In the real world, an image can be represented by various descriptors, such as SIFT, local binary pattern (LBP), HOG, etc. [1]. A webpage can be described by the images, links, texts, etc. [2]. The values of blood test and magnetic resonance images can be viewed as two references for disease diagnosing [3]. The above phenomena indicate that almost all things can be represented from different perspectives (views). These features acquired from different views are regarded as the multiview features [4], [5]. Generally, multiview features can represent data more comprehensive because features of different views provide many complementary information. Thus, exploiting the multiview features has the potential to improve the performance of different machine learning tasks [6]–[12].

Multiview clustering is one of the hottest research directions in this filed, which focuses on adaptively partitioning data points into their respective groups without any label information [13]–[16]. In the past few years, many multiview clustering methods have been proposed, such as the multiview k-means clustering [17], canonical correlation analysis (CCA)-based method [18], co-regularized multiview spectral clustering [19], low-rank tensor-based method [20], the deep matrix factorization-based method [4], etc. For multiview clustering, ensuring the clustering agreement among all views is the key to achieving good performance [21]. To this end, Wang et al. [21] introduced a views-agreement constraint to guarantee that the data-cluster graphs learned from all views are consistent to each other. In [14], a structured low-rank matrix factorization-based method is proposed to learn more consistent low-dimensional data-cluster representations for all views by exploiting the manifold structures and introducing the divergence constraint term jointly. Besides, in [22], based on the multiview spectral clustering, a novel angular-based regularizer is imposed on the sparse representation vectors of all views to learn the consensus similarity graph shared by all views for clustering. In [8], a more robust multiview spectral clustering model is proposed, which can learn the optimal similarity graphs and data-cluster representations with views agreement. For these multiview clustering methods, they commonly require that all views of data are complete. However, the requirement is often impossible to satisfy because it is often the case that some views of samples are missing in the real-world applications, especially in the applications of disease diagnosing [3] and webpage clustering [23]. This incomplete problem of views leads to the failure of the conventional multiview methods [24].

To address this issue, many efforts have been made in recent years, which can be generally categorized into two groups in terms of the exploited techniques, i.e., matrix factorization-based incomplete multiview clustering (MFIMC) and graph-based incomplete multiview clustering (GIMC). MFIMC focuses on learning a consensus representation with low dimensionality for all views directly via the matrix factorization technique. For example, partial multiview clustering (PMVC) seeks to learn a common latent subspace for all views, in which the instances of different views are enforced to have the same representation [3]. Different from PMVC,
multi-incomplete-view clustering first fills in the missing views with the average of all instances in the corresponding views, and then uses the weighted non-negative matrix factorization technique to jointly learn the latent representations of different views and the consensus representation for all views [25]. For these matrix-based methods, the common problem is that they only focus on learning the consensus representation while ignoring the intrinsic structure of data, which cannot guarantee the compactness and discriminability of the learned representation.

GIMC focuses on learning the low-dimensional representation from different graphs that reveal the relationships of all samples. Compared with the matrix factorization-based methods, GIMC can effectively exploit the geometric structures of data. For GIMC, the graph construction is very crucial because it directly influences the clustering performance. However, since some views are missing, it is impossible to construct such graphs to connect all samples completely. To address the issue, Trivedi et al. [23] proposed to complete the incomplete graph of the view with missing instances referring to the Laplacian matrix of the complete view, and then learned the low-dimensional representations for different views via the kernel CCA. The biggest shortcoming of this method is that it requires at least one complete view. Gao et al. [26] proposed to fill in the missing view with the average of instances in the corresponding view for graph construction and subspace learning. However, when the multiview data has large number of the missing instances in all views, this approach will fail because the filled missing views will dominate the subspace learning [25]. Thus, completing the graph by filling in the missing views is not a good choice for incomplete multiview clustering.

Zhao et al. [27] proposed a more novel graph learning method for incomplete multiview clustering, in which a robust consensus graph is adaptively learned from the low-dimensional consensus representation. Different from the above graph-based methods, the method proposed by Zhao et al. does not need to complete the graph or fill in the missing instances of any view.

Although many methods have been proposed to address the incomplete problem of multiview clustering, they still have many problems to address. For example, a limitation of these methods is that they require few samples to have features of all views. In other words, for data with more than three views, these methods fail to deal with the case that no sample contains all of the three views. The second limitation is that these graph-based methods cannot learn the global optimal consensus representation for clustering because the subspace learning and graph construction are treated independently in two separated steps. To solve the above issues, we, in this paper, propose a more general framework for multiview scenarios, which is suitable to all kinds of the multiview data, including arbitrary incomplete cases and the complete case. The proposed method aims to jointly learn the low-dimensional consensus representation and similarity graphs for all views, which enables our method to obtain the global optimal consensus representation and thus has the potential to perform better. To this end, we exploit the low-rank representation technique to adaptively learn the similarity graph of each view owing to its good performance in discovering the intrinsic relationships of data [28], [29]. Meanwhile, considering that graphs constructed from different incomplete views generally have great differences in magnitude, structures, and dimensions owing to the diversity of missing views, it is impossible to directly learn a unified graph or representation for all views [30]. To tackle this problem, we provide an ingenious approach, which indirectly learns the consensus representation from the low-dimensional representations of all views rather than the graphs by introducing a co-regularization term. Meanwhile, the proposed method is robust to noise by introducing the sparse error term to compensate the noise. Overall, this paper has the following contributions.

1) In this paper, we provide a more general multiview clustering framework, which can deal with both complete and incomplete multiview cases.

2) The proposed method is a pioneering work that integrates the graph construction and consensus representation learning into a joint optimization framework for incomplete multiview clustering. Compared with the other methods, the proposed method is able to learn the optimal similarity graph of each view and the consensus cluster representation for all views and, thus, has the potential to obtain a better performance.

3) The proposed method is robust to noise to some extent. Especially, by introducing the sparse error term to compensate the noise, the proposed method can greatly reduce the negative influence of noise, which is beneficial to discover the intrinsic structure of the noisy data.

The rest of this paper is organized as follows. Section II briefly introduces some related works to the proposed method. In Section III, we present the proposed method and its optimization algorithm in detail. Section IV gives a deep analysis to the proposed method. In Section V, several experiments are conducted to prove the effectiveness of the proposed method. Section VI offers the conclusions of this paper.

II. RELATED WORKS

A. Single-View Spectral Clustering

For a dataset \( X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m \times n} \) with \( n \) samples, spectral clustering tries to solve the following problem to learn the low-dimensional representation \( F \in \mathbb{R}^{d \times c} \) for clustering [31], [32]:

\[
\min_{F: F^T I F = I} \text{Tr}(F^T LF)
\]

where \( \text{Tr}(\cdot) \) denotes the trace operation. \( L \in \mathbb{R}^{n \times n} \) is the Laplacian graph, which is generally calculated as \( L = D - W \) in the ratio cut method [33] and \( L = I - D^{-1/2} WD^{-1/2} \) in the normalized cut method [31], where \( D \in \mathbb{R}^{n \times n} \) is a diagonal matrix and its \( i \)th diagonal element is calculated as \( D_{i,i} = \sum_j (W_{i,j} + W_{j,i})/2 \), \( W(W \geq 0) \) is the symmetric similarity graph with non-negative elements constructed from the data, \( I \) is the identity matrix.

