Multiview Subspace Clustering by an Enhanced Tensor Nuclear Norm

Wei Xia, *Graduate Student Member, IEEE*, Xiangdong Zhang, Quanxue Gao, Xiaochuang Shu, Jungong Han, and Xinbo Gao, *Senior Member, IEEE*

***Abstract*—Despite the promising preliminary results, tensor-singular value decomposition (t-SVD)-based multiview subspace is incapable of dealing with real problems, such as noise and illumination changes. The major reason is that tensor-nuclear norm minimization (TNNM) used in t-SVD regularizes each singular value equally, which does not make sense in matrix completion and coefficient matrix learning. In this case, the singular values represent different perspectives and should be treated differently. To well exploit the significant difference between singular values, we study the weighted tensor Schatten *p*-norm based on t-SVD and develop an efficient algorithm to solve the weighted tensor Schatten *p*-norm minimization (WTSNM) problem. After that, applying WTSNM to learn the coefficient matrix in multiview subspace clustering, we present a novel multiview clustering method by integrating coefficient matrix learning and spectral clustering into a unified framework. The learned coefficient matrix well exploits both the cluster structure and high-order information embedded in multiview views. The extensive experiments indicate the efficiency of our method in six metrics.**

***Index Terms*—Multiview clustering, spectral clustering, tensor-singular value decomposition (t-SVD), weighted nuclear norm.**

I. Introduction

MULTIVIEW data are ubiquitous in machine learning and artificial intelligence, and help provide complementary information embedded in multiviews for multiview clustering. Multiview clustering aims to separate multiview data into several meaningful groups and has become an active topic in artificial intelligence and data analysis [5], [16], [22], [25], [28]. Yang and Wang [39] provided a comprehensive review of multiview clustering. We, herein, center on multiview subspace clustering (MVSC) that is one of the most representative clustering techniques.

Subspace clustering aims to learn a robust coefficient matrix or affinity matrix, which is usually used for spectral clustering. Low-rank representation (LRR) well characterizes the relationship between data and has become one of the most representative techniques of learning the affinity matrix in subspace clustering [12]. For clustering of imaging data, Wu and Bajwa [33] considered imaging data as lateral slices of the tensor and proposed the structure-constrained low-rank submodule clustering (SCLRSmC) method, which models them as lying near a union of free submodules (UoFS) [1]. For multiview clustering, Zhang *et al.* [43] viewed affinity matrices, which are learned by different views via the self-representation technique, as lateral slices of tensor, and presented the low-rank tensor-constrained multiview subspace clustering (LT-MSC) method. LT-MSC captures the high-order information underlying multiview data by minimizing the nuclear norm of the tensor-unfolding matrix. However, the nuclear norm of the tensor-unfolding matrix is not a tight convex relaxation of both the Tucker rank and ℓ_1-norm [32], [37], [47]. To handle this problem, Lu *et al.* [17] proposed the tensor-singular value decomposition (t-SVD)-based tensor nuclear norm. This new norm is a convex relaxation of ℓ_1-norm. Motivated by this, Xie *et al.* [37] proposed a t-SVD-based multiview subspace clustering (t-SVD-MSC), method which well-characterizes high-order information embedded in multiview data.

Although the new tensor nuclear norm minimization (TNNM) achieves impressive results for multiview clustering, existing TNNM still exists the following shortcomings.

1) It neglects the significant difference between all singular values of a matrix due to the fact that tensor nuclear norm minimization leverages the same parameter to shrink all singular values. In real applications, there has a significant difference between nonsingular values of a matrix, and the first several largest singular values usually characterize the salient structure information embedded in the matrix. This significant difference, which is called prior information, is very important for image denoising, matrix completion, and so on, but similar investigations for multiview clustering

Manuscript received July 29, 2020; accepted January 12, 2021. This work was supported in part by the National Natural Science Foundation of China under Grant 61773302; in part by the Natural Science Basic Research Plan in Shaanxi Province under Grant 2020JZ-19 and Grant 2020JQ-327; and in part by the Innovation Fund of Xidian University, Special Projects for Key Fields in Higher Education of Guangdong, China, under Grant 2020ZDZX3077. This article was recommended by Associate Editor A. F. S. Gómez. *(Corresponding author: Quanxue Gao.)*

Wei Xia, Xiangdong Zhang, Quanxue Gao, and Xiaochuang Shu are with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China (e-mail: qxgao@xidian.edu.cn).

Jungong Han is with the Computer Science Department, Aberystwyth University, Aberystwyth SY23 3FL, U.K.

Xinbo Gao is with the State Key Laboratory of Integrated Services Networks, Xidian University, Xi'an 710071, China, also with the School of Electronic Engineering, Xidian University, Xi'an 710071, China, and also with the Chongqing Key Laboratory of Image Cognition, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCYB.2021.3052352.

Digital Object Identifier 10.1109/TCYB.2021.3052352

have been found lacking so far, which is one of the motivations behind this work. Thus, to well exploit the salient structure information embedded in affinity matrices, we should make the larger singular values shrink less, while t-SVD-MSC does not.

2) TPSSV (tensor partial sum of singular values) minimization [44], which is an extensor of TNNM, only shrinks the last several smaller singular values. Doing so implies that the small singular values characterize the unimportant structure information in data, while the larger singular values do not carry the information, which has nothing to do with the content of data. However, this assumption is very strict and does not makes sense in real applications. For example, given an image with large illumination variation in contents, the first three largest singular values contain somewhat illumination information [2] that has nothing to do with the content of the image.
3) t-SVD-MSC executes the coefficient representation and spectral clustering in two separated steps, which limits its performance.

To well exploit both the salient structure information and high-order information embedded in multiview data, inspired by the t-SVD-based nuclear norm, we present the t-SVD-based weighted tensor Schatten p-norm (WTSN) and study the minimization problem of WTSN (WTSNM). After that, we apply it to MVSC to exploit high-order correlation and propose an efficient algorithm that has good convergence. The main contributions of our work are summarized as follows.

1) We study the weighted tensor Schatten p-norm minimization (WTSNM) based on t-SVD, and propose an efficient algorithm to solve WTSNM, which has good convergence. The existing weighted tensor nuclear norm based on t-SVD can be considered as a special case of our model.
2) Applying WTSNM to MVSC, we propose a novel tensor low-rank constraint MVSC method. Our method attains a good affinity matrix, which well characterizes both the relationship between data and high-order information embedded in different views.
3) Our method integrates the coefficient matrix learning and spectral clustering into a unified framework. Thus, the learned coefficient representation well characterizes the cluster structure and encodes discriminant information.

II. Multiview Subspace Clustering

Multiview clustering has become an active topic in pattern analysis and artificial intelligence due to the ubiquitous multiview data in real applications [5], [11], [21], [23], [27], [35]. Being the efficiency of learning affinity matrix, which well characterizes the relationship in data, MVSC has become one of the most representative clustering techniques. It learns a unified coefficient matrix or affinity matrix from all views. The affinity matrix well exploits the relationship in multiview data. Then, clustering is performed on this affinity matrix. Self-representation is one of the successful subspace techniques and has been widely used in MVSC. The general self-representation multiview subspace model is

$$\min_{\mathbf{Z}\in\mathbb{C}} \sum_{v=1}^{m} \left\| \mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} \right\|_l + \lambda\Omega(\mathbf{Z}^1, \mathbf{Z}^2, \ldots, \mathbf{Z}^m) \quad (1)$$

where $\|\cdot\|_l$ denotes the metric of a matrix, and m is the number of views $\mathbf{X}^{(v)}$ and $\mathbf{Z}^{(v)}$ $(v = 1, \ldots, m)$ denote the data matrix and self-expression coefficient matrix of the vth view, respectively. $\mathbb{C}$ is a set of constraints on $\mathbf{Z}^{(v)}$. Parameter λ balances the error loss and regularized term $\Omega(\mathbf{Z}^{(v)})$. Applying different metrics to the first term and second term in the model (1), the researcher developed many impressive subspace clustering methods. For instance, Nie *et al.* [22] leveraged F-norm to characterize the self-representation error and presented a new MVSC. MVSC integrates self-representation subspace learning and spectral clustering into a unified framework to learn a common indicator matrix that preserves the cluster structure shared by different views. To enhance complementary information, Belhumeur *et al.* [2] leveraged the Hilbert Schmidt independence criterion (HSIC) to measure the diversity between $\mathbf{Z}^{(v)}$ $(v = 1, 2, \ldots, m)$, and proposed the diversity-induced MVSC method. To well exploit the local structure, which characterizes the relationship between data, ℓ_1-norm regularization is usually imposed on the coefficient representation to improve clustering performance [3], [9], [30], [31], [35], [48]. Inspired by this, Yin *et al.* [41] employed the ℓ_1-norm to characterize both the sparseness of coefficient representations $\mathbf{Z}^{(v)}$ and similarity between them. To well exploit the complementary information, Wang *et al.* [26] enforced the different coefficient representations $\mathbf{Z}^{(v)}$ to be diverse by minimizing the ℓ_1-norm of the Hadamard product between them, and proposed the exclusivity-consistency multiview subspace clustering (ECMVSC) method, which is robust to the magnitude of element values.

Although the impressive clustering performance have been obtained by the above methods, all of them are 1-D, element-based coefficient representation model, element by element. Thus, they neglect the spatial structure embedded in $\mathbf{Z}^{(v)}$ [4], [28]. To handle this problem, many low-rank constraint MVSC methods have been developed [7], [27]. For example, Ding and Fu [7] proposed a low-rank common subspace method. It imposes the nuclear norm constraint on both the projection matrix and common representation. To well exploit both the local structure and spatial structure, Wang *et al.* [29] imposed the low-rank constraint on each coefficient matrix, and then leveraged both the Laplacian regularization and view-agreement constraint to characterize the correlation consensus among multiview data. To well improve robustness of coefficient representations to noise, Wu and Bajwa [33] learned the affinity matrix from the latent subspace and presented the latent multiview subspace clustering (LMSC) method with low-rank constraint. Inspired by LMSC, Xie *et al.* [35] added the Laplacian regularization on the latent representation and developed a new clustering method, which well preserves the local geometric structure. To well exploit complementary information and high-order information, Luo *et al.* [18] proposed consistency-specificity multiview clustering by dividing the self-representation coefficient matrix of each view into consistency and specificity, where the consistency has a low-rank structure and is shared different views, and the specificity characterizes the inherent difference in each view.

However, the aforementioned low-rank constraint multiview clustering methods neglect the correlation consistency among coefficient representations [6], [45] due to the fact that they only impose a low-rank constraint on each view's coefficient representation directly. Thus, they cannot well exploit the high-order information and complementary information. To tackle this problem, Xie *et al.* [38] developed LT-MSC for multiview clustering. LT-MSC obtained the affinity matrix by minimizing the nuclear norm of the tensor-unfolding matrix. However, the nuclear norm is not a tight convex relaxation of the Tucker rank [12]. Motivated by the new t-SVD based tensor nuclear norm [8], [14], [37], recently, many new multiview clustering methods have been developed [12], [13], [17], [38], [39] and obtain impressive clustering performance. Two of the most representative methods are ETLMSC [12] and t-SVD-MSC [13]. ETLMSC leverages the tensor robust principal component analysis (TRPCA) model [17] to learn a robust 3-way clean graph, which well exploits both the high-order information and complementary information embedded in multiview data. t-SVD-MSC constructs a 3-way tensor whose frontal slices are composed of $\mathbf{Z}^{(v)}$, and then learns a 3-way affinity matrix by minimizing the new tensor nuclear norm. To well preserve the local geometric structure, Xie *et al.* [38] added hyper-Laplacian regularization in t-SVD-MSC and proposed a novel clustering method. However, all of them execute the coefficient matrix learning and spectral clustering in two separated steps, resulting in suboptimal performance. Moreover, they regularize each singular value equally due to the fact that they use the same parameter to shrink all singular values in optimizing the tensor nuclear norm minimization. So the aforementioned methods cannot exploit the significant difference between singular values, resulting in the instability of algorithms. To handle these problems, motivated by t-SVD-MSC and the new tensor nuclear norm, we studied the WTSNM and proposed a novel tensor low-rank-constrained MVSC method. Our method integrates affinity matrix learning and spectral clustering into a unified framework. Moreover, our method explicitly exploits the prior information embedded in singular values in solving tensor nuclear norm minimization. Thus, the learned coefficient representation, which is shared by different views, captures the high-order correlation and complementary information underlying multiview data.

