Exclusivity-Consistency Regularized Multi-view Subspace Clustering

Xiaobo Wang[1,2] Xiaojie Guo[2,3] Zhen Lei[1,2*] Changqing Zhang[4] Stan Z. Li[1,2]
[1]Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences
[2] University of Chinese Academy of Sciences
[3] State Key Laboratory of Information Security, IIE, Chinese Academy of Sciences
[4] School of Computer Science and Technology, Tianjin University
{xiaobo.wang, zlei, szli}@nlpr.ia.ac.cn xj.max.guo@gmail.com zhangchangqing@tju.edu.cn

Abstract

Multi-view subspace clustering aims to partition a set of multi-source data into their underlying groups. To boost the performance of multi-view clustering, numerous subspace learning algorithms have been developed in recent years, but with rare exploitation of the representation complementarity between different views as well as the indicator consistency among the representations, let alone considering them simultaneously. In this paper, we propose a novel multi-view subspace clustering model that attempts to harness the complementary information between different representations by introducing a novel position-aware exclusivity term. Meanwhile, a consistency term is employed to make these complementary representations to further have a common indicator. We formulate the above concerns into a unified optimization framework. Experimental results on several benchmark datasets are conducted to reveal the effectiveness of our algorithm over other state-of-the-arts.

1. Introduction

Clustering data points into different groups such that the objects in the same group are highly similar to each other, is one of the most fundamental topics in computer vision and pattern recognition [1, 30, 17, 22, 23]. In the past decades, a number of clustering approaches have been developed, such as the iteration based methods [28, 13], the factorization based methods [6, 14], and the spectral clustering based approaches [24, 8, 20]. Among them, spectral clustering based ones have become popular and dominant. For example, Standard Spectral Clustering (SPC) proposed in [24] aims to perform clustering by learning the similarity matrix based on the data locality. Sparse Subspace Clustering (SSC) [8] is to seek a sparse representation for each instance over the whole data. Having the representation calculated, the spectral clustering algorithm (*e.g.*, Normalized Cuts [27]) is performed to obtain the clustering result. The Low-Rank Representation (LRR) approach proposed in [20] tries to find a low-rank representation. Additionally, the method proposed in [18], called Structured Sparse Subspace Clustering (S3C), achieves promising results by integrating the sparse representation learning and the spectral clustering into one framework. However, these methods mainly focus on advancing the clustering performance for single source features. For multi-view ones, they are difficult to find good clusters due to the potential presence of view insufficiency or the high-dimensionality of data. Typically, they cannot be directly applied to multi-view cases. This paper devotes to boost the clustering performance by recovering the subspace structure of the data set with multi-view features.

In practice, we often face data in multiple views. Different views characterize different and partly independent information about the data. For instance, images and videos are described by different kinds of features, such as color, texture and edge. Web pages contain texts, hyperlinks and possibly existing visual information. In general, these multi-view representations can seamlessly capture the rich information from multiple data cues as well as the complementary information among different cues, thus will be beneficial to clustering. To integrate different features, much progress has been made in developing effective multi-view clustering methods [7, 15, 5, 2, 4, 9, 31]. The work [7] utilizes a bipartite similarity matrix to connect two types of features and adopts the standard spectral clustering to generate the final result. The co-regularized multi-view spectral clustering introduced in [15] is to perform clustering on different views simultaneously with a co-regularization constraint. The method designed in [5] learns a common representation under the spectral clustering framework by combining Laplacians of different views. By considering the complementary information between representations, the

*Corresponding Author

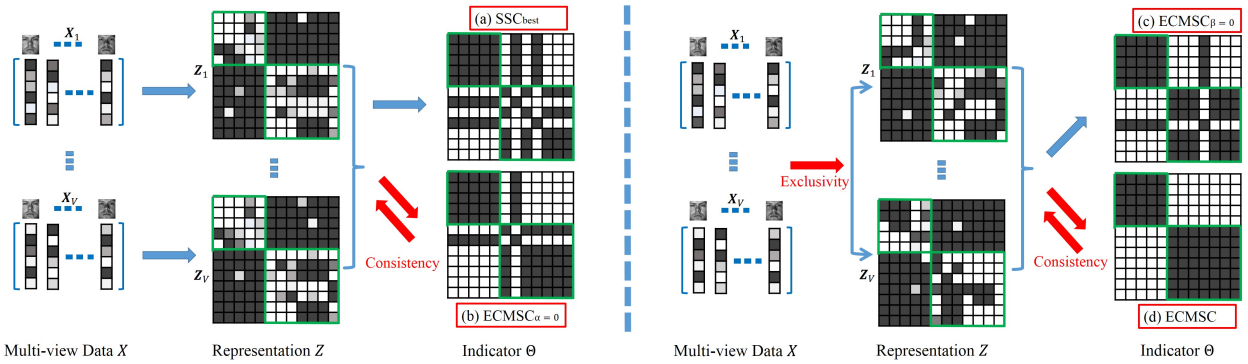


Figure 1. Different viewpoints of our proposed ECMSC model. The green rectangle indicates the ground-truth clustering while the red rectangle denotes the different version of the ECMSC. With the multi-view input, our ECMSC algorithm simultaneously exploits the representation exclusivity (parameter α) and indicator consistency (parameter β) in one framework.

work [2] employs the HSIC criterion [10] to pursue the diverse representations. Gao et al. [9] unified the representation learning and spectral clustering into one framework. Zhang *et al.* [31] proposed to cluster multi-view features with a low-rank tensor constraint for capturing the high-order cross information among multiple views.

Although the above multi-view clustering approaches generally provide more promising results than most single view methods like [8, 20], they have two main shortcomings: 1) Many works [4, 5, 21] prefer to learn a common representation, ignoring the complementary information between different views; 2) Existing works [2, 31, 5, 29] tend to execute the subspace learning and spectral clustering in two separated steps without consideration of the fact that these two pipelined steps highly depend on each other.

