This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2020.3013408, IEEE Transactions on Multimedia

Adaptive Graph Completion Based Incomplete Multi-view Clustering

Jie Wen, Ke Yan, Zheng Zhang, Yong Xu*, *Senior Member*, *IEEE*, Junqian Wang, Lunke Fei, Bob Zhang

***Abstract*—In real-world applications, it is often that the collected multi-view data are incomplete, *i.e.*, some views of samples are absent. Existing clustering methods for incomplete multi-view data all focus on obtaining a common representation or graph from the available views but neglect the hidden information of missing views and information imbalance of different views. To solve these problems, a novel method, called adaptive graph completion based incomplete multi-view clustering (AGC_IMC), is proposed in this paper. Specifically, AGC_IMC develops a joint framework for graph completion and consensus representation learning, which mainly contains three components, *i.e.*, within-view preservation, between-view inferring, and consensus representation learning. To reduce the negative influence of information imbalance, AGC_IMC introduces some adaptive weights to balance the importance of different views during the consensus representation learning. Importantly, AGC_IMC has the potential to recover the similarity graphs of all views with the optimal cluster structure, which encourages it to obtain a more discriminative consensus representation. Experimental results on five well-known datasets show that AGC_IMC significantly outperforms the state-of-the-art methods.**

***Index Terms*—Incomplete multi-view clustering, common representation, graph completion, similarity graph.**

I. Introduction

As a machine learning paradigm, multi-view clustering (MVC) has received a lot of attention from researchers and engineers in recent years [1-4]. MVC aims to partition the given subjects into different groups in an unsupervised way by combining feature information from multiple views collected from different domains. Since features collected from different views contain much complementary information, multi-view learning methods has the potential to achieve a better performance than the single-view based methods using either one of these views [3, 5-10]. In past years, many MVC methods have been proposed, where some representative works are summarized in [11]. In [12], a bipartite matching constrained based clustering method is proposed for multi-view video summarization. Chen et al. proposed a joint graph learning framework to learn a consensus graph from the tensor space for multi-view data clustering [13]. For these MVC methods, a common assumption is that all views of samples exist. In other words, these methods can only cluster the data with complete views while failing to handle the data with missing views. In our work, we refer to the data whose all views of samples are fully observed as complete multi-view data and refer to the data with missing views as incomplete multi-view data.

In practical applications, owing to some uncontrollable factors, the collected multi-view data are usually incomplete [14-17]. For example, in visual-audio based speaker grouping task, some speakers may only have either audio or visual information [18]. Owing to the view missing, some problems occur in the multi-view learning. First, it is difficult to explore the complementary information of these incomplete multiple views. Second, the balanced information of multiple views is seriously broken since these views may have different numbers of instances and features. The above problems make the view missing challenging in MVC tasks. For convenience, we refer to the clustering problem with missing views as incomplete multi-view clustering (IMC).

To the best of our knowledge, the first method to handle the IMC problem is proposed by Trivedi et al. in 2010 [19]. This method exploits the kernel canonical correlation analysis (KCCA) to recover the kernel matrix with respect to the missing view and then extract the features for clustering. However, this method is inflexible since it can only handle the two-view data and requires the data to have one complete view [20]. In recent years, researchers proposed some advanced IMC methods. For example, Li et al. proposed non-negative matrix factorization based partial multi-view clustering (P-MVC), which decomposes two views from the same sample into the same latent representation [18]. On the base of PMVC, Zhao et al. proposed the incomplete multi-modality grouping (IMG) method which further introduced an adaptive manifold constraint to learn a common graph for spectral clustering [21]. Compared with the KCCA based method in [19], PMVC and IMG are more flexible since they do not have the strict requirement of one complete view. However, these methods

This work is partially supported by Shenzhen Fundamental Research Fund under Grant JCYJ20190806142416685, Guangdong Basic and Applied Basic Research Foundation under Grants 2019A1515110582 & 2019A1515110475, National Postdoctoral Program for Innovative Talent under Grant BX20190100, Establishment of Key Laboratory of Shenzhen Science and Technology Innovation Committee under Grant ZDSYS20190902093015527, Natural Science Foundation of Guangdong Province under Grant 2019A1515011811, and University of Macau (File no. MYRG2019-00006-FST).

Jie Wen, Ke Yan and Junqian Wang are with the Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China, are also with the Shenzhen Key Laboratory of Visual Object Detection and Recognition, Shenzhen 518055, China (Email: jiewen_pr@126.com; yanke401@163.com; wangjunqian@stu.hit.edu.cn).

Zheng Zhang and Yong Xu are with Bio-Computing Research Center, Harbin Institute of Technology, Shenzhen, Shenzhen 518055, China; with the Shenzhen Key Laboratory of Visual Object Detection and Recognition, Shenzhen 518055, China; are also with the Pengcheng Laboratory, Shenzhen 518055, China. (Email: darrenzz219@gmail.com; yongxu@ymail.com)

Lunke Fei is with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China (Email: flksxm@126.com).

Bob Zhang is with the PAMI Research Group, Department of Computer and Information Science, University of Macau, Taipa, Macau, China (Email: bobzhang@um.edu.mo).

Corresponding author: Yong Xu (Email: yongxu@ymail.com)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2020.3013408, IEEE Transactions on Multimedia

are inapplicable to the data whose incomplete samples have more than one view [22]. To improve the flexibility, some weighted matrix factorization based IMC methods have been proposed, where the most representative works are multi-incomplete-view clustering (MIC) [23], online multi-view clustering (OMVC) [24], doubly aligned incomplete multi-view clustering (DAIMC) [25], and one-pass incomplete multi-view clustering (OPIMC) [26], etc. These methods commonly introduce the view present and absent information as a weighted matrix to regularize the matrix factorization models of all views jointly. Compared with PMVC and IMG, these weighted matrix factorization based methods are more superior since they can handle all kinds of incomplete multi-view data. In recent years, many graph based methods have also been proposed for the difficult IMC tasks [27-29]. For instance, Wen et al. proposed a low-rank representation based graph and consensus representation joint learning framework [27]. Wang et al. proposed a perturbation-oriented IMC method, which produced the consensus representation from the fixed similarity graphs pre-constructed from data [28]. Recently, deep learning has made impressive achievements in many applications [30-34]. Owing to its superiority in the high-level representation learning, many deep learning based IMC methods have been proposed, such as Adversarial IMC [35] and PMVC via consistent generative adversarial networks (GANs) [36]. However, the two methods are only applicable to the case with large amounts of paired-samples.

From the learning models of the existing IMC methods, we can observe that learning a consensus representation or graph shared by all views is one of the most promising approaches for IMC. However, the existing methods suffer from the following two issues: 1) The information of missing views are ignored. 2) The information imbalance factor hiding in these incomplete views is not considered since they treat all views equally. As a result, these methods cannot obtain the optimal common representation or graph, which limits their performance. In this paper, we propose a novel graph completion based IMC method to solve the above problems. Specifically, the proposed method seeks to recover the graphs of all incomplete views by fully exploring the within-view information of every view and the between-view information borrowed from the other views. In this way, the hidden connections of the missing instances and available instances can be recovered and in turn used to enhance the common representation learning. To guarantee the global optima of the latent representation and graphs of all views, we integrate the graph completion and common representation learning into a joint optimization framework. Moreover, considering different views carry different degrees of discriminant information, we impose a scale penalty vector on the learning models of all views to balance the effectiveness of these views. Experimental results show that the proposed method can improve the clustering performance significantly in comparison with the state-of-the-art IMC methods. Compared with the existing works, our work has the following superior properties:

1) A novel graph completion based method is proposed to address the IMC problem. To the best of our knowledge, it is of the first work that handles the incomplete multi-view clustering problem from the aspect of incomplete multi-graph recovery. By recovering the graphs, the proposed method has the potential to exploit the hidden information of missing instances and available instances to enhance the consensus representation learning.

2) The proposed method integrates the graph recovering and common representation learning into a joint optimization framework, which is beneficial to obtain the optimal graphs with exact cluster structure and the optimal discriminative latent representation, such that a better clustering performance can be obtained.

3) The proposed method imposes an adaptive scale vector on the learning models of all views, which can effectively reduce the negative influence of information imbalance of multiple views caused by view missing.

We organize the remainder of the paper as follows: In Section II, the spectral clustering and multi-view spectral clustering are briefly introduced as two related works to the proposed method. In Section III, we first describe the learning model, optimization process, and then discuss the computational complexity of the proposed method. Several experiments are conducted in Section VI. Section V offers a brief conclusion to the paper.

II. Related works

A. Spectral clustering

In view of the fact that spectral clustering mainly extracts a graph that reveals the intrinsic relationships of samples for clustering, it can be viewed as the graph based clustering method [37]. For a dataset $X = [x_1, x_2, \ldots, x_n] \in R^{m \times n}$, spectral clustering generally first constructs a similarity graph $W \in R^{n \times n}$ with non-negative elements and symmetric structure from the data, where m and n represent the feature dimension and number of samples, respectively. In the similarity graph, each element can be regarded as the probability of the corresponding two samples come from the same class to some extent. Then it minimizes the following objective function to obtain the new representation of all samples:

$$\min_U Tr\left(U^T L U\right) \ \ s.t. \ U^T U = I \tag{1}$$

where $U \in R^{n \times c}$ is the new representation (each row vector of U is the new representation of the corresponding sample). c is the feature dimension of the new representation, which is generally chosen as the cluster number of the data. I denotes an identity matrix. $L \in R^{n \times n}$ is the Laplacian matrix of graph W, which is calculated as $L = D - W$ in ratio cut [38] and $L = I - D^{-1/2} W D^{-1/2}$ in normalized cut [39], where $D \in R^{n \times n}$ is a diagonal matrix whose ith diagonal element is computed as the sum of the ith row vector of graph W.