Equation (1) is the typical eigenvalue decomposition problem and its solution is the eigenvector set corresponding to the first \( c \) minimum eigenvalues of \( L \). For \( F \), its each row
can be viewed as the new representation of the corresponding original sample. Then spectral clustering utilizes the $k$-means algorithm to partition the new representation $F$ into several clusters.

### B. Multiview Subspace Clustering

Generally, due to the different feature distributions of different views, graphs constructed from different views will have large differences. In this case, it is difficult to find the intrinsic consensus graphs of multiple views for clustering. Fortunately, it is possible to find the optimal and unique representation for all views. Inspired by this motivation, Gao et al. [30] proposed the multiview subspace clustering (MVSC), which focuses on learning a consensus cluster indicator matrix for clustering. For a dataset with $k$ views whose $v$th view is represented as $X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \ldots, x_n^{(v)}] \in \mathbb{R}^{m \times n}$, the learning model of MVSC is designed as follows:

$$
\begin{align*}
\min_{Z^{(v)}, E^{(v)}, F} & \sum_v \|X^{(v)} - X^{(v)}Z^{(v)} - E^{(v)}\|_F^2 \\
+ & \sum_v \left( \lambda_1 \text{Tr}(F^T L^{(v)} F) + \lambda_2 \|E^{(v)}\|_1 \right) \\
\text{s.t.} & \quad Z^{(v)T} I = I, \quad Z^{(v)}_{ij} = 0, \quad F^T F = I
\end{align*}
$$

where $I$ is a column vector with all ones, $Z^{(v)}_{ij} = 0$ means that all diagonal elements of matrix $Z^{(v)}$ are 0, $\lambda_1$ and $\lambda_2$ are penalty parameters, $L^{(v)}$ is the Laplacian graph of the $v$th view and is calculated as $L^{(v)} = D^{(v)} - W^{(v)}$, where $W^{(v)} = (||Z^{(v)}|| + ||Z^{(v)}||^T)/2$ and $D^{(v)}_{ij} = \sum_j W^{(v)}_{ij}$. $E^{(v)}$ is the error matrix used to model different noises. In (2), $\|\cdot\|_F$ and $\|\cdot\|_1$ are the “Frobenius” norm and “$l_1$” norm, respectively [34], [35]. $\text{Tr}()$ denotes the trace operator.

MVSC integrates the graph construction and low-dimensional representation learning into a joint learning framework, which enables it to find the global optimal indicator matrix for clustering. By alternatively solving all variables, MVSC can find the local optimum of the consensus representation $F$. Then MVSC performs $k$-means on the consensus representation to obtain the final clustering results.

### III. PROPOSED METHOD

Fig. 1 shows two cases of the incomplete multiview data. In the first case, only few samples contain features of all views and the other samples contain only one view. In the second case, all views are arbitrarily missing. It includes the case that no samples contain the features of all views. For the first case, many methods, such as PMVC [3] and incomplete multimodel grouping (IMG) [27], have been proposed based on the matrix factorization. However, these methods fail to deal with the second incomplete case. To address this issue, in this section, we will propose a more general incomplete multiview clustering framework that can handle all kinds of incomplete cases.

#### A. Incomplete Multiview Spectral Clustering With Adaptive Graph Learning

From the previous presentation, we can find that all conventional graph-based multiview clustering methods, including MVSC, learn the subspace based on the similarity graph of each view. These methods require that graphs constructed from all views are complete, which not only cannot reveal the similarity relationships of all samples but also have the same size of $n \times n$ for the multiview data with $n$ samples. However, as shown in Fig. 1, since all views are incomplete, graphs constructed from these incomplete views will have different sizes and also cannot reveal the relationships of all samples, which leads to the failure of the existing GIMC methods.

Some researchers propose to fill in the missing views with the average of instances of the corresponding view, and then construct the similarity graph of each view independently [26]. Although this approach can make the graphs of different views have the same size, it cannot produce the correct representation for each view and thus is harmful to the multiview clustering. This is mainly because that these missing instances will be regarded as the same class and be connected with the same weight (weight of 1 in the binary nearest neighbor graph), which in turn pulls these missing instances together in the low-dimensional subspace whether they are from the same cluster or not. Therefore, this graph completion approach is unreasonable for incomplete multiview clustering, especially for the case with large number of missing views. A more reasonable way to avoid this issue is to set the connected weights
corresponding to these missing instances as 0 in the similarity graph of the corresponding view. In this way, the uncertain similarity information corresponding to the missing views will not play a negative role in learning the data-cluster representation. In contrast, only the authentic similarity information of the available instances are exploited to guide the representation learning, which is beneficial to achieve a more reliable data-cluster representation and reduces the negative influence of the missing views. Based on the above analysis, we rewrite MVSC as follows for incomplete multiview learning:

\[
\begin{align*}
\min_{Z^{(v)}, E^{(v)}, F} & \sum_{v} \left\| Y^{(v)} - Y^{(v)} Z^{(v)} - E^{(v)} \right\|_F^2 \\
& + \sum_{v} \left( \lambda_1 \text{Tr} \left( F^T \bar{L}^{(v)} F \right) + \lambda_2 \left\| E^{(v)} \right\|_1 \right) \\
\text{s.t.} & \quad Z^{(v)^T} I = I, Z_{i,j}^{(v)} = 0, F^T F = I
\end{align*}
\]  

where \( Y^{(v)} = [y_1^{(v)}, y_2^{(v)}, \ldots, y_n^{(v)}] \in \mathbb{R}^{m_v \times n_v} \) denotes the set of the un-missing instances in the \( v \)-th view, \( m_v \) and \( n_v \) are the number of the features and un-missing instances of the \( v \)-th view, respectively. The original instance sets including the missing and un-missing instances in the \( v \)-th view are still represented as \( X^{(v)} = [x_1^{(v)}, x_2^{(v)}, \ldots, x_n^{(v)}] \in \mathbb{R}^{m_v \times n} \) (\( n_v < n \)). \( \bar{L}^{(v)} = \bar{D}^{(v)} - \bar{W}^{(v)}, \bar{W}^{(v)} = [(\bar{Z}^{(v)}) + (\bar{Z}^{(v)^T})]/2, \bar{D}^{(v)} = \sum_j \bar{W}_{i,j}, \bar{Z}^{(v)} \in \mathbb{R}^{n_v \times n_v} \) is the complete graph that connects all instances including the missing and un-missing instances of the \( v \)-th view. In our method, we can exploit the following formula to obtain the completed graph \( \tilde{Z}^{(v)} \) based on the graph \( Z^{(v)} \in \mathbb{R}^{n_v \times n_v} \) learned from the un-missing instances:

\[
\tilde{Z}^{(v)} = G^{(v)^T} Z^{(v)} G^{(v)}
\]

where \( G^{(v)} \in \mathbb{R}^{n_v \times n} \) is an index matrix used to complete the graph, whose elements related to the missing instances are enforced to zero. Especially, matrix \( G^{(v)} \) is defined as follows:

\[
G_{i,j}^{(v)} = \begin{cases} 1, & \text{if } y_i^{(v)} \text{ is the original instance } x_j^{(v)} \\ 0, & \text{otherwise} \end{cases}
\]

It is simple to prove that the following equation is satisfied:

\[
\bar{L}^{(v)} = G^{(v)^T} \bar{L}^{(v)} G^{(v)}
\]

where \( \bar{L}^{(v)} \) is the Laplacian matrix of graph \( Z^{(v)} \).

