, vol. 65, no. 23, pp. 6276–6291, Dec. 2017.

[9] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Honolulu, HI, USA, 2017, pp. 4279–4287.

[10] F. Schroff, D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 815–823.

[11] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 1582–1590.

[12] Y. Wang, L. Wu, X. Lin, M. Fang, and S. Pan, "Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering," in *Proc. 25th Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 2153–2159.

[13] E. Arias-Castro, G. Lerman, and T. Zhang, "Spectral clustering based on local PCA," *J. Mach. Learn. Res.*, vol. 18, no. 9, pp. 1–57, 2017.

[14] U. Von Luxburg, "A tutorial on spectral clustering," *Stat. Comput.*, vol. 17, no. 4, pp. 395–416, 2007.

[15] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, 2009, pp. 2790–2797.

[16] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Adv. Neural Inf. Process. Syst.*, Vancouver, BC, Canada, 2002, pp. 849–856.

[17] 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.

[18] C.-G. Li and R. Vidal, "Structured sparse subspace clustering: A unified optimization framework," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 277–286.

[19] T. Zhou, K.-H. Thung, X. Zhu, and D. Shen, "Effective feature learning and fusion of multimodality data using stage-wise deep neural network for dementia diagnosis," *Human Brain Map.*, vol. 40, no. 3, pp. 1001–1016, 2019.

[20] C. Gong, D. Tao, S. J. Maybank, W. Liu, G. Kang, and J. Yang, "Multi-modal curriculum learning for semi-supervised image classification," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3249–3260, Jul. 2016.

[21] C. Gong, "Exploring commonality and individuality for multi-modal curriculum learning," in *Proc. AAAI Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 1926–1933.

[22] T. Zhou, M. Liu, K.-H. Thung, and D. Shen, "Latent representation learning for Alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data," *IEEE Trans. Med. Image*, to be published.

[23] T. Zhou, K.-H. Thung, M. Liu, F. Shi, C. Zhang, and D. Shen, "Multi-modal neuroimaging data fusion via latent space learning for Alzheimer's disease diagnosis," in *Proc. Int. Workshop Predict. Intell. Med.*, 2018, pp. 76–84.

[24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[25] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[26] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

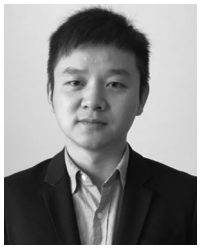
[27] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 1. San Diego, CA, USA, 2005, pp. 886–893.

[28] D. Zhou and C. J. C. Burges, "Spectral clustering and transductive learning with multiple views," in *Proc. Int. Conf. Mach. Learn.*, Corvallis, OR, USA, 2007, pp. 1159–1166.

[29] L. Shu and L. J. Latecki, "Integration of single-view graphs with diffusion of tensor product graphs for multi-view spectral clustering," in *Proc. Asian Conf. Mach. Learn.*, Hamilton, New Zealand, 2016, pp. 362–377.

[30] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Proc. Int. Joint Conf. Artif. Intell.*, Austin, TX, USA, 2015, pp. 2750–2756.

