

Multiview Subspace Clustering With Grouping Effect

Man-Sheng Chen^{ID}, Ling Huang^{ID}, Member, IEEE, Chang-Dong Wang^{ID}, Member, IEEE,
Dong Huang^{ID}, Member, IEEE, and Philip S. Yu^{ID}, Life Fellow, IEEE

Abstract—Multiview subspace clustering (MVSC) is a recently emerging technique that aims to discover the underlying subspace in multiview data and thereby cluster the data based on the learned subspace. Though quite a few MVSC methods have been proposed in recent years, most of them cannot explicitly preserve the locality in the learned subspaces and also neglect the subspacewise grouping effect, which restricts their ability of multiview subspace learning. To address this, in this article, we propose a novel MVSC with grouping effect (MvSCGE) approach. Particularly, our approach simultaneously learns the multiple subspace representations for multiple views with smooth regularization, and then exploits the subspacewise grouping effect in these learned subspaces by means of a unified optimization framework. Meanwhile, the proposed approach is able to ensure the cross-view consistency and learn a consistent cluster indicator matrix for the final clustering results. Extensive experiments on several benchmark datasets have been conducted to validate the superiority of the proposed approach.

Index Terms—Cross-view consistency, multiview clustering, subspace clustering, subspacewise grouping effect.

I. INTRODUCTION

MULTIVIEW data collected from different fields or obtained from multiple feature extractors are extremely universal in many real-world applications [1]–[10]. For instance, news can be reported by different languages, such as Chinese, English, and Spanish. A person can be described by face, fingerprint, iris, and so on. The webpage can be

Manuscript received 26 July 2020; revised 20 October 2020; accepted 25 October 2020. Date of publication 7 December 2020; date of current version 19 July 2022. This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0809700; in part by NSFC under Grant 61876193 and Grant 61976097; in part by the Guangdong Natural Science Funds for Distinguished Young Scholar under Grant 2016A030306014; and in part by NSF under Grant III-1763325, Grant III-1909323, and Grant SaTC-1930941. This article was recommended by Associate Editor S. Ventura. (Corresponding author: Chang-Dong Wang.)

Man-Sheng Chen and Chang-Dong Wang are with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou 511400, China, also with the Guangdong Province Key Laboratory of Computational Science, Sun Yat-sen University, Guangzhou 511400, China, and also with the Key Laboratory of Machine Intelligence and Advanced Computing, Ministry of Education, Sun Yat-sen University, Guangzhou 511400, China (e-mail: chenmsh27@mail2.sysu.edu.cn; changdongwang@hotmail.com).

Ling Huang and Dong Huang are with the College of Mathematics and Informatics, South China Agricultural University, Guangzhou 510635, China (e-mail: huanglinghl@hotmail.com; huangdonghere@gmail.com).

Philip S. Yu is with the Department of Computer Science, University of Illinois at Chicago, Chicago, IL 60607 USA, and also with the Institute for Data Science, Tsinghua University, Beijing 10085, China (e-mail: psyu@cs.uchicago.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/TCYB.2020.3035043>.

Digital Object Identifier 10.1109/TCYB.2020.3035043

represented by multiview features based on text, image, and video. Different views or feature subsets comprehensively describe different aspects of the data [11]–[14]. In general, it is infeasible to just consider single view rather than multiple views [15]–[20]. Thus, how to integrate multiple views or the underlying subspace structures is considerably crucial. For traditional multiview clustering, a naive idea for integrating multiview data is to concatenate all multiview features into a new feature vector, and then perform a single view clustering method on the new feature vector to obtain the final clustering results. However, such a naive idea ignores the correlation between different views that will result in the suboptimal clustering results. Therefore, it still remains a challenging problem of how to well integrate the multiview features to obtain better clustering results.

In the past few years, many methods have been proposed for multiview learning in different areas [1], [21]–[25], and the existing methods can be classified into three main categories: 1) co-training-based methods; 2) multiple kernel learning-based methods; and 3) subspace learning-based methods. Among them, the co-training-based methods aim to maximize the agreement on two different views of the data by training alternately [26]–[28]. The multiple kernel learning-based methods regard each kernel as a view, and linearly or nonlinearly combine those kernels to improve their performance [29]–[31]. Assuming that multiple views are generated from a latent subspace, the subspace learning-based methods aim to find this latent subspace shared by different views [32], [33].

Many multiview clustering methods have also been developed to improve the clustering performance by capturing the correlation between multiple views, such as [27] and [28]. For high-dimensional multiview data, studies on multiview subspace clustering (MVSC) have attracted an increasing amount of attention recently [34]–[36]. For instance, Zhang *et al.* [34] focused on searching for such an underlying latent subspace of multiple views. In [35], each subspace representation corresponding to different views is learned and exploited to obtain the final clustering results. Despite significant success, most of these multiview subspace learning methods lack the ability to directly explore the multiple graph structures among data points in the learned subspaces, and also neglect the subspacewise grouping effect in these subspaces, which restricts their ability to well capture the rich and diverse subspace information in multiview high-dimensional data.

To deal with multiple views, it is vital to explore the multiple graph structures to acquire the consistent and robust clustering results [37]. Recently, some efforts have been devoted to exploiting multiple graph structures. For instance, Zhan *et al.* introduced a graph learning method to address multiple views [37]. Furthermore, a multiview consensus graph clustering method was proposed to explore the consensus graph between different views with constraint [38]. Though these methods [37], [38] can mine the information of multiple graphs from multiple views, they are still not able to explore multiple subspaces in the multiview data and meanwhile exploit subspace learning and graph fusion in a unified optimization framework.

Aiming to address the aforementioned problem, in this article, we propose a novel MVSC with grouping effect (MvSCGE) method, in which the locality in each learned subspace is explicitly preserved, and the cross-view consistency in multiview learning is emphasized. First, each subspace representation for different views can be acquired by minimizing the reconstruction error of subspace clustering with a smooth regularization. Then, the subspacewise grouping effect is exploited in these learned subspaces by means of a unified optimization framework. In order to ensure the final consistent clustering results, the common cluster structure will be obtained by learning each subspace representation, respectively. Thus, the proposed method jointly considers the subspace locality and the cross-view consistency.

The main contributions of this article are summarized as follows.

- 1) We propose a novel subspace learning method called MvSCGE, which simultaneously learns a more stable subspace representation for each view and a consistent cluster indicator matrix in a unified optimization framework.
- 2) Our framework explicitly exploits the subspacewise grouping effect in the learned subspaces where the locality is ensured. Meanwhile, an alternate minimizing strategy is developed to deal with our optimization problem.
- 3) Extensive experiments conducted on both image and document datasets have demonstrated the superiority of the proposed method.

The remainder of this article is organized as follows. In Section II, some related work is reviewed. The proposed MVSC approach with grouping effect is described in Section III in which the optimization algorithm is proposed. In Section IV, the experimental results are reported where five benchmark datasets are used and 13 state-of-the-art methods are compared. Finally, this article is concluded in Section V.

II. RELATED WORK

In the past few years, many multiview clustering methods have been proposed most of which are based on the graph representation to explore the multiview features [33], [39]–[42]. For instance, De Sa proposed to construct a bipartite graph to connect the features in two views [39], and the final clustering results could be obtained by using the standard spectral

clustering (SC) algorithm. The work in [27] co-regularized the clustering hypotheses so that different graphs could agree with each other in the SC framework. Kumar and Daumé [28] proposed to seek for the clusters which agree across the multiple views. Xia *et al.* [33] proposed a Markov chain method whose input is a shared low-rank transition probability matrix associated with multiple views. Without the postprocessing step, Zhan *et al.* [37] developed a graph learning method for multiview clustering in which initial graph for each view is learned first with a rank constraint, and then a global graph is learned by integrating these initial graphs with an optimization procedure. Notice that all the multiview clustering methods mentioned above emphasize the importance of cross-view consistency.

In multiview learning, either the consistency principle or the complementary principle is mainly exploited to ensure the learning performance [1]. To be specific, the consistency principle aims to maximize the agreement on multiple different views. As for the complementary principle, it states that each view of the data may consist of some information the other views do not have. Thus, we can employ multiple views to describe the data comprehensively. In order to achieve better performance in multiview learning, we should take into account both the consistency and the complementary principles to take full advantage of multiple views.

Recently, many subspace clustering methods based on the self-representation property have been proposed to explore the relationships between data points, e.g., sparse subspace clustering (SSC) [43], low-rank representation segmentation (LRR) [44], [45], and smooth representation clustering (SMR) [46]. In particular, the work in [43] was to find a sparse representation corresponding to the data points from the same subspace. In [46], a subspace representation with the subspacewise grouping effect was learned. However, these methods are just applicable for the single view setting. To solve the problem in the case of multiple views, some MVSC methods have been proposed, in which each sample is described with information from multiple feature extractors. For instance, Gao *et al.* [35] proposed an MVSC framework to learn the subspace representation corresponding to each view, and the final common clustering results were obtained by using these subspace representation. Zhang *et al.* [34] aimed to find the underlying latent representation of multiple views, and the data reconstruction was performed on the learned latent representation rather than the original dataset. Furthermore, based on the neural networks, Zhang *et al.* [47] provided a more generalized framework to settle general correlations between different samples. The work in [48] proposed to simultaneously explore the complementary representations and consistent indicator among multiple views. Based on the intact space learning, Wang *et al.* [49] proposed to compute an informative intactness-aware affinity. Tang *et al.* [50] attempted to learn a shared affinity presentation for MVSC by simultaneously considering the diversity regularization and a rank constraint. The high-order correlations among multiple subspace representations were explored by considering all these subspace representations as a tensor in [51]. The work in [52] proposed to explore a common similarity matrix from multiple

views for kernel MVSC in one step. Li *et al.* [53] proposed to discover a latent common space from multiple views by mapping the original features into a kernel space. The work in [54] attempted to acquire consistent and diverse characteristics from multiple views by learning general and specific representations. Zhou *et al.* [55] aimed to jointly capture correlations between the common information from different views, and discover diverse properties for different views by using diverse information. Luo *et al.* [56] proposed to simultaneously learn consistency and specificity in subspace representation for MVSC. Xing *et al.* [57] explored dense correlations between different samples in the same subspace and deployed the correntropy-induced metric (CIM) to depict the noise from each view. The work in [36] learned each subspace representation by extending the standard subspace clustering into the multiview case, and employed the Hilbert Schmidt independence criterion (HSIC) to explore the diversity of multiple subspace representations.

As analyzed in [46], the subspacewise grouping effect is proved to be important for subspace clustering. However, most of the MVSC methods do not explicitly take the subspacewise grouping effect into consideration. Thus, in the proposed model, not only the subspacewise grouping effect is considered as a smooth regularization term but also the consistency principle is regarded as a cross-view consistency regularization term.

III. MULTIVIEW SUBSPACE CLUSTERING WITH GROUPING EFFECT

In this section, we will describe in detail the proposed MvSCGE model. First, the proposed model will be described in Section III-A, where the standard subspace clustering method will also be revisited. Then, an alternate minimizing strategy will be introduced to solve the proposed optimization problem, and the entire algorithm will be summarized in Section III-B. Finally, the convergence analysis and the time complexity analysis are provided in Sections III-C and III-D, respectively.

A. Proposed Model

According to the assumption of subspace clustering [43], [58], [59], a dataset does not distribute uniformly in the entire space, instead, it usually lies in an underlying low-dimensional subspace. Therefore, we can use a low-dimensional subspace to represent the data points. After acquiring the subspace structure of the original dataset, clustering can be performed based on the subspace instead of the entire space.