Considering that the data is generally drawn from several low-rank subspaces, thus the learned graph should better discover the low-rank structure of data [28]. Moreover, inspired by the motivation that learning a non-negative graph is beneficial to improve the clustering performance and make the learned graph more interpretable [36], we rewrite model (3) as follows to learn multiple non-negative graphs for multiview subspace learning:

\[
\begin{align*}
\min_{Z^{(v)}, E^{(v)}, F} & \sum_{v} \left( \left\| Z^{(v)} \right\|_1 + \lambda_2 \left\| E^{(v)} \right\|_1 \right) \\
& + \lambda_1 \sum_{v} \text{Tr} \left( F^T G^{(v)^T} \bar{L}^{(v)} G^{(v)} F \right) \\
\text{s.t.} & \quad Y^{(v)} = Y^{(v)} Z^{(v)} + E^{(v)}, Z^{(v)} I = I \\
& \quad 0 \leq Z^{(v)} \leq 1, Z_{i,j}^{(v)} = 0, F^T F = I
\end{align*}
\]  

where \( \left\| Z^{(v)} \right\|_1 \) is the nuclear norm of matrix \( Z^{(v)} \), which is calculated as the summation of all singular values of matrix \( Z^{(v)} \) [37]-[39]. Constraint \( Z^{(v)} I = I \) is used to avoid that any sample has no contribution in the joint representation, \( I \) is a column vector with all elements as 1 [40]. In model (7), \( F \) is an \( n \times c \) matrix whose each row denotes the representation of the corresponding sample.

In model (7), the third term is equivalent to \((1/2)\sum_{j=1}^{c} \sum_{h=1}^{n} (\left\| F_{i,j} - F_{i,h} \right\|_2^2 \sum_y \bar{W}_{i,j}^{(v)})\). This demonstrates that the regularization weights to the target cluster indicator matrix \( F \) are the summation of all similarity weights of multiple graphs. For complete data, this approach may be effective to learn the optimal cluster indicator for each sample. However, for incomplete multiview data, this approach will fail because the regularization weights to the missing instances and un-missing instances are too unfair. It may occur the case that through the summation of the weights of multiple graphs, the regularization weights of
samples from different classes may larger than those of the samples from the same class, which will lead to the incorrect cluster indicators of samples. To solve this problem, we propose to learn the consensus representation from those cluster indicator matrices of all views as follows:

$$\min_{Z^{(v)}, E^{(v)}, F^{(v)}} \sum_v \left( \left\| Z^{(v)} \right\|_* + \lambda_1 \text{Tr} \left( F^{(v)T} G^{(v)} T L^{(v)} G^{(v)} F^{(v)} \right) \right)$$

$$+ \sum_v \left( \lambda_2 \left\| E^{(v)} \right\| + \lambda_3 \frac{1}{2} \Gamma \left( F^{(v)}, U \right) \right)$$

s.t. $Y^{(v)} = Y^{(v)} Z^{(v)} + E^{(v)}, Z^{(v)} I = I$

$$0 \leq Z^{(v)} \leq 1, Z^{(v)} I = 0, F^{(v)T} F^{(v)} = I, U^T U = I$$

(8)

where $\lambda_3$ is also a penalty parameter, matrix $U \in \mathbb{R}^{n \times c}$ is the target cluster indicator matrix (or consensus representation) to learn. Function $\Gamma (F^{(v)}, U)$ measures the disagreement of the consensus representation $U$ and representation $F^{(v)}$. $\Gamma (F^{(v)}, U)$ is defined as follows [19]:

$$\Gamma (F^{(v)}, U) = \frac{K_U}{\| K_U \|_F^2} - \frac{K_{F^{(v)}}}{\| K_{F^{(v)}} \|_F^2}$$

(9)

where $K_U$ and $K_{F^{(v)}}$ are the similarity matrices of $U$ and $F^{(v)}$, respectively. For simplicity, we also choose the linear kernel, i.e., $K_U = U U^T$ as the similarity measure metric [19]. Based on the facts that $\| K_U \|_F^2 = c$ and $\| K_{F^{(v)}} \|_F^2 = c$, (9) can be rewritten as follows:

$$\Gamma (F^{(v)}, U) = \frac{2 (c - \text{Tr} (F^{(v)T} F^{(v)} U U^T))}{c^2}$$

(10)

As a result, our final model is expressed as follows:

$$\min_{Z^{(v)}, E^{(v)}, F^{(v)}, U} \sum_v \left( \left\| Z^{(v)} \right\|_* + \lambda_3 \left( c - \text{Tr} (F^{(v)T} F^{(v)} U U^T) \right) \right)$$

$$+ \sum_v \lambda_1 \text{Tr} \left( F^{(v)T} G^{(v)} T L^{(v)} G^{(v)} F^{(v)} \right) + \sum_v \lambda_2 \left\| E^{(v)} \right\|_1$$

s.t. $Y^{(v)} = Y^{(v)} Z^{(v)} + E^{(v)}, Z^{(v)} I = I$

$$0 \leq Z^{(v)} \leq 1, Z^{(v)} I = 0, F^{(v)T} F^{(v)} = I, U^T U = I$$

(11)

Since the proposed IMC framework is based on the adaptive graph learning and spectral clustering, we refer to the proposed method as incomplete multiview spectral clustering with adaptive graph learning (IMSC_AGL). The framework of our method is briefly outlined in Fig. 2.