To overcome the above shortcomings, this paper proposes a novel multi-view clustering algorithm namely ECMSC, to simultaneously exploit the ***representation exclusivity*** and ***indicator consistency*** in a unified manner. Specifically, to exploit the complementary information between different representations, we introduce a novel *position-aware* exclusivity term to enforce the representations of different views to be as exclusive as possible. Meanwhile, an indicator consistent term is employed to advocate the label consistency among these complementary representations. Therefore, we have integrated the two pipelined steps, *i.e.* subspace learning and spectral clustering, into one optimization framework. Figure 1 gives an illustration of our proposed ECMSC. Moreover, to efficiently and effectively seek the solution of the associated optimization problem, an alternative based algorithm is designed. Extensive experiments on benchmark datasets are conducted to demonstrate the superiority of our method over state-of-the-art alternatives.

Notation: Throughout the paper, all the matrices are written as uppercase. For a matrix $\mathbf{G}$, the i-th row, the j-th column and the ij-th element of $\mathbf{G}$ are denoted by $\mathbf{g}^i$, $\mathbf{g}_j$ and g_{ij}, respectively. $\mathbf{0}$, $\mathbf{1}$ and $\mathbf{I}$ represent the all-zeros, all-ones and identity matrices with appropriate sizes, respectively.

2. Problem Statement

This section first briefly introduces the general procedure of subspace clustering based approaches, then provides the formulation of our multi-view subspace clustering method[1].

2.1. Subspace Clustering

For a data set, it usually lies in an underlying low-dimensional subspace rather than distributing uniformly in the entire space [18, 9]. Thus, the data points can be represented by a low-dimensional subspace. After obtaining the subspace structure of the data set, the clustering is accomplished based on the recovered subspace instead of on the entire space.

Suppose $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n] \in \mathbb{R}^{d\times n}$ is the matrix of data vectors, each column of which is a sample vector, d is the dimensionality of the feature space and n is the total amount of data points. The subspace clustering tries to find the self-representation by solving the following optimization problem:

$$\min_{\mathbf{Z},\mathbf{E}} ||\mathbf{E}||_k + \lambda||\mathbf{Z}||_l \ \text{s.t.}\ \mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E},\ \text{diag}(\mathbf{Z}) = \mathbf{0}, \tag{1}$$

where $\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}$ is the self-representation model. $\mathbf{Z} = [\mathbf{z}_1, \ldots, \mathbf{z}_n] \in \mathbb{R}^{n\times n}$ is the subspace representation matrix, each $\mathbf{z}_i$ is the coding coefficient of the original data point $\mathbf{x}_i$ over the observed $\mathbf{X}$, while $\mathbf{E}$ is the error matrix. $||\cdot||_k$ and $||\cdot||_l$ are two properly chosen norms, λ is the trade-off parameter, and the constraint $\text{diag}(\mathbf{Z}) = \mathbf{0}$ is optionally used to rule out the trivial solution of $\mathbf{Z}$ being an identity.

After obtaining the subspace structure $\mathbf{Z}$ by solving (1), the similarity matrix is formed via $\mathbf{S} = (|\mathbf{Z}|^T + |\mathbf{Z}|)/2$. Then, we can perform spectral clustering on such a subspace similarity matrix through optimizing the following:

$$\min_{\mathbf{F}} \text{tr}(\mathbf{F}^T(\mathbf{D} - \mathbf{S})\mathbf{F}) \ \text{s.t.}\ \ \mathbf{F} \in \mathcal{C}, \tag{2}$$

[1]Our code is released at `http://www.cbsr.ia.ac.cn/users/xiaobowang/`.

where $\mathbf{D}$ is a diagonal matrix whose diagonal elements are defined as $d_{ii} = \sum_j s_{ij}$. The constraint $\mathcal{C}$ can be detailed as $\{\mathbf{F} \in \{0,1\}^{n\times k} : \mathbf{F1} = \mathbf{1}, rank(\mathbf{F}) = k.\}$. Specifically, $\mathbf{F} = [\mathbf{q}_1, \ldots, \mathbf{q}_k] \in \mathbb{R}^{n\times k}$ is a binary matrix indicating the membership of each data point to a specific group. $\mathbf{F1} = \mathbf{1}$ means that each data point appears in one and only one subspace. The constraint $rank(\mathbf{F}) = k$ is to ensure that the number of subspaces is equal to the desired k. For simplicity, the above problem (2) can be solved approximately via spectral clustering by relaxing the constraint $\mathbf{F} \in \mathcal{C}$ to $\mathbf{F}^T\mathbf{F} = \mathbf{I}$. Furthermore, note that the objective $\mathrm{tr}(\mathbf{F}^T(\mathbf{D} - \mathbf{S})\mathbf{F})$ can be rewritten as:

$$\begin{aligned}\mathrm{tr}(\mathbf{F}^T(\mathbf{D} - \mathbf{S})\mathbf{F}) &= \sum_{i,j} \frac{1}{2} s_{ij}(||\mathbf{f}^i - \mathbf{f}^j||_2^2) \\ &= \sum_{i,j} |z_{ij}|(\frac{1}{2}||\mathbf{f}^i - \mathbf{f}^j||_2^2) = ||\mathbf{Z} \odot \mathbf{\Theta}||_1\end{aligned} \tag{3}$$

where $\theta_{ij} = \frac{1}{2}||\mathbf{f}^i - \mathbf{f}^j||_2^2$. Therefore, the final step spectral clustering reads:

$$\min_{\mathbf{F}} \ ||\mathbf{Z} \odot \mathbf{\Theta}||_1 \ \mathrm{s.t.} \ \ \mathbf{F}^T\mathbf{F} = \mathbf{I}. \tag{4}$$

Considering the ubiquitousness of multi-view representation in practice, extending the single-view clustering to the multi-view setting is of vital importance.