Suppose there are n data points $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, in which each column is a sample vector and the dimension of the feature space is d . The standard subspace clustering method is to acquire the subspace structure of the original dataset and subsequently perform SC on such a subspace representation of the data. In order to cluster the data, a similarity matrix encoding the pairwise similarities between data points is needed to be computed, which is constructed from the subspace representation. To be specific, the subspace clustering generally employs the self-expression property of the dataset

to represent itself [43], which can be written as

$$X = XZ + E \quad (1)$$

where $Z = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n] \in \mathbb{R}^{n \times n}$ is the subspace representation matrix with each \mathbf{z}_i being the new representation of the original data point \mathbf{x}_i . $E \in \mathbb{R}^{d \times n}$ is the error matrix.

To minimize the error matrix, the problem in (1) can be converted into the following optimization problem [46]:

$$\min_Z f(Z) = \|X - XZ\|_F^2 + \lambda \Omega(Z) \quad (2)$$

where λ is the tradeoff parameter, and $\Omega(Z)$ means the certain regularization term. After solving the optimization problem above, the new representation \mathbf{z}_i of the data point \mathbf{x}_i will be obtained. Therefore, the subspace structure Z for the original dataset can be acquired finally.

With the self-representation matrix Z , we can construct the similarity matrix as [46]

$$W = \frac{|Z| + |Z^T|}{2} \quad (3)$$

where $|\cdot|$ is the absolute operator. After that, to obtain the final clustering results, the SC algorithm can be performed on such a similarity matrix

$$\min_F \text{Tr}(F^T L F), \quad \text{s.t. } F^T F = I \quad (4)$$

where F is the cluster indicator matrix, and $L = D - W$ in which D is a diagonal matrix with its i th diagonal element being $d_{ii} = \sum_j w_{ij}$.

The subspace clustering methods have shown their powerful performance in many fields, but most of them are just applicable for single view setting. Therefore, some efforts have been made recently in further improving the clustering performance by extending the subspace clustering to the multiview case, such as MVSC [35] and latent MVSC (LMSC) [34]. The main idea is to learn the corresponding subspace representation, and then perform SC algorithm on the subspace representations. However, there are two non-negligible drawbacks.

The first drawback is that they do not explicitly enforce to preserve the locality in the learned new subspace. If there are arbitrarily two data points \mathbf{x}_i and \mathbf{x}_j that are close to each other, the corresponding subspace representations \mathbf{z}_i and \mathbf{z}_j should also be explicitly enforced to be close. Nevertheless, most of the MVSC methods do not take the problem above into consideration, which may lead to the final misclassification in multiview clustering, just as illustrated in Fig. 1. To preserve the locality in the learned new subspace, in this article, we will exploit the subspacewise grouping effect in the learned subspaces by means of a unified optimization framework.

The definition of the subspacewise grouping effect is introduced as follows [46].

Definition 1 (Subspacewise Grouping Effect): Given a set of d -dimensional data points $X = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n] \in \mathbb{R}^{d \times n}$, a self-representation matrix $Z = [\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n] \in \mathbb{R}^{n \times n}$ has the grouping effect if $\|\mathbf{x}_i - \mathbf{x}_j\|_2 \rightarrow 0 \Rightarrow \|\mathbf{z}_i - \mathbf{z}_j\|_2 \rightarrow 0 \forall i \neq j$.

To explicitly enforce the subspace representation to satisfy the subspacewise grouping effect, the graph regularization technique should be exploited. Specifically, we can enforce the

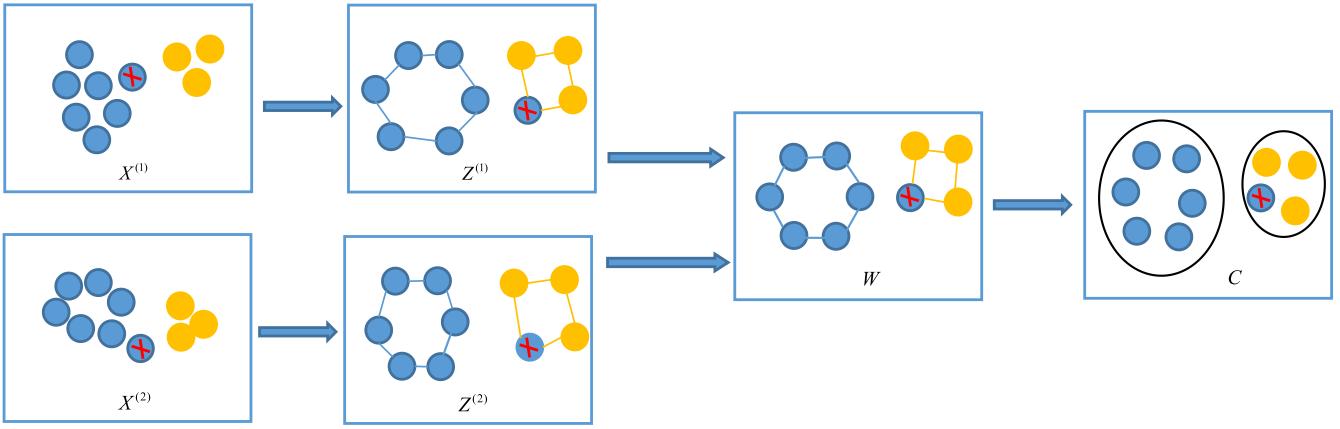


Fig. 1. Illustration of subspacewise grouping effect. $X^{(1)}$ and $X^{(2)}$, respectively, represent the data representations in the first view and the second view. Correspondingly, $Z^{(1)}$ and $Z^{(2)}$ are the learned subspace representations. W is constructed by the acquired subspace representations. The final clustering results are C . In the original data space, all the blue data points are close to each other, and all the orange data points are close to each other. Without preserving the locality (the subspacewise grouping effect) in the learned subspace, the data points which are close in the original space may be not close in the learned subspace, e.g., the point with red cross which should be close to those blue ones which, however, turns out to be close to orange ones. In consequence, there is an error in the final clustering results.

subspacewise grouping effect explicitly by the affinity of data points, which can be defined as follows:

$$\begin{aligned}\Omega(Z^{(v)}) &= \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n s_{ij}^{(v)} \|z_i^{(v)} - z_j^{(v)}\|_2^2 \\ &= \text{tr}(Z^{(v)} L^{(v)} Z^{(v)T})\end{aligned}\quad (5)$$

where $\text{tr}(\cdot)$ means the trace of the matrix. $S^{(v)} = (s_{ij}^{(v)})$ is the affinity matrix that measures the spatial closeness of the data points, and is generated as the inner product of the original data matrix in the v th view. $L^{(v)} = \bar{D}^{(v)} - S^{(v)}$ is the graph Laplacian matrix of the v th view where $\bar{D}^{(v)}$ denotes the diagonal matrix with its diagonal element being $\bar{d}_{ii}^{(v)} = \sum_j s_{ij}^{(v)}$. It is well known that there are many construction strategies for the affinity matrix. One typical method is the k -nearest-neighborhood, in which the subspacewise grouping effect can contribute to a well balanced affinity matrix. In order to exploit the merit of subspacewise grouping effect, inner product is utilized to measure the affinity in this article, which is easier to be achieved and performs well in practice.

With the help of the subspacewise grouping effect, the learned subspace representation will be more stable. To be specific, the reconstruction error in the subspace clustering methods based on the self-representation idea can be regarded as a first-order energy that encodes the whole subspace structure of the data. The subspacewise grouping effect of a representation matrix shows that $\Omega(Z^{(v)})$ consists of a second-order energy that can penalize the discontinuities in the subspace representation. With this second-order energy, the new representation can be more stable. Meanwhile, spatially close data points may be beneficial for each other to prevent over-fitting in the samples reconstruction. Therefore, the subspacewise grouping effect should be explicitly enforced in the data self-representation subspace clustering model.

As stated above, the subspacewise grouping effect is extremely significant for subspace clustering, and should be explicitly taken into consideration in subspace clustering. However, most of the existing MVSC methods do not explicitly take it into account, which may lead to the suboptimal results eventually.

The second drawback is that the existing MVSC methods usually ignore the importance of cross-view consistency. Thus, integrating the consistency of the multiview feature in subspace clustering is necessary. Intuitively, there is a straightforward strategy to concatenate all the multiview features together, and then perform clustering algorithm on the concatenated feature vector. Nevertheless, this straightforward method does not consider the correlation between views, and it treats all views equally no matter which view is more informative. Therefore, the optimal solution cannot be obtained. Furthermore, in the previous computer vision research, subspace representation can be learned simultaneously in each view, and then the unified subspace representation is able to be learned from the subspace representation of each view [35]. After that, we can perform SC based on the unified subspace representation. As a matter of fact, this framework does not work well for MVSC, which can be demonstrated in Fig. 2 due to the dramatic difference in magnitude of element values in $Z^{(v)}$. Thus, it is not trivial to integrate the subspace clustering using multiview features.

To address the problems mentioned above, we propose a novel MvSCGE algorithm in this article. First, based on the self-expression property, multiview subspace learning is performed in each view. Then, the subspacewise grouping effect is explicitly enforced in the learned subspaces by means of a unified optimization framework. In particular, we propose to integrate the clustering results instead of the unified subspace representation by using different subspace representation $Z^{(v)}$, which is directly comprised in the objective function rather than separately using the SC as the postprocessing step.

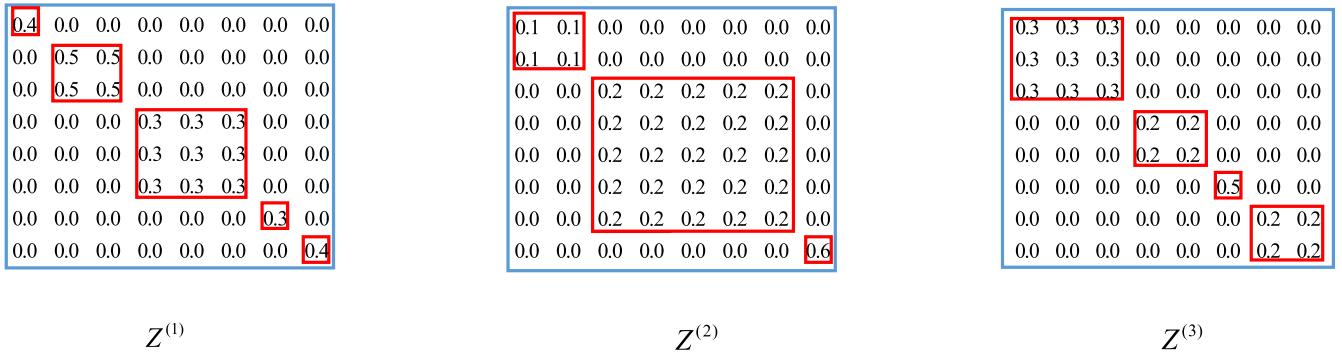


Fig. 2. Illustration of subspace representation. $Z^{(1)}$, $Z^{(2)}$, and $Z^{(3)}$ are the learned subspace representation for three views, respectively. The magnitude of element values in $Z^{(v)}$ is different despite of their similar block-diagonal structure. For instance, the first block in each subspace representation (i.e., [0.4], [0.1, 0.1; 0.1, 0.1], [0.3, 0.3, 0.3; 0.3, 0.3, 0.3; 0.3, 0.3, 0.3]) varies in magnitude of element values from one view to another.