B. Multi-view spectral clustering

Spectral clustering based MVC method, referred to as multi-view spectral clustering, is one of the most representative methods in fields of multi-view clustering. Generally, multi-view spectral clustering seeks to learn a consensus representation from multiple similarity graphs constructed from all views, followed by implementing k-means on the consensus representation to achieve the final clustering results [5, 40]. As one of the well-known methods, co-regularized multi-view spectral clustering designs the following model to learn the consensus representation agreed by all views [41]:

$$\max_{U^{(v)},U^*} \sum_{v=1}^{l} \left(Tr\left(U^{(v)T}L^{(v)}U^{(v)}\right) + \lambda_v Tr\left(U^{(v)}U^{(v)T}U^*U^{*T}\right)\right) \\ s.t.\, U^{(v)T}U^{(v)} = I, U^{*T}U^* = I \tag{2}$$

where λ_v denotes the penalty parameter of the vth view. $L^{(v)}$ is the normalized graph of the vth view and is calculated as $L^{(v)} = \left(D^{(v)}\right)^{-1/2} W^{(v)} \left(D^{(v)}\right)^{-1/2}$, $D^{(v)}$ is a diagonal matrix and calculated as $D_{i,i}^{(v)} = \sum_{j=1}^{n} W_{i,j}^{(v)}$ for the ith diagonal element, $W^{(v)} \in R^{n\times n}$ is the pre-constructed similarity graph with symmetrical structure of the vth view, n denotes the number of samples. $U^{(v)} \in R^{n\times d}$ can be viewed as the new representation of data in the vth view, d is the feature dimension of new representation. $U^* \in R^{n\times d}$ denotes the consensus representation shared by all views.

III. The proposed method

As presented in the previous section, multi-view spectral clustering constructs some graphs from all views for consensus representation learning. It strictly requires that all constructed graphs are complete. In other words, all graphs constructed from the multi-view data with n instances should have the same dimension of $n \times n$. However, for the incomplete multi-view data, it is obviously impossible to construct such complete graphs, which results in the failure of the traditional multi-view spectral clustering methods. Therefore, if we can complete the graphs, the issue that perplexes the IMC can be naturally solved. Inspired by this motivation, we propose a novel IMC method based on the graph completion in this section. The framework of the proposed method is shown in Fig.1. The proposed method focuses on recovering the incomplete graphs and calculating the consensus representation simultaneously in a joint framework.

A. Learning model of the proposed method

For any incomplete multi-view data with l views and n samples, let $Y^{(v)} \in R^{m_v\times n_v}$ be the set of available instances from the vth view, where m_v and n_v ($n_v \le n$) are the feature dimension and number of available instances of the vth view. $\bar{S}^{(v)} \in R^{n_v\times n_v}$ denotes the symmetric graph pre-constructed from the available instances of the vth view, where all elements of $\bar{S}^{(v)}$ are non-negative. Due to the view missing, every graph $\bar{S}^{(v)}$ is incomplete, which cannot reveal the comprehensive relationships of all samples. Our goal is to complete these flawed graphs such that the complementary information of different views can be better explored for the consensus representation learning. The learning model of our proposed method is mainly composed of three components: within-view preservation, between-view inferring, and consensus representation learning.

Within-view preservation: Let $S^{(v)} \in R^{n\times n}$ be the referred (completed) graph of the vth view. Understandably, for the vth view, the similarity information of the available instances in $\bar{S}^{(v)}$ should be preserved in the referred graph $S^{(v)}$. To this end, the following within-view preservation model is designed:

$$\min_{S^{(v)}} \sum_{v=1}^{l} \left\| \left(S^{(v)}\right)_A - \bar{S}^{(v)} \right\|_F^2 \tag{3}$$

where $\left(S^{(v)}\right)_A \in R^{n_v\times n_v}$ denotes the subgraph of $S^{(v)}$ whose every element represents the similarity information of the corresponding two available instances as that in $\bar{S}^{(v)}$.

We assume that matrix $E \in R^{l\times n}$ records the index information of the missing views, where $E_{i,j} = 0$ means that the jth instance is missing in the ith view, otherwise $E_{i,j} = 1$. Based on the view-missing information recorded in E, we can transform (3) into the following equivalent formula:

$$\min_{S^{(v)}} \sum_{v=1}^{l} \left\| \left(S^{(v)} - \widetilde{S}^{(v)}\right) \odot \left(E_{v,:}^T E_{v,:}\right) \right\|_F^2 \tag{4}$$

where $E_{v,:}$ represents the vth row vector of matrix E, $\odot$ denotes the element-wise based multiplication operation. If we define $W^{(v)} = E_{v,:}^T E_{v,:}$, then $W_{i,j}^{(v)} = 1$ means that the ith sample and the jth sample all have the instances of the vth view, otherwise $W_{i,j}^{(v)} = 0$. $\widetilde{S}^{(v)} \in R^{n\times n}$ is an extended graph filled by graph $\bar{S}^{(v)}$, where the elements in $\widetilde{S}^{(v)}$ related to the missing instances are set as 0. Mathematically, $\bar{S}^{(v)}$ and $\widetilde{S}^{(v)}$ have the following connections:

$$\widetilde{S}^{(v)} = G^{(v)} \bar{S}^{(v)} G^{(v)T} \tag{5}$$

where $G^{(v)} \in R^{n\times n_v}$ is defined as follows according to the view-missing information:

$$G_{i,j}^{(v)} = \begin{cases} 1, & if\ y_j^{(v)}\ is\ the\ vth\ view\ of\ the\ ith\ sample \\ 0, & otherwise \end{cases} \tag{6}$$

where $y_j^{(v)}$ is the jth instance of the available instance set $Y^{(v)}$ from the vth view.

Between-view inferring: For the incomplete data, due to the lack of similarity information of the missing instances and available instances, it is obvious impossible to obtain the complete graphs by only exploring the within-view information. Fortunately, the multi-view data contains many complementary information among views. Moreover, there must be some connected relationships between the sample and the other samples in at least one view for the incomplete multi-view data. This demonstrates that it is possible to infer the missing rows and columns corresponding to the missing views in graphs by borrowing the similarity information from the other views [42]. Inspired by this motivation, we design the following sparse representation based between-view inferring model for recovering the incomplete graphs:

$$\min_{S^{(v)},B} \sum_{v=1}^{l} \left\| S^{(v)} - \sum_{i=1,i\neq v}^{l} S^{(i)} B_{i,v} \right\|_F^2 \\ s.t.\, 0 \le S^{(v)} \le 1, S^{(v)T}\mathit{1} = \mathit{1}, S_{i,i}^{(v)} = 0, \\ 0 \le B_{i,v} \le 1, \sum_{i=1,i\neq v}^{l} B_{i,v} = 1, B_{v,v} = 0 \tag{7}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2020.3013408, IEEE Transactions on Multimedia

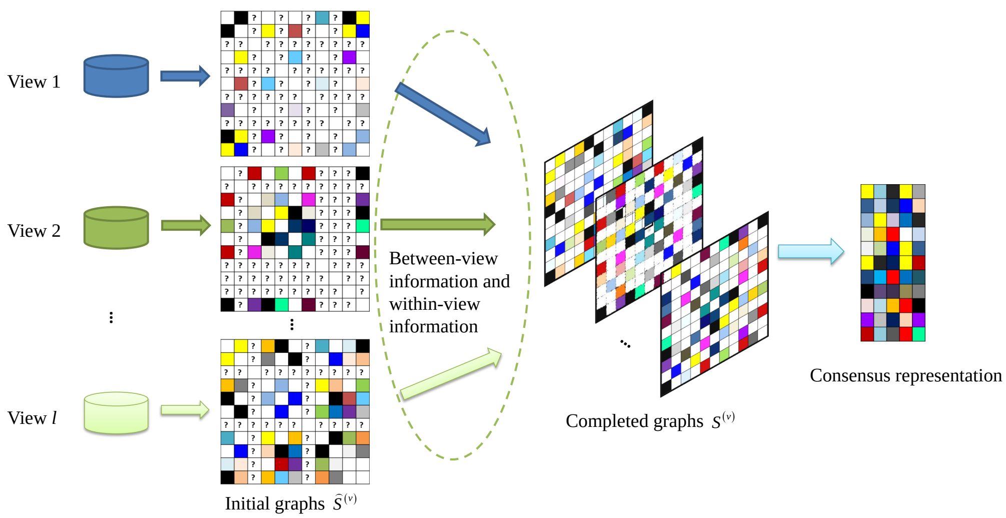


Fig. 1. The framework of the proposed method. Our method simultaneously predicts the missing graph rows/columns and learns the consensus representation by exploring the between-view information and within-view information of multiple views.

where $B \in R^{l\times l}$ can be viewed as the linear regression (or self-representation) matrix. $\mathit{1} \in R^{n\times 1}$ is a column vector with all elements as one. $S_{i,i}^{(v)} = 0$ and $B_{v,v} = 0$ represent that all diagonal elements of matrix $S^{(v)}$ and B are zero. $\{0 \le B_{i,v} \le 1, \sum_{i=1,i\neq v}^{l} B_{i,v} = 1, B_{v,v} = 0\}$ and $\{0 \le S^{(v)} \le 1, S^{(v)T}\mathit{1} = \mathit{1}, S_{i,i}^{(v)} = 0\}$ are boundary constraints. Model (7) is able to adaptively select the most reliable information of the other views for graph completion such that more precise similarity graphs can be achieved.