### B. Solution to IMSC_AGL

For the optimization problem (11), we choose the alternating direction method of multipliers (ADMMs) to calculate its local optimal solution [41]. At the beginning, we introduce several variables to make problem (11) separable as follows:

$$\min_{Z^{(v)}, E^{(v)}, F^{(v)}, U, S^{(v)}, W^{(v)}} \sum_v \left( \left\| S^{(v)} \right\|_* - \lambda_3 \text{Tr} (F^{(v)T} F^{(v)} U U^T) \right)$$

$$+ \sum_v \lambda_1 \text{Tr} \left( F^{(v)T} G^{(v)} T L^{(v)} G^{(v)} F^{(v)} \right) + \sum_v \lambda_2 \left\| E^{(v)} \right\|_1$$

s.t. $Y^{(v)} = Y^{(v)} Z^{(v)} + E^{(v)}, Z^{(v)} = S^{(v)}, Z^{(v)} = W^{(v)}$

$W^{(v)} I = I, 0 \leq W^{(v)} \leq 1, W^{(v)} I = 0, F^{(v)T} F^{(v)} = I, U^T U = I$

(12)

where $L^{(v)}$ denotes the Laplacian graph of matrix $W^{(v)}$. Note: Since $c$ is a constant, we can ignore it in (11). The augmented Lagrangian function of (12) is formulated as follows:

$$L = \sum_v \left( \left\| S^{(v)} \right\|_* + \lambda_1 \text{Tr} \left( F^{(v)T} G^{(v)} T L^{(v)} G^{(v)} F^{(v)} \right) \right)$$

$$- \sum_v \lambda_3 \text{Tr} (F^{(v)T} U U^T) + \sum_v \lambda_2 \left\| E^{(v)} \right\|_1$$

$$+ \frac{\mu}{2} \sum_v \left( \left\| Z^{(v)} - S^{(v)} + \frac{C^{(v)}}{\mu} \right\|_F^2 \right)$$

$$+ \left\| Z^{(v)} - W^{(v)} + \frac{C^{(v)}}{\mu} \right\|_F^2$$

$$+ \frac{\mu}{2} \sum_v \left( \left\| Y^{(v)} - Y^{(v)} Z^{(v)} - E^{(v)} + \frac{C^{(v)}}{\mu} \right\|_F^2 \right)$$

$$- \sum_v \psi^{(v)} (W^{(v)} I - I)$$

(13)

where matrices $C^{(v)}_1, C^{(v)}_2$, and $C^{(v)}_3$, and vector $\psi^{(v)}$ are Lagrange multipliers, $\mu$ is a penalty parameter.

Then we can iteratively solve all unknown variables one-by-one as follows.

**Step 1 (Update Variable $Z^{(v)}$):** Fixing all of the other variables, the problem to solve variable $Z^{(v)}$ is degraded to minimize the following problem:

$$L(Z^{(v)}) = \left\| Z^{(v)} - S^{(v)} + \frac{C^{(v)}}{\mu} \right\|_F^2 + \left\| Z^{(v)} - W^{(v)} + \frac{C^{(v)}}{\mu} \right\|_F^2$$

$$+ \left\| Y^{(v)} - Y^{(v)} Z^{(v)} - E^{(v)} + \frac{C^{(v)}}{\mu} \right\|_F^2$$

(14)

Then we can obtain variable $Z^{(v)}$ by setting the derivative of $L(Z^{(v)})$ with respect to $Z^{(v)}$ to zero as follows:

$$\frac{\partial L(Z^{(v)})}{\partial Z^{(v)}} = 2 Z^{(v)} - 2 M^{(v)} + 2 Z^{(v)} - 2 M^{(v)}$$

$$+ 2 Y^{(v)T} (Y^{(v)} Z^{(v)} - M^{(v)}) = 0$$

$$\Leftrightarrow Z^{(v)} = (Y^{(v)T} Y^{(v)} + 2 I)^{-1} (Y^{(v)T} M^{(v)} + M^{(v)} + M^{(v)}$$

(15)

where $M^{(v)} = S^{(v)} - (C^{(v)}_3 / \mu), M^{(v)} = W^{(v)} - (C^{(v)}_3 / \mu)$, and $M^{(v)} = Y^{(v)} - E^{(v)} + (C^{(v)}_1 / \mu)$.

**Step 2 (Update Variable $S^{(v)}$):** Fixing the other variables, the subproblem to calculate variable $S^{(v)}$ is degraded to the following formula:

$$\min_{S^{(v)}} \left\| S^{(v)} \right\|_* + \frac{\mu}{2} \left\| Z^{(v)} - S^{(v)} + \frac{C^{(v)}}{\mu} \right\|_F^2$$

(16)

Equation (16) can be computed by the singular value thresholding (SVT) shrinkage operator as follows [42–44]:

$$S^{(v)} = \Theta_1 \mu \left( Z^{(v)} + \frac{C^{(v)}}{\mu} \right)$$

(17)

where $\Theta$ denotes the SVT shrinkage operator.
Step 3 (Update Variables $W^{(v)}$ and $\psi^{(v)}$): Fixing the other variables, the optimization problem to calculate variable $W^{(v)}$ is formulated as follows:

$$
\min_{W^{(v)} \geq 0, W_{i,j}^{(v)} = 0} \lambda_1 \lambda_i (F^{(v)T} G^{(v)T} L_{w}^{(v)} G^{(v)} F^{(v)}) + \mu 2 \left\| Z^{(v)} - W^{(v)} + C_3 \right\|^2_F - \psi^{(v)}(W^{(v)} I - I) \right). (18)
$$

Define $P^{(v)} = G^{(v)} F^{(v)}, R^{(v)} = Z^{(v)} + (C_3 / \mu), (18)$ is equivalent to the following optimization problem:

$$
\min_{W^{(v)} \geq 0, W_{i,j}^{(v)} = 0} \frac{\lambda_1}{2} \sum_{i,j} P_{i,j}^{(v)} - P_{j,i}^{(v)} W_{i,j}^{2} + \frac{\mu}{2} \left\| W^{(v)} - R^{(v)} \right\|^2_F - \psi^{(v)}(W^{(v)} I - I) \right). (19)
$$

Step 5 (Update Variable $F^{(v)}$): Fixing the other variables, the cluster indicator matrix $F^{(v)}$ of each view can be calculated by minimizing the following formula:

$$
\min_{F^{(v)}T F^{(v)} = I} \lambda_1 \lambda_i (F^{(v)T} G^{(v)T} L_{w}^{(v)} G^{(v)} F^{(v)}) - \lambda_3 \lambda_i (F^{(v)T} U U^{T}) \right) \Rightarrow \max \left( F^{(v)T} (\lambda_3 U U^{T} - \lambda_1 G^{(v)T} L_{w}^{(v)} G^{(v)}) F^{(v)} \right). (25)
$$

Problem (25) can be solved by the eigenvalue decomposition, where the first $c$ eigenvectors corresponding to the first $c$ largest eigenvalues of matrix $(\lambda_3 U U^{T} - \lambda_1 G^{(v)T} L_{w}^{(v)} G^{(v)})$ are chosen as the optimal solution to variable $F^{(v)}$.

Step 6 (Update Variable $U$): Fixing the other variables, the problem to obtain the consensus cluster representation $U$ is degraded to the following problem:

$$
\min_{U^{T} U = I} - \sum_v \lambda_3 \lambda_i (F^{(v)T} U U^{T}) \right) \Rightarrow \max \left( U^{T} \left( \sum_v F^{(v)T} F^{(v)} \right) U \right). (26)
$$

Problem (26) can also be simply computed by the eigenvalue decomposition. The optimal solution to variable $U$ is the eigenvector set corresponding to the first $c$ largest eigenvalues of matrix $(\sum_v F^{(v)T} F^{(v)})$.