2.2. Exclusivity-Consistency Regularized Multi-view Subspace Clustering

In this part, we detail our proposed multi-view subspace clustering model. Given the data set $\mathbf{X}_v \in \mathbb{R}^{d_v\times n}$, which denotes the features in the v-th view ($v = 1, \ldots, V$). Performing the subspace learning on each single view gives the subspace representation $\mathbf{Z}_v$ for the v-th view. The nonzero elements in $\mathbf{Z}_v$ correspond to the data points from the same subspace. In fact, how to combine multi-view features in subspace clustering is challenging. The naive method is to concatenate all the features together and execute the clustering on the concatenated features. However, in this way, each view will be treated equally, which is not always proper. In order to combine the multi-view subspace learning results, existing works [4, 9] prefer to learn a common representation or a common indicator matrix among all the views. Unfortunately, they can not guarantee the complementarity across different $\mathbf{Z}_v$'s. Recently, to exploit such complementary information, the work [2] enforces different representations to be diverse by adopting the Hilbert-Schmidt Independence Criterion (HSIC). The drawback of using HSIC is that HSIC is a value-aware criterion (*i.e.*, the diversity between $\mathbf{Z}_v$ and $\mathbf{Z}_w$ is related to their values). Due to the scale issue of $\mathbf{Z}_v$[2], the value-aware regularization may greatly degrade the performance. Moreover, the work [2] executes the representation learning and spectral clustering in two separated steps, which also limits its performance. Based on these observations, we introduce a novel *position-aware* exclusivity term to exploit the complementary information and an indicator consistency term to unify the processing of subspace clustering.

2.2.1 Representation Exclusivity

Definition 1. ***(Exclusivity [12])*** *Exclusivity between two matrix $\mathbf{U} \in \mathbb{R}^{n\times n}$ and $\mathbf{V} \in \mathbb{R}^{n\times n}$ is defined as $\mathcal{H}(\mathbf{U}, \mathbf{V}) = ||\mathbf{U} \odot \mathbf{V}||_0 = \sum_{i,j}(u_{ij} \cdot v_{ij} \neq 0)$, where $\odot$ designates the Hadamard product (i.e., element-wise product), and $||\cdot||_0$ is the ℓ_0-"norm"*[3].

From the definition, we can observe that the exclusivity encourages two matrix to be as diverse as possible. Ideally, if the position (i, j) of $\mathbf{U}$ is not equal to zero, then the exclusivity term enforces the same position (i, j) of $\mathbf{V}$ to be zero. Consequently, we can say that the defined exclusivity term is *position-aware*. Naturally, for the diversity between different representations, we have:

$$\mathcal{H}(\mathbf{Z}_v, \mathbf{Z}_w) = ||\mathbf{Z}_v \odot \mathbf{Z}_w||_0. \tag{5}$$

Furthermore, owning to the non-convexity and discreteness of ℓ_0 norm, we relax the ℓ_0 norm into ℓ_1-norm to make the objective (5) computationally tractable. In the sequel, we reach the following relaxed exclusivity as:

$$\mathcal{H}(\mathbf{Z}_v, \mathbf{Z}_w) = ||\mathbf{Z}_v \odot \mathbf{Z}_w||_1. \tag{6}$$

Extending the model (1) into multi-view case by incorporating the relaxed exclusivity term (6) leads to the following objective function:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{Z}_v} \sum_{v=1}^{V} \left(||\mathbf{E}_v||_1 + \lambda_1||\mathbf{Z}_v||_1 + \lambda_2 \sum_{w\neq v} ||\mathbf{Z}_v \odot \mathbf{Z}_w||_1 \right) \\ \mathrm{s.t.} \ \ \forall\, v,\ \mathbf{X}_v = \mathbf{X}_v\mathbf{Z}_v + \mathbf{E}_v,\ \mathrm{diag}\,(\mathbf{Z}_v) = \mathbf{0}.\end{aligned} \tag{7}$$

Here we adopt the ℓ_1-norm for pursuing sparse representation $||\mathbf{Z}||_1$. As pointed out in [9], the low-rank constraint is sometimes bad for the low dimension features due to the rigor rank restriction. For the error term $\mathbf{E}$, its norm $||\cdot||_l$ depends upon the prior knowledge about the pattern of noise or corruptions [32]. This work simply adopts the ℓ_1-norm to handle with the sparse corruptions. Moreover, we do not use the HSIC for measuring the diversity mainly for two reasons. One is that HSIC is *value-aware* while the proposed exclusivity term is *position-aware*, expecting that the position-aware criterion could be better to handle the case of the magnitude of element values. The other is that the relaxed exclusivity term can be seamlessly incorporated with the SSC [8] framework.

[2] Although the diagonal block in different $\mathbf{Z}_v$ are similar, the magnitude of element values in $\mathbf{Z}_v$ can be dramatically different.

[3] ℓ_0-norm is not a real norm.

2.2.2 Indicator Consistency

The work [18] has validated that integrating the subspace learning and spectral clustering into one framework can improve the performance for single view cases. As for multi-view cases, knowing that the goal of clustering is to classify a point with multi-view features into only one cluster, we can introduce the label consistency term naturally as:

$$\min \sum_{v=1}^{V} ||\mathbf{Z}_v \odot \mathbf{\Theta}||_1 \text{ s.t. } \mathbf{F}^T\mathbf{F} = \mathbf{I}, \tag{8}$$

where $\mathbf{F}$ is the common indicator matrix for all the views. Consequently, the clustering is expected to be consistent for all the views, *i.e.* the corresponding points with exclusive representations should be in the same cluster.