Consequently, the novel MvSCGE method is proposed, which can be reformulated as

$$\begin{aligned}
 \min_{Z^{(1)}, Z^{(2)}, \dots, Z^{(V)}, F} & \mathcal{O}(Z^{(1)}, Z^{(2)}, \dots, Z^{(V)}, F) \\
 &= \underbrace{\sum_{v=1}^V \|X^{(v)} - X^{(v)}Z^{(v)}\|_F^2}_{\text{reconstruction error}} \\
 &\quad + \underbrace{\lambda \sum_{v=1}^V \text{tr}(Z^{(v)}L^{(v)}Z^{(v)T})}_{\text{smooth regularization}} \\
 &\quad + \underbrace{\beta \sum_{v=1}^V \text{tr}(F^T(D^{(v)} - W^{(v)})F)}_{\text{cross-view consistency regularization}} \\
 \text{s.t. } & F^T F = I, W^{(v)} = Z^{(v)T}Z^{(v)}, D_{ii}^{(v)} = \sum_{j=1}^n w_{ij}^{(v)}
 \end{aligned} \tag{6}$$

where λ and β are the tradeoff factors, respectively, corresponding to the smooth regularization and cross-view consistency regularization term, $W^{(v)}$ is the subsequent affinity matrix based on the learned subspace representation $Z^{(v)}$, $D^{(v)}$ is the corresponding diagonal matrix, and $F \in \mathbb{R}^{n \times c}$ is the cluster indicator matrix with c being the number of clusters. Assuming that data points are drawn from different subspaces, the first term of the objective function makes sure the relationships are constructed in the same subspaces. The second term is to enforce each learned subspace representation to satisfy the subspacewise grouping effect. In the last term, the common cluster indicator matrix F which fully considers different subspace structures is utilized for all the views. Therefore, the cluster structure will be consistent for all the views, which means that the corresponding data points in different views will be categorized into the same cluster. In conclusion, the proposed method explicitly enforce the subspacewise grouping effect in subspace learning, and the common cluster indicator matrix is learned by using different subspace representations, which fully considers the structure of different views.

B. Optimization

In this section, an alternate minimizing strategy is proposed to solve the problem in (6), where in one iteration one variable is updated when fixing the others.

1) *Computation of Cluster Indicator Matrix F*: With all $Z^{(v)}$ fixed, the cluster indicator matrix F is updated. Thus, the problem in (6) can be reduced to the following problem with respect to F :

$$\begin{aligned}
 \min_F & \sum_v \text{tr}(F^T(D^{(v)} - Z^{(v)T}Z^{(v)})F) \\
 \text{s.t. } & F^T F = I.
 \end{aligned} \tag{7}$$

Specifically, it can be further rewritten as

$$\begin{aligned}
 \min_F & \text{tr}(F^T M F) \\
 \text{s.t. } & F^T F = I
 \end{aligned} \tag{8}$$

where $M = \sum_v (D^{(v)} - Z^{(v)T}Z^{(v)})$ is the Laplacian matrix. The problem in (8) can be solved by calculating the eigenvectors corresponding to the smallest c eigenvalues of the Laplacian matrix M .

2) *Computation of Subspace Representation Matrix $Z^{(v)}$* : With F fixed, the subspace representation matrix $Z^{(v)}$ can be updated. When fixing F , it is equivalent to solving

$$\begin{aligned}
 \min_{Z^{(v)}} & \sum_v \|X^{(v)} - X^{(v)}Z^{(v)}\|_F^2 + \lambda \sum_v \text{tr}(Z^{(v)}L^{(v)}Z^{(v)T}) \\
 &\quad + \beta \sum_v \text{tr}(F^T(D^{(v)} - W^{(v)})F) \\
 \text{s.t. } & W^{(v)} = Z^{(v)T}Z^{(v)}, D_{ii}^{(v)} = \sum_{j=1}^n w_{ij}^{(v)}.
 \end{aligned} \tag{9}$$

To be specific, $Z^{(v)}$ can be updated, respectively, for each view v . Thus, the problem in (9) can be written as

$$\begin{aligned}
 \min_{Z^{(v)}} & \|X^{(v)} - X^{(v)}Z^{(v)}\|_F^2 + \lambda \text{tr}(Z^{(v)}L^{(v)}Z^{(v)T}) \\
 &\quad + \beta \text{tr}(F^T(D^{(v)} - W^{(v)})F) \\
 \text{s.t. } & W^{(v)} = Z^{(v)T}Z^{(v)}, D_{ii}^{(v)} = \sum_{j=1}^n w_{ij}^{(v)}.
 \end{aligned} \tag{10}$$

The problem in (10) needs to be reformulated so as to make it easier to be optimized, and we have the following theorem.

Theorem 1: For Laplacian matrix $L \in \mathbb{R}^{n \times n}$ and the matrix $F \in \mathbb{R}^{n \times c}$, we will have the following equation:

$$\text{tr}(F^T L F) = \frac{1}{2} \text{tr}(WP) \quad (11)$$

where $P_{ij} = \|\mathbf{f}_i^T - \mathbf{f}_j^T\|_2^2$ with \mathbf{f}_i^T being the i th row of matrix F .

The above theorem is proved in [60]. According to Theorem 1, the problem in (10) can be formulated as

$$\begin{aligned} \min_{Z^{(v)}} & \|X^{(v)} - X^{(v)} Z^{(v)}\|_F^2 + \lambda \text{tr}(Z^{(v)} L^{(v)} Z^{(v)T}) \\ & + \frac{1}{2} \beta \text{tr}(Z^{(v)T} Z^{(v)} P) \\ \text{s.t. } & P_{ij} = \|\mathbf{f}_i^T - \mathbf{f}_j^T\|_2^2. \end{aligned} \quad (12)$$

The above (12) is equivalent to solving

$$\begin{aligned} \min_{Z^{(v)}} & \|X^{(v)} - X^{(v)} Z^{(v)}\|_F^2 + \lambda \text{tr}(Z^{(v)} L^{(v)} Z^{(v)T}) \\ & + \frac{1}{2} \beta \text{tr}(Z^{(v)T} P Z^{(v)}) \\ \text{s.t. } & P_{ij} = \|\mathbf{f}_i^T - \mathbf{f}_j^T\|_2^2. \end{aligned} \quad (13)$$

Actually, the problem in (13) is a smooth convex program. The optimal solution $Z^{(v)*}$ can be acquired by differentiating (13) with respect to $Z^{(v)}$ and setting it to zero, which satisfies

$$X^{(v)T} X^{(v)} Z^{(v)*} + Z^{(v)*} \left(\lambda L^{(v)} + \frac{1}{2} \beta P \right) = X^{(v)T} X^{(v)}. \quad (14)$$

Equation (14) is a standard Sylvester equation that has a unique solution and can be solved by the Bartels–Stewart algorithm [61]. Thus, the similar method to the smooth subspace clustering can be utilized to optimize our subspace representation $Z^{(v)}$.

Before iteration, the variable $Z^{(v)}$ is initialized by solving the following problem:

$$\min_{Z^{(v)}} \|X^{(v)} - X^{(v)} Z^{(v)}\|_F^2 + \lambda \text{tr}(Z^{(v)} L^{(v)} Z^{(v)T}). \quad (15)$$

Then, by using the alternate minimizing algorithm, the variables F and $Z^{(v)}$ can be updated iteratively in an interplay manner until convergence (i.e., the difference between the current objective value and the previous one is less than 10^{-6}). After obtaining F , the classical k -means clustering algorithm is performed on F to obtain the final clustering results \mathcal{C} . For clarity, the overall algorithm of the proposed MvSCGE method is summarized in Algorithm 1.

C. Convergence Analysis

In this section, the convergence of Algorithm 1 is analyzed as follows.

Theorem 2: The objective function in (6) is guaranteed to converge by Algorithm 1.

Proof: With the initialization of each $Z^{(v)}$, the objective function in (6) can be optimized for each iteration. Let $[\cdot]_t$

Algorithm 1 MVSCGE

Input: Multiview data $X = \{X^{(1)}, \dots, X^{(V)}\}$, cluster number c , parameters λ and β .

```

1: for  $\forall v = 1, \dots, V$  do
2:   Initialize  $Z^{(v)}$  by solving the objective function (15).
3: end for
4: while not converged do
5:   Update  $F$  according to (8).
6:   for  $\forall v = 1, \dots, V$  do
7:     Obtain  $Z^{(v)}$  by solving the objective function (10).
8:   end for
9: end while
10: Perform  $k$ -means algorithm on  $F$ .

```

Output: Clustering result \mathcal{C} .

denote the updated value in the t th iteration. The proof is as follows.

The first step is to fix the others and update F . According to [62], the Hessian matrix of the Lagrangian function of (8) is positive semidefinite. Thus, the objective function in (8) is a convex problem, and we have

$$\begin{aligned} & \mathcal{O}\left([F]_{t+1}; [Z^{(1)}]_t, [Z^{(2)}]_t, \dots, [Z^{(V)}]_t\right) \\ & \leq \mathcal{O}\left([F]_t; [Z^{(1)}]_t, [Z^{(2)}]_t, \dots, [Z^{(V)}]_t\right) \end{aligned} \quad (16)$$

where F is computed by $[F]_{t+1} = \arg \min_{F^T F = I} \text{tr}(F^T M F)$ when fixing the other variables.

The second step is to update $Z^{(v)}$ $\forall v$ while fixing the others. Due to the standard Sylvester equation ((14)), the unique solution can be obtained. For notation, \mathcal{F} and $\bar{\mathcal{F}}$ are, respectively, corresponding to the objective function of the v th view and all the other views in (10). Given the updated value $[Z^{(v)}]_t$ in the t th iteration, and then we have

$$\begin{aligned} & \mathcal{F}\left([Z^{(v)}]_t; [F]_t, [Z^{(1)}]_t, \dots, [Z^{(v-1)}]_t, [Z^{(v+1)}]_t, \dots, \right) \\ & \geq \mathcal{F}\left([Z^{(v)}]_{t+1}; [F]_t, [Z^{(1)}]_t, \dots, [Z^{(v-1)}]_t \right. \\ & \quad \left. [Z^{(v+1)}]_t, \dots, \right). \end{aligned} \quad (17)$$

Specifically, the objective function in (10) can be decomposed into two parts, that is, the following \mathcal{F} and $\bar{\mathcal{F}}$:

$$\begin{aligned} & \mathcal{O}\left([Z^{(v)}]_t; [F]_t, [Z^{(1)}]_t, \dots, [Z^{(v-1)}]_t, [Z^{(v+1)}]_t, \dots, \right) \\ & = \mathcal{F}\left([Z^{(v)}]_t; [F]_t, [Z^{(1)}]_t, \dots, [Z^{(v-1)}]_t, [Z^{(v+1)}]_t, \dots, \right) \\ & \quad + \bar{\mathcal{F}}\left([F]_t, [Z^{(1)}]_t, \dots, [Z^{(v-1)}]_t, [Z^{(v+1)}]_t, \dots, \right). \end{aligned} \quad (18)$$

Subsequently, we can make a difference of the objective function in (10) between the t th and $(t+1)$ th iteration. Then, we have

$$\begin{aligned} & \mathcal{O}\left([Z^{(v)}]_t; [F]_t, [Z^{(1)}]_t, \dots, [Z^{(v-1)}]_t, [Z^{(v+1)}]_t, \dots, \right) \\ & - \mathcal{O}\left([Z^{(v)}]_{t+1}; [F]_t, [Z^{(1)}]_t, \dots, [Z^{(v-1)}]_t \right. \\ & \quad \left. [Z^{(v+1)}]_t, \dots, \right) \end{aligned}$$

$$\begin{aligned}
&= \mathcal{F}\left(\left[Z^{(v)}\right]_t; [F]_t, \left[Z^{(1)}\right]_t, \dots, \left[Z^{(v-1)}\right]_t \right. \\
&\quad \left. \left[Z^{(v+1)}\right]_t, \dots\right) \\
&- \mathcal{F}\left(\left[Z^{(v)}\right]_{t+1}; [F]_t, \left[Z^{(1)}\right]_t, \dots, \left[Z^{(v-1)}\right]_t \right. \\
&\quad \left. \left[Z^{(v+1)}\right]_t, \dots\right) \geq 0. \tag{19}
\end{aligned}$$

Due to the same $\bar{\mathcal{F}}$ in the objective function of the t -th and $(t+1)$ -th iteration, it can be proved that the equality above holds.