Step 7 (Update Variables $C_1^{(v)}, C_2^{(v)}, C_3^{(v)},$ and $\mu$): These four variables are updated as follows, respectively,

$$
C_1^{(v)} = C_1^{(v)} + \mu (Y^{(v)} - Y^{(v)} Z^{(v)} - E^{(v)}) \right) (27)
$$

$$
C_2^{(v)} = C_2^{(v)} + \mu (Z^{(v)} - S^{(v)}) \right) (28)
$$

$$
C_3^{(v)} = C_3^{(v)} + \mu (Z^{(v)} - W^{(v)}) \right) (29)
$$

$$
\mu = \min(\rho \mu, \mu_0) \right) (30)
$$

where $\rho$ and $\mu_0$ are constants.

Algorithm 1 summarizes the computation procedures presented above. After obtaining the consensus representation $U$, we perform the $k$-means algorithm on it to obtain the final clustering results.

IV. ANALYSIS OF THE PROPOSED METHOD

A. Computational Complexity

For simplicity, we do not take into account the computational costs of matrix multiplication, elements-based matrix division, matrix addition and subtraction, etc., since these operations are very simple in comparison with the other matrix operations. For the proposed algorithm listed in Algorithm 1, the major computational costs are the operations like matrix inverse, singular value decomposition, and eigenvalue decomposition. Generally, the computational complexities of the above three operations are $O(n^3)$ for an $n \times n$ matrix, $O(m^2)$ for an $m \times m$ matrix, and $O(n^3)$ for an $n \times n$ matrix, respectively. Therefore, the computational complexities of steps 2, 5, and 6 are about $O(n^3)$, $O(n^2)$, and $O(n^3)$, respectively. For step 1, although it needs to calculate the inverse operation,
Algorithm 1 IMSC_AGL [Solving (12)]

**Input:** Incomplete multi-view data $Y^{(v)}$, index matrix $G^{(v)}$, $v \in [1, k]$, parameters $\lambda_1, \lambda_2$, and $\lambda_3$.

**Initialization:** Initialize $Z^{(v)}$ with the $k$-nearest neighbor graph of each view; initialize $F^{(v)}$ via the eigenvalue decomposition on the Laplacian graph of each view; using (26) to initialize the cluster matrix $U$, $\mu_0 = 10^6$, $\rho = 1.1$, $\mu = 0.01$.

**while** not converged **do**

  **for** $v$ from 1 to $k$

  1. Update variable $Z^{(v)}$ via (15);
  2. Update variable $S^{(v)}$ via (17);
  3. Update variables $\psi^{(v)}$ and $W^{(v)}$ via (22) and (21), respectively, and then implement $W^{(v)} = \max(W^{(v)}, 0)$;
  4. Update variable $E^{(v)}$ via (24);
  5. Update variable $F^{(v)}$ by solving (25);
  6. Update variables $C_1^{(v)}$, $C_2^{(v)}$, and $C_3^{(v)}$ via (27), (28), and (29), respectively.

**end**

7. Update $U$ by solving (26);
8. Update $\mu$ via (30).

**end while**

**Output:** $U$

---

In this section, we conduct experiments to prove the effectiveness of the proposed method. The MATLAB code of our IMSC_AGL is released at: https://sites.google.com/view/jerry-wen/-publications.

**A. Experimental Settings**

1) **Baseline Algorithms and Evaluation Metrics:** The following six methods are selected to compare with the proposed method.

   1. **Best Single View (BSV) [27]:** For BSV, the missing instances in every view are first filled in the average of instances in the corresponding view. Then it performs $k$-means on all views independently and reports their best clustering results.

   2. **Concat [27]:** Concat adopts the same approach as BSV to fill in the missing instances. Their difference is that it concatenates all views into a single view with long dimensions, and then performs $k$-means to obtain the clustering results.

   3. **PMVC [3]:** Based on the non-negative matrix factorization, PMVC learns a common latent representation for all views, then performs $k$-means to obtain the clustering results.

   4. **IMG [27]:** IMG also uses the matrix factorization technique to learn a common latent representation for all views. Different from PMVC, IMG simultaneously learns a graph from the common representation and finally exploits the spectral clustering algorithm to obtain the clustering results.

   5. **Double Constrained Non-Negative Matrix Factorization (DCNMF) [50]:** DCNMF uses local geometric structure of each view to guide the representation learning.

   6. **Graph Regularized PMVC (GPMVC) [51]:** GPMVC can be viewed as a variant of DCNMF, which learns the common representation from the normalized individual representations of all views.

IMG, DCNMF, and GPMVC have many tunable parameters (more than two parameters). In our experiments, we implement these methods with a wide candidate parameter set, such as $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 10^{0}, 10^{1}, 10^{2}, 10^{3}, 10^{4}\}$ and report their best clustering results for fair comparison.

There are many criterions to evaluate the clustering performances of different methods [52]. In our experiments, we choose three most well-known evaluation criterions, i.e., clustering accuracy (ACC), normalized mutual information (NMI), and purity to compare the above methods [17]. Generally, we expect the values of these evaluation criterions as big as possible.

2) **Databases:** Seven real-world databases listed in Table I are adopted for evaluation. Their detailed information are as follows.

1) **Handwritten Digit Database [53]:** Following the experimental settings in [53], we conduct the experiments on the multiview handwritten database\(^1\) with 2000 samples in total.

---

\(^1\)https://www.dropbox.com/s/b000dxc2s96o1to/handwritten.mat
There are ten digits, i.e., 0–9, in the used handwritten digit database. In our experiments, we only use two views, i.e., pixel average features and Fourier coefficient features, to conduct experiments, in which the pixel average view has 240 features per sample and the Fourier coefficient view contains 76 features for one sample.

2) BUAA-VisNir Face Database (BUAA) [54]: Following the experimental settings in [27], we implement the experiments on the subset of the BUAA face database, which contains 90 visual images and 90 near infrared images of the first 10 classes. The used BUAA face database is available at https://github.com/hdzhao/IMG/tree/master/data. For the used BUAA face database, two types of images, i.e., visual images and near infrared images are regarded as the two views of persons. All images were resized into a 10 × 10 matrix and vectorized in advance.

3) Cornell Database [2], [55]: Cornell database is one of the popular WebKB databases, which is composed of 195 documents over the five labels, i.e., student, project, course, staff, and faculty. Each document is described by two views, i.e., 1703 content features and 195 citation features.

4) Caltech101 Database [56]: The Caltech101 database is one of the popular object databases which contains 101 objects in total. Each object provides 40–800 images. In our experiments, a multiview subset which contains seven classes and 1474 images is adopted [53]. For convenience, we refer to the subset as Caltech7. The original multiview Caltech7 database contains five types of features. In our experiments, we only select two views to implement the experiments, in which the one is GIST features [57] with 512 dimensions per sample and the other one is the LBPs features [57] with 928 dimensions per sample.

5) ORL Database. The ORL face database is composed of 400 faces taken from 40 individuals. In the experiments, we first preresized all images into the size of 32 × 32, and then extract the features of LBP, GIST, and pyramid of histogram of oriented gradients [58]. Combining the original pixel features of each image, we form the multiview ORL dataset with four views, in which their feature dimensions are 1024, 512, 1024, and 1024, respectively.