III. Notations and Preliminaries

For convenience, we first introduce the notations and definitions used throughout this article. We use bold calligraphy letters for third-order tensors, for example, $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, bold uppercase letters for matrices, for example, $\mathbf{Z}$, bold lowercase letters for vectors, for example, $\mathbf{z}$, and lowercase letters such as z_{ijk} for the entries of $\boldsymbol{\mathcal{Z}}$. Moreover, we denote $\boldsymbol{\mathcal{Z}}^{(i)}$ by the ith frontal slice of $\boldsymbol{\mathcal{Z}}$ and $\overline{\boldsymbol{\mathcal{Z}}}$ by the discrete fast Fourier transform (FFT) of $\boldsymbol{\mathcal{Z}}$ along the third dimension, that is, $\overline{\boldsymbol{\mathcal{Z}}} = \mathit{fft}(\boldsymbol{\mathcal{Z}}, [], 3)$. Thus, $\boldsymbol{\mathcal{Z}} = \mathit{ifft}(\overline{\boldsymbol{\mathcal{Z}}}, [], 3)$.

Definition 1 [47]: Given tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, denote by $\boldsymbol{\mathcal{Z}}^T \in \mathbb{R}^{n_2\times n_1\times n_3}$ the conjugate transpose of $\boldsymbol{\mathcal{Z}}$, and $\|\boldsymbol{\mathcal{Z}}\|_F = \sqrt{\sum_{ijk} |z_{ijk}|^2}$.

Definition 2 [47]: Given tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, the block-diagonal matrix of tensor $\overline{\boldsymbol{\mathcal{Z}}}$ is

$$b\mathrm{diag}(\overline{\boldsymbol{\mathcal{Z}}}) = \mathrm{diag}(\overline{\mathbf{Z}}^{(1)}; \overline{\mathbf{Z}}^{(2)}; \cdots ; \overline{\mathbf{Z}}^{(n_3)}). \tag{2}$$

Definition 3 [47]: The block circulant matrix of the tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is defined as

$$bcirc(\boldsymbol{\mathcal{Z}}) = \begin{bmatrix} \boldsymbol{\mathcal{Z}}^{(1)} & \boldsymbol{\mathcal{Z}}^{(n_3)} & \cdots & \boldsymbol{\mathcal{Z}}^{(2)} \\ \boldsymbol{\mathcal{Z}}^{(2)} & \boldsymbol{\mathcal{Z}}^{(1)} & \cdots & \boldsymbol{\mathcal{Z}}^{(3)} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\mathcal{Z}}^{(n_3)} & \boldsymbol{\mathcal{Z}}^{(n_3-1)} & \cdots & \boldsymbol{\mathcal{Z}}^{(1)} \end{bmatrix}. \tag{3}$$

According to Definitions 2 and 3, we have the following theorem.

Theorem 1 [14], [47]: Given tensor $\boldsymbol{\mathcal{Z}}$, the relationship between $bcirc(\boldsymbol{\mathcal{Z}})$ and $\mathrm{bdiag}(\overline{\boldsymbol{\mathcal{Z}}})$ is

$$(\mathbf{F_{n_3}} \otimes \mathbf{I_{n_1}})\cdot bcirc(\boldsymbol{\mathcal{Z}})\cdot((\mathbf{F_{n_3}})^{-1} \otimes \mathbf{I_{n_2}}) = \mathbf{bdiag}(\overline{\boldsymbol{\mathcal{Z}}}) \tag{4}$$

where $\otimes$ is the Kronecker product, $\mathbf{I}_{n_1} \in \mathbb{R}^{n_1\times n_1}$ and $\mathbf{I}_{n_2} \in \mathbb{R}^{n_2\times n_2}$ are identity matrices, respectively, and $\mathbf{F}_{n_3} \in \mathbb{R}^{n_3\times n_3}$ is the discrete Fourier transform (DFT) matrix.

Definition 4 [47]: Given tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, then

$$\begin{aligned} \mathrm{unfold}(\boldsymbol{\mathcal{Z}}) &= \left[\boldsymbol{\mathcal{Z}}^{(1)}; \boldsymbol{\mathcal{Z}}^{(2)}; \cdots ; \boldsymbol{\mathcal{Z}}^{(n_3)}\right] \\ \mathrm{fold}(\mathrm{unfold}(\boldsymbol{\mathcal{Z}})) &= \boldsymbol{\mathcal{Z}}. \end{aligned} \tag{5}$$

Definition 5 [14]: Given tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ and $\boldsymbol{\mathcal{G}} \in \mathbb{R}^{n_2\times l\times n_3}$, the t-product between $\boldsymbol{\mathcal{Z}}$ and $\boldsymbol{\mathcal{G}}$ is $\boldsymbol{\mathcal{H}} \in \mathbb{R}^{n_1\times l\times n_3}$, that is

$$\boldsymbol{\mathcal{H}} = \boldsymbol{\mathcal{Z}} * \boldsymbol{\mathcal{G}} = \mathrm{fold}(bcirc(\boldsymbol{\mathcal{Z}})\cdot\mathrm{unfold}(\boldsymbol{\mathcal{G}})). \tag{6}$$

The model (6) can be efficiently calculated by two steps. First, obtain $\overline{\boldsymbol{\mathcal{H}}}^{(i)}=\overline{\boldsymbol{\mathcal{Z}}}^{(i)} \cdot \overline{\boldsymbol{\mathcal{G}}}^{(i)}$, $i = 1, 2, \ldots, n_3$. Second, obtain $\boldsymbol{\mathcal{H}} = \mathit{ifft}(\overline{\boldsymbol{\mathcal{H}}}, [], 3)$.

Definition 6 [14]: Given tensor $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, if $\boldsymbol{\mathcal{D}}^{(i)}$ $(i = 1, 2, \ldots, n_3)$ are diagonal matrices, then $\boldsymbol{\mathcal{D}}$ is an f-diagonal tensor.

Theorem 2 [14]: Given tensor $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$, then t-SVD of $\boldsymbol{\mathcal{Z}}$ is

$$\boldsymbol{\mathcal{Z}} = \boldsymbol{\mathcal{U}} * \boldsymbol{\mathcal{D}} * \boldsymbol{\mathcal{V}}^T \tag{7}$$

where $\boldsymbol{\mathcal{U}} \in \mathbb{R}^{n_1\times n_1\times n_3}$ and $\boldsymbol{\mathcal{V}} \in \mathbb{R}^{n_2\times n_2\times n_3}$ are orthogonal, and $\boldsymbol{\mathcal{D}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is an f-diagonal tensor.

Definition 7: The nuclear norm of $\boldsymbol{\mathcal{Z}} \in \mathbb{R}^{n_1\times n_2\times n_3}$ is defined as [14] and [47]

$$\|\boldsymbol{\mathcal{Z}}\|_{\circledast} = \sum_{i=1}^{n_3} \left\|\overline{\boldsymbol{\mathcal{Z}}}^{(i)}\right\|_* = \sum_{i=1}^{n3} \sum_{j=1}^{h} \sigma_j\left(\overline{\boldsymbol{\mathcal{Z}}}^{(i)}\right) \tag{8}$$

where $\sigma_j(\overline{\boldsymbol{\mathcal{Z}}}^{(i)})$ is the jth singular value of $\overline{\boldsymbol{\mathcal{Z}}}^{(i)}$, $h = \min(n_1, n_2)$.

IV. Weighted Tensor Schatten p-Norm Minimization

A. Problem Formulation and Objective

Recently, TNNM has been widely used in many applications, such as multiview clustering, color image denosing, matrix completion, and so on [13], [17], [32], [37]–[40]. A general TNNM is

$$\underset{\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{A}}\|_F^2+\tau\|\boldsymbol{\mathcal{X}}\|_{\circledast}. \tag{9}$$

According to Definition 7 and the relationship between the time domain and Fourier domain, the model (9) can be converted to Fourier domain to solve. Moreover, according to Definitions 1 and 7, we have that the model (9) can be finally divided into the following n_3 independent models:

$$\arg\min_{\overline{\boldsymbol{\mathcal{X}}}^{(i)}} \frac{1}{2}\left\|\overline{\boldsymbol{\mathcal{X}}}^{(i)}-\overline{\boldsymbol{\mathcal{A}}}^{(i)}\right\|_F^2+\sum_{j=1}^{h}\tau\cdot\sigma_j(\overline{\boldsymbol{\mathcal{X}}}^{(i)}) \tag{10}$$

where $h=\min(n_1,n_2)$, $i=1,2,\ldots,n_3$.

The optimal solution $\overline{\boldsymbol{\mathcal{X}}}^{(i)*}$ in (10) can be obtained by shrinking the singular values $\sigma_j(\overline{\boldsymbol{\mathcal{A}}}^{(i)})$ via $\sigma_j{}^*(\overline{\boldsymbol{\mathcal{A}}}^{(i)}) = \max(\sigma_j(\overline{\boldsymbol{\mathcal{A}}}^{(i)})-\tau,0)$. It can be seen that all singular vales are considered to be equally important. However, this is unreasonable in real applications due to the fact that there exists a significant difference between singular values of a matrix, and large singular values characterize the main structure of the matrix. It means that the model (9) neglects this prior information. To well exploit the salient structure information embedded in data, we should make the larger singular values shrink less in tensor nuclear norm minimization. Before introducing our model, we first introduce the definition of WTSN as follows.

Definition 8: Given $\boldsymbol{\mathcal{X}}\in\mathbb{R}^{n_1\times n_2\times n_3}$, $h=\min(n_1,n_2)$, WTSN $\|\boldsymbol{\mathcal{X}}\|_{\boldsymbol{\omega},S_p}$ of $\boldsymbol{\mathcal{X}}$ is defined as

$$\begin{aligned}\|\boldsymbol{\mathcal{X}}\|_{\boldsymbol{\omega},S_p} &= \left(\sum_{i=1}^{n3}\left\|\overline{\boldsymbol{\mathcal{X}}}^{(i)}\right\|_{\boldsymbol{\omega},S_p}^p\right)^{\frac{1}{p}}\\ &= \left(\sum_{i=1}^{n3}\sum_{j=1}^{h}\omega_j\cdot\sigma_j(\overline{\boldsymbol{\mathcal{X}}}^{(i)})^p\right)^{\frac{1}{p}}\end{aligned} \tag{11}$$

where ω_j denotes the jth element of the weighted vector $\boldsymbol{\omega}$, p is a parameter of power, and $\sigma_j(\overline{\boldsymbol{\mathcal{X}}}^{(i)})$ denotes the jth singular value of $\overline{\boldsymbol{\mathcal{X}}}^{(i)}$. For the sake of description, we assume all singular value are in nonincreasing order in our paper. Obviously, when $p=1$, (11) reduces to the weighted tensor nuclear norm [11], [19].

Then, we propose the WTSNM problem whose objective function is

$$\underset{\boldsymbol{\mathcal{X}}}{\operatorname{argmin}} \frac{1}{2}\|\boldsymbol{\mathcal{X}}-\boldsymbol{\mathcal{A}}\|_F^2+\tau\|\boldsymbol{\mathcal{X}}\|_{\boldsymbol{\omega},S_p}^p. \tag{12}$$

According to Definition 8, we have that the model (12) explicitly considers the significant difference between singular values by choosing p and $\boldsymbol{\omega}$.

Algorithm 1 Generalized Soft-Thresholding

Input: σ,ω,p,T
1. $\tau_p^{GST}(\omega)=(2\omega\cdot(1-p))^{\frac{1}{2-p}}+\omega\cdot p\cdot(2\omega\cdot(1-p))^{\frac{p-1}{2-p}}$
if $|\sigma|\le\tau_p^{GST}(\omega)$ **then**
  $T_p^{GST}(\sigma,\omega)$=0
else
  $k=0$, $\delta^{(k)}=|\sigma|$;
  for $k=0,1,\ldots,T$ **do**
    $\delta^{(k+1)}=|\sigma|-\omega\cdot p\cdot(\delta^{(k)})^{p-1}$
  end
  $T_p^{GST}(\sigma,\omega)=\operatorname{sign}(\sigma)\cdot\delta(k)$
end
Return $T_p^{GST}(\sigma,\omega)$

B. Optimization

For solving the WTSNM, that is, the model (12), we first introduce the following lemmas and theorems.