Consensus representation learning: When the graphs are recovered, many conventional methods can be chosen to learn the consensus representation [43-45]. For instance, co-regularized multi-view spectral clustering indirectly learns a consensus representation from the similarity graphs of all views [41]. Robust auto-weighted multi-view clustering tries to learn a consensus graph from all pre-constructed graphs [45]. In our work, we adopt the following model to learn the consensus representation shared by all views directly [44]:

$$\min_{U} \sum_{v=1}^{l} Tr\left(U^T L_{S^{(v)}} U\right) \ s.t.\, U^T U = I \tag{8}$$

where $U \in R^{n\times c}$ is the consensus representation, c is the manual selected dimension which is generally chosen as the cluster number. $L_{S^{(v)}}$ is the Laplacian matrix of graph $S^{(v)}$. In our method, since $S^{(v)}$ computed in our work is not a symmetric matrix, $L_{S^{(v)}}$ is computed as $L_{S^{(v)}} = D^{(v)} - \left(S^{(v)} + S^{(v)}\right)/2$, where $D^{(v)}$ is a diagonal matrix whose ith diagonal element is computed as $D_{i,i}^{(v)} = \sum_{j=1}^{n} \left(S_{i,j}^{(v)} + S_{j,i}^{(v)}\right) \Big/ 2$.

Overall objective function: Both of the within-view information and between-view information are important to the restoration of the similarity graphs of all views. Therefore, we combine the above two kinds of information for graph completion and consensus representation learning. Considering that different views may contain different degrees of useful information, the following adaptively weighted learning framework is developed to integrate the above three components, *i.e.*, (4), (7), and (8):

$$\begin{aligned}
&\min_{S^{(v)},U,B,\alpha^{(v)}} \lambda_1 \sum_{v=1}^{l} \left(\alpha^{(v)}\right)^r \underbrace{\left\| S^{(v)} - \sum_{i=1,i\neq v}^{l} S^{(i)} B_{i,v} \right\|_F^2}_{within-view\ preservation} \\
&+ \sum_{v=1}^{l} \left(\alpha^{(v)}\right)^r \left(\underbrace{\left\| \left(S^{(v)} - \widetilde{S}^{(v)}\right) \odot W^{(v)} \right\|_F^2}_{between-view\ inferring} + \lambda_2 Tr\left(U^T L_{S^{(v)}} U\right) \right) \\
&s.t.\ 0 \le S^{(v)} \le 1, S^{(v)T}\mathit{1} = \mathit{1}, S_{i,i}^{(v)} = 0, U^T U = I, \\
&\quad 0 \le B_{i,v} \le 1, \sum_{i=1,i\neq v}^{l} B_{i,v} = 1, B_{v,v} = 0, \\
&\quad \sum_{v=1}^{l} \alpha^{(v)} = 1, 0 \le \alpha^{(v)} \le 1
\end{aligned} \tag{9}$$

where $W^{(v)} = E_{v,:}^T E_{v,:}$ is defined in (4). λ_1 and λ_2 are the penalty parameters to balance the importance of the corresponding constraints. $\alpha^{(v)}$ is the weight to balance the importance of the vth view in the joint learning model. Smooth parameter $r > 1$ controls the distribution of weights $\alpha^{(1)}, \ldots, \alpha^{(l)}$.

Since learning model (9) adaptively recovers the incomplete graphs to address the incomplete multi-view clustering problem, we refer to the proposed method as Adaptive Graph Completion based Incomplete Multi-view Clustering (AGC_IMC). From the objective function (9), we can discover the following proposition.

Proposition 1: By optimization problem (9), the recovered graphs $S = \left\{S^{(1)}, \ldots, S^{(l)}\right\}$ of all views have the potential to possess exactly c connected components.

Proof: Please refer to the supplementary material for the detailed proof process of proposition 1.

As proved in many previous works, the optimal similarity graph should have exactly c connected components for the data with c clusters [46, 47]. Therefore, proposition 1 indicates that the proposed method is able to recover the graph with the optimal structure of every view, and thus has the potential to obtain a better clustering performance.

B. Solution to AGC_IMC

It is difficult to obtain the analytical solution of problem (9) since it has $3l+1$ variables to compute. In this section, we provide an alternative iterative optimization approach to find its local optimal solution [48-50]. Detailed optimization steps are presented as follows:

$S^{(v)}$***-Step***: Fixing the other variables, the optimization problem of (9) with respect to variable $S^{(v)}$ is degraded as:

$$\begin{aligned}&\min_{0\le S^{(v)}\le 1, S^{(v)T}\mathit{1}=\mathit{1}, S^{(v)}_{i,i}=0} \lambda_1 \sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r \left\|S^{(v)}-\sum_{i=1,i\neq v}^{l} S^{(i)}B_{i,v}\right\|_F^2\\ &+\left(\alpha^{(v)}\right)^r\left(\left\|\left(S^{(v)}-\widetilde{S}^{(v)}\right)\odot W^{(v)}\right\|_F^2+\lambda_2 Tr\left(U^T L_{S^{(v)}}U\right)\right)\end{aligned} \tag{10}$$

The first term of problem (10) can be transformed as:

$$\begin{aligned}&\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r\left\|S^{(v)}-\sum_{i=1,i\neq v}^{l}S^{(i)}B_{i,v}\right\|_F^2\\ &=\sum_{i=1,i\neq v}^{l}\left(\alpha^{(i)}\right)^r\left\|B_{v,i}S^{(v)}-\left(S^{(i)}-\sum_{j=1,j\neq v,j\neq i}^{l}S^{(j)}B_{j,i}\right)\right\|_F^2\\ &\quad+\left(\alpha^{(v)}\right)^r\left\|S^{(v)}-\sum_{i=1,i\neq v}^{l}S^{(i)}B_{i,v}\right\|_F^2\end{aligned} \tag{11}$$

Defining $N^{(i)}=S^{(i)}-\sum_{j=1,j\neq v,j\neq i}^{l}S^{(j)}B_{j,i}$, $M^{(v)}=\sum_{i=1,i\neq v}^{l}S^{(i)}B_{i,v}$, $Z_{i,j}=\|U_{i,:}-U_{j,:}\|_2^2$, and $\psi\left(S^{(v)}\right)=\left\{0\le S^{(v)}\le 1, S^{(v)T}\mathit{1}=\mathit{1}, S^{(v)}_{i,i}=0\right\}$, problem (10) can be simplified as follows:

$$\begin{aligned}&\min_{\psi\left(S^{(v)}\right)}(\alpha^{(v)})^r\left(\left\|\left(S^{(v)}-\widetilde{S}^{(v)}\right)\odot W^{(v)}\right\|_F^2+\frac{\lambda_2}{2}\sum_{i,j}^{n}Z_{i,j}S^{(v)}_{i,j}\right)\\ &\quad+\sum_{i=1,i\neq v}^{l}(\alpha^{(i)})^r(B_{v,i})^2\lambda_1\left\|S^{(v)}-N^{(i)}\big/B_{v,i}\right\|_F^2\\ &\quad+\lambda_1\left(\alpha^{(v)}\right)^r\left\|S^{(v)}-M^{(v)}\right\|_F^2\\ \Leftrightarrow&\min_{0\le S^{(v)}\le 1, S^{(v)T}\mathit{1}=\mathit{1}, S^{(v)}_{i,i}=0}\sum_{i,j=1}^{n}\left(S^{(v)}_{i,j}-T^{(v)}_{i,j}\right)^2\end{aligned} \tag{12}$$

where $T^{(v)}_{i,j}=\frac{P^{(v)}_{i,j}}{\left(\alpha^{(v)}\right)^r W^{(v)}_{i,j}+\sum_{k=1,k\neq v}^{l}\left(\alpha^{(k)}\right)^r(B_{v,k})^2\lambda_1+\lambda_1\left(\alpha^{(v)}\right)^r}$, $P^{(v)}_{i,j}=\sum_{k=1,k\neq v}^{l}\left(\alpha^{(k)}\right)^r B_{v,k}\lambda_1 N^{(k)}_{i,j}+\lambda_1\left(\alpha^{(v)}\right)^r M^{(v)}_{i,j}+\left(\alpha^{(v)}\right)^r\left(\widetilde{S}^{(v)}_{i,j}W^{(v)}_{i,j}-\frac{\lambda_2}{4}Z_{i,j}\right)$.

Problem (12) is independent with respect to all columns, and thus we can optimize problem (12) column by column. The optimal solution of (12) can be expressed as follows [27, 47]:

$$S^{(v)}_{i,j}=\begin{cases}\left(T^{(v)}_{i,j}+\eta_i\right)_+, & i\neq j\\ 0, & i=j\end{cases} \tag{13}$$

where function $(A)_+=\max(A,0)$ ensures all elements of A to be non-negative. According to the constraints $S^{(v)T}\mathit{1}=\mathit{1}$ and $S^{(v)}_{i,i}=0$, we can obtain the optimal solution of η_j as follows:

$$\eta_j=\left(1-\sum_{i=1,i\neq j}^{n}T^{(v)}_{i,j}\right)\Big/(n-1) \tag{14}$$

U***-Step***: Fixing variables $\{S^{(v)},B,\alpha^{(v)}\}$ in problem (9), the sub-optimization problem with respect to variable U can be expressed as:

$$\min_{U^TU=I}\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r\left(Tr\left(U^TL_{S^{(v)}}U\right)\right) \tag{15}$$

Problem (15) is a typical eigenvalue decomposition problem. Suppose $u_1,u_2,\ldots,u_c$ are the eigenvectors corresponding to the first c minimum eigenvalues of matrix $\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^rL_{S^{(v)}}$, the optimal solution to problem (15) is expressed as $U=[u_1,u_2,\ldots,u_c]\in R^{n\times c}$.