Consequently, the overall objective function in (6) is non-increasing for each iteration. Thus, Theorem 2 can be proved. The overall objective function in (6) is guaranteed to converge by Algorithm 1. ■

D. Computational Complexity Analysis

In this section, the computational complexity analysis will be provided. In the overall algorithm of the proposed MvSCGE method, the first step is to initialize each $Z^{(v)}$ by the similar method as [46], which can be optimized into a standard Sylvester equation. For Sylvester equation, the classical Bartels–Stewart algorithm transforms the coefficient matrices into Schur forms by means of QR decomposition, and achieves the resulting triangular system by backsubstitution. Therefore, initializing each $Z^{(v)}$ takes $O(n^3)$. With the alternate minimizing strategy, the second step is to update F according to (8), which is an eigen-decomposition procedure. To be specific, the c eigenvectors corresponding to the smallest c eigenvalues of the Laplacian matrix M need to be computed. Thus, updating F takes $O(cn^2)$ with $c \ll n$. The third step is to update each $Z^{(v)}$ by solving the objective function in (10). In particular, (10) is also a standard Sylvester equation after optimization. Therefore, updating each $Z^{(v)}$ takes $O(n^3)$. Last but not least, the final step is to obtain the clustering results by k -means whose computational complexity is $O(nct_0)$ with t_0 being the number of iterations of k -means. So the computational complexity of the overall algorithm is $O(Vn^3 + (cn^2 + Vn^3)t + nct_0)$ in total with t being the number of iterations of alternate minimizing. Since $c \ll n$, $V \ll n$, $t_0 \ll n$, and $t \ll n$, the main computational complexity of the model in (6) depends on solving the Sylvester equation to initialize and acquire each subspace representation matrix $Z^{(v)}$.

IV. EXPERIMENTS

In this section, extensive experiments are conducted on several real-world datasets to validate the superiority of the proposed method. First, we will introduce in detail the datasets and evaluation measures used in our experiments. Then, the parameter analysis will be conducted. After that, the comparison results will be reported and analyzed. Finally, convergence analysis will be conducted.

A. Datasets and Evaluation Measures

In our experiments, five widely used benchmark datasets, namely, UCI digits, MSRCv1, BBCSport, ORL, and Notting-Hill are used.

- 1) *UCI Digits*: This dataset is a handwritten digits dataset obtained from UCI machine learning repository [33].

TABLE I
STATISTIC OF FIVE REAL-WORLD DATASETS

	UCI digits	MSRCv1	BBCSport	ORL	Notting-Hill
View1	240-D	24-D	3183-D	4096-D	2000-D
View2	76-D	512-D	3203-D	3304-D	3304-D
View3	6-D	256-D	-	6750-D	6750-D
View4	-	254-D	-	-	-
#Size	2000	210	544	400	550
#Class	10	7	5	40	5

The handwritten digits are, respectively, “0,” “1,” …, and “9,” which are taken as the ground-truth class labels. The dataset contains 2000 samples and each digit is one class consisting of 200 samples. In our experiment, the UCI digits dataset contains three views, which are the intensity-averaged feature (view 1), the Fourier-coefficient feature (view 2), and the morphological feature (view 3).

- 2) *MSRCv1*: The dataset is an image dataset [63], which consists of 210 objects belonging to seven classes. The seven classes are tree, building, airplane, cow, face, car, and bicycle, which are taken as the ground-truth class labels. In our experiment, four kinds of features, namely, CM feature (view 1), GIST feature (view 2), LBP feature (view 3), and GENT feature (view 4) are used to represent the images.
- 3) *BBCSport*: The dataset is a document dataset consisting of 544 documents from the BBC Sport website corresponding to the sports news in five topical areas between 2004 and 2005 [33]. The five topical areas are business, entertainment, politics, sport, and tech, which are taken as the ground-truth class labels. In our experiments, two views are used, whose dimensions are 3183 and 3203, respectively.
- 4) *ORL*: The dataset consists of 400 face images belonging to 40 distinct subjects, with each subject containing ten images. For each subject, images were taken at different times, lights, facial expression (open or closed eyes, and smiling or not smiling), and facial details (with glasses or not). With the subjects in an upright and frontal position, all of the images were taken against a dark homogeneous background. In our experiments, the ORL dataset contains three views, which are the intensity feature (view 1), the LBP feature (view 2), and the Gabor feature (view 3). The subject identity information of each image is taken as ground-truth label.
- 5) *Notting-Hill*: This dataset is a video face dataset obtained from the movie Notting-Hill [64]. The Notting-Hill dataset consists of 4660 facial images in which the faces of five main casts are selected. In our experiments, 110 facial images of each cast are sampled randomly. Three views are used, which are the intensity feature (view 1), the LBP feature (view 2), and the Gabor feature (view 3). The subject identity information of each image is taken as ground-truth label.

The statistic of these five real-world datasets is shown in Table I.

For evaluation measures, seven widely used evaluation measures, namely, ACC (accuracy), normalized mutual

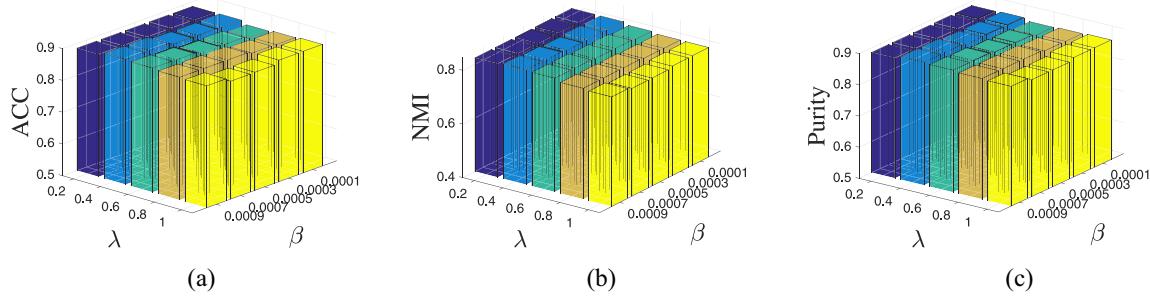


Fig. 3. Parameter analysis on λ and β on UCI digits. (a) ACC. (b) NMI. (c) Purity.

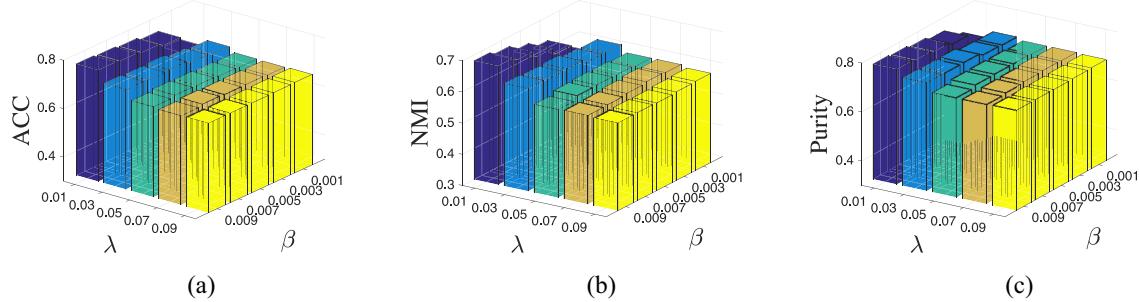


Fig. 4. Parameter analysis on λ and β on MSRCv1. (a) ACC. (b) NMI. (c) Purity.

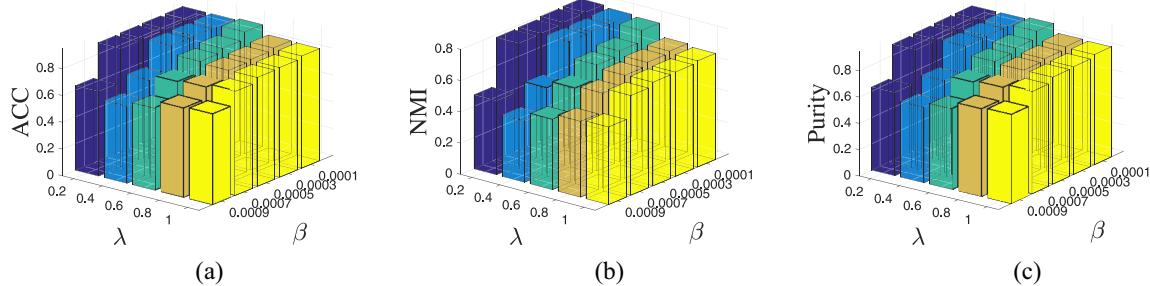


Fig. 5. Parameter analysis on λ and β on BBCSport. (a) ACC. (b) NMI. (c) Purity.

information (NMI), Purity, F -measure, Precision, Recall, and adjusted rand (AR) are used comprehensively to evaluate the performance. Specially, each measure penalizes or favors distinct property in the clustering results. Therefore, the results are reported by these diverse measures. For all the measures, higher values indicate better clustering performance.

B. Parameter Analysis

In this section, we conduct parameter analysis on the two parameters λ and β , which balance the first two regularization terms in our method. Due to the different properties of different datasets, we apply different ranges of λ and β between different datasets. As illustrated in Figs. 3–7, we show the influence of different parameter values in terms of ACC, NMI, and Purity, respectively, on the five benchmark datasets. It can be observed that for the benchmark datasets, our model is generally insensitive to λ and β over the corresponding ranges of values except for the BBCSport dataset. Therefore, our method is relatively robust to the parameters. In the meantime, best results in different datasets can be obtained from the figures. For instance, there are the best results for MSRCv1 dataset

when λ is 0.01 and β is 0.009. Accordingly, best results of the other datasets can also be acquired from the figures.

C. Comparison Experiments

1) *Compared Methods:* In our comparison experiments, 13 existing methods are compared, including SC [65], FeatConcat-SC, ConcatPCA-SC, co-regularized SC (Co-Reg) [27], co-training multiview clustering (Co-Training) [28], SC with two views (Min-Disagreement) [39], robust multiview SC (RMS) [33], MVSC [35], LMSC [34], graph learning for multiview clustering (MVGL) [37], naive MVSC (NaMSC) [36], diversity-induced MVSC (DiMSC) [36], and joint affinity graph for MVSC (JAMSC) [50].

- 1) *SC* [65]: It is a classical SC method designed for single-view clustering. The reason for adopting single-view clustering method is to show the performance improvements achieved by multiview clustering over single-view clustering.
- 2) *FeatConcat-SC*: The method first concatenates the features of all views, and then the standard SC algorithm is adopted to obtain the final clustering results.