Lemma 1 (Generalized Soft-Thresholding) [42]: For the following optimization problem:

$$\min_{\delta\ge0}\quad f(\delta)=\frac{1}{2}(\delta-\sigma)^2+\omega\cdot\delta^p \tag{13}$$

with the given p and ω, there exists a specific threshold

$$\tau_p^{GST}(\omega)=(2\omega\cdot(1-p))^{\frac{1}{2-p}}+\omega\cdot p\cdot(2\omega\cdot(1-p))^{\frac{p-1}{2-p}}. \tag{14}$$

We have the following conclusion.

1) When $\sigma\le\tau_p^{\mathrm{GST}}(\omega)$, the optimal solution $T_p^{\mathrm{GST}}(\sigma,\omega)$ of (13) is 0.
2) When $\sigma>\tau_p^{\mathrm{GST}}(\omega)$, the optimal solution of (13) is $T_p^{\mathrm{GST}}(\sigma,\omega)=\operatorname{sign}(\sigma)\cdot S_{\mathrm{p}}^{\mathrm{GST}}(\sigma,\omega)$, where $S_p^{\mathrm{GST}}(\sigma,\omega)$ can be obtain by solving $S_p^{\mathrm{GST}}(\sigma,\omega)-\sigma+\omega\cdot p\cdot(S_p^{\mathrm{GST}}(\sigma,\omega))^{p-1}=0$.

We summarize the pseudocode of the generalized soft-thresholding (GST) in Algorithm 1.

Theorem 3 [36]: Let $\mathbf{Y}=\mathbf{U}_Y\cdot\mathbf{D}_Y\cdot\mathbf{V}_Y^T$ be the SVD of $\mathbf{Y}\in\mathbb{R}^{m\times n}$, $\tau>0$ and $l=\min(m,n)$, $0\le\omega_1\le\omega_2\le\cdots\le\omega_l$, a global optimal solution of the following model:

$$\arg\min_{\mathbf{X}}\frac{1}{2}\|\mathbf{X}-\mathbf{Y}\|_F^2+\tau\|\mathbf{X}\|_{\boldsymbol{\omega},S_p}^p \tag{15}$$

is

$$\Gamma_{\tau\cdot\boldsymbol{\omega}}[\mathbf{Y}]=\mathbf{U}_Y\cdot\mathbf{P}_{\tau\cdot\boldsymbol{\omega}}\cdot(\mathbf{Y})\cdot\mathbf{V}_Y^T \tag{16}$$

where $\mathbf{P}_{\tau\cdot\boldsymbol{\omega}}(\mathbf{Y}) = \operatorname{diag}(\gamma_1,\gamma_2,\ldots,\gamma_l)$ and $\gamma_i = T_p^{GST}(\sigma_i(\mathbf{Y}),\tau\cdot\omega_i)$, which can be obtained by Algorithm 1.

The fact that a closed-form global minimizer can be found comes from von Neumann's trace inequality [20]: $\{\sigma_i(\mathbf{Y})\}$ is in the nonincreasing order while $\{\ \omega_i\}$ is in the nondecreasing order.

Theorem 4: Suppose $\boldsymbol{\mathcal{A}}\in\mathbb{R}^{n_1\times n_2\times n_3}$, $l=\min(n_1,n_2)$, $0\le\omega_1\le\omega_2\le\cdots\le\omega_l$, let $\boldsymbol{\mathcal{A}}=\boldsymbol{\mathcal{U}}*\boldsymbol{\mathcal{S}}*\boldsymbol{\mathcal{V}}^T$. For our model (12), the optimal solution is

$$\boldsymbol{\mathcal{X}}^*=\Gamma_{\tau\cdot n_3\cdot\boldsymbol{\omega}}(\boldsymbol{\mathcal{A}})=\boldsymbol{\mathcal{U}}*\mathit{ifft}(\mathbf{P}_{\tau\cdot n_3\cdot\boldsymbol{\omega}}(\overline{\boldsymbol{\mathcal{A}}}))*\boldsymbol{\mathcal{V}}^T \tag{17}$$

where $\mathbf{P}_{\tau\cdot n_3\cdot\boldsymbol{\omega}}(\overline{\mathcal{A}})$ is a tensor, and $\mathbf{P}_{\tau\cdot n_3\cdot\boldsymbol{\omega}}(\overline{\mathcal{A}}^{(i)})$ is the ith frontal slice of $\mathbf{P}_{\tau\cdot n_3\cdot\boldsymbol{\omega}}(\overline{\mathcal{A}})$.

Proof: In the Fourier domain and combining the definition of WTSN, the model (12) can be reformulated as

$$\overline{\mathcal{X}}^* = \underset{\overline{\mathcal{X}}}{\operatorname{argmin}} \frac{1}{2}\left\|\overline{\mathcal{X}} - \overline{\mathcal{A}}\right\|_F^2 + \sum_{i=1}^{n3} \tau\cdot n_3 \cdot \left\|\overline{\mathcal{X}}^{(i)}\right\|_{\boldsymbol{\omega},S_p}^p. \tag{18}$$

According to Definition 1, we have

$$\underset{\overline{\mathcal{X}}}{\operatorname{argmin}} \sum_{i=1}^{n_3}\left(\frac{1}{2}\left\|\overline{\mathcal{X}}^{(i)} - \overline{\mathcal{A}}^{(i)}\right\|_F^2 + \tau\cdot n_3\cdot\left\|\overline{\mathcal{X}}^{(i)}\right\|_{\boldsymbol{\omega},S_p}^p\right) \tag{19}$$

where $\overline{\mathcal{X}}^{(i)}$ is the ith frontal slice of $\overline{\mathcal{X}}$.

In (19), each variable $\overline{\mathcal{X}}^{(i)}$ is independent. Thus, it can be divided into n_3 independent subproblems. For the ith ($i = 1, 2, \ldots, n_3$) subproblem, we have

$$\overline{\mathcal{X}}^{(i)*} = \underset{\overline{\mathcal{X}}^{(i)}}{\arg\min} \frac{1}{2}\left\|\overline{\mathcal{X}}^{(i)} - \overline{\mathcal{A}}^{(i)}\right\|_F^2 + \tau\cdot n_3\cdot\left\|\overline{\mathcal{X}}^{(i)}\right\|_{\boldsymbol{\omega},S_p}^p. \tag{20}$$

According to Theorem 3, the solution of (20) is $\overline{\mathcal{X}}^{(i)*} = \Gamma_{\tau\cdot n_3\cdot\boldsymbol{\omega}}[\overline{\mathcal{A}}^{(i)}] = \overline{\mathcal{U}}^{(i)}\mathbf{P}_{\tau\cdot n_3\cdot\boldsymbol{\omega}}(\overline{\mathcal{A}}^{(i)})\overline{\mathcal{V}}^{(i)T}$, which is the ith frontal slice of $\overline{\mathcal{X}}^*$. Since we obtain global solutions of all subproblems, according to Definition 5, we can easily obtain the global solution of the optimization problem (12), that is

$$\mathcal{X}^* = \Gamma_{\tau\cdot n_3\cdot\boldsymbol{\omega}}[\mathcal{A}] = \mathcal{U} * \mathit{ifft}\left(\mathbf{P}_{\tau\cdot n_3\cdot\boldsymbol{\omega}}\left(\overline{\mathcal{A}}\right)\right) * \mathcal{V}^T \tag{21}$$

where $\mathcal{U} = \mathit{ifft}(\overline{\mathcal{U}}, [], 3)$, $\mathcal{V} = \mathit{ifft}(\overline{\mathcal{V}}, [], 3)$. ■

V. Multiview Clustering Based on WTSNM

A. Problem Formulation

Given multiview dataset $\{\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(m)}\}$, $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v*N}$ denotes the data matrix of the vth ($v = 1, 2, \ldots, m$) view; d_v and N denote the dimension and number of samples in the vth view, respectively; and m is the number of views. Inspired by LRR, the coefficient matrix or affinity matrix $\mathbf{Z}^{(v)}$, which is learned by LRR in the vth view, has a low-rank structure, and the low-rank structures between $\mathbf{Z}^{(v)}$ ($v = 1, 2, \ldots, m$) are similar. Thus, tensor $\mathcal{Z}$, which consists of $\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots, \mathbf{Z}^{(m)}$, has a good tensor low-rank structure. To well exploit this structure and high-order information embedded in $\mathcal{Z}$, the t-SVD based tensor low-rank constraint has been widely used in multiview clustering and obtains impressive experimental results [13], [32], [37], [38].

The existing tensor low-rank constraint methods consider each singular value equally and shrink all singular values via the same parameter. However, in real applications, there has been a significant difference between nonzero singular values of a matrix, and the first several largest singular values usually characterize the salient structure information embedded in the matrix. The significant difference, which is called prior information, is very important for image denoising, matrix completion, and so on, but similar investigations for multiview clustering have been found to be lacking so far, which

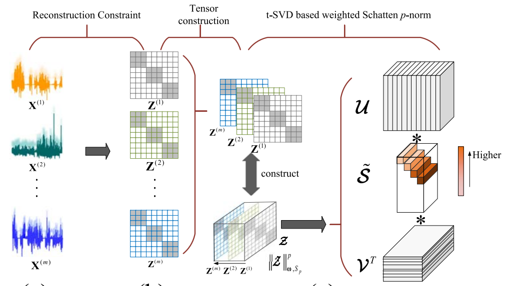


Fig. 1. Framework of the proposed method. (a) Multi-view data. (b) Multi-view subspace representation. (c) Tensor construction and the weighted tensor Schatten p-norm minimization.

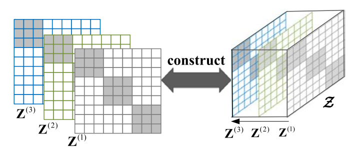


Fig. 2. Tensor $\mathcal{Z}$ construction.

is one of the motivations behind this work. This degrades the performance of clustering algorithms significantly in the existence of noise such as illumination variation. Moreover, the existing tensor low-rank multiview methods execute the affinity matrices learning and spectral clustering in two separated steps. Thus, the learned affinity matrices cannot well characterize the cluster structure. This limits the multiview clustering performance.

To handle these limitations, we propose a new multiview subspace clustering by using our proposed WTSNM. Fig. 1 shows the framework of the proposed model. We learn the self-representation coefficient matrix for each view and employ a tensor low-rank constraint to obtain a robust self representation, which well exploits both the high-order information and complementary information, by solving a WTSNM problem, and then incorporates the spectral clustering into a unified framework. It helps make the final fusion similarity matrix characterize the cluster structure and be prominent for clustering. The objective function is formulated as

$$\begin{aligned}\min_{\mathcal{Z},\mathbf{E}^{(v)},\mathbf{F}} \quad & \|\mathcal{Z}\|_{\boldsymbol{\omega},S_p}^p + \lambda\|\mathbf{E}\|_{2,1} + 2\alpha tr(\mathbf{F}^T\mathbf{L}_{\hat{\mathbf{Z}}}\mathbf{F}) \\ \text{s.t.} \quad & \mathbf{X}^{(v)} = \mathbf{X}^{(v)}\mathbf{Z}^{(v)} + \mathbf{E}^{(v)}, v = 1, 2, \ldots, m\end{aligned} \tag{22}$$

where the lateral slices of tensor $\mathcal{Z} \in \mathbb{R}^{N\times m\times N}$ of $\mathbf{Z}^{(v)}$, that is, $\mathcal{Z}(:, v, :) = \mathbf{Z}^{(v)}$ (see Fig. 2). $\mathbf{E}^{(v)} \in \mathbb{R}^{d_v*N}$ is the error matrix of the vth view, and $\mathbf{E} = [\mathbf{E}^{(1)}; \cdots; \mathbf{E}^{(m)}]$ can enforce the column of $\mathbf{E}^{(v)}$ in each view to have jointly consistent magnitude values. $\mathbf{L}_{\hat{\mathbf{Z}}} = \mathbf{D}_{\hat{\mathbf{Z}}} - \hat{\mathbf{Z}}$ is the Laplacian matrix, $\hat{\mathbf{Z}} = (1)/(m)\sum_{v=1}^m [(|\mathbf{Z}^{(v)}| + |\mathbf{Z}^{(v)}|^T)/(2)]$ and $\mathbf{D}_{\hat{\mathbf{Z}}}$ is a diagonal matrix, whose diagonal entries are $\mathbf{D}_{\hat{\mathbf{Z}}}(i, i) = \sum_j (\hat{\mathbf{Z}}_{ij} + \hat{\mathbf{Z}}_{ji})$. $\mathbf{F} \in \mathbb{R}^{c*N}$ denotes the cluster indicator matrix, and c is the number of clusters. α and λ are two balance parameters.