B***-Step***: From (9), the optimization problem with respect to variable B can be expressed as follows:

$$\min_{0\le B_{i,v}\le 1, \sum_{i=1,i\neq v}^{l}B_{i,v}=1, B_{v,v}=0}\sum_{v=1}^{l}\left\|S^{(v)}-\sum_{i=1,i\neq v}^{l}S^{(i)}B_{i,v}\right\|_F^2 \tag{16}$$

Let $G\in R^{n^2\times l}$ be a matrix formed by $\left\{S^{(v)}\right\}_{v=1}^{l}$, whose vth column is the vector stacked by all columns of matrix $S^{(v)}$, then (16) can be transformed into the following problem:

$$\min_{0\le B_{i,v}\le 1, \sum_{i=1,i\neq v}^{l}B_{i,v}=1, B_{v,v}=0}\sum_{v=1}^{l}\|G_{:,v}-GB_{:,v}\|_2^2 \tag{17}$$

Obviously, problem (17) can be transformed into l independent optimization sub-problems. Therefore, we can calculate the vth column of matrix B by solving the following problem:

$$\min_{0\le B_{i,v}\le 1, \sum_{i=1,i\neq v}^{l}B_{i,v}=1, B_{v,v}=0}\|G_{:,v}-GB_{:,v}\|_2^2 \tag{18}$$

Problem (18) is a typical simplex representation based optimization problem and can be fast solved via the accelerated projected gradient method [51]. For detailed optimization process of problem (18), please refer to [51].

$\alpha^{(v)}$***-Step***: Fixing variables $\{S^{(v)},B,U\}$ and defining $d^{(v)}=\left\|\left(S^{(v)}-\widetilde{S}^{(v)}\right)\odot W^{(v)}\right\|_F^2+\lambda_1\left\|S^{(v)}-\sum_{i=1,i\neq v}^{l}S^{(i)}B_{i,v}\right\|_F^2+\lambda_2Tr\left(U^TL_{S^{(v)}}U\right)$, problem (9) can be degraded into the following minimization problem with respect to variable $\alpha^{(v)}$:

$$\begin{aligned}&\min_{\alpha^{(v)}}\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^r d^{(v)}\\ &s.t.\ \sum_{v=1}^{l}\alpha^{(v)}=1, 0\le\alpha^{(v)}\le 1\end{aligned} \tag{19}$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2020.3013408, IEEE Transactions on Multimedia

The Lagrange function of problem (19) is formulated as follows [3]:

$$\Psi\left(\alpha^{(v)},\beta\right)=\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^{r}d^{(v)}-\beta\left(\sum_{v=1}^{l}\alpha^{(v)}-1\right)\tag{20}$$

where β is the Lagrange multiplier. The partial derivative of Ψ with respect to variable $\alpha^{(v)}$ is:

$$\frac{\partial\Psi}{\partial\alpha^{(v)}}=r\left(\alpha^{(v)}\right)^{r-1}d^{(v)}-\beta\tag{21}$$

Let $\partial\Psi/\partial\alpha^{(v)}=0$, we have:

$$\alpha^{(v)}=\left(\beta\Big/\left(rd^{(v)}\right)\right)^{1/(r-1)}\tag{22}$$

According to the constraint $\sum_{v=1}^{l}\alpha^{(v)}=1$, we can obtain the optimal solution of $\alpha^{(v)}$ as follows:

$$\alpha^{(v)}=\left(d^{(v)}\Big/\sum_{v=1}^{l}d^{(v)}\right)^{1/(1-r)}\tag{23}$$

The above optimization processes are summarized in **Algorithm 1**, where the convergence condition is set as $|Obj(t)-Obj(t-1)|<10^{-6}$ and the maximum iteration is set as 50. t denotes the iteration step and the objective function value Obj is calculated as:

$$Obj=\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^{r}\left(\begin{array}{l}\left\|\left(S^{(v)}-\widetilde{S}^{(v)}\right)\odot W^{(v)}\right\|_F^2\\+\lambda_1\left\|S^{(v)}-\sum\limits_{i=1,i\neq v}^{l}S^{(i)}B_{i,v}\right\|_F^2\\+\lambda_2Tr\left(U^TL_{S^{(v)}}U\right)\end{array}\right).$$

Algorithm 1 : AGC_IMC (solving (9))

Input: Multi-view data $\left\{X^{(v)}\right\}_{v=1}^{l}$, index matrix $E\in R^{l\times n}$, parameters λ_1, λ_2, and r.
Initialization: Construct the similarity graph $\left\{\bar{S}^{(v)}\right\}_{v=1}^{l}$ from every view, then exploit it to achieve graph $\left\{\widetilde{S}^{(v)}\right\}_{v=1}^{l}$ based on the view missing information. $U\in R^{n\times c}$ is an orthogonal matrix initialized by (15). $\alpha^{(v)}=1/l$.
while not converged **do**
1. Update variable $\left\{S^{(v)}\right\}_{v=1}^{l}$ via (13);
2. Update variable U by solving (15);
3. Update variable B by solving (18);
4. Update variable $\left\{\alpha^{(v)}\right\}_{v=1}^{l}$ via (23);

end while
Output: U

C. Computational complexity analysis

In the previous section, we have provided an alternating iterative optimization approach to solve objective problem (9). For the first step, *i.e.*, $S^{(v)}$-step, we can find that this step only contains some element-wise based operations, and thus the computational cost of this step can be ignored. The second step computes the representation U, where the most computational cost is consumed by the eigenvalue decomposition. Generally, the computational complexity of the eigenvalue decomposition on an $n\times n$ matrix is about $O\left(n^3\right)$. Fortunately, for problem (15), we only need to obtain c eigenvectors corresponding to the first c minimum eigenvalues of matrix $\sum_{v=1}^{l}\left(\alpha^{(v)}\right)^{r}L_{S^{(v)}}$ rather than calculating its all eigenvalues and eigenvectors. Therefore, we can adopt a more efficient function 'eigs' [52] to speed up the computational efficiency, which only costs $O\left(cn^2\right)$. Hence, the computational complexity of the second step is about $O\left(cn^2\right)$. For the third step, *i.e.*, B-step, an efficient projected gradient algorithm proposed in [51] is adopted to solve the optimization sub-problem (18). As presented in [51], the projected gradient algorithm also contains only the element-wise based vector addition and subtraction operations. So we can also ignore the computational cost of this step. For the last step, *i.e.*, $\alpha^{(v)}$-step, it is obvious that the optimal solution of the corresponding problem (19) can be simply computed via the numerical division operation. Therefore, the computational complexity of the fourth step can be also ignored. Based on the above analysis, the computational complexity of the exploited optimization approach in Algorithm 1 is about $O\left(\tau cn^2\right)$, where τ denotes the iteration number.

IV. Experiments and Analyses

In this section, we aim to verify the effectiveness of the proposed AGC_IMC through the comparison with state-of-the-art IMC methods. Moreover, we also conduct several experiments to analyze the parameter sensitivity and convergence property of AGC_IMC. For the proposed method, we simply initialize the similarity graph $\left\{\bar{S}^{(v)}\right\}_{v=1}^{l}$ as k-nearest-neighbor (KNN) graph for every view. *The code of our AGC_IMC is released at*: https://sites.google.com/view/jerry-wen-hit/publications.

A. Dataset description and incomplete multi-view data construction

Five multi-view datasets listed in Table I are chosen to validate the proposed method:

(i) **BBCSport** [53]: BBCSport is a document dataset which comprises of 737 news articles related to five sports (*i.e.*, athletics, cricket, football, rugby and tennis) from the BBC Sport website in 2004-2005. Following the experimental settings in [27], we adopt a subset[1] with four views of BBC sport multi-view datasets to evaluate different IMC methods. The subset includes 116 samples and the feature dimensions of different views are 1991, 2063, 2113, and 2158, respectively [54].

(ii) **3Sources**[2]: 3Sources is a well-known multi-view text dataset. It collects 416 distinct news stories covering six topical areas from three online news sources, *i.e.*, BBC, Reuters, and The Guardian, from the period of February-April, 2009, where 169 stories are simultaneously reported in the above three news sources. Article from each news source can be regarded as one view for the story. In our experiments, the subset with 169 stories reported in all three views is chosen to evaluate different IMC methods.

[1] https://github.com/GPMVCDummy/GPMVC/tree/master/partialMV/PVC/recreateResults/data.

[2] http://erdos.ucd.ie/datasets/3sources.html.

TABLE I
DESCRIPTION OF THE MULTI-VIEW DATASETS.

Dataset	# Class	# View	# Samples	# Features
BBCSport	5	4	116	1991/2063/2113/2158
3Sources	6	3	169	3560/3631/3068
Handwritten	10	6	2000	240/76/216/47/64/6
Caltech20	20	6	2386	48/40/254/1984/512/928
Animal	50	2	10158	4096/4096

(iii) **Handwritten**[3] [55]: This dataset contains 10 digits, *i.e.*, 0-9, where each digit has 200 handwritten images. Six kinds of features, *i.e.*, pixel averages, Fourier coefficients, profile correlations, Zernike moment, Karhunen-love coefficient, and morphological, are extracted from every sample as six views, where the feature dimensions are 240, 76, 216, 47, 64, and 6, respectively.

(iv) **Caltech-101** [56]: The original Caltech-101 dataset contains 101 objects and each object has about 40-800 images. In our experiments, a subset, referred to as **Caltech20** which contains 20 objects and 2386 samples is adopted [55]. Following [55], six kinds of features, *i.e.*, Gabor, wavelet moments, CENTRIST, HOG, GIST, and LBP, are extracted from all images as six views. The feature dimensions of the above views for each instance are 48, 40, 254, 1984, 512, and 928, respectively.

(v) **Animal** [56]: The multi-view Animal dataset released by Zhang et al. [57] is chosen to evaluate the proposed method. There are 10158 images provided by 50 classes in the dataset, where each image is represented by two kinds of features extracted by DECAF [58] and VGG19 [32], respectively.