Based on the above analysis, we prefer to simultaneously harness the exclusivity of different representations and the consistency of indicator. Now, putting all concerns, say (7) and (8), together results in our final Exclusivity-Consistency Regularized Multi-view Subspace Clustering (ECMSC) model as follows:

$$\begin{aligned}\min_{\mathbf{F},\mathbf{Z}_1,\ldots,\mathbf{Z}_V} \sum_{v=1}^{V} ||\mathbf{E}_v||_1 + \lambda_1||\mathbf{Z}_v||_1 + \lambda_2 \underbrace{\sum_{w\neq v} ||\mathbf{Z}_v \odot \mathbf{Z}_w||_1}_{\textbf{Exclusivity}} \\ + \lambda_3 \underbrace{||\mathbf{Z}_v \odot \mathbf{\Theta}||_1}_{\textbf{Consistency}} \\ \text{s.t. } \mathbf{X}_v = \mathbf{X}_v\mathbf{Z}_v + \mathbf{E}_v,\ \text{diag}(\mathbf{Z}_v) = \mathbf{0},\ \mathbf{F}^T\mathbf{F} = \mathbf{I},\end{aligned} \tag{9}$$

where λ_1, λ_2 and λ_3 are the tradeoffs corresponding to the sparsity, exclusivity and consistency terms, respectively.

3. Optimization

It is difficult to jointly solve all the variables in (9) at the same time. In this section, we propose a solution based on solving the following two subproblems alternatively:

- Given $\mathbf{F}$, find the exclusive representation $\mathbf{Z}_v$ and the residual $\mathbf{E}_v$ ($v = 1, \ldots, V$) by ADMM algorithm.
- With $\mathbf{Z}_v$ and $\mathbf{E}_v$ fixed, compute the consistent indicator $\mathbf{F}$ by spectral clustering.

3.1. Update Representation $\mathbf{Z}_v$ and Residual $\mathbf{E}_v$

Given the clustering indicator matrix $\mathbf{F}$, we solve for $\mathbf{Z}_v$ and $\mathbf{E}_v$ by solving the following sub-problem:

$$\begin{aligned}\min_{\mathbf{Z}_v,\mathbf{E}_v} ||\mathbf{E}_v||_1 + \lambda_1||\mathbf{Z}_v||_1 + \lambda_2 \sum_{w\neq v} ||\mathbf{Z}_v \odot \mathbf{Z}_w||_1 \\ + \lambda_3||\mathbf{Z}_v \odot \mathbf{\Theta}||_1 \\ \text{s.t. } \mathbf{X}_v = \mathbf{X}_v\mathbf{Z}_v + \mathbf{E}_v,\ \text{diag}(\mathbf{Z}_v) = \mathbf{0}.\end{aligned} \tag{10}$$

Algorithm 1: ADMM for solving problem (11)

Input: Data matrix $\mathbf{X}$, $\mathbf{\Theta}_0$, $\mathbf{Z}_w$, where $w \in \{1, \ldots, V\}$ and $w \neq v$. λ_1, λ_2 and λ_3.
Initialize: $\mathbf{\Theta} = \mathbf{\Theta}_0$, $\mathbf{E} = \mathbf{0}$, $\mathbf{C} = \mathbf{Z} = \mathbf{Z}_0$, $\mathbf{Q}_1 = \mathbf{0}$, $\mathbf{Q}_2 = \mathbf{0}$, $\rho = 1.1$, $\epsilon = 10^{-6}$.
while *not convergence* **do**
  Update $\mathbf{Z}$ via (14);
  Update $\mathbf{C}$ via (16);
  Update $\mathbf{E}$ via (18);
  Update $\mathbf{Q}_1$ and $\mathbf{Q}_2$ via (19);
  Update $\mu = \mu\rho$;
  Check the convergence condition $||\mathbf{X} - \mathbf{XC} - \mathbf{E}||_\infty \leq \epsilon$.
end
Output: $\mathbf{Z}$ and $\mathbf{E}$.

For convenience, ignoring the subscript tentatively, the above problem is equivalent to:

$$\begin{aligned}\min_{\mathbf{Z},\mathbf{E}} ||\mathbf{E}||_1 + \lambda_1||\mathbf{Z}||_1 + \lambda_2 \sum_{w\neq v} ||\mathbf{Z} \odot \mathbf{Z}_w||_1 \\ + \lambda_3||\mathbf{Z} \odot \mathbf{\Theta}||_1 \\ \text{s.t. } \mathbf{X} = \mathbf{XC} + \mathbf{E},\ \mathbf{C} = \mathbf{Z} - \text{diag}(\mathbf{Z}).\end{aligned} \tag{11}$$

We solve this problem using the Alternating Direction Method of Multipliers (ADMM) [19]. The augmented Lagrangian can be given by:

$$\begin{aligned}&\mathcal{L}(\mathbf{Z}, \mathbf{C}, \mathbf{E}, \mathbf{Q}_1, \mathbf{Q}_2) = \\ &||\mathbf{E}||_1 + \lambda_1||\mathbf{Z}||_1 + \lambda_2 \sum_{w\neq v} ||\mathbf{Z} \odot \mathbf{Z}_w||_1 + \lambda_3||\mathbf{Z} \odot \mathbf{\Theta}||_1 \\ &+ \Phi(\mathbf{Q}_1, \mathbf{X} - \mathbf{XC} - \mathbf{E}) + \Phi(\mathbf{Q}_2, \mathbf{C} - \mathbf{Z} + \text{diag}(\mathbf{Z})),\end{aligned} \tag{12}$$

with the definition $\Phi(\mathbf{Q}, \mathbf{Y}) = \frac{\mu}{2}||\mathbf{Y}||_F^2 + \langle \mathbf{C}, \mathbf{Y} \rangle$, where $\langle \cdot, \cdot \rangle$ denotes the matrix inner product, μ is a positive penalty scalar, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are the Lagrangian multipliers. To find a minimal point for $\mathcal{L}$, we update each of $\mathbf{Z}$, $\mathbf{C}$, $\mathbf{E}$, $\mathbf{Q}_1$ and $\mathbf{Q}_2$ alternatively while keeping the other variables fixed.