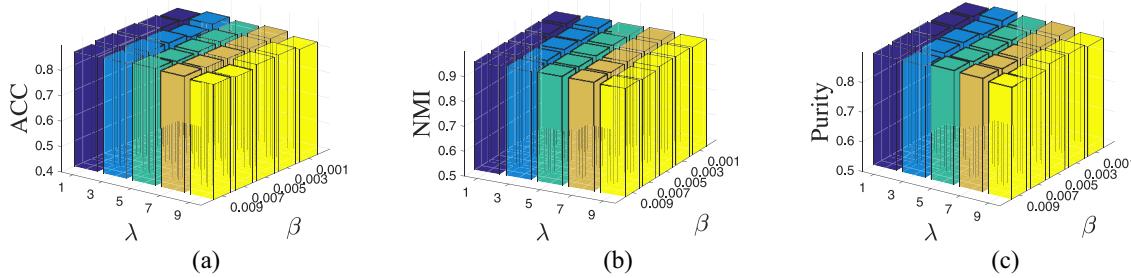
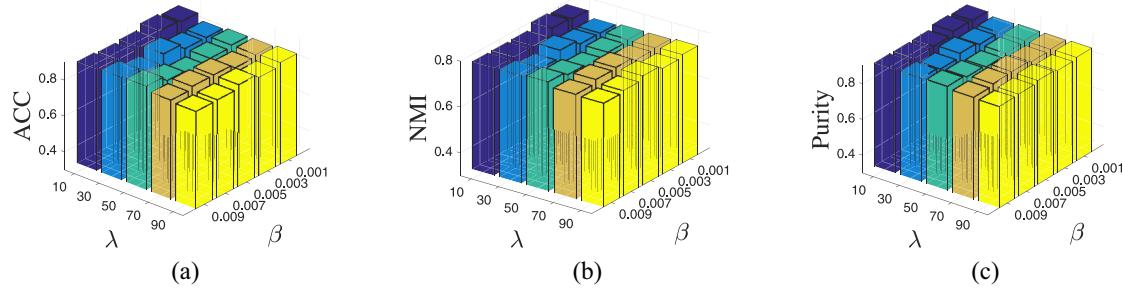
Fig. 6. Parameter analysis on λ and β on ORL. (a) ACC. (b) NMI. (c) Purity.Fig. 7. Parameter analysis on λ and β on Notting-Hill. (a) ACC. (b) NMI. (c) Purity.

TABLE II
COMPARISON RESULTS: THE MEAN AND STANDARD DEVIATIONS ACHIEVED BY ALL THE METHODS ON THE UCI DIGITS DATASET

Method	ACC	NMI	Purity	F-measure	Precision	Recall	AR
SC 1	0.6154 ± 0.0083	0.5891 ± 0.0066	0.6324 ± 0.0091	0.5096 ± 0.0078	0.5036 ± 0.0079	0.5158 ± 0.0078	0.4547 ± 0.0087
SC 2	0.6779 ± 0.0018	0.6116 ± 0.0056	0.6780 ± 0.0020	0.5502 ± 0.0019	0.5454 ± 0.0014	0.5551 ± 0.0023	0.5000 ± 0.0020
SC 3	0.6269 ± 0.0038	0.5949 ± 0.0034	0.6449 ± 0.0081	0.5302 ± 0.0053	0.5203 ± 0.0064	0.5406 ± 0.0042	0.4772 ± 0.0061
FeatConcate-SC	0.8284 ± 0.0325	0.7630 ± 0.0166	0.8289 ± 0.0321	0.7407 ± 0.0308	0.7367 ± 0.0319	0.7448 ± 0.0297	0.7118 ± 0.0343
ConcatPCA-SC	0.8184 ± 0.0327	0.7577 ± 0.0165	0.8216 ± 0.0295	0.7329 ± 0.0293	0.7277 ± 0.0314	0.7382 ± 0.0275	0.7031 ± 0.0327
Co-Reg	0.8444 ± 0.0078	0.7781 ± 0.0040	0.8538 ± 0.0064	0.7576 ± 0.0066	0.7489 ± 0.0080	0.7681 ± 0.0046	0.7302 ± 0.0075
Co-Training	0.8401 ± 0.0140	0.7961 ± 0.0053	0.8548 ± 0.0106	0.7799 ± 0.0095	0.7648 ± 0.0131	0.7985 ± 0.0056	0.7549 ± 0.0107
Min-Disagreement	0.7839 ± 0.0122	0.7483 ± 0.0061	0.8069 ± 0.0098	0.7141 ± 0.0095	0.6866 ± 0.0124	0.7444 ± 0.0061	0.6809 ± 0.0107
RMSC	0.6401 ± 0.0179	0.6090 ± 0.0187	0.6235 ± 0.0186	0.6125 ± 0.0206	0.6167 ± 0.0181	0.6181 ± 0.0236	0.6059 ± 0.0232
MVSC	0.7924 ± 0.0314	0.7050 ± 0.0244	0.7926 ± 0.0312	0.7072 ± 0.0278	0.6566 ± 0.0341	0.7664 ± 0.0189	0.6720 ± 0.0317
LMSM	0.8566 ± 0.0377	0.7837 ± 0.0273	0.8581 ± 0.0334	0.7628 ± 0.0396	0.7590 ± 0.0423	0.7668 ± 0.0372	0.7364 ± 0.0442
MVGL	0.8390 ± 0.0000	0.8213 ± 0.0000	0.8390 ± 0.0000	0.7562 ± 0.0000	0.7309 ± 0.0000	0.7834 ± 0.0000	0.7283 ± 0.0000
NaMSC	0.8639 ± 0.0035	0.7881 ± 0.0017	0.8639 ± 0.0035	0.7707 ± 0.0029	0.7690 ± 0.0032	0.7724 ± 0.0027	0.7453 ± 0.0033
DiMSC	0.8661 ± 0.0030	0.7920 ± 0.0015	0.8661 ± 0.0030	0.7753 ± 0.0021	0.7733 ± 0.0022	0.7773 ± 0.0023	0.7504 ± 0.0023
JAMSC	0.8709 ± 0.0002	0.7877 ± 0.0002	0.8709 ± 0.0002	0.7737 ± 0.0003	0.7696 ± 0.0003	0.7780 ± 0.0003	0.7486 ± 0.0003
MvSCGE	0.8941 ± 0.0026	0.8283 ± 0.0021	0.8941 ± 0.0026	0.8162 ± 0.0033	0.8111 ± 0.0031	0.8213 ± 0.0036	0.7957 ± 0.0036

- 3) *ConcatPCA-SC*: The method concatenates the features of all views in the first step. For the second step, the PCA method is applied to extract the low-dimensional subspace representation. Then, it applies the standard SC algorithm on the subspace representation to obtain the final clustering results.
- 4) *Co-Reg* [27]: The pairwise multiview SC method co-regularizes the clustering hypotheses to enforce the same cluster membership for the corresponding data points in each view. Gaussian kernel is used for constructing the similarity matrix and the default setting is used in the experiment as suggested by the authors.
- 5) *Co-Training* [28]: The co-training-based multiview SC method assumes that a point would be assigned by the true underlying clustering to the same category regardless of the views. Gaussian kernel is also used for constructing the similarity matrix. As suggested by the

authors, its parameter λ is searched in the range of [1 1.5], which gives reasonable results.

- 6) *Min-Disagreement* [39]: Based on the SC algorithm, the method creates a bipartite graph with the “minimizing-disagreement” strategy. For this method, the same graph construction method is used as SC, Co-Reg, and Co-Training.
- 7) *RMSC* [33]: The method uses the standard Markov chain for clustering. Gaussian kernel is also utilized for constructing the similarity matrix. As suggested by the authors, its parameter λ is searched in the range of [0.005 100].
- 8) *MVSC* [35]: The method first learns the subspace representation for each view by the self-expression property, and then the final common indicator matrix is learned by these subspace representations. In the experiment, the indicator matrix F is initialized with the result of SC,

TABLE III
COMPARISON RESULTS: THE MEAN AND STANDARD DEVIATIONS ACHIEVED BY ALL THE METHODS ON THE MSRCv1 DATASET

Method	ACC	NMI	Purity	F-measure	Precision	Recall	AR
SC 1	0.4105 \pm 0.0151	0.3229 \pm 0.0222	0.4595 \pm 0.0164	0.3212 \pm 0.0194	0.3184 \pm 0.0198	0.3241 \pm 0.0190	0.2107 \pm 0.0227
SC 2	0.6845 \pm 0.0446	0.5961 \pm 0.0344	0.7262 \pm 0.0297	0.5873 \pm 0.0374	0.5790 \pm 0.0367	0.5958 \pm 0.0383	0.5197 \pm 0.0435
SC 3	0.6167 \pm 0.0045	0.5103 \pm 0.0103	0.6500 \pm 0.0045	0.4842 \pm 0.0075	0.4773 \pm 0.0069	0.4913 \pm 0.0083	0.3997 \pm 0.0086
SC 4	0.6945 \pm 0.0186	0.5355 \pm 0.0151	0.6945 \pm 0.0186	0.5355 \pm 0.0140	0.5276 \pm 0.0144	0.5437 \pm 0.0143	0.4594 \pm 0.0164
FeatConcate-SC	0.6152 \pm 0.0043	0.5058 \pm 0.0177	0.6462 \pm 0.0123	0.4810 \pm 0.0140	0.4735 \pm 0.0145	0.4887 \pm 0.0135	0.3958 \pm 0.0164
ConcatePCA-SC	0.6155 \pm 0.0067	0.5087 \pm 0.0092	0.6490 \pm 0.0069	0.4837 \pm 0.0074	0.4766 \pm 0.0072	0.4910 \pm 0.0077	0.3991 \pm 0.0860
Co-Reg	0.6233 \pm 0.0057	0.5104 \pm 0.0036	0.6448 \pm 0.0048	0.4931 \pm 0.0041	0.4773 \pm 0.0050	0.5138 \pm 0.0036	0.4077 \pm 0.0050
Co-Training	0.6918 \pm 0.0099	0.6156 \pm 0.0064	0.7179 \pm 0.0071	0.5900 \pm 0.0072	0.5767 \pm 0.0080	0.6046 \pm 0.0066	0.5221 \pm 0.0085
Min-Disagreement	0.5923 \pm 0.0071	0.5174 \pm 0.0046	0.6075 \pm 0.0067	0.4825 \pm 0.0061	0.4728 \pm 0.0060	0.4928 \pm 0.0063	0.3972 \pm 0.0071
RMSC	0.2998 \pm 0.0189	0.2819 \pm 0.0138	0.2826 \pm 0.0163	0.2883 \pm 0.0185	0.2675 \pm 0.0158	0.2942 \pm 0.0156	0.2581 \pm 0.0219
MVSC	0.5405 \pm 0.0048	0.4431 \pm 0.0057	0.5595 \pm 0.0048	0.4289 \pm 0.0055	0.3717 \pm 0.0071	0.5069 \pm 0.0047	0.3200 \pm 0.0072
LMSC	0.6743 \pm 0.0591	0.5776 \pm 0.0606	0.6900 \pm 0.0624	0.5454 \pm 0.0663	0.5340 \pm 0.0648	0.5573 \pm 0.0684	0.4703 \pm 0.0773
MVGL	0.6714 \pm 0.0000	0.5775 \pm 0.0000	0.7048 \pm 0.0000	0.5550 \pm 0.0000	0.5072 \pm 0.0000	0.6128 \pm 0.0000	0.4754 \pm 0.0000
NaMSC	0.6869 \pm 0.0185	0.6027 \pm 0.0126	0.7117 \pm 0.0175	0.5767 \pm 0.0170	0.5738 \pm 0.0172	0.5797 \pm 0.0168	0.5081 \pm 0.0198
DiMSC	0.7300 \pm 0.0197	0.6395 \pm 0.0220	0.7526 \pm 0.0228	0.6310 \pm 0.0240	0.6265 \pm 0.0273	0.6357 \pm 0.0207	0.5710 \pm 0.0284
JAMSC	0.7402 \pm 0.0402	0.6346 \pm 0.0368	0.7402 \pm 0.0402	0.6102 \pm 0.0435	0.6021 \pm 0.0450	0.6184 \pm 0.0419	0.5463 \pm 0.0508
MvSCGE	0.7610 \pm 0.0518	0.6810 \pm 0.0374	0.7702 \pm 0.0329	0.6506 \pm 0.0479	0.6408 \pm 0.0498	0.6607 \pm 0.0461	0.5932 \pm 0.0561