For the BBCSport and 3Sources datasets, under the condition that each sample contains at least one instance, we randomly remove 10%, 30%, and 50% instances from every view to construct the incomplete multi-view dataset with different missing-view rates. Similarly, for Handwritten and Caltech20 datasets, 30%, 50%, and 70% instances are randomly removed from every view to construct the incomplete multi-view datasets. For the Animal dataset, $p\%$ ($p = \{30, 50, 70\}$) samples are randomly selected as the paired-samples whose views are fully observed. Then we randomly remove the first view for half of the remaining samples and remove the second view for the other half of samples. In this way, the incomplete Animal dataset with $p\%$ paired-samples is constructed.

B. Compared methods and evaluation metric

The following methods that can handle the incomplete multi-view cases are selected as baselines:

(i) Best single view (BSV) [21]: BSV implements k-means on all views separately, and then reports the best clustering result of these views. For BSV, all missing instances are filled in the average instance of every view.

(ii) Concat [21]: Concat stacks multiple views into a single view by integrating features of all views into a long feature vector, then implements the k-means on the stacked single view and reports the clustering result. For Concat, the missing instances of every view are also filled in the average of available instances in the corresponding view as BSV.

[3]https://archive.ics.uci.edu/ml/datasets/Multiple+Features.

(iii) Graph regularized partial multi-view clustering (GPMVC) [54]: By integrating the instance-missing information into the multi-view matrix factorization model, GPMVC obtains a common representation shared by all views indirectly from the representations derived from different views. Specially, GPMVC introduces the graph constraint to exploit the local geometric structure of data for representation learning.

(iv) MIC [23]: MIC designs a weighted multi-view matrix factorization framework to learn the consensus representation for all views, where the instance-missing information are constrained as the weight to avoid the negative influence of missing views.

(v) DAIMC [25]: DAIMC learns a consensus representation for all views by adopting two main techniques, *i.e.*, instance information alignment based weighted matrix factorization and basis matrices alignment based sparse regression.

(vi) OMVC [24]: Similar to MIC, OMVC also designs a weighted non-negative matrix factorization framework to learn the consensus representation for all incomplete views. As an extension of MIC, it provides a chunk by chunk training approach to improve the efficiency on large-scale datasets.

(vii) OPIMC [26]: OPIMC tries to learn a consensus representation with binary elements (0 and 1) via the weighted joint matrix factorization model. It also provides an one-pass based chunk by chunk training approach to improve the clustering efficiency.

In our experiment, we adopt seven well-known indicators, *i.e.*, accuracy (Acc), normalized mutual information (NMI), purity, adjusted Rand index (AR), F-score, precision, and recall as the evaluation metrics to compare these IMC methods [59-61]. For the above seven metrics, a higher value means relative better clustering performance. For fairly comparing, we run the above methods several times with respect to different view missing groups, and then collect their average values (%). In addition, all compared methods are implemented with a wide parameter ranges and their best performances are reported.

C. Experimental results and analysis

Table II-Table III, and Fig. 2-Fig. 6 show the experimental results of different IMC methods on the five multi-view datasets with different missing-view or paired-sample rates. The comparison of the clustering results of different methods in these figures and tables reflects the following points:

1) Obviously, the proposed method outperforms all the other methods on the five multi-view datasets in terms of all the seven clustering evaluation metrics. For instance, from Table II, on the BBCsport dataset, the proposed method obtains about 13% improvement of Acc in comparison with the second best method. The experimental results listed in Table III show that on the Caltech20 dataset, the NMI obtained by the proposed method is about 4 percent higher than that of the second best method.

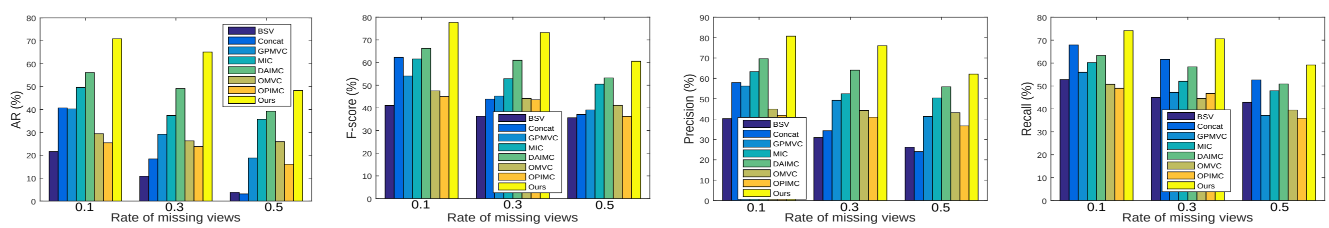
2) DAIMC and the proposed method perform better than the other methods on the five datasets in most cases. Among these methods, the proposed method and DAIMC commonly focus on exploiting more information from data to guide the consensus representation learning. In particular, DAIMC tries

TABLE II
ACCS (%), NMIS (%), PURITIES (%) OF DIFFERENT IMC METHODS ON THE BBCSPORT AND 3SOURCE DATASETS WITH DIFFERENT MISSING RATES OF VIEWS. BOLD NUMBERS DENOTE THE BEST RESULT.

		Acc (%)			NMI (%)			Purity (%)		
Dataset	Method\Rate	0.1	0.3	0.5	0.1	0.3	0.5	0.1	0.3	0.5
BBCSport	BSV	58.62±3.94	51.31±5.33	44.03±3.78	43.73±7.43	31.03±2.08	21.40±2.61	65.79±5.52	55.07±1.51	47.59±2.28
	Concat	70.62±3.76	58.72±5.42	33.21±2.19	61.69±6.72	38.92±7.87	18.61±1.44	80.59±4.59	63.24±5.82	37.00±1.54
	GPMVC	51.44±8.20	46.89±5.01	43.91±6.31	28.23±10.31	20.04±7.39	15.48±4.54	58.39±8.58	52.76±5.60	45.29±5.41
	MIC	51.21±4.21	46.21±4.71	46.03±5.19	29.90±6.25	25.84±3.24	24.01±5.39	55.00±4.15	51.72±4.27	52.41±6.23
	DAIMC	68.62±4.59	63.45±10.97	56.89±5.59	56.62±4.60	50.17±9.91	37.89±6.22	76.90±5.89	71.72±10.76	61.03±5.08
	OMVC	53.33±3.21	51.38±3.06	48.79±3.10	30.64±2.00	41.57±2.79	40.63±2.45	56.49±2.81	59.20±2.12	57.47±2.80
	OPIMC	54.14±4.78	52.93±4.93	45.69±6.00	35.66±4.71	31.56±6.10	21.75±6.44	58.28±4.82	56.72±5.76	50.86±6.87
	Ours	**83.10±5.74**	**80.17±3.19**	**70.86±6.14**	**73.19±4.73**	**67.79±4.88**	**52.41±5.92**	**86.03±2.08**	**83.79±3.83**	**76.03±4.54**
3Sources	BSV	56.90±3.69	47.38±3.07	39.24±3.08	50.07±1.22	34.46±4.07	22.34±1.91	68.14±1.67	57.63±1.32	48.99±0.63
	Concat	53.54±3.00	46.79±3.99	37.68±2.91	51.98±1.37	37.87±3.66	18.32±3.25	69.78±1.09	58.51±3.18	46.48±2.82
	GPMVC	48.24±6.73	44.50±9.65	42.01±9.97	34.82±9.55	30.44±10.63	28.15±6.09	60.47±5.32	58.58±7.13	57.40±4.64
	MIC	49.11±3.60	47.69±7.61	42.49±8.63	37.23±6.13	38.62±3.81	26.08±7.42	57.28±3.36	61.30±4.28	52.31±4.96
	DAIMC	56.33±4.23	52.43±6.63	50.73±3.87	52.98±3.65	49.07±5.78	41.64±2.43	68.99±4.26	67.21±4.89	63.56±3.38
	OMVC	43.95±7.35	41.11±4.31	39.53±3.63	36.48±10.77	28.42±3.41	24.34±1.50	59.37±8.26	48.76±5.44	45.44±3.10
	OPIMC	55.73±2.85	54.20±4.48	43.08±6.98	40.62±2.28	38.83±3.86	22.69±3.83	64.73±1.70	64.26±2.03	53.61±4.36
	Ours	**77.63±0.87**	**71.60±5.48**	**68.16±4.09**	**68.84±1.71**	**61.53±3.84**	**51.59±3.92**	**83.33±0.49**	**77.87±3.68**	**73.73±1.85**

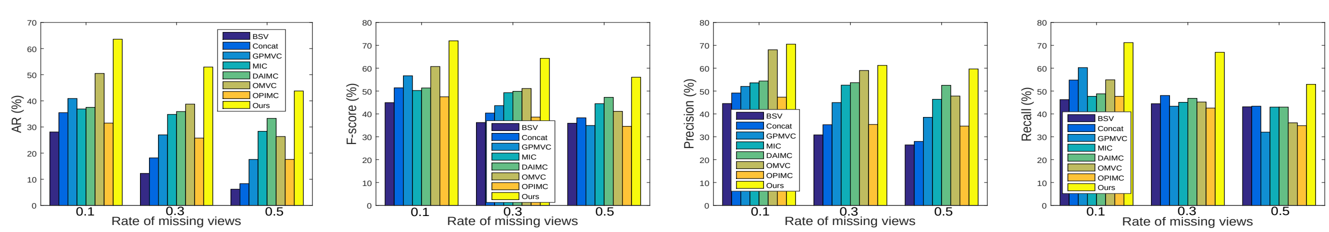
TABLE III
ACCS (%), NMIS (%), PURITIES (%) OF DIFFERENT IMC METHODS ON THE HANDWRITTEN, CALTECH20, AND ANIMAL DATASETS WITH DIFFERENT MISSING RATES OF VIEWS. BOLD NUMBERS DENOTE THE BEST RESULT.