B. Optimization

Inspired by the inexact augmented lagrange multiplier (ALM), we introduce an auxiliary tensor variable $\mathcal{J}$ and rewrite the model (22) as minimizing the following unconstrained problem:

$$\begin{aligned}&\mathcal{L}(\mathbf{Z}^{(1)},\dots,\mathbf{Z}^{(m)},\mathcal{J},\mathbf{E}^{(1)},\dots,\mathbf{E}^{(m)},\mathbf{F})\\&=\|\mathcal{J}\|^p_{\boldsymbol{\omega},S_p}+\lambda\|\mathbf{E}\|_{2,1}+2\alpha tr(\mathbf{F}^T\mathbf{L}_{\hat{\mathbf{Z}}}\mathbf{F})\\&\quad+\sum_{v=1}^m\left(\left\langle\mathbf{Y}^{(v)},\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\right\rangle\right.\\&\qquad\left.+\frac{\mu}{2}\left\|\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\right\|_F^2\right)\\&\quad+\langle\mathcal{Q},\mathcal{Z}-\mathcal{J}\rangle+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{J}\|_F^2\end{aligned}\tag{23}$$

where the matrix $\mathbf{Y}_v$ and tensor $\mathcal{Q}$ represent two Lagrange multipliers, and μ and ρ are actually the penalty parameters.

Model (23) can be divided into four subproblems as follows.

$\mathbf{Z}^{(v)}$-Subproblem (Variables $\mathbf{E}$, $\mathcal{J}$, and $\mathbf{F}$ Are Fixed):

$$\begin{aligned}&\arg\min_{\mathbf{Z}^{(v)}}2\alpha tr(\mathbf{F}^T\mathbf{L}_{\hat{\mathbf{Z}}}\mathbf{F})+\langle\mathcal{Q},\mathcal{Z}-\mathcal{J}\rangle\\&\quad+\sum_{v=1}^m\left(\left\langle\mathbf{Y}^{(v)},\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\right\rangle\right.\\&\qquad\left.+\frac{\mu}{2}\left\|\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\right\|_F^2\right)\\&\quad+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{J}\|_F^2\\&=\arg\min_{\mathbf{Z}^{(v)}}2\alpha tr(\mathbf{F}^T\mathbf{L}_{\hat{\mathbf{Z}}}\mathbf{F})+\langle\mathcal{Q},\mathcal{Z}-\mathcal{J}\rangle\\&\quad+\sum_{v=1}^m\left(\frac{\mu}{2}\left\|\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}+\frac{\mathbf{Y}^{(v)}}{\mu}\right\|_F^2\right.\\&\qquad\left.+\frac{\rho}{2}\left\|\mathbf{Z}^{(v)}-\mathbf{J}^{(v)}+\frac{\mathbf{W}^{(v)}}{\rho}\right\|_F^2\right)+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{J}\|_F^2.\end{aligned}\tag{24}$$

Denote by $\mathbf{P}=[\mathbf{P}_1,\dots,\mathbf{P}_j,\dots,\mathbf{P}_N]$, where $\mathbf{P}_j=[\|\mathbf{F}_1-\mathbf{F}_j\|_2^2;\cdots;\|\mathbf{F}_N-\mathbf{F}_j\|_2^2]$, $\mathbf{F}_j$ is the jth row of $\mathbf{F}$. Then, we have

$$\begin{aligned}2tr(\mathbf{F}^T\mathbf{L}_{\hat{\mathbf{Z}}}\mathbf{F})&=tr(\mathbf{P}^T\hat{\mathbf{Z}})\\&=tr\left(\mathbf{P}^T\left(\frac{1}{m}\sum_{v=1}^m\frac{\left|\mathbf{Z}^{(v)}\right|+\left|\mathbf{Z}^{(v)}\right|^T}{2}\right)\right)\\&=\frac{1}{2m}\sum_{v=1}^m tr\left(\mathbf{P}^T\left|\mathbf{Z}^{(v)}\right|+\mathbf{P}^T\left|\mathbf{Z}^{(v)}\right|^T\right).\end{aligned}\tag{25}$$

Then, the model (24) becomes

$$\begin{aligned}&\arg\min_{\mathbf{Z}^{(v)}}\frac{\alpha}{2m}tr\left(\mathbf{P}^T\left|\mathbf{Z}^{(v)}\right|+\mathbf{P}^T\left|\mathbf{Z}^{(v)}\right|^T\right)\\&\quad+\frac{\mu}{2}\left\|\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}+\frac{\mathbf{Y}^{(v)}}{\mu}\right\|_F^2\\&\quad+\frac{\rho}{2}\left\|\mathbf{Z}^{(v)}-\mathbf{J}^{(v)}+\frac{\mathbf{W}^{(v)}}{\rho}\right\|_F^2.\end{aligned}\tag{26}$$

The solution of (26) is

$$\begin{aligned}\mathbf{Z}_t^{(v)*}&=\left(\mu(\mathbf{X}^{(v)})^T\mathbf{X}^{(v)}+\rho\mathbf{I}\right)^{-1}\\&\quad\left(\mu(\mathbf{X}^{(v)})^T\mathbf{X}^{(v)}+(\mathbf{X}^{(v)})^T\mathbf{Y}^{(v)}+\rho J^{(v)}\right.\\&\qquad-\mu(\mathbf{X}^{(v)})^T\mathbf{E}^{(v)}-\mathbf{W}^{(v)}\\&\qquad\left.-\frac{\alpha}{2m}\left(\mathbf{P}\odot sign(\mathbf{Z}_{t-1}^{(v)})+\mathbf{P}^T\odot sign(\mathbf{Z}_{t-1}^{(v)})^T\right)\right).\end{aligned}\tag{27}$$

$\mathbf{E}^{(v)}$-Subproblem: In this case, variables $\mathbf{Z}^{(v)}$, $\mathbf{F}$, and $\mathcal{J}$ are fixed. Thus, we have

$$\begin{aligned}&\arg\min_{\mathbf{E}}\lambda\|\mathbf{E}\|_{2,1}+\sum_{v=1}^m\left\langle\mathbf{Y}_v,\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\right\rangle\\&\quad+\sum_{v=1}^m\frac{\mu}{2}\left\|\mathbf{X}^{(v)}-\mathbf{X}^{(v)}\mathbf{Z}^{(v)}-\mathbf{E}^{(v)}\right\|_F^2\\&=\arg\min_{\mathbf{E}}\frac{\lambda}{\mu}\|\mathbf{E}\|_{2,1}+\frac{1}{2}\|\mathbf{E}-\mathbf{D}\|_F^2.\end{aligned}\tag{28}$$

The optimal solution is [7]

$$\mathbf{E}^*_{:,i}=\begin{cases}\dfrac{\left\|\mathbf{D}_{:,i}\right\|_2-\frac{\lambda}{\mu}}{\left\|\mathbf{D}_{:,i}\right\|_2}\mathbf{D}_{:,i} & \left\|\mathbf{D}_{:,i}\right\|_2>\dfrac{\lambda}{\mu}\\ \mathbf{0} & \text{otherwise}\end{cases}\tag{29}$$

where $\mathbf{D}_{:,i}$ denotes the ith column of $\mathbf{D}=[\mathbf{D}^1;\cdots;\mathbf{D}^m]$, $\mathbf{D}^j=\mathbf{X}^{(j)}-\mathbf{X}^{(j)}\mathbf{Z}^{(j)}+(1)/(\mu)\mathbf{Y}_j$, $j=1,\dots,m$.

$\mathcal{J}$-Subproblem (Variables $\mathbf{Z}^{(v)}$, $\mathbf{E}^{(v)}$ and $\mathbf{F}$ Are Fixed):

$$\begin{aligned}\mathcal{J}^*&=\arg\min_{\mathcal{J}}\|\mathcal{J}\|^p_{\boldsymbol{\omega},S_p}+\langle\mathcal{Q},\mathcal{Z}-\mathcal{J}\rangle+\frac{\rho}{2}\|\mathcal{Z}-\mathcal{J}\|_F^2\\&=\arg\min_{\mathcal{J}}\|\mathcal{J}\|^p_{\boldsymbol{\omega},S_p}+\frac{\rho}{2}\left\|\mathcal{Z}-\mathcal{J}+\frac{\mathcal{Q}}{\rho}\right\|_F^2.\end{aligned}\tag{30}$$

According to Theorem 4, the solution of the model (30) is

$$\mathcal{J}^*=\Gamma_{\frac{1}{\rho}\cdot n_3\cdot\boldsymbol{\omega}}\left(\mathcal{Z}+\frac{1}{\rho}\mathcal{Q}\right).\tag{31}$$

$\mathbf{F}$-Subproblem (Other Variables $\mathbf{Z}^{(v)}$, $\mathbf{E}$ and $\mathcal{J}$ Are Fixed): In this case, the model (23) becomes

$$\begin{aligned}\mathbf{F}=\arg\min_{\mathbf{F}}&\ tr\left(\mathbf{F}^T\mathbf{L}_{\hat{\mathbf{Z}}}\mathbf{F}\right)\\\text{s.t.}&\ \mathbf{F}^T\mathbf{F}=\mathbf{I},\mathbf{F}\in\mathbb{R}^{N\times c}.\end{aligned}\tag{32}$$

The optimal solution $\mathbf{F}$ consists of the eigenvectors corresponding to the c smallest eigenvalues of $\mathbf{L}_{\hat{\mathbf{Z}}}$.

Finally, we summarize the pseudocode in Algorithm 2.

C. Convergence Analysis

The existing works [8] and [12] have demonstrated that when the number of block variables is more than 2, it is still an open problem to prove the convergence of inexact ALM. In Algorithm 2, we have four block variables $\mathbf{E}$, $\mathcal{J}$, $\mathbf{F}$, and $\mathcal{Z}$. Thus, we cannot prove the convergence of our proposed algorithm. However, inspired by the theoretical results in [8], our proposed algorithm is considered to be converged if the following two conditions are satisfied. First, in the model (22),

Algorithm 2 WTSNM for Multiview Clustering

Input: Given Multi-view data: $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \ldots, \mathbf{X}^{(m)}$, λ, $\boldsymbol{\omega}$, and cluster number K.
Output: Clustering result C.
Initialized: $\mathbf{Z}^{(v)} = \mathbf{0}$, $\mathbf{E}^{(v)} = \mathbf{0}$, $\mathbf{Y}^{(v)} = \mathbf{0}$, $i = 1, \ldots, m$, $\mathcal{J} = \mathbf{0}$, $\mathcal{Q} = \mathbf{0}$, $\mu = 10^{-5}$, $\rho = 10^{-4}$, $\eta = 2$, $\mu_{\max} = \rho_{\max} = 10^{10}$, $\varepsilon = 10^{-7}$.
while not converge **do**
(1) Update $\mathbf{Z}^{(v)}$, $(v = 1, 2, \cdots, m)$ by (27);
(2) Update $\mathbf{E}$ by (29);
(3) Update $\mathbf{Y}^{(v)}$, $(v = 1, 2, \cdots, m)$ by $\mathbf{Y}^{(v)} = \mathbf{Y}^{(v)} + \mu(\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)})$;
(4) Obtain $\mathcal{Z} = \Phi(\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \ldots, \mathbf{Z}^{(m)})$;
(5) Update $\mathcal{J}$ by (31);
(6) Update $\mathcal{Q}$ by $\mathcal{Q} = \mathcal{Q} + \rho(\mathcal{Z} - \mathcal{J})$;
(7) Update $\mathbf{F}$ by (32);
(8) Update parameters μ and ρ : $\mu = \min(\eta\mu, \mu_{\max})$, $\rho = \min(\eta\rho, \rho_{\max})$;
(9) Obtain $(\mathbf{J}^{(1)}, \mathbf{J}^{(2)}, \ldots, \mathbf{J}^{(m)}) = \Phi^{-1}(\mathcal{J})$;
(10) Check the convergence conditions:
$\left\|\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)}\right\|_\infty < \varepsilon$ & $\left\|\mathbf{Z}^{(v)} - \mathbf{J}^{(v)}\right\|_\infty < \varepsilon$;
end
(11) Obtain the affinity matrix by
$\mathbf{S} = \frac{1}{m}\sum_{v=1}^{m}(\left|\mathbf{Z}^{(v)}\right| + \left|\mathbf{Z}^{(v)^T}\right|)$;
(12) Output C via performing spectral clustering on $\mathbf{S}$.