Z-subproblem: Dropping the unrelated terms with respect to $\mathbf{Z}$ yields:

$$\begin{aligned}\min_{\mathbf{Z}} \lambda_1||\mathbf{Z}||_1 + \lambda_2 \sum_{w\neq v} ||\mathbf{Z} \odot \mathbf{Z}_w||_1 + \lambda_3||\mathbf{Z} \odot \mathbf{\Theta}||_1 \\ + \Phi(\mathbf{Q}_2, \mathbf{C} - \mathbf{Z} + \text{diag}(\mathbf{Z})).\end{aligned} \tag{13}$$

The closed-form solution for $\mathbf{Z}$ can be computed by

$$\begin{aligned}&\mathbf{Z} = \hat{\mathbf{Z}} - \text{diag}(\hat{\mathbf{Z}}), \\ &\hat{\mathbf{Z}} = \mathcal{S}_{\frac{\lambda_1 \mathbf{1} + \lambda_2 \sum_{w\neq v} |\mathbf{Z}_w| + \lambda_3 |\Theta|}{\mu}} [\mathbf{C} + \frac{\mathbf{Q}_2}{\mu}],\end{aligned} \tag{14}$$

where $\mathcal{S}_\tau[\cdot]$ is the shrinkage thresholding operator.

C-subproblem: Similar to the Z-subproblem, we obtain:

$$\min_{\mathbf{C}} \ \Phi(\mathbf{Q}_1, \mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}) + \Phi(\mathbf{Q}_2, \mathbf{C} - \mathbf{Z} + \mathrm{diag}\,(\mathbf{Z})). \tag{15}$$

Taking the derivative of the objective with respect to $\mathbf{C}$ and setting it to zero lead to the following closed-form solution:

$$\begin{aligned}\mathbf{C} =&(\mathbf{X}^T\mathbf{X} + \mathbf{I})^{-1}\bigg(\mathbf{X}^T(\mathbf{X} - \mathbf{E} + \frac{\mathbf{Q}_1}{\mu}) \\ &+ \mathbf{Z} - \mathrm{diag}\,(\mathbf{Z}) - \frac{\mathbf{Q}_2}{\mu}\bigg).\end{aligned} \tag{16}$$

E-subproblem: The associated optimization problem with respect to $\mathbf{E}$ can be written as follows:

$$\min_{\mathbf{E}} \ ||\mathbf{E}||_1 + \Phi(\mathbf{Q}_1, \mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}), \tag{17}$$

whose solution can be calculated by :

$$\mathbf{E} = \mathcal{S}_{\frac{1}{\mu}}[\mathbf{X} - \mathbf{X}\mathbf{C} + \frac{\mathbf{Q}_1}{\mu}]. \tag{18}$$

Multiplier: Besides, the multiplier $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are also needed to be updated, which can be simply done through:

$$\begin{aligned}\mathbf{Q}_1 &= \mathbf{Q}_1 + \mu(\mathbf{X} - \mathbf{X}\mathbf{C} - \mathbf{E}), \\ \mathbf{Q}_2 &= \mathbf{Q}_2 + \mu(\mathbf{C} - \mathbf{Z} + \mathrm{diag}\,(\mathbf{Z})).\end{aligned} \tag{19}$$

For clarity, the entire ADMM algorithm of solving the problem (11) is summarized in Algorithm 1.

3.2. Update Indicator $\mathbf{F}$

Given all the exclusive self-representations $\mathbf{Z}_v$ ($v = 1, \ldots, V$) and error matrices $\mathbf{E}_v$, the second step is to update the consistent indicator matrix $\mathbf{F}$. The associated problem with respect to $\mathbf{F}$ in Eq. (9) can be simplified as:

$$\begin{aligned}\min_{\mathbf{F}} \sum_v \ ||\mathbf{Z}_v \odot \mathbf{\Theta}||_1 = \min_{\mathbf{F}} \sum_v \mathrm{tr}(\mathbf{F}^T(\mathbf{D}_v - \mathbf{S}_v)\mathbf{F}) \\ \mathrm{s.t.} \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}.\end{aligned} \tag{20}$$

It can be further reformulated in the following shape:

$$\min_{\mathbf{F}} \ \mathrm{tr}(\mathbf{F}^T\mathbf{M}\mathbf{F}) \ \ \mathrm{s.t.} \quad \mathbf{F}^T\mathbf{F} = \mathbf{I}, \tag{21}$$

where we define $\mathbf{M} = \sum_v(\mathbf{D}_v - \mathbf{S}_v)$. The solution to (21) are the eigenvectors corresponding to the smallest k eigenvalue of the Laplacian matrix $\mathbf{M}$. The rows of $\mathbf{F}$ are then used as input to the k-means algorithm, which produces a clustering of the rows of $\mathbf{F}$ that can be used to produce a binary matrix $\mathbf{F} \in \{0, 1\}^{n\times k}$ such that $\mathbf{F}\mathbf{1} = \mathbf{1}$.

For clarity and completeness, we summarize the whole scheme to solving problem in Algorithm 2. The algorithm alternates between solving for the matrices of exclusive representations and the error $(\mathbf{Z}_v, \mathbf{E}_v)$ given the indicator $\mathbf{F}$ using Algorithm 1, and solving for the consistent indicator $\mathbf{F}$ given $(\mathbf{Z}_v, \mathbf{E}_v)$ using spectral clustering.