		Acc (%)			NMI (%)			Purity (%)		
Dataset	Method\Rate	0.3	0.5	0.7	0.3	0.5	0.7	0.3	0.5	0.7
Handwritten	BSV	60.80±9.94	41.18±5.05	26.57±2.31	51.44±8.47	35.06±5.22	19.93±2.62	57.76±8.91	42.11±4.71	27.81±2.26
	Concat	61.40±1.10	45.25±0.94	30.85±0.67	55.52±0.92	39.83±1.04	25.13±0.68	61.72±0.76	45.92±0.93	31.74±0.59
	GPMVC	47.03±2.92	34.39±4.82	25.70±1.43	38.59±4.92	26.06±3.39	15.84±1.39	48.55±5.69	35.40±3.21	27.22±2.28
	MIC	53.34±5.99	41.01±2.19	24.87±2.07	48.37±3.89	33.66±3.11	16.64±1.71	55.10±3.84	41.86±2.39	25.90±1.87
	DAIMC	82.79±2.25	78.39±1.11	55.89±5.37	71.80±2.84	64.05±1.89	41.03±3.08	82.79±2.25	78.39±1.11	56.03±5.27
	OMVC	54.53±3.72	39.46±4.97	31.32±2.06	45.51±1.66	30.45±4.03	22.08±2.35	55.23±3.50	40.97±.80	33.34±2.40
	OPIMC	76.22±3.82	72.25±7.18	63.53±8.06	72.67±1.70	66.61±4.77	55.03±4.32	78.08±2.80	74.02±6.61	65.79±6.89
	Ours	**85.73±1.93**	**83.88±1.63**	**82.25±4.03**	**85.64±0.80**	**82.91±2.09**	**73.14±2.53**	**86.75±0.56**	**84.82±1.52**	**82.64±3.46**
Caltech20	BSV	34.66±1.34	34.69±1.18	34.42±1.02	43.99±1.09	32.60±0.95	19.82±0.43	62.43±0.85	54.22±0.95	46.29±0.58
	Concat	38.89±0.97	29.46±0.66	23.72±0.59	45.57±0.85	36.28±0.67	26.51±0.83	68.37±0.69	60.36±0.45	52.06±0.59
	GPMVC	33.89±1.77	25.69±1.85	19.98±2.14	38.58±1.11	31.14±1.67	23.28±1.88	61.21±2.21	57.35±1.86	51.33±1.44
	MIC	35.37±2.32	28.15±2.09	24.08±0.96	44.41±0.70	36.07±1.14	26.68±0.74	69.08±0.62	61.81±0.93	52.24±1.38
	DAIMC	47.44±1.28	44.77±2.28	37.01±1.63	54.20±0.56	50.39±1.55	36.24±0.94	72.04±0.88	70.93±1.10	63.56±1.48
	OMVC	35.20±3.92	34.59±1.07	38.09±2.76	35.86±1.50	35.00±0.94	39.59±1.27	62.75±1.76	49.57±4.30	48.90±2.28
	OPIMC	56.05±7.50	53.13±4.21	38.59±6.79	36.42±4.37	32.00±3.52	21.88±4.36	60.11±2.93	57.53±1.81	51.28±3.05
	Ours	**59.74±1.50**	**56.97±1.15**	**48.73±2.41**	**58.28±0.79**	**54.20±1.45**	**45.69±1.32**	**74.55±0.84**	**72.48±1.25**	**66.77±0.60**
Animal	BSV	42.05±1.20	48.63±1.89	56.22±1.20	48.16±0.44	55.91±0.58	63.99±0.38	45.20±0.88	52.26±1.19	60.31±0.78
	Concat	42.79±0.67	49.34±1.39	53.99±0.99	55.46±0.16	59.31±0.38	63.88±0.35	48.12±0.45	53.24±0.88	59.26±0.81
	GPMVC	43.10±1.79	49.24±1.42	54.77±1.24	47.02±0.57	52.23±0.53	58.48±0.17	46.78±1.24	52.77±0.81	59.31±0.74
	MIC	43.38±0.63	45.88±0.34	49.15±0.88	52.79±0.77	55.69±0.36	59.30±0.54	49.21±0.78	52.31±0.34	55.33±0.64
	DAIMC	50.18±2.18	53.87±1.36	56.42±1.37	55.03±1.03	59.36±1.16	62.76±0.46	54.82±1.57	59.51±1.65	62.12±1.04
	OMVC	42.51±0.89	43.98±0.77	46.39±1.02	50.77±0.63	53.11±0.83	55.38±0.46	47.33±0.66	50.42±0.91	52.97±0.76
	OPIMC	46.33±2.14	53.14±1.38	53.88±1.26	52.34±0.69	58.51±0.46	62.04±0.26	49.49±1.41	56.23±1.20	57.91±0.43
	Ours	**57.19±0.89**	**60.11±2.01**	**62.59±1.72**	**63.97±0.58**	**66.66±0.74**	**68.26±1.06**	**63.66±0.52**	**66.39±0.93**	**67.70±0.77**



(a) AR (b) F-score (c) Precision (d) Recall

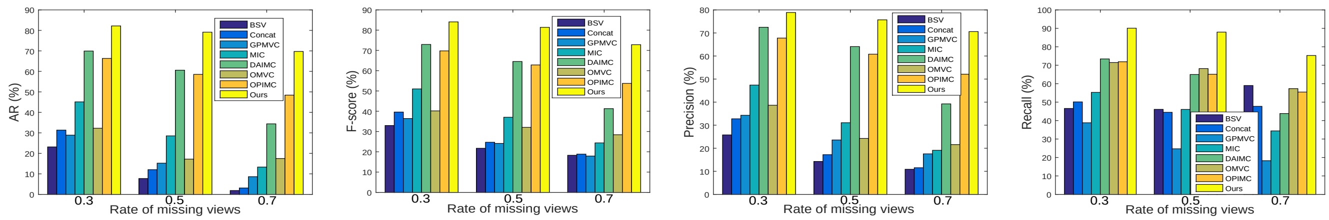
Fig. 2. Experimental results with respect to (a) AR (%), (b) F-score (%), (c) precision (%), and (d) Recall (%) of different IMC methods on the BBCSport dataset with different missing rates of views.



(a) AR (b) F-score (c) Precision (d) Recall

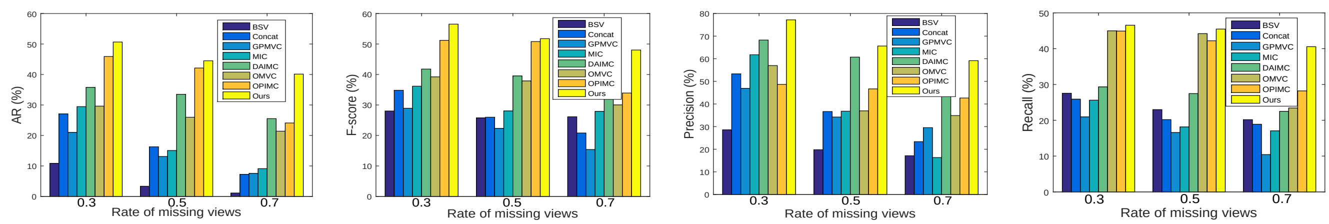
Fig. 3. Experimental results with respect to (a) AR (%), (b) F-score (%), (c) precision (%), and (d) Recall (%) of different IMC methods on the 3Sources dataset with different missing rates of views.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2020.3013408, IEEE Transactions on Multimedia



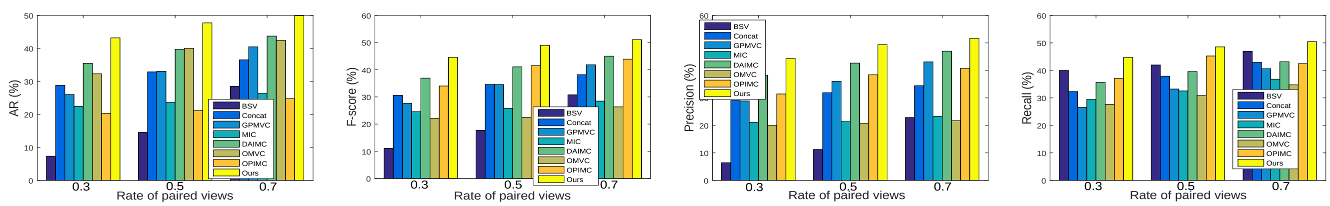
(a) AR (b) F-score (c) Precision (d) Recall

Fig. 4. Experimental results with respect to (a) AR (%), (b) F-score (%), (c) precision (%), and (d) Recall (%) of different IMC methods on the Handwritten dataset with different missing rates of views.



(a) AR (b) F-score (c) Precision (d) Recall

Fig. 5. Experimental results with respect to (a) AR (%), (b) F-score (%), (c) precision (%), and (d) Recall (%) of different IMC methods on the Caltech20 dataset with different missing rates of views.



(a) AR (b) F-score (c) Precision (d) Recall

Fig. 6. Experimental results with respect to (a) AR (%), (b) F-score (%), (c) precision (%), and (d) Recall (%) of different IMC methods on the Animal dataset with different rates of paired-samples.

to explore more complementary information of multiple views by introducing a novel alignment constraint. The proposed method tries to recover yet exploit the similarity information of the missing instances and available instances for the consensus representation learning. However, the other methods, such as MIC, OMVC and OPIMC, only capture the consistency information shared by all views via the simple weighted matrix factorization technique. Therefore, the experimental results demonstrate the effectiveness of capturing more complementary and consistency information of data in multi-view clustering cases.

3) From the experimental results shown in these tables and figures, it can be observed that when the missing rate of views increases, the clustering performance of all methods in terms of all evaluation metrics commonly decreases. This phenomenon illustrates that it is difficult to learn the reasonable consensus representation shared by all views from the multi-view data with a high missing-view rate. This is mainly because that under the case of large missing rate, the multi-view data losses much consistence information and complementary information which are harmful to the MVC.