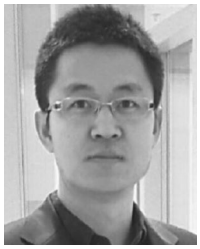
[31] W. Tang, Z. Lu, and I. S. Dhillon, "Clustering with multiple graphs," in *Proc. ICDM*, Miami, FL, USA, 2009, pp. 1016–1021.
[32] A. Banerjee, S. Basu, and S. Merugu, "Multi-way clustering on relation graphs," in *Proc. SIAM Int. Conf. Data Min.*, Minneapolis, MS, USA, 2007, pp. 145–156.
[33] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *Proc. Int. Conf. Mach. Learn.*, Bellevue, WA, USA, 2011, pp. 393–400.
[34] A. Kumar, P. Rai, and H. Daumé, "Co-regularized multi-view spectral clustering," in *Proc. Adv. Neural Inf. Process. Syst.*, Granada, Spain, 2011, pp. 1413–1421.
[35] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, Santiago, Chile, 2015, pp. 4238–4246.
[36] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 586–594.
[37] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. Int. Joint Conf. Artif. Intell.*, Quebec City, QC, Canada, 2014, pp. 2149–2155.
[38] T. Zhou, C. Zhang, C. Gong, H. Bhaskar, and J. Yang, "Multiview latent space learning with feature redundancy minimization," *IEEE Trans. Cybern.*, to be published.
[39] C. Zhang *et al.*, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, to be published.
[40] A. Gretton, O. Bousquet, A. Smola, and B. Schölkopf, "Measuring statistical dependence with Hilbert–Schmidt norms," in *Proc. Int. Conf. Algorithmic Learn. Theory*, vol. 16. Singapore, 2005, pp. 63–78.
[41] M. B. Blaschko and C. H. Lampert, "Correlational spectral clustering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Anchorage, AL, USA, 2008, pp. 1–8.
[42] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proc. Int. Conf. Mach. Learn.*, Montreal, QC, Canada, 2009, pp. 129–136.
[43] Y. Wang, X. Lin, L. Wu, W. Zhang, Q. Zhang, and X. Huang, "Robust subspace clustering for multi-view data by exploiting correlation consensus," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3939–3949, Nov. 2015.
[44] X. Xue, F. Nie, S. Wang, X. Chang, B. Stantic, and M. Yao, "Multi-view correlated feature learning by uncovering shared component," in *Proc. Int. Joint Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 2810–2816.
[45] L. Zhang, Q. Zhang, B. Du, D. Tao, and J. You, "Robust manifold matrix factorization for joint clustering and feature extraction," in *Proc. Int. Joint Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 1662–1668.
[46] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *Proc. IEEE Int. Conf. Data Min.*, Shenzhen, China, 2014, pp. 110–119.
[47] W. 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.
[48] Z. Lin, R. Liu, and Z. Su, "Linearized alternating direction method with adaptive penalty for low-rank representation," in *Proc. Adv. Neural Inf. Process. Syst.*, Granada, Spain, 2011, pp. 612–620.
[49] J. Huang, F. Nie, and H. Huang, "Spectral rotation versus *k*-means in spectral clustering," in *Proc. Int. Joint Conf. Artif. Intell.*, Bellevue, WA, USA, 2013, pp. 431–437.
[50] Z. Lin, M. Chen, and Y. Ma, "The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices," *arXiv preprint arXiv:1009.5055*, 2010.
[51] 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, 2010.
[52] E. Kim, M. Lee, and S. Oh, "Elastic-net regularization of singular values for robust subspace learning," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Boston, MA, USA, 2015, pp. 915–923.
[53] X. Fang, S. Teng, Z. Lai, Z. He, S. Xie, and W. K. Wong, "Robust latent subspace learning for image classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 6, pp. 2502–2515, Jun. 2018.
[54] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.
[55] J. Xu, J. Han, and F. Nie, "Discriminatively embedded *k*-means for multi-view clustering," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 5356–5364.
[56] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proc. Int. Joint Conf. Artif. Intell.*, San Francisco, CA, USA, 2017, pp. 2408–2414.
[57] I. T. Jolliffe, "Principal component analysis," in *International Encyclopedia of Statistical Science*. New York, NY, USA: Springer, 2011, pp. 1094–1096.
[58] M. D. Collins, J. Liu, J. Xu, L. Mukherjee, and V. Singh, "Spectral clustering with a convex regularizer on millions of images," in *Proc. Eur. Conf. Comput. Vis.*, Zürich, Switzerland, 2014, pp. 282–298.
[59] V. R. De Sa, "Spectral clustering with two views," in *Proc. Int. Conf. Mach. Learn. Workshop Learn. Multiple Views*, Lille, France, 2005, pp. 20–27.
[60] L. V. D. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.
[61] Y. Yang, H. T. Shen, F. Nie, R. Ji, and X. Zhou, "Nonnegative spectral clustering with discriminative regularization," in *Proc. Int. Joint Conf. Artif. Intell.*, San Francisco, CA, USA, 2011, pp. 2–4.
[62] J. Wang, H. Liu, X. Qian, Y. Jiang, Z. Deng, and S. Wang, "Cascaded hidden space feature mapping, fuzzy clustering, and nonlinear switching regression on large datasets," *IEEE Trans. Fuzzy Syst.*, vol. 26, no. 2, pp. 640–655, Apr. 2018.
[63] C. Zhang, Y. Liu, and H. Fu, "AE^2-NETs: Autoencoder in autoencoder networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2019, pp. 1–9.
[64] L. Liu, F. Nie, A. Wiliem, Z. Li, T. Zhang, and B. C. Lovell, "Multi-modal joint clustering with application for unsupervised attribute discovery," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4345–4356, Sep. 2018.
[65] Y. Zhang *et al.*, "Strength and similarity guided group-level brain functional network construction for MCI diagnosis," *Pattern Recognit.*, vol. 88, pp. 421–430, Apr. 2019.
[66] Y. Zhang, C. S. Nam, G. Zhou, J. Jin, X. Wang, and A. Cichocki, "Temporally constrained sparse group spatial patterns for motor imagery BCI," *IEEE Trans. Cybern*, to be published.
[67] C. Zhang, Z. Yu, H. Fu, P. Zhu, L. Chen, and Q. Hu, "Hybrid noise-oriented multilabel learning," *IEEE Trans. Cybern.*, to be published.