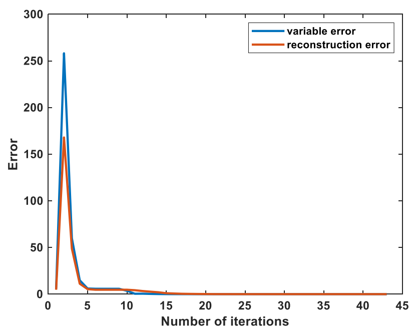


Fig. 3. Convergence curves on the Caltech-101 database.

the dictionary matrix should be of full common rank. Second, $\text{error}_k = \|(\mathcal{Z}_t, \mathbf{E}_t, \mathbf{F}_t) - \mathcal{L}_{\mathcal{Z},\mathbf{E},\mathbf{F}}\|_F^2$ is monotonically decreasing, where $\mathcal{Z}_t, \mathbf{E}_t, \mathbf{F}_t$ is the solution at the tth iteration in Algorithm 2, and $\mathcal{L}_{\mathcal{Z},\mathbf{E},\mathbf{F}}$ denotes the solution by minimizing the model (23) with respect to (w.r.t.) $\mathcal{Z}$, $\mathbf{E}$, and $\mathbf{F}$ simultaneously. The first condition can be satisfied by orhogonalizing the columns of $\mathbf{X}^{(v)^T}$ [21]. For the second condition, the convexity of the Lagrangian function could guarantee its validity to some extent according to the work of Eckstein and Bertsekas [8]. Therefore, Algorithm 2 should have good convergence properties. To further show the convergence, Fig. 3 lists the reconstruction error $\|\mathbf{X}^{(v)} - \mathbf{X}^{(v)}\mathbf{Z}^{(v)} - \mathbf{E}^{(v)}\|_\infty$ and variable error $\|\mathbf{Z}^{(v)} - \mathbf{J}^{(v)}\|_\infty$ versus the number of iterations on the Caltech-101 database. It can be seen that our proposed algorithm has good convergence in the experiments.

D. Complexity Analysis

The computational complexity mainly focuses on four unknown variables ($\mathbf{Z}^{(v)}$, $\mathcal{J}$, $\mathbf{E}$ and $\mathbf{F}$). For solving $\mathbf{Z}^{(v)}$, $\mathcal{J}$, $\mathbf{E}$ and $\mathbf{F}$, the complexities are $\mathcal{O}(mN^2d_v)$, $\mathcal{O}(mN^2\log(mN) + m^2N^2)$, $\mathcal{O}(mN^2)$ and $\mathcal{O}(N^3)$, respectively, where m is the number of views, N is the number of samples in each view. Considering number of iteration T and the fact $m \ll N$, the computational complexity of our proposed method is $\mathcal{O}(T(N^3 + mN^2d_v + mN^2\log(mN)))$.

VI. Experimental Results and Analysis

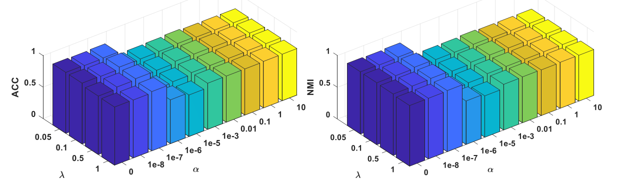
A. Database and Competitors

1) Database: We leverage the different clustering tasks to evaluate the performance of our method in the experiments. These different tasks involve the following five databases.

1) *Yale Database*[1]*:* It includes 15 persons with 165 gray images. Eleven images per person have different lighting, expressions, and occlusion changes. In the experiments, we leverage the way as in [18] to select three types of features as different views. They are LBP features with 3304 dimension, intensity features with 4096 dimension, and Gabor features with 6750 dimension.
2) *Caltech-101 Database [15]:* It includes 101 categories with 8677 images. Each class contains about 40–800 images. In the clustering experiments, the gallery includes 1474 images belonging to seven classes. These classes are Face, Garfield, Stop-sign, Motorbikes, Snoopy, Windsor-Chair, and Dolla-Bill. We extract three types of features as different views. They are HOG features with 620 dimension, sift features with 2560, and LBP features with 1160 dimension.
3) *Scene-15 Database [14]:* This database includes 15 scene categories with 4485 images. All images are derived from a wide range of indoor and outdoor environments, such as industrial, bedroom, kitchen, office, store, etc. In the experiments, we extract three types of image features via the way in [37], and consider them as different views.
4) *Notting-Hill (NH) Database [46]:* NH is derived from the movie "*Notting Hill*" and has 4660 faces of five main cats in 76 tracks. In the experiments, we construct the gallery by randomly selecting 110 images of each cast, and then leverage the way as in [18] to select three types of features as different views. They are Gabor, LBP, and intensity features.
5) *ORL Database*[2]*:* This gallery contains 40 distinct subjects. Each subject has ten images sampled under different time with varying facial expression and lighting.

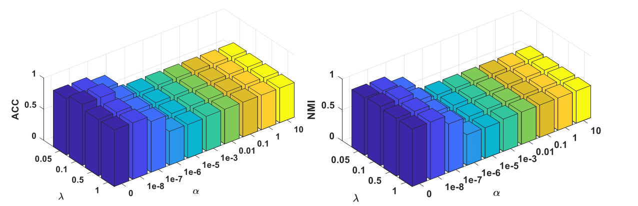
2) Competitors: To assess our method, we leverage six metrics to estimate the clustering performances on the aforementioned five databases. The six metrics are purity, accuracy (ACC), recall, normalized mutual information (NMI), adjusted rand index (AR), and F-score. In the following experiments,

[1]http://vision.ucsd.edu/content/yale-face-database
[2]http://www.uk.research.att.com/facedatabase.html



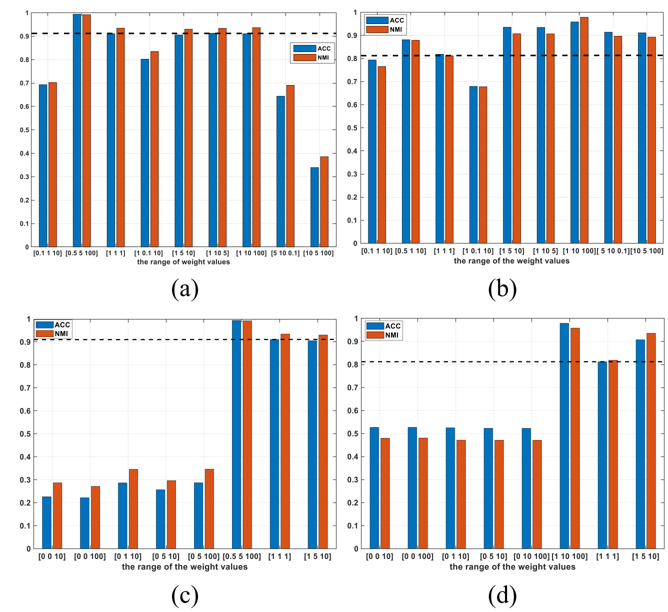
(a) (b)

Fig. 4. Results of our method versus α and λ on the Yale dataset ($p = 0.5$ and $\boldsymbol{\omega} = [0.5, 5, 100]$). (a) ACC. (b) NMI.



(a) (b)

Fig. 5. Results of our method versus α and λ on the Scene-15 dataset ($p = 1$ and $\boldsymbol{\omega} = [1, 10, 100]$). (a) ACC. (b) NMI.



(a) (b) (c) (d)

Fig. 6. Results of our method on the Yale and Scene-15 datasets, where $p = 0.5$, and λ and α are set to 0.1 and 10^{-7} on the Yale dataset. $p = 1$, and λ and α are set to 0.1 and 10^{-8} on the Scene-15 dataset, respectively. (a) Results on the Yale dataset. (b) Results on the Scene-15 dataset. (c) Results on the Yale dataset. (d) Results on the Scene-15 dataset.

we select spectral clustering, which is one of the representative single-view clustering methods, and six representative multiview clustering methods, such as MLAN [21], LT-MSC [43], CSMSC [18], ETLMSC [32], RMSC [34], and t-SVD-MSC [37]. For spectral clustering, we perform spectral clustering on all views, respectively, and list the best performance. This process is called SC in the following experiments. Moreover, we also show the performance of spectral clustering on the concatenated view features. This process is termed as feature in the following sections.

3) Parameter Setting and Analysis: In the model (22), parameter λ is used to balance the proportion of error $\mathbf{E}$, and α reflects the importance of the spectral clustering term. In the following experiments, we tune λ in the range of $[0.05, 0.1, 0.5, 1]$, α in the range of $[10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 0.001, 0.01, 0.1, 1, 10]$, $p \in (0, 1]$, and weights $\omega_i \in (0, 100]$ to obtain the best results. Specifically, p is set to 0.5, λ is set to 0.1, α is set to 10^{-7}, and weighted vector $\boldsymbol{\omega}$ is set to $[0.5, 5, 100]$ on the Yale dataset; p is set to 0.6, λ is set to 0.1, α is set to 10^{-8}, and weighted vector $\boldsymbol{\omega}$ is set to $[0.5, 1, 10]$ on the NH dataset; p is set to 0.9, λ is set to 0.05, α is set to 10^{-8}, and weighted vector $\boldsymbol{\omega}$ is set to $[5, 10, 100]$ on the Caltech-101 dataset; p is set to 0.9, λ is set to 0.5, α is set to 10^{-7}, and weighted vector $\boldsymbol{\omega}$ is set to $[5, 10, 100]$ on the ORL dataset; and p is set to 1.0, λ is set to 0.1, α is set to 10^{-8}, and weighted vector $\boldsymbol{\omega}$ is set to $[1, 10, 100]$ on the Scene-15 dataset. For all the compared methods, we follow the experiments settings in the corresponding papers.

Figs. 4 and 5 show the clustering performances (ACC and NMI) of our method versus λ and α on the Yale and Scene-15 datasets, respectively. It can be seen that when λ is fixed, performances of our method fluctuate remarkably with varying α, while our method fluctuates small with the fixed α. Our method obtains the best performance with $\alpha = 10^{-7}$ and $\lambda = 0.1$ on the Yale dataset and $\alpha = 10^{-8}$ and $\lambda = 0.1$ on the Scene-15 dataset, respectively. It indicates that α is important for improving clustering performance. When $\alpha = 0$, our method is inferior to the best performance with $\alpha = 10^{-7}$ on the Yale dataset and $\alpha = 10^{-8}$ on the Scene-15 database, but its performance is still good. When $\alpha \geq 10^{-6}$, the clustering performance of our method remarkably degrades. The reason may be that tensor low-rank constraint well-exploits high-order information and complementary information embedded in different views. Thus, the learned coefficient matrix well characterizes the relationship between data in itself. Spectral clustering is leveraged as a regularized term in our model, and it helps to further make the learned coefficient matrix exploit the cluster structure. Thus, we should assign a small value for α. Moreover, the value of spectral clustering term is much larger than other terms in our model (22), resulting in the unbalance penalty. So, in the following experiments, we set α as a small value such as 10^{-7} on the Yale and ORL databases and 10^{-8} on the other three databases.