In Fig. 7, the weight α obtained by the proposed method on the Handwritten dataset with different missing-view rates are plotted. It is obvious that different views are set with different weight values adaptively. This demonstrates that the proposed method can sufficiently consider the full information of multi-views and effectively reduce the negative influence of information imbalance problem.

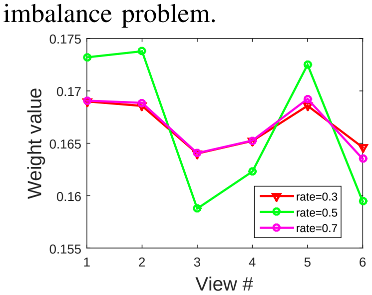


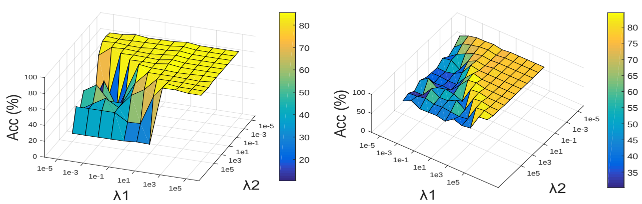
Fig. 7. Weight α with respect to different views of the proposed method on the Handwritten dataset with different missing-view rates.

D. Sensitivity analysis of the penalty parameters

The proposed learning model has two penalty parameters λ_1 and λ_2, and a smooth parameter r. In this section, we analyze the sensitivities of these parameters in terms of the clustering accuracy.

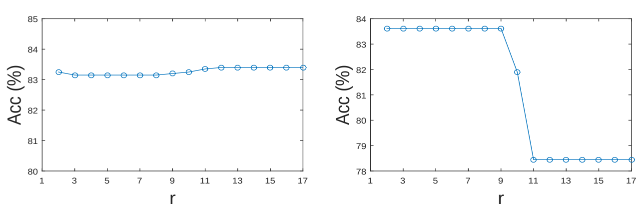
Parameters λ_1 and λ_2: We conduct experiments with different combinations of parameters λ_1 and λ_2 selected from a set $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4, 10^5\}$ on the Handwritten digit with 50% missing instances of every view and BBCSport dataset with 30% missing instance of every view. Experimental results of the proposed method on the above two datasets are shown in Fig.8. From this figure, we can find that the proposed method can obtain a relative good clustering performance when $\lambda_2 \leq 100\lambda_1$ on the Handwritten digit dataset. Specially, on the BBCSport dataset, the proposed method obtains the best performance when the values of the two parameters satisfy $\lambda_2 = \lambda_1 \geq 0.1$. The experimental results demonstrate that it is easy to choose the penalty parameters of the proposed method. In the experiments of the previous section, we experimentally select the value of parameter λ_1 and λ_2 from the set of $\{10^{-1}, 1, 10, 10^2\}$.

Parameter r: Fig.9 shows the Acc (%) of the proposed method versus smooth parameter r on the Handwritten dataset with 50% missing instances of every view and BBCSport dataset with 30% missing instances of every view. From Fig.9, we can find that the proposed method is insensitive to the values of parameter r on the Handwritten dataset to some extent and can obtain a relative good performance on the BBCSport dataset when parameter r is selected from the range of $[2, 9]$. According to the experimental results in Fig.9, we can simply select parameter r from $[2, 9]$ for clustering.



(a) Handwritten (b) BBCSport

Fig. 8. Acc (%) versus parameters λ_1 and λ_2 of the proposed method on the (a) Handwritten digit dataset with 50% missing instances of every view and (b) BBCSport dataset with 30% missing instances of every view.



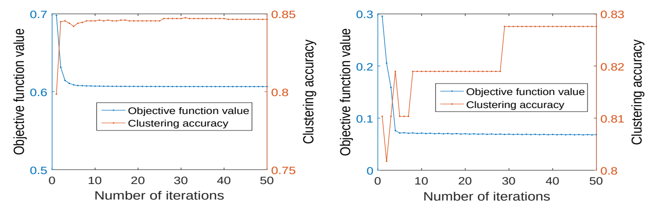
(a) Handwritten (b) BBCSport

Fig. 9. Acc (%) versus parameter r of the proposed method on the (a) Handwritten digit dataset with 50% missing instances of every view and (b) BBCSport dataset with 30% missing instances of every view.

E. Convergence analysis

Theoretical: From the optimization process presented in the previous section, objective function (9) with $3l+1$ variables is divided into four convex sub-problems and each sub-problem has the optimal solution. This indicates that the objective function value is monotonously non-increasing for each sub-problem. Thus, by optimizing all sub-problems, the objective function value of problem (9) is also monotonously non-increasing overall. In addition, we can find that the objective function value of problem (9) has a lower bound 0. Therefore, we can conclude that the objective function value of problem (9) can converge after some iteration steps via the utilized alternating iterative optimization approach [3].

Experimental: Fig.10 shows the objective function value and clustering accuracy versus the iteration on the BBCsport dataset with a missing-view rate of 30% and Handwritten dataset with a missing-view rate of 50%. From Fig.10, we can find the good convergence property of our provided optimization approach, where the objective function value fast decreases to the stationary point with the iteration increases.



(a) Handwritten (b) BBCSport

Fig. 10. The objective function value and clustering accuracy versus the number of iterations of the proposed method on the (a) Handwritten dataset with a missing-view rate of 50% and (b) BBCSport dataset with a missing-view rate of 30%.

V. Conclusion

In this paper, a novel method, called AGC_IMC is proposed for the MVC scenarios with missing views. Different from the existing methods, AGC_IMC borrows the idea of multi-view spectral clustering and jointly performs the graph completion and consensus representation learning in a unified framework. By dexterously fusing the within-view information and between-view information, AGC_IMC can infer the intrinsic connective information of the missing instances and the available instances, which is beneficial to obtain a more reasonable consensus representation for clustering. Extensive experimental results conducted on five multi-view datasets with different missing rates of instances show that AGC_IMC outperforms the compared state-of-the-art IMC methods.

References

[1] C. Tang, X. Zhu, X. Liu, M. Li, P. Wang, C. Zhang, and L. Wang, "Learning a joint affinity graph for multiview subspace clustering," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1724–1736, 2019.

[2] X. Wu, C. W. Ngo, and A. G. Hauptmann, "Multimodal news story clustering with pairwise visual near-duplicate constraint." *IEEE Transactions on Multimedia*, vol. 10, no. 2, pp. 188–199, 2008.

[3] Z. Zhang, L. Liu, F. Shen, H. T. Shen, and L. Shao, "Binary multi-view clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1774–1782, 2018.

[4] X. Liu, X. Zhu, M. Li, L. Wang, C. Tang, J. Yin, D. Shen, H. Wang, and W. Gao, "Late fusion incomplete multi-view clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 10, pp. 2410–2423, 2019.

[5] C. Tang, X. Zhu, X. Liu, and L. Wang, "Cross-view local structure preserved diversity and consensus learning for multi-view unsupervised feature selection," in *AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 5101–5108.

[6] J. Li, B. Zhang, G. Lu, and D. Zhang, "Generative multi-view and multi-feature learning for classification," *Information Fusion*, vol. 45, pp. 215–226, 2019.

[7] K. Yan, J. Wen, Y. Xu, and B. Liu, "Protein fold recognition based on auto-weighted multi-view graph embedding learning model," *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020.

[8] G. Luo, S. Dong, K. Wang, W. Zuo, S. Cao, and H. Zhang, "Multi-views fusion cnn for left ventricular volumes estimation on cardiac mr images," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, pp. 1924–1934, 2017.

[9] K. Yan, X. Fang, Y. Xu, and B. Liu, "Protein fold recognition based on multi-view modeling," *Bioinformatics*, vol. 35, no. 17, pp. 2982–2990, 2019.

[10] J. Li, G. Lu, B. Zhang, J. You, and D. Zhang, "Shared linear encoder-based multikernel gaussian process latent variable model for visual classification," *IEEE Transactions on Cybernetics*, 2019.

[11] G. Chao, S. Sun, and J. Bi, "A survey on multi-view clustering," *arXiv preprint arXiv:1712.06246*, 2017.

[12] S. K. Kuanar, K. B. Ranga, and A. S. Chowdhury, "Multi-view video summarization using bipartite matching constrained optimum-path forest clustering," *IEEE Transactions on Multimedia*, vol. 17, no. 8, pp. 1166–1173, 2015.

[13] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Transactions on Multimedia*, 2019.

[14] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multi-view clustering with flexible locality structure diffusion," *IEEE Transactions on Cybernetics*, 2020.

[15] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, and W. Gao, "Multiple kernel k-means with incomplete kernels," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 5, pp. 1191–1204, 2020.

[16] J. Guo and W. Zhu, "Partial multi-view outlier detection based on collective learning," in *AAAI Conference on Artificial Intelligence*, 2018, pp. 298–305.

[17] X. Liu, L. Wang, X. Zhu, M. Li, E. Zhu, T. Liu, L. Liu, Y. Dou, and J. Yin, "Absent multiple kernel learning algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 42, no. 6, pp. 1303–1316, 2020.

[18] S.-Y. Li, Y. Jiang, and Z.-H. Zhou, "Partial multi-view clustering," in *AAAI Conference on Artificial Intelligence*, 2014, pp. 1969–1974.

[19] A. Trivedi, P. Rai, H. Daumé III, and S. L. DuVall, "Multiview clustering with incomplete views," in *NIPS Workshop*, 2010, pp. 1–7.

[20] X. Liu, M. Li, C. Tang, J. Xia, J. Xiong, L. Liu, M. Kloft, and E. Zhu, "Efficient and effective regularized incomplete multi-view clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020.