Tao Zhou (M'17) received the M.S. degree in computer application technology from Jiangnan University, Wuxi, China, in 2012 and the Ph.D. degree in pattern recognition and intelligent system from the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China, in 2016.

He is currently a Research Scientist with the Inception Institute of Artificial Intelligence, Abu Dhabi, UAE. His current research interests include machine learning, computer vision, and medical image analysis.

Dr. Zhou is an Associate Editor of IEEE ACCESS.



Changqing Zhang (M'18) received the B.S. and M.S. degrees from the College of Computer, Sichuan University, Chengdu, China, in 2005 and 2008, respectively, and the Ph.D. degree in computer science from Tianjin University, Tianjin, China, in 2016.

He is an Assistant Professor with the College of Intelligence and Computing, Tianjin University. He has been a Post-Doctoral Research Fellow with the Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. His current research interests include machine learning, computer vision, and medical image analysis.

Xi Peng received the Ph.D. degree from Sichuan University, Chengdu, China, in 2013.

He is currently a Research Professor with the College of Computer Science, Sichuan University. From 2014 to 2017, he was a Research Scientist and a Co-PI with the Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore. His current research interests include machine learning, especially unsupervised representation learning, subspace learning, clustering, differentiable programming and their applications in image processing and computer vision. He has authored over 40 articles in the above areas.

Dr. Peng has served as an/a Associate/Guest Editor for several journals, such as the IEEE TRANSACTIONS ON NEURAL NETWORK AND LEARNING SYSTEMS; the Organization Co-Chair for a regular tutorial at the European Conference on Computer Vision (ECCV'16) and for a special session for IEEE Visual Communications and Image Processing (VCIP'17); the Session Chair for AAAI Conference on Artificial Intelligence (AAAI'16) and International Joint Conference on Artificial Intelligence (IJCAI'18); a Senior Program Committee Member for International Joint Conference on Artificial Intelligence (IJCAI'16); and a program committee member and a reviewer for over 30 international conferences and international journals. He has been invited to give a workshop talk at Vision and Learning Seminar (VALSE'18) and Tutorial talk at ECCV'16.

Harish Bhaskar received the M.Sc. degree in autonomous systems from the University of Exeter, Exeter, U.K., and the Ph.D. degree in computer science from Loughborough University, Loughborough, U.K.

He was a Chief Engineer with Samsung Electronics, Noida, India, and an Assistant Professor with Khalifa University, Abu Dhabi, UAE. He is the President of a consulting startup Zero One Infinity Consulting Services Ltd., Mississauga, ON, Canada. His current research interests include computer vision, image processing, visual data mining, medical imaging, quantum information, and machine vision.

Jie Yang received the Ph.D. degree from the Department of Computer Science, Hamburg University, Hamburg, Germany, in 1994.

He is currently a Professor with the Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, Shanghai, China. He has led many research projects, had one book published in Germany, and authored over 200 journal papers. His current research interests include object detection and recognition, data fusion and data mining, and medical image processing.