Fig. 6 lists ACC and NMI of our method versus weighted vector $\boldsymbol{\omega}$ on the Yale and Scene-15 databases, respectively. From Fig. 6(a) and (b), we have that that our method has a large fluctuation with varying weighted vector. When other variables are fixed, our method obtains the best performance with weighted vector $\boldsymbol{\omega} = [0.5, 5, 100]$ on the dataset, and $\boldsymbol{\omega} = [1, 10, 100]$ on the Scene-15 dataset, respectively. It indicates that weights are important for clustering. The reason is due to the fact that weights reflect the importance of singular values. When weighted vector $\boldsymbol{\omega}$ is set $\boldsymbol{\omega} = [1, 1, 1]$, it means that all singular values are equally important, in this case, our method is remarkably inferior to the best performance with $\boldsymbol{\omega} = [0.5, 5, 100]$ on the Yale database and $\boldsymbol{\omega} = [1, 10, 100]$ on the Scene-15 database. The reason is that there is a significant difference between all singular values, and the larger singular values are generally associated with some salient parts (main information) in the data. Thus, we should shrink large singular values less by assigning small weights. From Fig. 6(c) and (d), we have that when the first weight or the first two weights

TABLE I
EXPERIMENTAL RESULTS ON THE YALE, NH, CALTECH-101, ORL, AND SCENE-15 DATASETS

Method	Yale Database					
	ACC	NMI	Purity	F-score	Recall	AR
SC	0.556±0.044	0.592±0.034	0.567±0.044	0.403±0.042	0.427±0.038	0.362±0.045
Feature	0.539±0.043	0.588±0.026	0.551±0.038	0.391±0.035	0.413±0.034	0.349±0.038
MLAN	0.498±0.003	0.567±0.003	0.521±0.004	0.312±0.003	0.438±0.016	0.253±0.004
RMSC	0.703±0.037	0.723±0.020	0.709±0.035	0.563±0.029	0.592±0.029	0.533±0.031
CSMSC	0.750±0.002	0.783±0.002	0.750±0.002	0.639±0.001	0.671±0.004	0.615±0.001
t-SVD-MSC	0.874±0.013	0.918±0.010	0.883±0.012	0.834±0.020	0.865±0.018	0.823±0.022
ETLMSC	0.659±0.043	0.697±0.038	0.659±0.043	0.533±0.049	0.550±0.048	0.501±0.053
LT-MSC	0.733±0.000	0.753±0.002	0.733±0.000	0.608±0.005	0.630±0.004	0.581±0.005
our work	**0.994±0.000**	**0.992±0.000**	**0.994±0.000**	**0.987±0.000**	**0.988±0.000**	**0.986±0.000**

Method	Notting-Hill Database					
	ACC	NMI	Purity	F-score	Recall	AR
SC	0.837±0.057	0.715±0.016	0.852±0.025	0.770±0.047	0.766±0.037	0.706±0.061
Feature	0.725±0.080	0.615±0.065	0.756±0.049	0.675±0.076	0.673±0.072	0.585±0.099
MLAN	0.720±0.064	0.707±0.058	0.768±0.064	0.720±0.059	0.835±0.012	0.626±0.085
RMSC	0.818±0.027	0.763±0.034	0.842±0.014	0.807±0.046	0.805±0.054	0.754±0.058
CSMSC	0.927±0.000	0.832±0.000	0.927±0.000	0.883±0.000	0.880±0.000	0.850±0.000
t-SVD-MSC	0.965±0.000	0.919±0.000	0.965±0.000	0.935±0.000	0.922±0.000	0.918±0.000
ETLMSC	0.942±0.000	0.902±0.000	0.942±0.000	0.905±0.000	0.887±0.000	0.879±0.000
LT-MSC	0.836±0.000	0.732±0.000	0.836±0.000	0.729±0.000	0.707±0.000	0.657±0.000
our work	**0.993±0.000**	**0.976±0.000**	**0.993±0.000**	**0.988±0.000**	**0.987±0.000**	**0.985±0.000**

Method	Caltech-101 database					
	ACC	NMI	Purity	F-score	Recall	AR
SC	0.545±0.030	0.441±0.019	0.624±0.022	0.498±0.017	0.540±0.020	0.390±0.024
Feature	0.558±0.004	0.438±0.009	0.624±0.012	0.484±0.007	0.511±0.009	0.376±0.012
MLAN	0.587±0.000	0.492±0.000	0.655±0.000	0.475±0.000	0.592±0.000	0.347±0.000
RMSC	0.529±0.029	0.371±0.012	0.565±0.015	0.432±0.024	0.461±0.052	0.313±0.022
CSMSC	0.567±0.000	0.480±0.000	0.633±0.000	0.495±0.000	0.498±0.000	0.395±0.000
t-SVD-MSC	0.556±0.000	0.512±0.000	0.649±0.000	0.498±0.000	0.509±0.000	0.396±0.000
ETLMSC	0.642±0.000	0.610±0.000	0.739±0.000	0.617±0.000	0.638±0.000	0.539±0.000
LT-MSC	0.528±0.008	0.458±0.002	0.642±0.002	0.457±0.001	0.442±0.001	0.355±0.001
our work	**0.837±0.000**	**0.893±0.000**	**0.914±0.000**	**0.869±0.000**	**0.841±0.000**	**0.845±0.000**

Method	ORL database					
	ACC	NMI	Purity	F-score	Recall	AR
SC	0.693±0.040	0.848±0.017	0.731±0.030	0.603±0.040	0.651±0.040	0.593±0.038
Feature	0.638±0.029	0.810±0.012	0.676±0.026	0.528±0.032	0.576±0.032	0.516±0.033
MLAN	0.559±0.010	0.785±0.004	0.648±0.008	0.277±0.021	0.657±0.011	0.250±0.023
RMSC	0.747±0.017	0.872±0.011	0.776±0.015	0.670±0.020	0.710±0.024	0.662±0.020
CSMSC	0.857±0.019	0.942±0.005	0.882±0.012	0.819±0.019	0.864±0.014	0.813±0.020
t-SVD-MSC	0.962±0.009	0.990±0.003	0.973±0.008	0.960±0.009	0.979±0.006	0.959±0.010
ETLMSC	0.958±0.024	0.931±0.005	0.970±0.016	0.960±0.021	0.984±0.010	0.959±0.021
LT-MSC	0.805±0.002	0.921±0.006	0.845±0.008	0.757±0.016	0.817±0.016	0.751±0.016
our work	**0.990±0.000**	**0.998±0.000**	**0.993±0.000**	**0.990±0.000**	**0.996±0.000**	**0.990±0.000**

Method	Scene-15 database					
	ACC	NMI	Purity	F-score	Recall	AR
SC	0.680±0.040	0.456±0.012	0.534±0.020	0.374±0.015	0.372±0.014	0.328±0.017
Feature	0.312±0.016	0.288+±0.006	0.351±0.016	0.214±0.005	0.217±0.004	0.155±0.005
MLAN	0.340±0.031	0.486±0.029	0.3505±0.034	0.262±0.033	0.740±0.043	0.167±0.041
RMSC	0.519±0.000	0.488±0.000	0.559±0.000	0.402±0.000	0.399±0.000	0.358±0.000
CSMSC	0.509±0.000	0.564±0.000	0.611±0.000	0.433±0.000	0.489±0.000	0.385±0.000
t-SVD-MSC	0.892±0.000	0.919±0.000	0.922±0.000	0.883±0.000	0.892±0.000	0.874±0.000
ETLMSC	0.871±0.000	0.891±0.000	0.906±0.000	0.853±0.000	0.863±0.000	0.842±0.000
LT-MSC	0.714±0.002	0.669±0.003	0.759±0.000	0.608±0.002	0.622±0.001	0.586±0.001
our work	**0.979±0.000**	**0.958±0.000**	**0.979±0.000**	**0.960±0.000**	**0.959±0.000**	**0.958±0.000**

are 0, it means that we did not shrink the largest singular value or the first two largest singular values. In this case, our method degrades remarkably and is obviously inferior to the best performance with $\boldsymbol{\omega} = [0.5, 5, 100]$ on the Yale database and $\boldsymbol{\omega} = [1, 10, 100]$ on the Scene-15 database. The reason may be that the larger singular values may carry undesirable information, while we do not shrink them. Thus, the learned coefficient matrix cannot characterize the cluster structure of data.

B. Experimental Results and Analysis

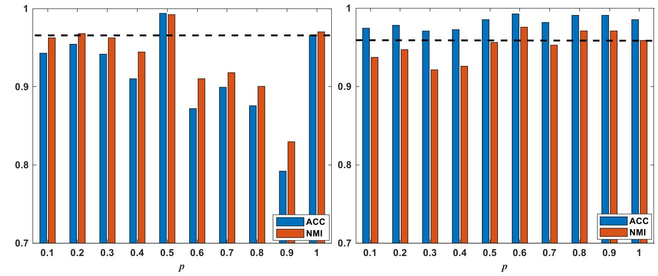
To well estimate the performance of our method for clustering, we list the experimental results of our method with six metrics, such as ACC, purity, recall, NMI, F-score, and AR in the aforementioned five databases. For each experiment, we repeat ten times and show the mean and corresponding standard deviation in Table I. Table I lists the results of all the eight algorithms on the five datasets. From Table I, we have the following interesting observation.

1) T-SVD-based tensor low-rank methods (t-SVD-MSC and our method) are remarkably superior to the classical tensor low-rank method LT-MSC. The reason may be that LT-MSC is based on the Tucker tensor decomposition, which is not a tight convex relaxation of the Tucker rank, while t-SVD-based tensor decomposition is an effective convex relaxation of ℓ_1-norm. Thus, the coefficient matrix, which is learned by our method and t-SVD-MSC, well characterizes the complementary information and high-order information embedded in multiview data.
2) Except for LT-MSC, tensor low-rank methods are superior to the other multiview clustering methods. This is probably because that tensor low-rank methods directly take into account the high-order correlation

embedded in multiview data. Moreover, the complementary information among different views can be explored more efficiently and thoroughly by the tensor low-rank methods.

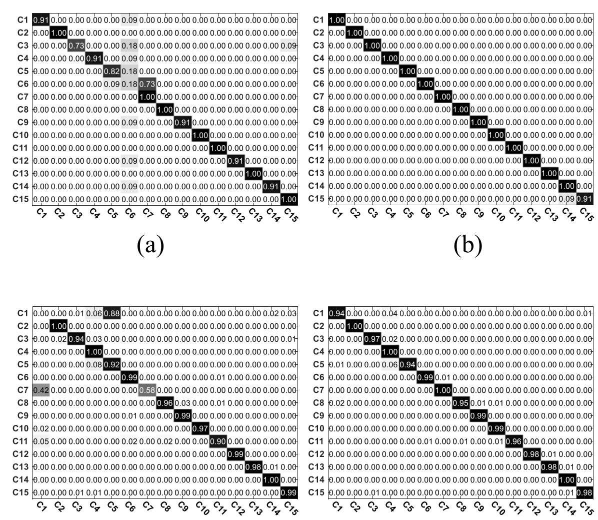
3) Our method is remarkably superior to the other seven methods on the five databases. For example, on the Yale dataset, our method indicates a significant increase of 12.0%, 7.4%, 11.1%, 15.3%, 12.3%, and 16.3% w.r.t. ACC, NMI, purity, F-score, recall, and AR, respectively, compared to the second best method t-SVD-MSC. On the Scene-15 dataset with 4485 images in three views, our method shows 8.7%, 3.9%, 5.7%, 7.7%, 6.7%, and 8.4% of relative improvement w.r.t. ACC, NMI, purity, F-score, recall, and AR over the second best method t-SVD-MSC. The reason may be that our method explicitly considers the contribution of each singular value, that is, the prior knowledge of matrix in solving the nuclear norm minimization problem. Moreover, our method integrates coefficient matrix learning and spectral clustering into a unified framework. Thus, the learned coefficient matrix well characterizes the cluster structure.
4) Single-view clustering methods are overall inferior to multiview clustering methods. The reason may be that multiview methods may leverage the complementary information embedded in multiview data, while single-view methods do not. The multiview method MLAN is overall inferior to best SC in all single-view data. This is probably due to the fact that multiview data are composed of heterogeneous features, but MLAN assumes that all-views data share a coefficient matrix, resulting in over fitting. Moreover, each view generally has different clustering performance, while MLAN does not take into account this in the learning coefficient matrix. The performances of SC on the concatenated multiview features are overall inferior to the other methods. The reason may be that heterogeneity in concatenated multiview features may cause scale issue, and each view has different role for improving clustering performance. Moreover, SC cannot well characterize the cluster structure due to the fact that the concatenated multiview features contain redundancy.
5) The samples in the Yale dataset include illumination changes, occlusion (such as sunglasses), etc. Obviously, the proposed method, respectively, improves by nearly 12.0% and 15.3% over the second best t-SVD-MSC on ACC and F-score on the Yale dataset. All these results clearly prove the superior effectiveness and robustness of our proposed method to illumination and occlusion.