[21] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping." in *International Joint Conference on Artificial Intelligence*, 2016, pp. 2392–2398.

[22] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, "Unified embedding alignment with missing views inferring for incomplete multi-view clustering," in *AAAI Conference on Artificial Intelligence*, 2019, pp. 5393–5400.

[23] W. Shao, L. He, and S. Y. Philip, "Multiple incomplete views clustering via weighted nonnegative matrix factorization with $l\_2,1$ regularization," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 2015, pp. 318–334.

[24] W. Shao, L. He, C.-t. Lu, and S. Y. Philip, "Online multi-view clustering with incomplete views," in *IEEE International Conference on Big Data*. IEEE, 2016, pp. 1012–1017.

[25] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," in *International Joint Conference on Artificial Intelligence*. AAAI Press, 2018, pp. 2262–2268.

[26] ——, "One-pass incomplete multi-view clustering," in *AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 3838–3845.

[27] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Transactions on Cybernetics*, vol. 50, no. 4, pp. 1418–1429, 2020.

[28] H. Wang, L. Zong, B. Liu, Y. Yang, and W. Zhou, "Spectral perturbation meets incomplete multi-view data," in *International Joint Conference on Artificial Intelligence*. AAAI Press, 2019, pp. 3677–3683.

[29] J. Wu, W. Zhuge, H. Tao, C. Hou, and Z. Zhang, "Incomplete multi-view clustering via structured graph learning," in *Pacific Rim International Conference on Artificial Intelligence*. Springer, 2018, pp. 98–112.

[30] S. Kuanar, K. Rao, M. Bilas, and J. Bredow, "Adaptive CU mode selection in hevc intra prediction: a deep learning approach," *Circuits, Systems, and Signal Processing*, vol. 38, no. 11, pp. 5081–5102, 2019.

[31] C. Tian, Y. Xu, and W. Zuo, "Image denoising using deep cnn with batch renormalization," *Neural Networks*, vol. 121, pp. 461–473, 2020.

[32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[33] S. Kuanar, C. Conly, and K. Rao, "Deep learning based hevc in-loop filtering for decoder quality enhancement," in *Picture Coding Symposium (PCS)*. IEEE, 2018, pp. 164–168.

[34] C. Tian, Y. Xu, Z. Li, W. Zuo, L. Fei, and H. Liu, "Attention-guided cnn for image denoising," *Neural Networks*, vol. 121, pp. 117–129, 2020.

[35] C. Xu, Z. Guan, W. Zhao, H. Wu, Y. Niu, and B. Ling, "Adversarial incomplete multi-view clustering," in *International Joint Conference on Artificial Intelligence*. AAAI Press, 2019, pp. 3933–3939.

[36] Q. Wang, Z. Ding, Z. Tao, Q. Gao, and Y. Fu, "Partial multi-view clustering via consistent gan," in *IEEE International Conference on Data Mining (ICDM)*. IEEE, 2018, pp. 1290–1295.

[37] J. Wen, B. Zhang, Y. Xu, J. Yang, and N. Han, "Adaptive weighted nonnegative low-rank representation," *Pattern Recognition*, vol. 81, pp. 326–340, 2018.

[38] L. Hagen and A. B. Kahng, "New spectral methods for ratio cut partitioning and clustering," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 11, no. 9, pp. 1074–1085, 1992.

[39] J. Shi and J. Malik, "Normalized cuts and image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 1997, pp. 731–737.

[40] Y. Yang and H. Wang, "Multi-view clustering: a survey," *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 83–107, 2018.

[41] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Advances in Neural Information Processing Systems*, 2011, pp. 1413–1421.

[42] S. Bhadra, S. Kaski, and J. Rousu, "Multi-view kernel completion," *Machine Learning*, vol. 106, no. 5, pp. 713–739, 2017.

[43] K. Zhan, F. Nie, J. Wang, and Y. Yang, "Multiview consensus graph clustering," *IEEE Transactions on Image Processing*, vol. 28, no. 3, pp. 1261–1270, 2018.

[44] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *IEEE International Conference on Computer Vision*, 2015, pp. 4238–4246.

[45] P. Ren, Y. Xiao, P. Xu, J. Guo, X. Chen, X. Wang, and D. Fang, "Robust auto-weighted multi-view clustering," in *International Joint Conference on Artificial Intelligence*, 2018, pp. 2644–2650.

[46] J. Wen, X. Fang, Y. Xu, C. Tian, and L. Fei, "Low-rank representation with adaptive graph regularization," *Neural Networks*, vol. 108, pp. 83–96, 2018.

[47] F. Nie, X. Wang, M. I. Jordan, and H. Huang, "The constrained laplacian rank algorithm for graph-based clustering," in *AAAI Conference on Artificial Intelligence*, 2016, pp. 1969–1976.

[48] S. Yi, Y. Liang, Z. He, Y. Li, W. Liu, and Y.-m. Cheung, "Dual pursuit for subspace learning," *IEEE Transactions on Multimedia*, vol. 21, no. 6, pp. 1399–1411, 2019.

[49] J. Wen, Z. Zhong, Z. Zhang, L. Fei, Z. Lai, and R. Chen, "Adaptive locality preserving regression," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 30, no. 1, pp. 75–88, 2020.

[50] Z. Zhang, Z. Lai, Z. Huang, W. K. Wong, G.-S. Xie, L. Liu, and L. Shao, "Scalable supervised asymmetric hashing with semantic and latent factor embedding," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 4803–4818, 2019.

[51] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in *International Joint Conference on Artificial Intelligence*, 2015.

[52] T. G. Wright and L. N. Trefethen, "Large-scale computation of pseudospectra using ARPACK and eigs," *SIAM Journal on Scientific Computing*, vol. 23, no. 2, pp. 591–605, 2001.

[53] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in *International Conference on Machine Learning*. ACM, 2006, pp. 377–384.

[54] N. Rai, S. Negi, S. Chaudhury, and O. Deshmukh, "Partial multi-view clustering using graph regularized nmf," in *International Conference on Pattern Recognition*. IEEE, 2016, pp. 2192–2197.

[55] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *AAAI Conference on Artificial Intelligence*, 2015, pp. 2750–2756.

[56] L. Fei-Fei, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 59–70, 2007.

[57] C. Zhang, Z. Han, H. Fu, J. T. Zhou, Q. Hu *et al.*, "CPM-Nets: Cross partial multi-view networks," in *Advances in Neural Information Processing Systems*, 2019, pp. 557–567.

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[59] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *IEEE international conference on computer vision*, 2015, pp. 1582–1590.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMM.2020.3013408, IEEE Transactions on Multimedia

[60] C. D. Manning, P. Raghavan, and H. Schütze, *Introduction to information retrieval*. Cambridge university press, 2008.
[61] L. Hubert and P. Arabie, "Comparing partitions," *Journal of classification*, vol. 2, no. 1, pp. 193–218, 1985.

Jie Wen received the Ph.D. degree in Computer Science and Technology at Harbin Institute of Technology, Shenzhen. His research interests include, biometrics, pattern recognition and machine learning. More information please refer to https://sites.google.com/view/jerry-wen-hit/home.

Ke Yan received the M.S. degree in computer science and technology from the Shenzhen Graduate School, Harbin Institute of Technology University, China, in 2013. He is currently pursuing the Ph.D. degree in computer science and technology with the Harbin Institute of Technology, Shenzhen, Shenzhen, China. His research interests include pattern recognition, machine learning and bioinformatics.

Zheng Zhang received his M.S. degree in Computer Science (2014) and Ph.D. degree in Computer Applied Technology (2018) from the Harbin Institute of Technology, China. Dr. Zhang was a Postdoctoral Research Fellow at The University of Queensland, Australia. He is currently an Assistant Professor at Harbin Institute of Technology, Shenzhen, China. He has published over 50 technical papers at prestigious international journals and conferences, including the IEEE TPAMI, IEEE TNNLS, IEEE TIP, IEEE T-CYB, IEEE CVPR, ECCV, AAAI, IJCAI, SIGIR, ACMM, etc. He serves/served as a (leading) Guest Editor of Information Processing & Management journal and Neurocomputing journal, a Publication Chair of the 16th International Conference on Advanced Data Mining and Applications (ADMA 2020), and an SPC/PC member of several top conferences. His current research interests include machine learning, computer vision and multimedia analytics.

Yong Xu received the Ph.D. degree in Pattern Recognition and Intelligence system at Nanjing University of Science and Technology, Nanjing, China, in 2005. He is currently a Professor in the School of Computer Science and Technology at Harbin Institute of Technology, Shenzhen. His current interests include pattern recognition, biometrics, machine learning and video analysis. More information please refer to http://www.yongxu.org/lunwen.html.

Junqian Wang is currently pursuing the Ph.D. degree in computer science and technology with the Harbin Institute of Technology, Shenzhen, Shenzhen, China. Her research interests include pattern recognition and machine learning.

Lunke Fei received the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, China, in 2016. Since 2017, he has been with the School of Computer Science and Technology, Guangdong University of Technology, Guangzhou, China. His research interests include pattern recognition, biometrics, image processing, and machine learning.

Bob Zhang received the B.A. degree in computer science from York University, Toronto, ON, Canada, in 2006, the M.A.Sc. degree in information systems security from Concordia University, Montreal, QC, Canada, in 2007, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011. After graduating from Waterloo, he remained with the Center for Pattern Recognition and Machine Intelligence, and later was a Postdoctoral Researcher in the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA, USA. He is currently an Assistant Professor in the Department of Computer and Information Science, University of Macau, Taipa, Macau. His research interests focus on biometrics, pattern recognition, and image processing. Dr. Zhang is a Technical Committee Member of the IEEE Systems, Man, and Cybernetics Society, an Associate Editor for the International Journal of Image and Graphics, as well as an Editorial Board member for the International Journal of INFORMATION.