Algorithm 2: ECMSC algorithm

Input: Unlabeled multi-view data matrix $\mathcal{D} = \{\mathbf{X}^1, \ldots, \mathbf{X}^V\}$, number of subspaces k.
Initialize: $\mathbf{\Theta} = \mathbf{0}, \lambda_1, \lambda_2, \lambda_3, t = 0$.
while *not converged* **do**
  for *each view* $v \in V$ **do**
    Given $\mathbf{F}$, solve problem (10) via Algorithm 1 to obtain ($\mathbf{Z}^v$,$\mathbf{E}^v$);
  end
  Given all the ($\mathbf{Z}^v$,$\mathbf{E}^v$), solve problem (21) via spectral clustering to obtain $\mathbf{F}$;
  Check the convergence condition $||\mathbf{\Theta}_{t+1} - \mathbf{\Theta}_t||_\infty < 1$;
  if not converged, set $t = t + 1$;
end
Output: Segmentation matrix $\mathbf{F}$.

4. Experiments

4.1. Experimental Settings

4.1.1 Dataset description

Three datasets adopted in the experiments are those widely used in works [2, 31] for face image clustering, including:
Extended Yale-B consists of 2414 face images of 38 individuals. Each individual has 64 near frontal images under different illuminations. Similar with [2, 31], we select the first 10 classes as the final dataset, which has 640 frontal face images in total.
Yale is composed of 165 gray-scale images of 15 individuals. Each individual has 11 images, with different facial expression and configuration.
ORL contains 400 face images of 40 distinct subjects. Each subject has 10 different face images, which were taken at different times, changing with the lighting, facial expressions and facial details.

For all the datasets, we employ the multi-view features provided by the work [31]. Specifically, three types of features, *i.e.* intensity, LBP [26] and Gabor [16] are extracted. The standard LBP features are extracted with the sampling density size of 8 and the blocking number of 7×8. Gabor wavelets are extracted with one scale $\gamma = 4$ at four orientations $\theta = \{0°, 45°, 90°, 135°\}$. So the dimensionalities of LBP and Gabor are 3304 and 6750, respectively.

4.1.2 Compared Methods

We compare our method with recently proposed state-of-the-art alternatives, including 3 single-view methods and 6 multi-view ones.
SPC [24]: The most informative view is selected to perform with the standard spectral clustering scheme.
SSC [8]: Each single view features are separately used. After the sparse subspace learning, the spectral clustering are then employed to obtain the final clustering results.

Table 1. Results (mean ± standard deviation) on Extended Yale-B. We set $\alpha = 0.3, \beta = 0.5$ in ECMSC.

	Method	NMI	ACC	ARI	F-score	Precision	Recall
Single	SPC_{best}	0.360±0.016	0.366±0.059	0.225±0.018	0.303±0.011	0.296±0.010	0.310±0.012
	SSC_{best}	0.534±0.003	0.587±0.003	0.430±0.005	0.487±0.004	0.451±0.002	0.509±0.007
	$S3C_{best}$	0.542±0.010	0.391±0.012	0.415±0.007	0.492±0.004	0.417±0.005	0.487±0.009
Multiple	$FeaCon_{PCA}$	0.152±0.003	0.232±0.005	0.069±0.002	0.161±0.002	0.158±0.001	0.64±0.002
	Min-Dis	0.186±0.003	0.242±0.018	0.088±0.001	0.181±0.001	0.174±0.001	0.189±0.002
	Co-Reg SPC	0.151±0.001	0.224±0.000	0.066±0.001	0.160±0.000	0.157±0.001	0.162±0.000
	ConReg SPC	0.163±0.022	0.216±0.019	0.072±0.012	0.164±0.010	0.163±0.010	0.165±0.011
	LT-MSC	0.637±0.003	0.626±0.010	0.459±0.030	0.521±0.006	0.485±0.001	0.539±0.002
	DiMSC	0.635±0.002	0.615±0.003	0.453±0.000	0.504±0.006	0.481±0.002	0.534±0.001
Proposed	$ECMSC_{\alpha=0}$	0.719±0.011	0.692±0.013	0.492±0.008	0.548±0.007	0.481±0.004	0.691±0.006
	$ECMSC_{\beta=0}$	0.708±0.009	0.678±0.010	0.482±0.011	0.530±0.009	0.487±0.004	0.672±0.011
	ECMSC	**0.759±0.012**	**0.783±0.011**	**0.544±0.008**	**0.597±0.010**	**0.513±0.009**	**0.718±0.006**

Table 2. Results (mean ± standard deviation) on Yale. We set $\alpha = 0.1, \beta = 0.3$ in ECMSC.

	Method	NMI	ACC	ARI	F-score	Precision	Recall
Single	SPC_{best}	0.654±0.009	0.616±0.030	0.440±0.011	0.475±0.011	0.457±0.011	0.495±0.010
	SSC_{best}	0.671±0.011	0.627±0.000	0.475±0.004	0.517±0.007	0.509±0.003	0.547±0.004
	$S3C_{best}$	0.678±0.013	0.634±0.016	0.471±0.005	0.508±0.012	0.512±0.005	0.568±0.025
Multiple	$FeaCon_{PCA}$	0.665±0.037	0.578±0.038	0.396±0.011	0.434±0.011	0.419±0.012	0.450±0.009
	Min-Dis	0.645±0.005	0.615±0.043	0.433±0.006	0.470±0.006	0.446±0.005	0.496±0.006
	Co-Reg SPC	0.648±0.002	0.564±0.000	0.436±0.002	0.466±0.000	0.455±0.004	0.491±0.003
	ConReg SPC	0.673±0.023	0.611±0.035	0.466±0.032	0.501±0.030	0.476±0.032	0.532±0.029
	LT-MSC	0.765±0.008	0.741±0.002	0.570±0.004	0.598±0.006	0.569±0.004	0.629±0.005
	DiMSC	0.727±0.010	0.709±0.003	0.535±0.001	0.564±0.002	0.543±0.001	0.586±0.003
Proposed	$ECMSC_{\alpha=0}$	0.731±0.009	0.742±0.008	0.576±0.010	0.578±0.005	0.534±0.009	0.604±0.006
	$ECMSC_{\beta=0}$	0.738±0.007	0.721±0.005	0.554±0.009	0.565±0.012	0.538±0.009	0.578±0.011
	ECMSC	**0.773±0.010**	**0.771±0.014**	**0.590±0.014**	**0.617±0.012**	**0.584±0.013**	**0.653±0.013**