TABLE IV
COMPARISON RESULTS: THE MEAN AND STANDARD DEVIATIONS ACHIEVED BY ALL THE METHODS ON THE BBCSPORT DATASET

Method	ACC	NMI	Purity	F-measure	Precision	Recall	AR
SC 1	0.8453 \pm 0.0012	0.6717 \pm 0.0018	0.8453 \pm 0.0012	0.7596 \pm 0.0011	0.7795 \pm 0.0009	0.7406 \pm 0.0014	0.6868 \pm 0.0015
SC 2	0.5114 \pm 0.0011	0.2345 \pm 0.0004	0.5717 \pm 0.0000	0.4206 \pm 0.0003	0.3326 \pm 0.0006	0.5720 \pm 0.0006	0.1703 \pm 0.0008
FeatConcate-SC	0.8952 \pm 0.0000	0.7416 \pm 0.0000	0.8952 \pm 0.0000	0.8271 \pm 0.0000	0.8608 \pm 0.0000	0.7959 \pm 0.0000	0.7757 \pm 0.0000
ConcatePCA-SC	—	—	—	—	—	—	—
Co-Reg	0.6928 \pm 0.0070	0.5375 \pm 0.0021	0.7348 \pm 0.0033	0.6001 \pm 0.0043	0.5470 \pm 0.0026	0.6823 \pm 0.0111	0.4569 \pm 0.0047
Co-Training	0.6979 \pm 0.0039	0.5657 \pm 0.0017	0.7601 \pm 0.0020	0.6087 \pm 0.0024	0.6057 \pm 0.0024	0.6545 \pm 0.0143	0.4849 \pm 0.0023
Min-Disagreement	0.8507 \pm 0.0087	0.7843 \pm 0.0055	0.8715 \pm 0.0046	0.8433 \pm 0.0076	0.8450 \pm 0.0055	0.8450 \pm 0.0110	0.7943 \pm 0.0100
RMSC	0.7737 \pm 0.0098	0.7645 \pm 0.0117	0.7597 \pm 0.0106	0.7799 \pm 0.0064	0.7570 \pm 0.0180	0.7607 \pm 0.0140	0.7527 \pm 0.0165
MVSC	0.7007 \pm 0.0011	0.4171 \pm 0.0019	0.7026 \pm 0.0011	0.5629 \pm 0.0015	0.4294 \pm 0.0015	0.8172 \pm 0.0009	0.3640 \pm 0.0025
LMSC	0.8512 \pm 0.1203	0.7448 \pm 0.1356	0.8560 \pm 0.1053	0.8288 \pm 0.1136	0.8129 \pm 0.1284	0.8485 \pm 0.0866	0.7704 \pm 0.1651
MVGL	0.4191 \pm 0.0000	0.0880 \pm 0.0000	0.4228 \pm 0.0000	0.3991 \pm 0.0000	0.2543 \pm 0.0000	0.9276 \pm 0.0000	0.0394 \pm 0.0000
NaMSC	0.8639 \pm 0.0032	0.6887 \pm 0.0060	0.8639 \pm 0.0032	0.7843 \pm 0.0047	0.7911 \pm 0.0061	0.7777 \pm 0.0037	0.7175 \pm 0.0063
DiMSC	0.8971 \pm 0.0000	0.7920 \pm 0.0000	0.8971 \pm 0.0000	0.8723 \pm 0.0000	0.8675 \pm 0.0000	0.8772 \pm 0.0000	0.8320 \pm 0.0000
JAMSC	0.9320 \pm 0.0000	0.8401 \pm 0.0000	0.9320 \pm 0.0000	0.9023 \pm 0.0000	0.9045 \pm 0.0000	0.9001 \pm 0.0000	0.8718 \pm 0.0000
MvSCGE	0.8952 \pm 0.0000	0.7998 \pm 0.0000	0.8952 \pm 0.0000	0.8751 \pm 0.0000	0.8662 \pm 0.0000	0.8841 \pm 0.0000	0.8354 \pm 0.0000

- and its parameter α is searched to infinity as suggested by the authors. In addition, the other two parameters λ_1 and λ_2 are searched in the range of [0.0001 100].
- 9) *LMSC* [34]: The method discovers the common latent structure shared by all different views at first. Then, the subspace representation is generated by the common latent structure. Finally, the SC algorithm is used to obtain the clustering results. As suggested by the authors, its parameter K is set as 100, and the parameter λ is searched in the range of {0.001, 0.01, 0.1, 1, 10, 100, 1000}.
 - 10) *MVGL* [37]: The method first learns the optimized (ideal) graphs of different views, and then a global graph can be obtained by integrating these optimized graphs. In the experiment, the parameter k is searched in the range of [1 100].
 - 11) *NaMSC* [36]: The method first performs the subspace representation learning for each view with the graph regularization technique. Then, SC is applied to the combination of the learned subspace representations. In the experiment, its parameter λ is searched in the range of [0.001 100].
 - 12) *DiMSC* [36]: The method explores the complementary information among multiview features by using the HSIC. After achieving the subspace representation of

each view, SC is performed on the combination of the learned subspace representations to get the final clustering results. In the experiment, its parameters λ_S and λ_V are searched in the range of [0.001 1000] to achieve reasonable performance.

- 13) *JAMSC* [50]: The method recovers a shared affinity presentation for MVSC by simultaneously considering the diversity regularization and a rank constraint, and normalized cuts are employed on the learned affinity representation to obtain the final clustering results. In the experiment, its parameters λ and β are searched in the range of [0.001, 0.01, 0.1, 1, 10, 100, 1000].

For all the 14 methods, including the above 13 compared methods and the proposed MvSCGE method, we run each method 20 times and compute the average performance as well as their standard deviation (std. dev.). In addition, k -means algorithm is needed to be perform as the final step in all the methods except MVGL. Thus, in each experiment, k -means clustering algorithm is run ten times to alleviate the impact of random initialization. If not stated otherwise, inner product kernel is utilized for calculating the similarity matrix in all experiments. In this article, the subspace representations $Z^{(v)}$ of $V - 1$ views are initialized by using the SMR method [46].

- 2) *Results and Analysis*: The clustering results obtained by different clustering methods on the five real-world datasets are

TABLE V
COMPARISON RESULTS: THE MEAN AND STANDARD DEVIATIONS ACHIEVED BY ALL THE METHODS ON THE ORL DATASET

Method	ACC	NMI	Purity	F-measure	Precision	Recall	AR
SC 1	0.6571 \pm 0.0242	0.8053 \pm 0.0109	0.6931 \pm 0.0200	0.5424 \pm 0.0206	0.5105 \pm 0.0190	0.5789 \pm 0.0245	0.5312 \pm 0.0210
SC 2	0.7735 \pm 0.0261	0.8910 \pm 0.0102	0.8025 \pm 0.0205	0.7101 \pm 0.0255	0.6714 \pm 0.0299	0.7539 \pm 0.0257	0.7030 \pm 0.0262
SC 3	0.6973 \pm 0.0319	0.8407 \pm 0.0169	0.7326 \pm 0.0263	0.6076 \pm 0.0381	0.5710 \pm 0.0397	0.6495 \pm 0.0369	0.5980 \pm 0.0391
FeatConcat-SC	0.6579 \pm 0.0274	0.8062 \pm 0.0126	0.6935 \pm 0.0210	0.5435 \pm 0.0274	0.5101 \pm 0.0305	0.5819 \pm 0.0249	0.5323 \pm 0.0282
ConcatPCA-SC	0.6607 \pm 0.0215	0.8069 \pm 0.0113	0.6932 \pm 0.0180	0.5452 \pm 0.0221	0.5128 \pm 0.0216	0.5823 \pm 0.0252	0.5341 \pm 0.0227
Co-Reg	0.6921 \pm 0.0037	0.8376 \pm 0.0017	0.7294 \pm 0.0028	0.6004 \pm 0.0038	0.5577 \pm 0.0044	0.6518 \pm 0.0036	0.5904 \pm 0.0039
Co-Training	0.7539 \pm 0.0058	0.8813 \pm 0.0031	0.7879 \pm 0.0050	0.6880 \pm 0.0072	0.6422 \pm 0.0079	0.7420 \pm 0.0067	0.6802 \pm 0.0074
Min-Disagreement	0.7259 \pm 0.0062	0.8617 \pm 0.0030	0.7625 \pm 0.0051	0.6499 \pm 0.0068	0.6032 \pm 0.0072	0.7052 \pm 0.0073	0.6412 \pm 0.0070
RMSC	0.7603 \pm 0.0259	0.7200 \pm 0.0209	0.7387 \pm 0.0169	0.7293 \pm 0.0219	0.7290 \pm 0.0227	0.7556 \pm 0.0199	0.7227 \pm 0.0291
MVSC	0.6932 \pm 0.0353	0.7924 \pm 0.0179	0.7125 \pm 0.0286	0.4943 \pm 0.0440	0.3875 \pm 0.0514	0.6901 \pm 0.0181	0.4792 \pm 0.0461
LMSC	0.8013 \pm 0.0333	0.9066 \pm 0.0204	0.8379 \pm 0.0293	0.7455 \pm 0.0461	0.6932 \pm 0.0481	0.8067 \pm 0.0449	0.7392 \pm 0.0472
MVGL	0.7350 \pm 0.0000	0.8651 \pm 0.0000	0.7950 \pm 0.0000	0.5580 \pm 0.0000	0.4281 \pm 0.0000	0.8011 \pm 0.0000	0.5447 \pm 0.0000
NaMSC	0.7574 \pm 0.0257	0.8768 \pm 0.0103	0.7889 \pm 0.0209	0.6861 \pm 0.0276	0.6442 \pm 0.0350	0.7345 \pm 0.0212	0.6784 \pm 0.0284
DiMSC	0.7970 \pm 0.0261	0.9045 \pm 0.0134	0.8294 \pm 0.0258	0.7439 \pm 0.0325	0.7010 \pm 0.0342	0.7926 \pm 0.0326	0.7376 \pm 0.0333
JAMSC	0.7950 \pm 0.0308	0.9004 \pm 0.0158	0.8236 \pm 0.0275	0.7375 \pm 0.0372	0.6980 \pm 0.0398	0.7820 \pm 0.0366	0.7310 \pm 0.0382
MvSCGE	0.8614 \pm 0.0226	0.9386 \pm 0.0103	0.8861 \pm 0.0207	0.8261 \pm 0.0267	0.7887 \pm 0.0288	0.8673 \pm 0.0257	0.8219 \pm 0.0273