As shown in Fig. 7, we analyze the impact of the power p in WTSNM on Yale and Scene-15 datasets. One can observe that the proposed method has the different clustering results (ACC and NMI) under the different power p, and when $p = 0.5$ and $p = 0.6$, we obtain the best clustering results in the Yale dataset and NH dataset, respectively. Meanwhile, we find that the power p has a great influence on the clustering performance. This is because we perform the power processing strategy on different singular values. By this strategy, we can make the proposed WTSNM preserve useful information in the multiview data, which in turn makes the proposed WTSNM more flexible and robust to noise information.



(a) (b)

Fig. 7. Results on Yale and NH datasets, where $\boldsymbol{\omega} = [0.5; 5; 100]$, and λ and α are set to 0.5 and 10^{-7} on Yale dataset, respectively. $\boldsymbol{\omega} = [0.5; 1; 10]$, and λ and α are set to 0.1 and 10^{-8} on NH dataset, respectively. (a) Results on Yale dataset. (b) Results on NH dataset.



(a) (b)

(c) (d)

Fig. 8. (a) Confusion matrices of t-SVD-MSC on Yale dataset. (b) Confusion matrices of our work on Yale dataset. (c) Confusion matrices of t-SVD-MSC on Scene-15 dataset. (d) Confusion matrices of our work on Scene-15 dataset.

To further evaluate the advantage of our method, we visualize the confusion matrices in Fig. 8, which are obtained by t-SVD-MSC and our method. In Fig. 8, the row and the column are true and predicted labels, respectively. Herein, the predicted cluster label calculates by performing the permutation mapping function in ACC. We can see that compared with t-SVD-MSC, our method wins in almost all categories in terms of clustering ACC. The reason may be that the learned representation well encodes the cluster structure in our method.

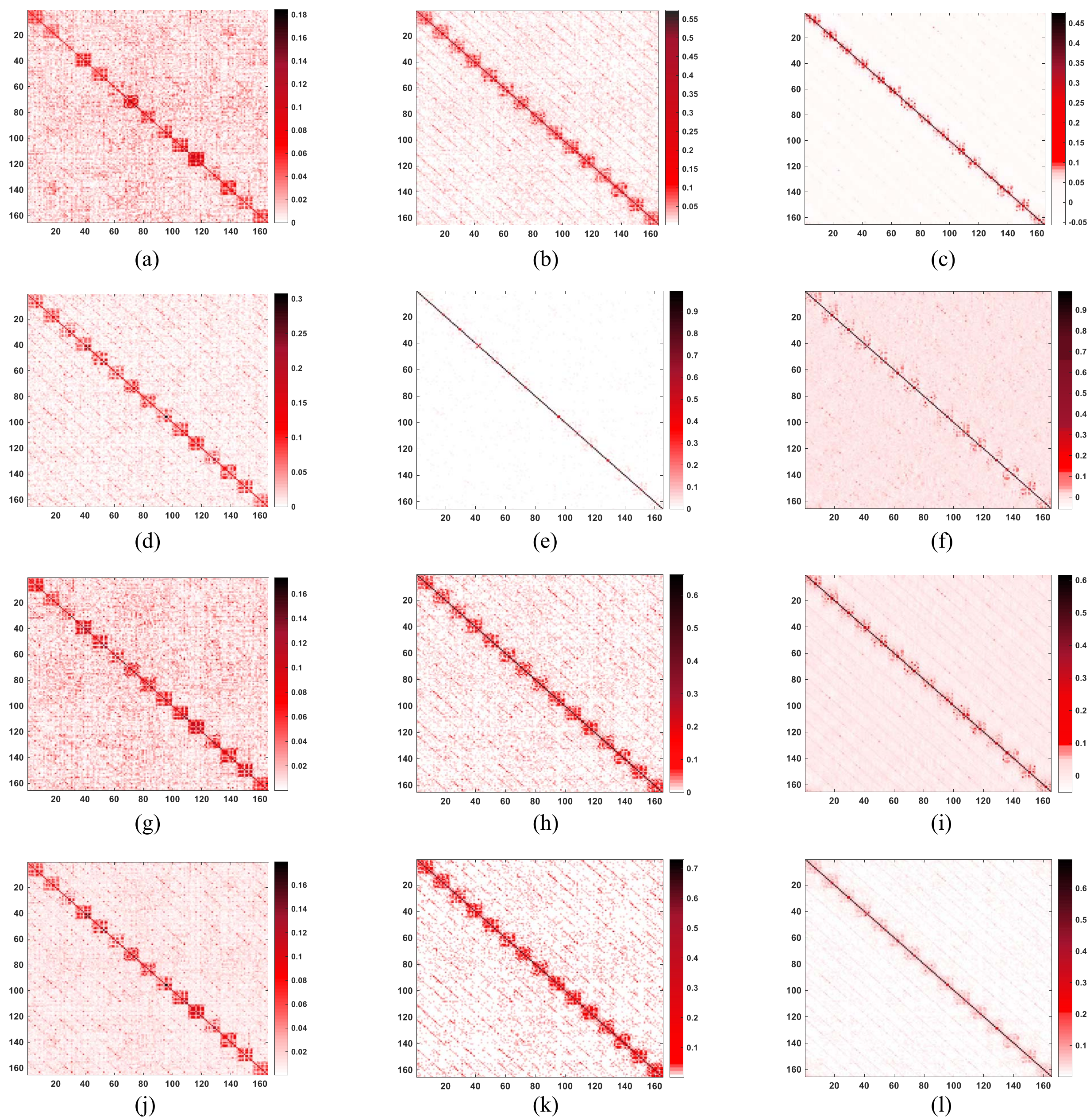


Fig. 9. (a)–(i) Comparison between LRR and t-SVD-MSC with each view and our model in terms of affinity matrix on the Yale dataset. (j)–(l) Final affinity matrix $\mathbf{S} = (1/m)\sum_{v=1}^{m} \left|\mathbf{Z}^{(v)}\right| + \left|\mathbf{Z}^{(v)^T}\right|$.

C. Contributes of Multiview Feature

We analyze the changes of the affinity matrix for all the views before and after the proposed optimization procedure. Figs. 9 and 10 present the view-specific affinity matrices, which are obtained by LRR on the corresponding view data, and the final affinity matrix on the Yale and Scene-15 datasets, respectively. We also show the view-specific affinity matrices and the final affinity matrix of our method on the Yale and Scene-15 datasets, respectively. Obviously, the affinity matrices of all the views, which are learned by our model, have the apparent block-diagonal structures, compared with the corresponding affinity matrix learned by LRR. This is an evidence that the complementary information and high-order information are important and can be propagated among all the views.

Figs. 9 and 10 also present the view-specific affinity matrices and the final affinity matrix of t-SVD-MSC and our method on the Yale and Scene-15 datasets, respectively. It can be seen that both t-SVD-MSC and our method have apparent block-diagonal structures for affinity matrices, but compared with t-SVD-MSC, the nonblock-diagonal elements in affinity matrices, which are learned by our method, are overall smaller in specific view and final affinity matrix. It indicates that our method well characterizes the cluster structure. Moreover, the block-diagonal structure in affinity matrices, which correspond to different views, is different. It means that each view has different roles for improving the clustering performance.

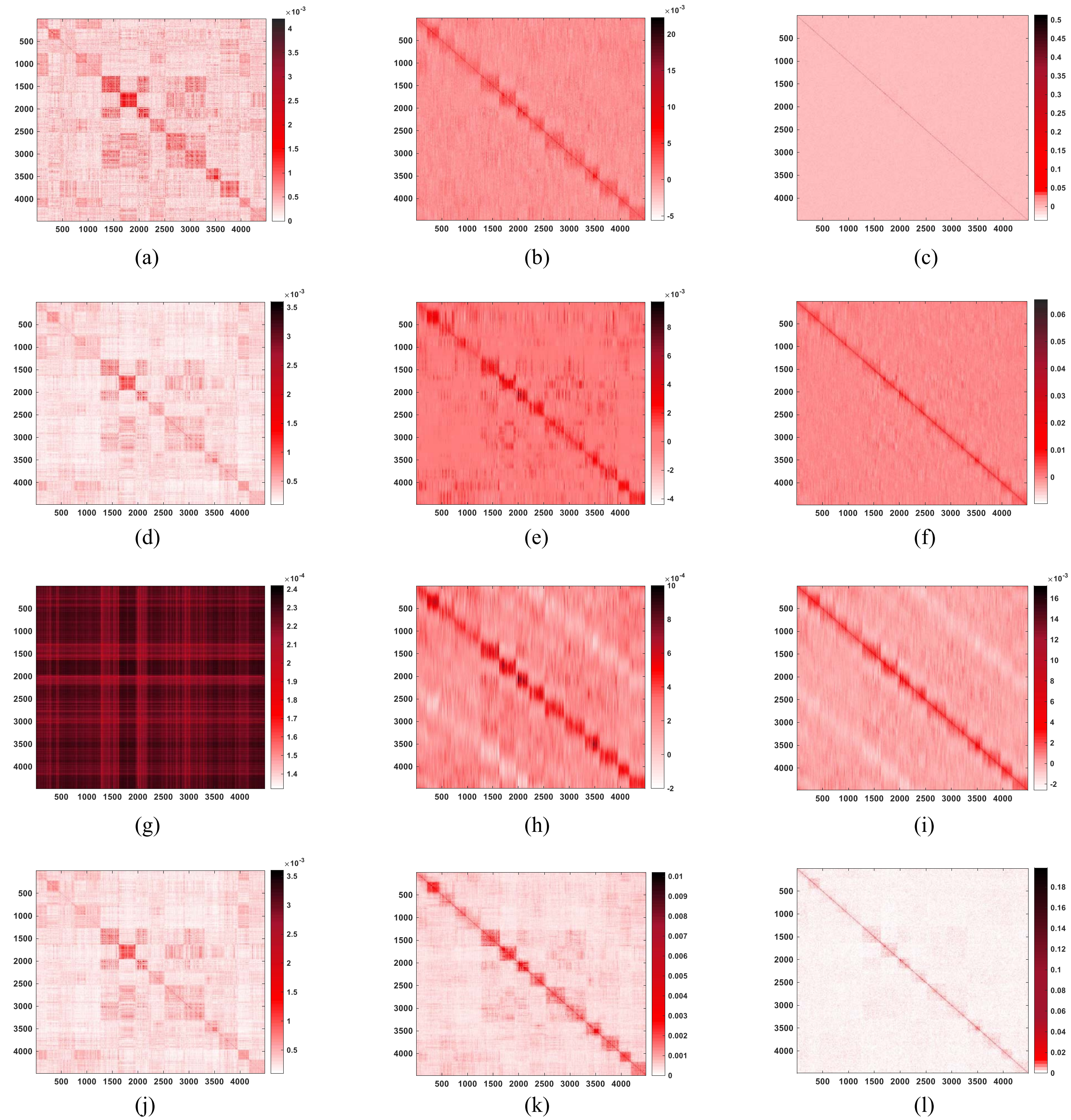


Fig. 10. (a)–(i) Comparison between LRR with each view and our model in terms of affinity matrix on the Scene-15 dataset. (j)–(l) Final affinity matrix $\mathbf{S} = (1/m)\sum_{v=1}^{m} \left|\mathbf{Z}^{(v)}\right| + \left|\mathbf{Z}^{(v)^T}\right|$.

VII. Conclusion

We studied the WTSNM based on t-SVD, and proposed an efficient iterative algorithm to solve it. As can be seen, the existing tensor nuclear norm based on t-SVD can be viewed as a special case of our method, and our WTSNM can also be applied to the standard matrix nuclear norm minimization. Applying WTSNM to MVSC, we developed a novel MVSC model, which obtains the self-representations and cluster indicator matrix simultaneously by well exploiting the high-order correlation embedded in multiview data. The extensive experimental results on five widely used benchmarks indicate that our method is superior to state-of-the-art multiview clustering methods. In our proposed method, for the sake of simplicity, we assigned equal weights for all frontal slices. In real applications, we should assign different weights to different frontal slices. We will study it our future work.