S3C [18]: Similarly, we carry out the clustering on each single view and report the best performance.

FeaCon$_{PCA}$: The method firstly concatenates all types of features and applies PCA to reduce the feature dimension to 300. Then, spectral clustering is performed on the low dimensional features.

Min-Dis [7]: This method creates a bipartite graph and tries to minimize the disagreement. Then the final result is obtained by spectral clustering.

Co-Reg SPC [15]: The method co-regularizes the clustering hypotheses to enforce that corresponding data points should be in the same cluster.

ConvexReg SPC [5]: A common representation for all views is first learned. Then the standard spectral clustering is carried out on the similarity matrix.

DiMSC [2]: The method firstly enforces the diversity of different views by the HSIC criterion [10], and then applies the spectral clustering to obtain the final result.

LT-MSC [31]: Low-rank tensor constraint is enforced to directly construct the similarity matrix and then performs the spectral clustering.

To illustrate the advantage of our representation exclusivity term and indicator consistency term, we add two extra methods, the difference between which comes from the way of parameter setting. Specifically, by setting $\alpha = 0$ (*i.e.*, without exclusivity), we report the result as $ECMSC_{\alpha=0}$[4]. If disabling the consistency term (*i.e.*, $\beta = 0$), we learn the exclusive representations, and then, the similarity is constructed as $\mathbf{S} = \sum_{v=1}^{V}(|\mathbf{Z}_v|^T + |\mathbf{Z}_v|)/2$. After that, the

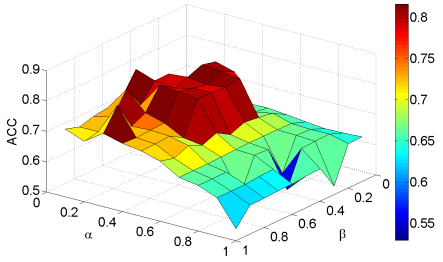


Figure 2. Parameters α and β tuning on Extended Yale-B.

[4]Since the code of [9] is not available, we do not explicitly compare with it. However, from the formulation, the work in [9] can be approximately regarded as $ECMSC_{\alpha=0}$.

Table 3. Results (mean ± standard deviation) on ORL. We set $\alpha = 0.1, \beta = 0.7$ in ECMSC.

	Method	NMI	ACC	ARI	F-score	Precision	Recall
Single	SPC_{best}	0.884±0.002	0.726±0.025	0.655±0.005	0.664±0.005	0.610±0.006	0.728±0.005
	SSC_{best}	0.893±0.007	0.765±0.008	0.694±0.013	0.682±0.012	0.673±0.007	0.764±0.005
	$S3C_{best}$	0.902±0.012	0.784±0.009	0.705±0.019	0.698±0.018	0.688±0.012	0.791±0.011
Multiple	$FeaCon_{PCA}$	0.835±0.004	0.675±0.028	0.564±0.010	0.574±0.010	0.532±0.011	0.624±0.008
	Min-Dis	0.876±0.002	0.748±0.051	0.654±0.004	0.663±0.004	0.615±0.004	0.718±0.003
	Co-Reg SPC	0.853±0.003	0.715±0.000	0.602±0.004	0.615±0.000	0.567±0.004	0.666±0.004
	ConReg SPC	0.883±0.013	0.734±0.031	0.668±0.032	0.676±0.035	0.628±0.041	0.731±0.030
	LT-MSC	0.930±0.002	0.795±0.007	0.750±0.003	0.768±0.007	0.766±0.009	0.837±0.004
	DiMSC	0.940±0.003	0.838±0.001	0.802±0.000	0.807±0.003	0.764±0.012	0.856±0.004
Proposed	$ECMSC_{\alpha=0}$	0.923±0.006	0.822±0.008	0.789±0.012	0.782±0.009	0.769±0.007	0.834±0.009
	$ECMSC_{\beta=0}$	0.934±0.009	0.841±0.012	0.801±0.011	0.810±0.009	0.767±0.006	0.848±0.010
	ECMSC	**0.947±0.009**	**0.854±0.011**	**0.810±0.012**	**0.821±0.015**	**0.783±0.008**	**0.859±0.012**

spectral clustering is performed on this similarity. Finally, we set α and β suitably, termed it as ECMSC. For fair comparison, following the protocol in [31], we report the average accuracy and standard derivation of all the competitors over 30 independent trials.

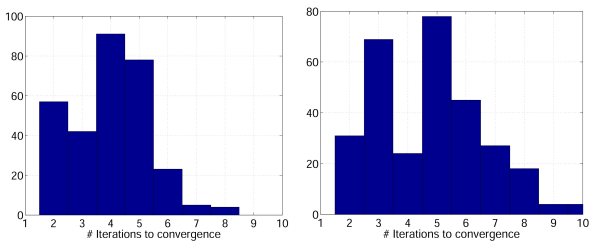
4.1.3 Evaluation Metrics

To assess the performance, six metrics including Normalized Mutual Information (**NMI**), Accuracy (**ACC**), Adjusted Rand index (**AR**), **F-score**, **Precision** and **Recall** are utilized, of which, the **ACC** and **NMI** are the most two popular metrics and have been adopted in many literatures like [3, 11]. The metrics **ARI**, **F-score**, **Precision** and **Recall** have been widely used in [31, 2] to measure the clustering quality. These six metrics favor different properties in the clustering. For all the metrics, a higher value indicates a better clustering quality.