TABLE VI
COMPARISON RESULTS: THE MEAN AND STANDARD DEVIATIONS ACHIEVED BY ALL THE METHODS ON THE NOTTING-HILL DATASET

Method	ACC	NMI	Purity	F-measure	Precision	Recall	AR
SC 1	0.7491 \pm 0.0080	0.6382 \pm 0.0069	0.7711 \pm 0.0051	0.6857 \pm 0.0129	0.6942 \pm 0.0085	0.6774 \pm 0.0173	0.5997 \pm 0.0158
SC 2	0.8000 \pm 0.0000	0.6160 \pm 0.0000	0.8000 \pm 0.0000	0.6931 \pm 0.0000	0.7200 \pm 0.0000	0.6681 \pm 0.0000	0.6118 \pm 0.0000
SC 3	0.6889 \pm 0.0017	0.6701 \pm 0.0024	0.7984 \pm 0.0017	0.6695 \pm 0.0026	0.6763 \pm 0.0031	0.6629 \pm 0.0023	0.5788 \pm 0.0033
FeatConcat-SC	0.7499 \pm 0.0031	0.6416 \pm 0.0050	0.7721 \pm 0.0020	0.6872 \pm 0.0035	0.6959 \pm 0.0023	0.6786 \pm 0.0047	0.6016 \pm 0.0043
ConcatPCA-SC	0.7525 \pm 0.0069	0.6420 \pm 0.0037	0.7736 \pm 0.0039	0.6917 \pm 0.0126	0.6980 \pm 0.0055	0.6857 \pm 0.0198	0.6070 \pm 0.0151
Co-Reg	0.7897 \pm 0.0047	0.7301 \pm 0.0051	0.8181 \pm 0.0049	0.7666 \pm 0.0063	0.7681 \pm 0.0056	0.7655 \pm 0.0075	0.7020 \pm 0.0079
Co-Training	0.8133 \pm 0.0084	0.7663 \pm 0.0065	0.8432 \pm 0.0034	0.8045 \pm 0.0092	0.7974 \pm 0.0086	0.8140 \pm 0.0079	0.7493 \pm 0.0119
Min-Disagreement	0.7970 \pm 0.0092	0.7181 \pm 0.0053	0.8230 \pm 0.0059	0.7727 \pm 0.0071	0.7690 \pm 0.0072	0.7777 \pm 0.0068	0.7091 \pm 0.0091
RMSC	0.7483 \pm 0.0549	0.7464 \pm 0.0492	0.7297 \pm 0.0664	0.7481 \pm 0.0535	0.7248 \pm 0.0517	0.7364 \pm 0.0604	0.7140 \pm 0.0563
MVSC	0.8021 \pm 0.0007	0.6752 \pm 0.0023	0.8057 \pm 0.0007	0.7665 \pm 0.0007	0.7013 \pm 0.0004	0.8450 \pm 0.0010	0.6936 \pm 0.0008
LMSC	0.7508 \pm 0.0573	0.6613 \pm 0.0509	0.7848 \pm 0.0339	0.7144 \pm 0.0599	0.7162 \pm 0.0503	0.7131 \pm 0.0707	0.6356 \pm 0.0747
MVGL	0.9055 \pm 0.0000	0.8122 \pm 0.0000	0.9055 \pm 0.0000	0.8554 \pm 0.0000	0.8140 \pm 0.0000	0.9012 \pm 0.0000	0.8125 \pm 0.0000
NaMSC	0.7575 \pm 0.0057	0.7208 \pm 0.0037	0.8017 \pm 0.0004	0.7141 \pm 0.0054	0.7377 \pm 0.0053	0.6920 \pm 0.0056	0.6378 \pm 0.0068
DiMSC	0.8363 \pm 0.0096	0.7810 \pm 0.0157	0.8454 \pm 0.0096	0.8210 \pm 0.0176	0.8126 \pm 0.0140	0.8295 \pm 0.0212	0.7705 \pm 0.0222
JAMSC	0.7483 \pm 0.0207	0.7188 \pm 0.0209	0.8289 \pm 0.0073	0.7303 \pm 0.0238	0.7414 \pm 0.0200	0.7196 \pm 0.0275	0.6568 \pm 0.0300
MvSCGE	0.8929 \pm 0.0005	0.7968 \pm 0.0046	0.8929 \pm 0.0005	0.8343 \pm 0.0029	0.8364 \pm 0.0009	0.8323 \pm 0.0049	0.7884 \pm 0.0036

reported in terms of ACC, NMI, Purity, F-measure, Precision, Recall, and AR in Tables II–VI, respectively. In this table, “SC 1” implies that we perform the SC algorithm in the first view of a dataset, and similarly for “SC 2,” “SC 3,” and “SC 4.”

As shown in the five tables, in general, our method almost achieves the best clustering results. Due to the efficient utilization of cross-view consistency, MvSCGE achieves better results than NaMSC that only cares about the graph regularization. Moreover, with the graph regularization (i.e., the subspacewise grouping effect), our MvSCGE method further outperforms MVSC that simply considers consistency among multiple views.

On the UCI digits dataset, the experimental results in Table II illustrate that considering multiview generally achieves a better performance than any single view case. At the same time, our method outperforms the other compared methods.

On the MSRCv1 dataset, as shown in Table III, though there is one view that is less informative about the clustering, our method still can achieve better results after fully considering multiview information to mine the stable underlying subspaces for consistent clustering. Specifically, the performance improvement over the Co-Training, NaMSC, and DiMSC is 6.6%, 8%, and 4.2%, respectively, in terms of NMI.

On the BBCSport dataset, as shown in Table IV, MVGL performs relatively worse (though its value of Recall is the best). Facing with sparse features, the performance of the JAMSC method obtains a little better clustering results than the proposed method, but the proposed method achieves better results on the whole. Note that ConcatPCA cannot be performed on the BBCSport dataset since the features on the BBCSport dataset are too sparse to run SVD.

On the ORL dataset, as shown in Table V, promising results can be achieved by many compared methods. Despite of this, our method still outperforms the other methods. Furthermore, Fig. 8 plots the visualization of some clustering results on the ORL dataset obtained by LMSC and MvSCGE. According to Fig. 8, the clustering results of MvSCGE are intuitively better than the second best method LMSC on the ORL dataset. Due to the space limit, only a portion of the results are plotted. Therefore, there is no quantitative results in the figure. However, the visualization results further confirm the superiority of MvSCGE over the second best method LMSC.

On the Notting-Hill dataset, as shown in Table VI, MVGL performs a little better than our method due to the optimized graph learning on this video face dataset, but does not perform that well on other datasets, especially on BBCSport dataset. In



Fig. 8. Visualization of some clustering results on ORL obtained by LMSC and the proposed MvSCGE method, respectively. Each row represents the output of a video face cluster. Due to the space limit, only five clusters from the ORL dataset are randomly selected and plotted. The mistakenly clustered face images are highlighted in the red rectangles. (a) Results on ORL obtained by LMSC. (b) Results on ORL obtained by MvSCGE.

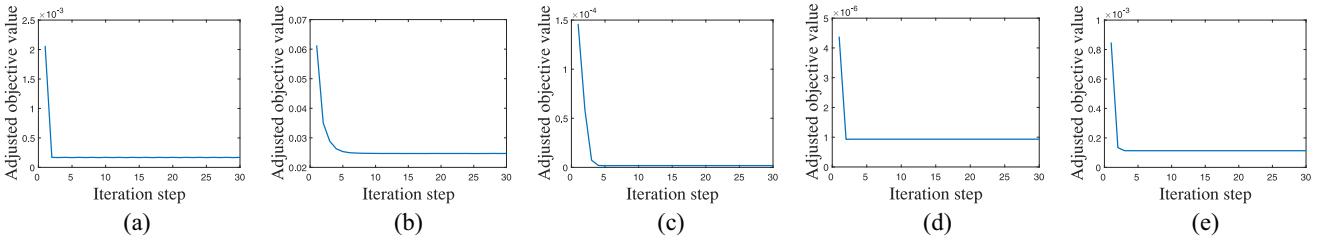


Fig. 9. Convergence analysis: the adjusted objective value as a function of the iteration step. (a) UCI digits. (b) MSRCv1. (c) BBCSport. (d) ORL. (e) Notting-Hill.

spite of that our method generally outperforms other compared methods, which ranks the second best.

According to the above analysis, the proposed method performs better on both image datasets and document dataset. With the graph regularization technique (i.e., the subspace-wise grouping effect), which is beneficial to the subspace clustering, our method can learn a more stable subspace representation. Meanwhile, it can ensure the consistent cluster structure among different views, which better satisfies the requirement of multiview learning.

D. Convergence Analysis

An alternate minimizing algorithm is developed to solve the optimization problem in (6). Since the objective function in (6) is nonincreasing with the iterations, the algorithm can be guaranteed to converge finally. In this section, convergence analysis is conducted to verify the convergence property of the proposed method. The convergence results on all the four real-world datasets are illustrated in Fig. 9. Notice that the objective value may be a very large value, which is not suitable to be plotted in the y-axis directly. Therefore, in this section, we use the *difference* between the current objective value (after a certain number of iterations) and the final objective value (after convergence) as the y-axis, and that is the adjusted objective value in Fig. 9. Obviously, we can see that within less than five iterations, the convergence can be reached.

V. CONCLUSION

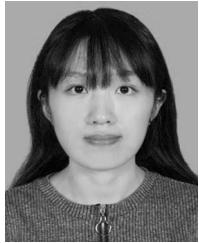
In this article, a new MvSCGE method is proposed to preserve the locality explicitly in subspaces and ensure the

cross-view consistency in multiple different views. With the framework, we can perform subspace learning in each view with a smooth regularization, and then exploit the subspace-wise grouping effect in these learned subspaces by means of a unified optimization framework. Furthermore, the consistent cluster indicator matrix can be obtained by considering each subspace structure with the cross-view consistency constraint. An alternate minimizing strategy is developed to deal with the proposed method. Extensive experiments on five real-world datasets have shown the superiority of the proposed method.

REFERENCES

- [1] C. Xu, D. Tao, and C. Xu, "A survey on multi-view learning," 2013. [Online]. Available: arXiv:1304.5634.
- [2] Y.-M. Xu, C.-D. Wang, and J.-H. Lai, "Weighted multi-view clustering with feature selection," *Pattern Recognit.*, vol. 53, pp. 25–35, May 2016.
- [3] C.-D. Wang, J.-H. Lai, and S. Y. Philip, "Multi-view clustering based on belief propagation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 4, pp. 1007–1021, Apr. 2016.
- [4] J.-H. Li, C.-D. Wang, P.-Z. Li, and J.-H. Lai, "Discriminative metric learning for multi-view graph partitioning," *Pattern Recognit.*, vol. 75, pp. 199–213, Mar. 2018.
- [5] G.-Y. Zhang, C.-D. Wang, D. Huang, and W.-S. Zheng, "Multi-view collaborative locally adaptive clustering with Minkowski metric," *Expert Syst. Appl.*, vol. 86, pp. 307–320, Nov. 2017.
- [6] G.-Y. Zhang, C.-D. Wang, D. Huang, W.-S. Zheng, and Y.-R. Zhou, "TW-CO- k -means: Two-level weighted collaborative k -means for multi-view clustering," *Knowl. Based Syst.*, vol. 150, pp. 127–138, Jun. 2018.
- [7] L. Huang, H.-Y. Chao, and C.-D. Wang, "Multi-view intact space clustering," *Pattern Recognit.*, vol. 86, pp. 344–353, Feb. 2019.
- [8] K.-Y. Lin, C.-D. Wang, Y.-Q. Meng, and Z.-L. Zhao, "Multi-view unit intact space learning," in *Proc. KSEM*, 2017, pp. 211–223.
- [9] K.-Y. Lin, L. Huang, C.-D. Wang, and H.-Y. Chao, "Multi-view proximity learning for clustering," in *Proc. DASFAA*, 2018, pp. 407–423.