Acknowledgment

The authors would like to thank C. Zhang for his opening code of LT-MSC and Y. Xie for the code of t-SVD-MSC. The authors would like to thank the anonymous reviewers and AE for their constructive comments and suggestions.

References

[1] S. Aeron and E. Kernfeld, "Group-invariant subspace clustering," in *Proc. Allerton Conf. Commun. Control Comput.*, 2015, pp. 666–671.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[3] D. Cai and X. Chen, "Large scale spectral clustering via landmark-based sparse representation," *IEEE Trans. Cybern.*, vol. 45, no. 8, pp. 1669–1680, Aug. 2015.

[4] X. Cao, X. Wei, Y. Han, and D. Lin, "Robust face clustering via tensor decomposition," *IEEE Trans. Cybern.*, vol. 45, no. 11, pp. 2546–2557, Nov. 2015.

[5] X. Cao, C. Zhang, H. Fu, H. Zhang, and S. Liu, "Diversity-induced multi-view subspace clustering," in *Proc. IEEE CVPR*, 2015, pp. 586–594.

[6] M. Cheng, L. Jing, and M. K. Ng, "Tensor-based low-dimensional representation learning for multi-view clustering," *IEEE Trans. Image Process.*, vol. 28, no. 5, pp. 2399–2414, May 2019.

[7] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *Proc. IEEE ICDM*, 2014, pp. 110–119.

[8] J. Eckstein and D. P. Bertsekas, "On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Math. Program.*, vol. 55, no. 1, pp. 293–318, 1992.

[9] E. Ehsan and V. Ren, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

[10] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," in *Proc. IEEE ICCV*, 2015, pp. 4238–4246.

[11] Q. Gao, W. Xia, Z. Wan, D. Xie, and P. Zhang, "Tensor-SVD based graph learning for multi-view subspace clustering," in *Proc. AAAI*, 2020, pp. 3930–3937.

[12] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[13] W. Hu, D. Tao, W. Zhang, Y. Xie, and Y. Yang, "The twist tensor nuclear norm for video completion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 2961–2973, Dec. 2017.

[14] M. E. Kilmer, K. Braman, N. Hao, and R. C. Hoover, "Third-order tensors as operators on matrices: A theoretical and computational framework with applications in imaging," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 1, pp. 148–172, 2013.

[15] F. Li, R. Fergus, and P. Perona, "Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories," in *Proc. IEEE CVPR Workshops*, 2004, p. 178.

[16] L. Liu, F. Nie, A. Wiliem, Z. Li, T. Zhang, and B. C. Lovell, "Multi-modal joint clustering with application for unsupervised attribute discovery," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4345–4356, Sep. 2018.

[17] C. Lu, J. Feng, Y. Chen, W. Liu, Z. Lin, and S. Yan, "Tensor robust principal component analysis: Exact recovery of corrupted low-rank tensors via convex optimization," in *Proc. IEEE CVPR*, 2016, pp. 5249–5257.

[18] S. Luo, C. Zhang, W. Zhang, and X. Cao, "Consistent and specific multi-view subspace clustering," in *Proc. AAAI*, 2018, pp. 3730–3737.

[19] B. Madathil and S. N. George, "Twist tensor total variation regularized-reweighted nuclear norm based tensor completion for video missing area recovery," *Inf. Sci.*, vol. 423, pp. 376–397, Jan. 2018.

[20] L. Mirsky, "A trace inequality of John von Neumann," *Monatshefte Mathematik*, vol. 79, no. 4, pp. 303–306, 1975.

[21] F. Nie, G. Cai, and X. Li, "Multi-view clustering and semi-supervised classification with adaptive neighbours," in *Proc. AAAI*, 2017, pp. 2408–2414.

[22] F. Nie, J. Li, and X. Li, "Self-weighted multiview clustering with multiple graphs," in *Proc. IJCAI*, 2017, pp. 2564–2570.

[23] F. Nie, L. Tian, and X. Li, "Multiview clustering via adaptively weighted procrustes," in *Proc. ACM SIGKDD*, 2018, pp. 2022–2030.

[24] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.

[25] H. Wang, F. Nie, and H. Huang, "Multi-view clustering and feature learning via structured sparsity," in *Proc. ICML*, 2013, pp. 352–360.

[26] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *Proc. IEEE CVPR*, 2017, pp. 1–9.

[27] Y. Wang, X. Lin, L. Wu, W. Zhang, Q. Zhang, and X. Huang, "Robust subspace clustering for multi-view data by exploiting correlation consensus," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3939–3949, Nov. 2015.

[28] Y. Wang, L. Wu, X. Lin, and J. Gao, "Multiview spectral clustering via structured low-rank matrix factorization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4833–4843, Oct. 2018.

[29] Y. Wang, W. Zhang, L. Wu, X. Lin, M. Fang, and S. Pan, "Iterative views agreement: An iterative low-rank based structured optimization method to multi-view spectral clustering," in *Proc. IJCAI*, 2016, pp. 2153–2159.

[30] J. Wen, Y. Xu, and H. Liu, "Incomplete multiview spectral clustering with adaptive graph learning," *IEEE Trans. Cybern.*, vol. 50, no. 4, pp. 1418–1429, Apr. 2020.

[31] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[32] J. Wu, Z. Lin, and H. Zha, "Essential tensor learning for multi-view spectral clustering," *IEEE Trans. Image Process.*, vol. 28, no. 12, pp. 5910–5922, Dec. 2019.

[33] T. Wu and W. U. Bajwa, "A low tensor-rank representation approach for clustering of imaging data," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1196–1200, Aug. 2018.

[34] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. AAAI*, 2014, pp. 2149–2155.

[35] D. Xie, X. Zhang, Q. Gao, J. Han, X. Gao, and S. Xiao, "Multi-view clustering by joint latent representation and similarity learning," *IEEE Trans. Cybern.*, vol. 50, no. 11, pp. 4848–4854, Nov. 2020, doi: 10.1109/TCYB.2019.2922042.

[36] Y. Xie, S. Gu, Y. Liu, W. Zuo, W. Zhang, and L. Zhang, "Weighted Schatten p-norm minimization for image denoising and background subtraction," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4842–4857, Oct. 2016.

[37] Y. Xie, D. Tao, W. Zhang, Y. Liu, L. Zhang, and Y. Qu, "On unifying multi-view self-representations for clustering by tensor multi-rank minimization," *Int. J. Comput. Vis.*, vol. 126, no. 11, pp. 1157–1179, 2018.

[38] Y. Xie, W. Zhang, Y. Qu, L. Dai, and D. Tao, "Hyper-laplacian regularized multilinear multiview self-representations for clustering and semisupervised learning," *IEEE Trans. Cybern.*, vol. 50, no. 2, pp. 572–586, Feb. 2020.

[39] Y. Yang and H. Wang, "Multiview clustering: A survey," *Bid Data Min. Anal.*, vol. 1, no. 2, pp. 83–107, 2018.

[40] M. Yin, J. Gao, S. Xie, and Y. Guo, "Multiview subspace clustering via tensorial t-product representation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 30, no. 6, pp. 851–964, Mar. 2019.

[41] Q. Yin, W. Shu, H. Ran, and W. Liang, "Multi-view clustering via pairwise sparse subspace representation," *Neurocomputing*, vol. 156, pp. 12–21, May 2015.

[42] C. Zhang *et al.*, "Generalized latent multi-view subspace clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 1, pp. 86–99, Jan. 2020.

[43] C. Zhang, H. Fu, S. Liu, G. Liu, and X. Cao, "Low-rank tensor constrained multiview subspace clustering," in *Proc. IEEE ICCV*, 2015, pp. 1582–1590.

[44] L. Zhang and Z. Peng, "Infrared small target detection based on partial sum of the tensor nuclear norm," *Remote Sens.*, vol. 11, no. 4, pp. 382–392, 2019.

[45] Y. Zhang, W. Yang, B. Liu, G. Ke, Y. Pan, and J. Yin, "Multi-view spectral clustering via tensor-SVD decomposition," in *Proc. IEEE 29th Int. Conf. Tools Artif. Intell. (ICTAI)*, 2017, pp. 493–497.

[46] Y.-F. Zhang, C. Xu, H. Lu, and Y.-M. Huang, "Character identification in feature-length films using global face-name matching," *IEEE Trans. Multimedia*, vol. 11, no. 7, pp. 1276–1288, Nov. 2009.

[47] Z. Zhang, G. Ely, S. Aeron, N. Hao, and M. Kilmer, "Novel methods for multilinear data completion and de-noising based on tensor-SVD," in *Proc. IEEE CVPR*, 2014, pp. 3842–3849.

[48] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. Zhang, "A survey of sparse representation: Algorithms and applications," *IEEE Access*, vol. 3, pp. 490–530, 2015.

Wei Xia (Graduate Student Member, IEEE) received the B.Eng. degree in communication engineering from the Lanzhou University of Technology, Lanzhou, China, in 2018. He is currently pursuing the Ph.D. degree in communication and information system with Xidian University, Xi'an, China.

His research interests include pattern recognition, machine learning, and deep learning.

Xiangdong Zhang received the B.Eng., M.S., and Ph.D. degrees from Xidian University, Xi'an, China, in 1992, 1995, and 1998, respectively.

He is currently an Associate Professor with the School of Telecommunications Engineering, Xidian University. His current research interests include pattern recognition and machine learning.

Quanxue Gao received the B.Eng. degree from Xi'an Highway University, Xi'an, China, in 1998, the M.S. degree from the Gansu University of Technology, Lanzhou, China, in 2001, and the Ph.D. degree from Northwestern Polytechnical University, Xi'an, in 2005.

He was an Associate Research with the Biometrics Center, Hong Kong Polytechnic University, Hong Kong, from 2006 to 2007. From 2015 to 2016, he was a Visiting Scholar with the Department of Computer Science, University of Texas at Arlington, Arlington, TX, USA. He is currently a Professor with the School of Telecommunications Engineering, Xidian University, Xi'an, and also a Key Member of State Key Laboratory of Integrated Services Networks. He has authored around 80 technical articles in refereed journals and proceedings, including IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, IEEE TRANSACTIONS ON IMAGE PROCESSING, IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS, IEEE TRANSACTIONS ON CYBERNETICS, CVPR, AAAI, and IJCAI. His current research interests include pattern recognition and machine learning.

Xiaochuang Shu received the B.E. degree in communication engineering from Zhengzhou University, Zhengzhou, China, in 2020. She is currently pursuing the master's degree in information and communication engineering with Xidian University, Xi'an, China, under the supervision of Prof. Q. Gao.

Her research interests include multiview learning and machine learning.

Jungong Han received the B.Eng., and Ph.D. degrees from Xidian University, Xi'an China, in 1999, and 2004, respectively.

He is currently a Full Professor and a Chair of Computer Science with Aberystwyth University, Aberystwyth, U.K. He has published over 180 papers, including 50+ IEEE Trans and 40+ A* conference papers. His research interests span the fields of video analysis, computer vision, and applied machine learning.

Xinbo Gao (Senior Member, IEEE) received the B.Eng., M.Sc., and Ph.D. degrees in electronic engineering, signal and information processing from Xidian University, Xi'an, China, in 1994, 1997, and 1999, respectively.

From 1997 to 1998, he was a Research Fellow with the Department of Computer Science, Shizuoka University, Shizuoka, Japan. From 2000 to 2001, he was a Postdoctoral Research Fellow with the Department of Information Engineering, Chinese University of Hong Kong, Hong Kong. Since 2001, he has been with the School of Electronic Engineering, Xidian University, where he is currently a Cheung Kong Professor of Ministry of Education, and a Professor of Pattern Recognition and Intelligent System, and a Professor of Computer Science and Technology with the Chongqing University of Posts and Telecommunications, Chongqing, China. He has published six books and around 300 technical articles in refereed journals and proceedings. His current research interests include Image processing, computer vision, multimedia analysis, machine learning, and pattern recognition.

Prof. Gao is on the editorial boards of several journals, including *Signal Processing* (Elsevier) and *Neurocomputing* (Elsevier). He served as the General Chair/Co-Chair, the Program Committee Chair/Co-Chair, or a PC Member for around 30 major international conferences. He is a Fellow of the Institute of Engineering and Technology and the Chinese Institute of Electronics.