6) 3 Sources Database: This database contains 948 news articles collected from three online news sources: 1) BBC;

7) BBCSport Database [59]: The original BBCSport database contains 737 documents about the sport news articles collected from the BBC Sport website. These documents are described by 2–4 views and categorized into five classes. In our experiments, we choose a subset with 116 samples described by all of the four views to validate the effectiveness of our method.

3) Incomplete Multiview Data Construction: In our experiments, we construct two types of incomplete multiview data.

1) Incomplete Case That Few Samples Have Complete Views: Following the experimental settings in [27], for handwritten, BUAA, Cornell, and Caltech7 databases, we randomly select 10%, 30%, 50%, 70%, and 90% samples from the database as the paired samples. For the half of corresponding remaining samples, we remove their first view, and for the other half of samples, we remove their second view to form the incomplete multiview scenarios. Similarly, for the ORL multiview database, we also randomly select 10%, 30%, and 50% samples as the paired samples, and then we follow the previous strategy to randomly select 25%, 25%, 25%, and 25% samples of the corresponding remaining samples as the single view samples.

2) Incomplete Case That All Samples Have Missing Views: In our experiments, we exploit the ORL, 3 Sources, and BBCSport databases to construct the incomplete multiview data with no paired samples, where about 53%, 55%, and 55% instances are randomly removed from each view of the three databases, respectively. For fairness, we repeatedly perform all compared methods five times on these databases and report their average clustering results.

B. Experimental Results and Analysis

Experimental results of different methods on the above two types of incomplete multiview databases are enumerated in Tables II–VII and shown in Fig. 3, respectively. From the experimental results, we can obtain the following points.

1) In most cases, Concat and BSV perform the first worst and second worst in comparison with the other methods. This demonstrates that concatenating all views into one long view is not a good approach in dealing with the multiview clustering tasks. This is mainly because that Concat not only ignores the differences of different views in the feature scales and distributions but also cannot exploit the complementary information across different views. While for BSV, the missing instances in each view will obviously be clustered into the same group
TABLE II
MEAN NMIS (%) AND ACCS (%) OF DIFFERENT METHODS ON THE HANDWRITTEN DIGIT DATABASE. BOLD NUMBERS DENOTE THE BEST RESULT

<table>
<thead>
<tr>
<th>Method \ PER</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSV</td>
<td>37.04</td>
<td>44.48</td>
<td>51.90</td>
<td>58.61</td>
<td>66.26</td>
<td>43.08</td>
<td>50.46</td>
<td>57.39</td>
<td>64.44</td>
<td>69.29</td>
</tr>
<tr>
<td>Concat</td>
<td>47.71</td>
<td>54.43</td>
<td>61.12</td>
<td>70.30</td>
<td>79.34</td>
<td>46.01</td>
<td>57.46</td>
<td>66.45</td>
<td>78.64</td>
<td>86.63</td>
</tr>
<tr>
<td>PMVC</td>
<td>55.13</td>
<td>60.85</td>
<td>64.88</td>
<td>68.54</td>
<td>72.83</td>
<td>63.81</td>
<td>70.90</td>
<td>73.44</td>
<td>75.20</td>
<td>77.82</td>
</tr>
<tr>
<td>IMG</td>
<td>58.04</td>
<td>62.38</td>
<td>64.91</td>
<td>68.21</td>
<td>73.57</td>
<td>69.22</td>
<td>75.41</td>
<td>76.36</td>
<td>77.54</td>
<td>81.78</td>
</tr>
<tr>
<td>DCNMF</td>
<td>54.23</td>
<td>65.56</td>
<td>74.41</td>
<td>78.14</td>
<td>80.90</td>
<td>51.21</td>
<td>76.63</td>
<td>80.61</td>
<td>86.16</td>
<td>89.16</td>
</tr>
<tr>
<td>GPMVC</td>
<td>60.99</td>
<td>63.99</td>
<td>72.23</td>
<td>73.68</td>
<td>75.24</td>
<td>65.60</td>
<td>74.04</td>
<td>76.94</td>
<td>79.06</td>
<td>81.08</td>
</tr>
<tr>
<td>IMSC_AGL</td>
<td>76.60</td>
<td>79.05</td>
<td>82.46</td>
<td>83.05</td>
<td>86.27</td>
<td>80.76</td>
<td>84.81</td>
<td>87.41</td>
<td>89.77</td>
<td>92.57</td>
</tr>
</tbody>
</table>

TABLE III
MEAN NMIS (%) AND ACCS (%) OF DIFFERENT METHODS ON THE BUAA DATABASE. BOLD NUMBERS DENOTE THE BEST RESULT

<table>
<thead>
<tr>
<th>Method \ PER</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSV</td>
<td>43.10</td>
<td>53.03</td>
<td>61.78</td>
<td>69.91</td>
<td>82.56</td>
<td>48.33</td>
<td>56.96</td>
<td>64.26</td>
<td>70.81</td>
<td>80.16</td>
</tr>
<tr>
<td>Concat</td>
<td>51.22</td>
<td>51.95</td>
<td>52.43</td>
<td>56.51</td>
<td>62.66</td>
<td>45.62</td>
<td>46.61</td>
<td>47.46</td>
<td>52.34</td>
<td>57.58</td>
</tr>
<tr>
<td>PMVC</td>
<td>61.35</td>
<td>67.07</td>
<td>71.97</td>
<td>78.70</td>
<td>84.22</td>
<td>57.41</td>
<td>66.46</td>
<td>70.01</td>
<td>75.92</td>
<td>80.73</td>
</tr>
<tr>
<td>IMG</td>
<td>54.72</td>
<td>67.53</td>
<td>76.74</td>
<td>82.83</td>
<td>85.90</td>
<td>53.95</td>
<td>67.39</td>
<td>76.14</td>
<td>79.36</td>
<td>80.78</td>
</tr>
<tr>
<td>DCNMF</td>
<td>61.78</td>
<td>68.75</td>
<td>72.05</td>
<td>79.66</td>
<td>86.42</td>
<td>58.36</td>
<td>67.58</td>
<td>72.15</td>
<td>76.58</td>
<td>82.42</td>
</tr>
<tr>
<td>GPMVC</td>
<td>62.12</td>
<td>70.25</td>
<td>74.33</td>
<td>81.63</td>
<td>86.78</td>
<td>58.98</td>
<td>68.75</td>
<td>74.28</td>
<td>78.28</td>
<td>84.24</td>
</tr>
<tr>
<td>IMSC_AGL</td>
<td>63.52</td>
<td>75.16</td>
<td>80.29</td>
<td>84.52</td>
<td>89.84</td>
<td>65.72</td>
<td>78.16</td>
<td>80.76</td>
<td>82.77</td>
<td>89.54</td>
</tr>
</tbody>
</table>