- [10] H. Tao, C. Hou, D. Yi, and J. Zhu, "Multiview classification with cohesion and diversity," *IEEE Trans. Cybern.*, vol. 50, no. 5, pp. 2124–2137, May 2020.
- [11] S. Huang, Z. Kang, I. W. Tsang, and Z. Xu, "Auto-weighted multi-view clustering via kernelized graph learning," *Pattern Recognit.*, vol. 88, pp. 174–184, Apr. 2019.
- [12] K. Zhan, J. Shi, J. Wang, and F. Tian, "Graph-regularized concept factorization for multi-view document clustering," *J. Vis. Commun. Image Represent.*, vol. 48, pp. 411–418, Oct. 2017.
- [13] G.-Y. Zhang, D. Huang, C.-D. Wang, and W.-S. Zheng, "Weighted multi-view on-line competitive clustering," in *Proc. BigDataService*, 2016, pp. 286–292.
- [14] B. Jiang, F. Qiu, S. Yang, and L. Wang, "Evolutionary multi-objective optimization for multi-view clustering," in *Proc. IEEE Congr. Evol. Comput. (CEC)*, 2016, pp. 3308–3315.
- [15] W. Tang, Z. Lu, and I. S. Dhillon, "Clustering with multiple graphs," in *Proc. IEEE ICDM*, 2009, pp. 1016–1021.
- [16] X. Cai, F. Nie, H. Huang, and F. Kamangar, "Heterogeneous image feature integration via multi-modal spectral clustering," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, 2011, pp. 1977–1984.
- [17] W. Shao, J. Zhang, L. He, and S. Y. Philip, "Multi-source multi-view clustering via discrepancy penalty," in *Proc. IEEE IJCNN*, 2016, pp. 2714–2721.
- [18] H. Liu and Y. Fu, "Consensus guided multi-view clustering," *ACM Trans. Knowl. Discov. Data*, vol. 12, no. 4, p. 42, 2018.
- [19] Y.-X. Ji *et al.*, "Multi-view outlier detection in deep intact space," in *Proc. ICDM*, 2019, pp. 1132–1137.
- [20] H. Tao, C. Hou, D. Yi, and J. Zhu, "Joint embedding learning and low-rank approximation: A framework for incomplete multi-view learning," 2018. [Online]. Available: arxiv.abs/1812.10012.
- [21] L. Zong, X. Zhang, X. Liu, and H. Yu, "Weighted multi-view spectral clustering based on spectral perturbation," in *Proc. AAAI*, 2018, pp. 4621–4629.
- [22] H. Yu, Y. Lian, L. Zong, and L. Tian, "Self-paced learning based multi-view spectral clustering," in *Proc. ICTAI*, 2017, pp. 6–10.
- [23] X. Zhang, X. Zhang, H. Liu, and X. Liu, "Multi-task multi-view clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 12, pp. 3324–3338, Dec. 2016.
- [24] H. Tao, C. Hou, X. Liu, T. Liu, D. Yi, and J. Zhu, "Reliable multi-view clustering," in *Proc. AAAI*, 2018, pp. 4123–4130.
- [25] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, Oct. 2018.
- [26] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. ACM COLT*, 1998, pp. 92–100.
- [27] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Proc. Neural Inf. Process. Syst.*, 2011, pp. 1413–1421.
- [28] A. Kumar and H. Daumé, "A co-training approach for multi-view spectral clustering," in *Proc. ICML*, 2011, pp. 393–400.
- [29] M. Gönen and E. Alpaydin, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, Jul. 2011.
- [30] C. Cortes, M. Mohri, and A. Rostamizadeh, "Learning non-linear combinations of kernels," in *Proc. Neural Inf. Process. Syst.*, 2009, pp. 396–404.
- [31] X. Liu *et al.*, "Absent multiple kernel learning algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 6, pp. 1303–1316, Jun. 2020.
- [32] K. Chaudhuri, S. M. Kakade, K. Livescu, and K. Sridharan, "Multi-view clustering via canonical correlation analysis," in *Proc. ACM ICML*, 2009, pp. 129–136.
- [33] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. AAAI*, 2014, pp. 2149–2155.
- [34] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 4279–4287.
- [35] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, 2015, pp. 4238–4246.
- [36] X. Cao, C. Zhang, H. Fu, S. Liu, and H. Zhang, "Diversity-induced multi-view subspace clustering," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2015, pp. 586–594.
- [37] K. Zhan, C. Zhang, J. Guan, and J. Wang, "Graph learning for multi-view clustering," *IEEE Trans. Cybern.*, vol. 48, no. 10, pp. 2887–2895, Oct. 2018.
- [38] K. Zhan, F. Nie, J. Wang, and Y. Yang, "Multiview consensus graph clustering," *IEEE Trans. Image Process.*, vol. 28, no. 3, pp. 1261–1270, Mar. 2019.
- [39] V. R. De Sa, "Spectral clustering with two views," in *Proc. ICML Workshop Learn. Multiple Views*, 2005, pp. 20–27.
- [40] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Proc. AAAI*, 2015, pp. 2750–2756.
- [41] M.-S. Chen, L. Huang, C.-D. Wang, and D. Huang, "Multi-view spectral clustering via multi-view weighted consensus and matrix-decomposition based discretization," in *Proc. DASFAA*, 2019, pp. 175–190.
- [42] D. Xie, Q. Gao, Q. Wang, and S. Xiao, "Multi-view spectral clustering via integrating global and local graphs," *IEEE Access*, vol. 7, pp. 31197–31206, 2019.
- [43] E. Elhamifar and R. Vidal, "Sparse subspace clustering," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2009, pp. 2790–2797.
- [44] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. ICML*, 2010, pp. 663–670.
- [45] 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.
- [46] H. Hu, Z. Lin, J. Feng, and J. Zhou, "Smooth representation clustering," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2014, pp. 3834–3841.
- [47] C. Zhang *et al.*, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.
- [48] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. Comput. Vis. Pattern Recognit. (CVPR)*, 2017, pp. 1–9.
- [49] X. Wang, Z. Lei, X. Guo, C. Zhang, H. Shi, and S. Z. Li, "Multi-view subspace clustering with intactness-aware similarity," *Pattern Recognit.*, vol. 88, pp. 50–63, Apr. 2019.
- [50] C. Tang *et al.*, "Learning a joint affinity graph for multiview subspace clustering," *IEEE Trans. Multimedia*, vol. 21, no. 7, pp. 1724–1736, Jul. 2019.
- [51] 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. (ICCV)*, 2015, pp. 1582–1590.
- [52] G.-Y. Zhang, Y.-R. Zhou, X.-Y. He, C.-D. Wang, and D. Huang, "One-step kernel multi-view subspace clustering," *Knowl. Based Syst.*, vol. 189, Feb. 2020, Art. no. 105126.
- [53] R. Li, C. Zhang, Q. Hu, P. Zhu, and Z. Wang, "Flexible multi-view representation learning for subspace clustering," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2019, pp. 2916–2922.
- [54] W. Zhu, J. Lu, and J. Zhou, "Structured general and specific multi-view subspace clustering," *Pattern Recognit.*, vol. 93, pp. 392–403, Sep. 2019.
- [55] T. Zhou, C. Zhang, X. Peng, H. Bhaskar, and J. Yang, "Dual shared-specific multiview subspace clustering," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3517–3530, Aug. 2020.
- [56] S. Luo, C. Zhang, W. Zhang, and X. Cao, "Consistent and specific multi-view subspace clustering," in *Proc. AAAI*, 2018, pp. 3730–3737.
- [57] L. Xing, B. Chen, S. Du, Y. Gu, and N. Zheng, "Correntropy-based multiview subspace clustering," *IEEE Trans. Cybern.*, early access, Nov. 28, 2019, doi: 10.1109/TCYB.2019.2952398.
- [58] X. Peng, S. Xiao, J. Feng, W. Yau, and Z. Yi, "Deep subspace clustering with sparsity prior," in *Proc. Int. Joint Conf. Artif. Intell. (IJCAI)*, 2016, pp. 1925–1931.
- [59] M. Brbic and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognit.*, vol. 73, pp. 247–258, Jan. 2018.
- [60] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Neural Inf. Process. Syst.*, 2002, pp. 585–591.
- [61] R. H. Bartels and G. W. Stewart, "Solution of the matrix equation $AX + XB = C$ [F4]," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, 1972.
- [62] B. Mohar, Y. Alavi, G. Chartrand, and O. Oellermann, "The laplacian spectrum of graphs," *Graph Theory Comb. Appl.*, vol. 2, nos. 871–898, p. 12, 1991.
- [63] J. Winn and N. Jojic, "LOCUS: Learning object classes with unsupervised segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, vol. 1, 2005, pp. 756–763.
- [64] Y.-F. Zhang, C. Xu, H. Lu, and Y.-M. Huang, "Character identification in feature-length films using global face-name matching," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1276–1288, Nov. 2009.
- [65] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Proc. Neural Inf. Process. Syst. (NeurIPS)*, 2002, pp. 849–856.



Man-Sheng Chen received the B.S. degree in computer science from South China Agricultural University, Guangzhou, China, in 2018. She is currently pursuing the master's degree with Sun Yat-sen University, Guangzhou.

She has published three papers in international journals and conferences, namely, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, AAAI, and DASFAA. Her research interest is multiview clustering.



Dong Huang (Member, IEEE) received the B.S. degree in computer science from the South China University of Technology, Guangzhou, China, in 2009, and the M.Sc. and Ph.D. degrees in computer science from Sun Yat-sen University, Guangzhou, in 2011 and 2015, respectively.

He joined South China Agricultural University, Guangzhou, in 2015, where he is currently an Associate Professor with the College of Mathematics and Informatics. His research interests include data mining and pattern recognition.



Ling Huang (Member, IEEE) received the Ph.D. degree in computer science from Sun Yat-sen University, Guangzhou, China, in 2020.

She joined South China Agricultural University, Guangzhou, in 2020, as an Associate Professor. She has published over ten papers in international journals and conferences, such as IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, ACM Transactions on Knowledge Discovery From Data, IEEE/ACM TRANSACTIONS ON COMPUTATIONAL BIOLOGY AND BIOINFORMATICS, Pattern Recognition, KDD, AAAI, IJCAI, and ICDM. Her research interest is data mining.



Philip S. Yu (Life Fellow, IEEE) received the B.S. degree in electrical engineering from National Taiwan University, Taipei, Taiwan, in 1972, the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, USA, in 1976 and 1978, respectively, and the M.B.A. degree from New York University, New York, NY, USA.

He is a Distinguished Professor of Computer Science with the University of Illinois at Chicago, Chicago, IL, USA, and holds the Wexler Chair of Information Technology. He was with IBM, Armonk, NY, USA, where he was a Manager of the Software Tools and Techniques Group, Watson Research Center. He has published more than 1000 papers in refereed journals and conferences. He holds or has applied for more than 300 U.S. patents. His research interest is on big data, including data mining, data stream, database, and privacy.

Dr. Yu was a recipient of ACM SIGKDD 2016 Innovation Award for his influential research and scientific contributions on mining, fusion and anonymization of big data, the IEEE Computer Society's 2013 Technical Achievement Award for *Pioneering and Fundamentally Innovative Contributions to the Scalable Indexing, Querying, Searching, Mining and Anonymization of Big Data*, and the Research Contributions Award from IEEE International Conference on Data Mining (ICDM) in 2003 for his pioneering contributions to the field of data mining. He also received the ICDM 2013 ten-year Highest-Impact Paper Award, and the EDBT Test of Time Award in 2014. He has received several IBM honors, including two IBM Outstanding Innovation Awards, an Outstanding Technical Achievement Award, two Research Division Awards, and the 94th plateau of Invention Achievement Awards. He was an IBM Master Inventor. He is on the Steering Committee of the ACM Conference on Information and Knowledge Management. He was a Member of the Steering Committee of IEEE Data Engineering and IEEE Conference on Data Mining. He was an Editor-in-Chief of ACM Transactions on Knowledge Discovery From Data from 2011 to 2017, and the IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING from 2001 to 2004. He is a Fellow of the ACM.



Chang-Dong Wang (Member, IEEE) received the Ph.D. degree in computer science from Sun Yat-sen University, Guangzhou, China, in 2013.

He was a visiting student with the University of Illinois at Chicago, Chicago, IL, USA, in 2012. He joined Sun Yat-sen University in 2013, where he is currently an Associate Professor with the School of Data and Computer Science. He has published over 70 scientific papers in international journals and conferences, such as IEEE TRANSACTIONS ON PATTERN ANALYSIS AND

MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, IEEE TRANSACTIONS ON CYBERNETICS, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, ACM Transactions on Knowledge Discovery from Data, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS, IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, KDD, AAAI, IJCAI, CVPR, ICDM, CIKM, and SDM. His current research interests include machine learning and data mining.

Dr. Wang's ICDM 2010 paper won the Honorable Mention for Best Research Paper Awards. He won 2012 Microsoft Research Fellowship Nomination Award. He was awarded the 2015 Chinese Association for Artificial Intelligence Outstanding Dissertation. He is an Associate Editor of *Journal of Artificial Intelligence Research*.