TABLE IV
MEAN NMIS (%) AND ACCS (%) OF DIFFERENT METHODS ON THE CORNELL DATABASE. BOLD NUMBERS DENOTE THE BEST RESULT

<table>
<thead>
<tr>
<th>Method \ PER</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSV</td>
<td>8.66</td>
<td>8.19</td>
<td>8.89</td>
<td>12.69</td>
<td>19.34</td>
<td>42.41</td>
<td>43.93</td>
<td>44.84</td>
<td>46.32</td>
<td>47.66</td>
</tr>
<tr>
<td>Concat</td>
<td>8.07</td>
<td>7.56</td>
<td>8.30</td>
<td>10.21</td>
<td>13.47</td>
<td>38.80</td>
<td>38.06</td>
<td>36.96</td>
<td>36.79</td>
<td>38.48</td>
</tr>
<tr>
<td>PMVC</td>
<td>15.76</td>
<td>16.00</td>
<td>18.21</td>
<td>19.76</td>
<td>21.03</td>
<td>42.56</td>
<td>42.56</td>
<td>43.79</td>
<td>42.56</td>
<td>43.03</td>
</tr>
<tr>
<td>IMG</td>
<td>12.36</td>
<td>16.62</td>
<td>19.24</td>
<td>20.89</td>
<td>22.98</td>
<td>45.13</td>
<td>45.79</td>
<td>47.08</td>
<td>45.51</td>
<td>44.76</td>
</tr>
<tr>
<td>DCNMF</td>
<td>13.59</td>
<td>17.72</td>
<td>19.17</td>
<td>21.09</td>
<td>23.98</td>
<td>39.94</td>
<td>43.29</td>
<td>43.18</td>
<td>43.74</td>
<td>45.52</td>
</tr>
<tr>
<td>GPMVC</td>
<td>13.90</td>
<td>16.07</td>
<td>18.99</td>
<td>15.03</td>
<td>17.07</td>
<td>40.39</td>
<td>43.86</td>
<td>46.53</td>
<td>44.56</td>
<td>44.35</td>
</tr>
<tr>
<td>IMSC_AGL</td>
<td>18.52</td>
<td>20.28</td>
<td>21.46</td>
<td>22.84</td>
<td>24.35</td>
<td>43.25</td>
<td>46.02</td>
<td>47.35</td>
<td>46.37</td>
<td>48.56</td>
</tr>
</tbody>
</table>

TABLE V
MEAN NMIS (%) AND ACCS (%) OF DIFFERENT METHODS ON THE CALTECH7 DATABASE. BOLD NUMBERS DENOTE THE BEST RESULT

<table>
<thead>
<tr>
<th>Method \ PER</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
<th>0.1</th>
<th>0.3</th>
<th>0.5</th>
<th>0.7</th>
<th>0.9</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSV</td>
<td>29.04</td>
<td>32.11</td>
<td>35.13</td>
<td>38.83</td>
<td>44.16</td>
<td>42.66</td>
<td>40.97</td>
<td>39.83</td>
<td>42.92</td>
<td>46.99</td>
</tr>
<tr>
<td>Concat</td>
<td>33.82</td>
<td>34.44</td>
<td>34.56</td>
<td>38.15</td>
<td>45.44</td>
<td>36.83</td>
<td>31.74</td>
<td>36.36</td>
<td>43.38</td>
<td>47.08</td>
</tr>
<tr>
<td>PMVC</td>
<td>38.99</td>
<td>40.26</td>
<td>40.17</td>
<td>41.60</td>
<td>41.94</td>
<td>43.46</td>
<td>43.96</td>
<td>44.46</td>
<td>44.76</td>
<td>44.34</td>
</tr>
<tr>
<td>IMG</td>
<td>32.38</td>
<td>33.29</td>
<td>35.05</td>
<td>35.96</td>
<td>37.64</td>
<td>42.05</td>
<td>42.36</td>
<td>42.23</td>
<td>41.17</td>
<td>43.23</td>
</tr>
<tr>
<td>DCNMF</td>
<td>33.86</td>
<td>38.19</td>
<td>41.40</td>
<td>41.24</td>
<td>44.04</td>
<td>40.63</td>
<td>44.53</td>
<td>45.62</td>
<td>48.50</td>
<td>50.74</td>
</tr>
<tr>
<td>GPMVC</td>
<td>40.05</td>
<td>40.96</td>
<td>41.83</td>
<td>42.61</td>
<td>46.02</td>
<td>45.57</td>
<td>47.19</td>
<td>46.99</td>
<td>49.69</td>
<td>49.10</td>
</tr>
<tr>
<td>IMSC_AGL</td>
<td>40.71</td>
<td>42.02</td>
<td>43.35</td>
<td>45.15</td>
<td>47.28</td>
<td>50.89</td>
<td>51.20</td>
<td>52.14</td>
<td>53.05</td>
<td>56.30</td>
</tr>
</tbody>
</table>

since they are filled in the same average instance, which makes BSV achieve very bad performance especially for the case with large number of missing instances in every view. The experimental results of Concat and BSV in the tables and figures also prove that filling in the missing instances with the corresponding average instances is not a good choice to address the incomplete problem of multiview clustering.

2) From Tables II–V and Fig. 3, we can find that PMVC, IMG, DCNMF, GPMVC, and the proposed method can achieve much better performance than BSV and Concat in most cases, which proves that exploiting the complementary information of multiview features to learn a common representation is an effective approach in dealing with the incomplete problem.

3) From Tables II–V and Fig. 3, we can find that DCNMF and GPMVC perform better than PMVC on the handwritten digit database and BUAA database, while perform worse than PMVC in some cases especially on the Caltech7 and ORL databases. Compared with PMVC, DCNMF, and GPMVC all try to exploit the geometric structure of each view to guide the common representation learning. Thus, these experimental results indicate that exploiting the intrinsic geometric structure of data has the potential to learn a more discriminative and compact common representation for clustering. However, if the constructed geometric structure is not the intrinsic structure, it will lead to the opposite effect. Thus, capture the intrinsic structure of each view is crucial for these methods.
Fig. 3. Purities (%) of different methods on the (a) handwritten digit database, (b) BUAA database, (c) Cornell database, and (d) Caltech7 database with different rates of paired samples.

Table VI

<table>
<thead>
<tr>
<th>Method</th>
<th>NMI</th>
<th>ACC</th>
<th>Purity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSV</td>
<td>35.55</td>
<td>41.73</td>
<td>55.55</td>
</tr>
<tr>
<td>Concat</td>
<td>35.44</td>
<td>35.03</td>
<td>46.93</td>
</tr>
<tr>
<td>PMVC</td>
<td>64.50</td>
<td>75.50</td>
<td>82.36</td>
</tr>
<tr>
<td>IMG</td>
<td>65.75</td>
<td>76.76</td>
<td>83.41</td>
</tr>
<tr>
<td>DCNMF</td>
<td>65.63</td>
<td>76.76</td>
<td>83.25</td>
</tr>
<tr>
<td>GPMVC</td>
<td>70.87</td>
<td>82.83</td>
<td>86.65</td>
</tr>
</tbody>
</table>

Table VII

<table>
<thead>
<tr>
<th>Method</th>
<th>NMI</th>
<th>ACC</th>
<th>Purity</th>
</tr>
</thead>
<tbody>
<tr>
<td>BSV</td>
<td>45.04</td>
<td>57.59</td>
<td>60.73</td>
</tr>
<tr>
<td>Concat</td>
<td>55.78</td>
<td>60.56</td>
<td>63.60</td>
</tr>
<tr>
<td>IMSC AGL</td>
<td>85.41</td>
<td>74.23</td>
<td>76.53</td>
</tr>
</tbody>
</table>

4) The proposed method significantly outperforms the other methods on the above multiview databases with all kinds of incomplete cases. For instance, on the handwritten digit database (Table II), the proposed method achieves 3% and 6% improvement of ACC and NMI in comparison with the second best method, i.e., DCNMF, respectively. In Table VII, the NMI, ACC, and purity of the proposed method are about 33%, 26%, and 31% higher than those of BSV and Concat on the BBCSport database. These good experimental results strongly prove the effectiveness of the proposed method in dealing with all kinds of the incomplete multiview clustering tasks.

C. Analysis of the Parameter Sensitivity

In this section, we mainly focus on analyzing the sensitivity of the three tunable parameters, i.e., $\lambda_1$, $\lambda_2$, and $\lambda_3$, in model (11). We first define a candidate set, i.e., $\{10^{-5}, 10^{-4}, 10^{-3}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4, 10^5\}$, for the three parameters, and then perform the proposed method with different combinations of the three parameters [60]. In Figs. 4 and 5, we show the relationships of NMI (%) and the three parameters on the ORL and BUAA face datasets with 50% and 70% paired samples, respectively. From these figures, it is obvious that the proposed method can obtain the stable and satisfied NMIIs when the three parameters are located in some feasible areas. For example, on the ORL dataset, when parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$ locate in the range of $[10^{-5}, 10^{-2}]$, $[10^1, 10^5]$, and $[10^{-1}, 10^2]$, respectively, a satisfactory clustering performance can be obtained. This demonstrates that the proposed method is insensitive to the three parameters to some extent.

For different databases, it is still an open problem to adaptively select the optimal values for these parameters to our best knowledge. In this section, we provide a simple strategy to find the optimal combination of the three parameters for experiments. From Figs. 4 and 5, we can find that the proposed method is relatively insensitive to the selection of parameter $\lambda_1$ in the small parameter range $[10^{-5}, 10^{-2}]$ to some extent, thus we can set $\lambda_1$ with a fixed value like $10^{-3}$ at first, and then focus on finding the best combination of parameters $\lambda_2$ and $\lambda_3$. According to the experimental results shown in Figs. 4 and 5, we define two candidate sets $\{10^1, 10^2, 10^3, 10^4, 10^5\}$ and $\{10^{-2}, 10^{-1}, 10^0, 10^1, 10^2\}$ for parameters $\lambda_2$ and $\lambda_3$, respectively. Then perform the proposed method with different values...
of the two parameters. In this way, we can find the best combination of parameters $\lambda_2$ and $\lambda_3$ in the 2-D space formed by their candidate parameters. To obtain the optimal value of parameter $\lambda_1$, we can fix $\lambda_2$ and $\lambda_3$ with the obtained optimal combination and then perform the proposed method with different values of $\lambda_1$. As a result, the optimal combination of these three values can be achieved. Then we use the selected parameters to conduct experiments and report the results for comparison.

D. Convergence Analysis Based on Experiments

In this section, two experiments are conduct to prove the convergence property of the proposed optimization approach listed in Algorithm 1. In Fig. 6, we show the objective function value versus the iteration steps, in which the objective function is calculated as $\text{obj} = \sum_v (\gamma^{(v)}_1 + \gamma^{(v)}_2) / \sum_v \|X^{(v)}\|_F^2$ according to the original model (11), where $\gamma^{(v)}_1 = \|Z^{(v)}\|_2^2 + \lambda_1 \text{Tr}(F^{(v)T}G^{(v)}T_L^{(v)}G^{(v)}F^{(v)}) + \lambda_2 \|E^{(v)}\|_1 + \lambda_3 (c - \text{Tr}(F^{(v)T}F^{(v)T}UUT^{(v)}))$ and $\gamma^{(v)}_2 = \|Y^{(v)} - Z^{(v)}\|_F^2 + \|Z^{(v)} - W^{(v)}\|_F^2$. From Fig. 6, it is obvious that the objective function curve is monotonicly decreasing till the stable level, which demonstrates that our provided optimization approach monotonically decreases the objective problem. This also proves that the proposed method will converge to the local optimal point after a few iterations. Moreover, from the objective function value, we can find that the proposed method has very promising convergence efficiency.

VI. CONCLUSION

In this paper, a novel incomplete multiview clustering framework is proposed. Different from the conventional methods that exploit the matrix factorization technique for incomplete multiview clustering, the proposed method is the first one that integrates the spectral clustering and adaptive graph learning technique to tackle this problem. Compared with the conventional methods, the proposed method is more flexible since it is able to handle all kinds of incomplete cases. We have conducted several experiments on the two types of incomplete multiview data. Experimental results commonly show that the proposed method can achieve a better performance than some state-of-the-art methods, which proves its effectiveness in dealing with the incomplete multiview clustering tasks.


J. Candes, X. Li, Y. Ma, and J. Wright, “Robust principal component analysis?” *J. ACM*, vol. 58, no. 3, p. 11, 2011.


Jie Wen received the M.S. degree in detection technology & automatic equipment from Harbin Engineering University, Harbin, China, in 2015. He is currently pursuing the Ph.D. degree in computer science and technology from the Harbin Institute of Technology, Shenzhen, China. His current research interests include image and video processing, pattern recognition, data clustering, and machine learning.

Yong Xu (M’06–SM’15) received the Ph.D. degree in pattern recognition and intelligence system from the Nanjing University of Science and Technology, Nanjing, China, in 2005. He is currently the Harbin Institute of Technology, Shenzhen, Shenzhen, China. His current research interests include pattern recognition, biometrics, machine learning, and video analysis. More information please refer to http://www.yongxu.org/lanwen.html.

Hong Liu (M’05) received the Ph.D. degree in electrical control and automation engineering from the Harbin Institute of Technology, Harbin, China, in 1996. He was a Postdoctoral Fellow in computer science and technology from the Peking University, Beijing, China, in 1996. He is currently the Supervisor of doctoral students, and the Director of the Research Department of Shenzhen Graduate School and the Director of the Intelligent Robot Laboratory, Peking University, Shenzhen, China. He has done exchange visits in many famous universities and research institutions in several countries and regions, including the U.S., Canada, France, The Netherlands, Japan, South Korea, Singapore, and Hong Kong. He has published over 100 papers in the important scholarly journals and international conferences. His current research interests include image processing and pattern recognition, intelligent robots and computer vision, and intelligent micro-systems hardware and software co-design.

Dr. Liu was a recipient of the Korean Academy of an Jung-geun Awards, the Department of Space Science and Technology Progress Award, the Peking University Teaching Excellence Award, and the Aeta Award and candidates for Peking University Top Ten Teachers. He is an Executive Director and the Vice Secretary of Intelligent Automation Committee of Chinese Automation Association.