4.2. Experimental Results

4.2.1 Parameter Setting

In our ECMSC, the parameters λ_1, λ_2 and λ_3 need to be set properly. Inspired by the works [25, 18], the convergence of ECMSC can also be improved by using $\lambda_3 \leftarrow \lambda_3\eta$. Therefore, we set $\lambda_1 = \eta^{1-t}$, $\lambda_2 = \alpha$ and $\lambda_3 = \beta\eta^{t-1}$ to balance the corresponding terms, where $\eta = 1.2$ and $t = \{1, 2, \ldots, T\}$ is the iteration index. Finally, we have two free parameters α and β to tune, which emphasize the importance of our exclusivity term and consistency term, respectively. Specifically, we empirically tune the parameters α and β in $\{0.1, 0.2, \ldots, 1\}$. Due to the limitation of the space, we only show the parameter effect on the Extended Yale-B dataset, which is plotted in Figure 2. We can see that the promising performance can be expected when the parameters α and β are chosen in a certain range (*e.g.*, $\alpha \in [0.1, 0.5]$, $\beta \in [0.2, 0.7]$). Moreover, the detailed parameter settings for each dataset are reported in their corresponding tables.



(a) Yale (b) ORL

Figure 3. Convergence analysis on Yale and ORL dataset.

4.2.2 Convergence Analysis

The whole algorithm alternates between solving for the matrices of exclusive representations and the error $(\mathbf{Z}_v, \mathbf{E}_v)$ given the segmentation $\mathbf{F}$ using Algorithm 1 (*i.e.*, ADMM) and solving for $\mathbf{F}$ given $(\mathbf{Z}_v, \mathbf{E}_v)$ using spectral clustering. Although we can not provide the theoretical proof of convergence for ECMSC, our experiments show that the algorithm has a very stable convergence behavior in practice. Similar to the work [18], we show the convergence of our ECMSC empirically. Specifically, we randomly sample $\{2, 3, 5, 8, 10\}$ classes of Yale and ORL datasets. For each number, we randomly sample 60 times and show the histogram of the numbers of iterations (*i.e.*, the statistics of how many iterations has been taken when the algorithm meets the convergence condition.) for ECMSC to converge. In Figure 3, we can observe that the proposed ECMSC algorithm converges in $2 \sim 8$ iterations on average.

4.2.3 Representation Visualization

To validate the effectiveness of our proposed representation exclusivity term and the indicator consistency term, due to

the limited space, we only show the visualization results on the Extended Yale-B dataset. As observed in Figure 4, we can see that the representations without exclusivity term (*i.e.*, the first row) are less diverse than the representations employed with exclusivity (*i.e.*, the second and third row). Moreover, from the last column (*i.e.*, the indicator matrix $\mathbf{\Theta}$ of different versions of ECMSC), we can clearly see that simultaneously exploiting representation exclusivity and indicator consistency will be beneficial for multi-view subspace clustering task.

4.2.4 Performance Comparison

We report the detailed clustering results on three face image datasets in Tables 1-3. In each table, the bold values represent the best performance.

Table 1 provides the quantitative comparison among the competitors on the Extended Yale-B dataset. It can be observed that most of the comparisons have relatively low performances. The major reason is the large variation of illumination of this dataset. However, our proposed ECMSC algorithm still achieves significant improvements around 12.2%, 15.7%, 8.5%, 7.6%, 2.8% and 17.9% over the most competitive method LT-MSC in terms of NMI, ACC, AR, F-score, Precision and Recall, respectively. Compared the work in [2] (DiMSC) with ECMSC$_{\beta=0}$, the main difference between these two methods is the diversity evaluation criteria, the former adopts the value-aware HSIC criterion while the latter is the proposed position-aware exclusivity criterion. From the results, we can see that the position-aware exclusivity criterion is much better than value-aware HSIC criterion for multi-view subspace clustering task.

Table 2 displays the clustering results on the Yale dataset. Similar trend to Table 1, most of the competitors achieve lower performance than the proposed ECMSC. From the values, we can see that ECMSC excels all the baselines, both single-view and multi-view methods. The main reason is the exploitation of representation exclusivity and indicator consistency simultaneously. The most two competitive multi-view subspace clustering methods LT-MSC and DiMSC, have achieved a relatively promising results. However, compared ECMSC$_{\beta=0}$ with DiMSC, we find that the exclusivity is better than HSIC. Further, compared ECMSC$_{\beta=0}$ with ECMSC, it can be concluded the indicator consistency (*i.e.*, unify the process of subspace clustering) can further gain improvements.

Table 3 shows the performance comparison on the ORL dataset, from which we notice that all of SPC, SSC, S3C, FeaCon, Min-Dis, Co-Reg SPC, ConReg SPC perform relatively poorly. The methods LT-MSC produce more promising results on this dataset. It seems that the low-rank representation model are more suitable for constructing the similarity matrix on this dataset. And the improvement of ECMSC over LT-MSC is not very significant. Compared our ECMSC$_{\beta=0}$ with DiMSC, due to the case of magnitude of element values is not obvious, our position-aware term do not shows its advantages, and the improvement is slight. However, our intact ECMSC draws the merits of complementary representation and consistent indicator. Therefore, it could achieve higher performance in general.

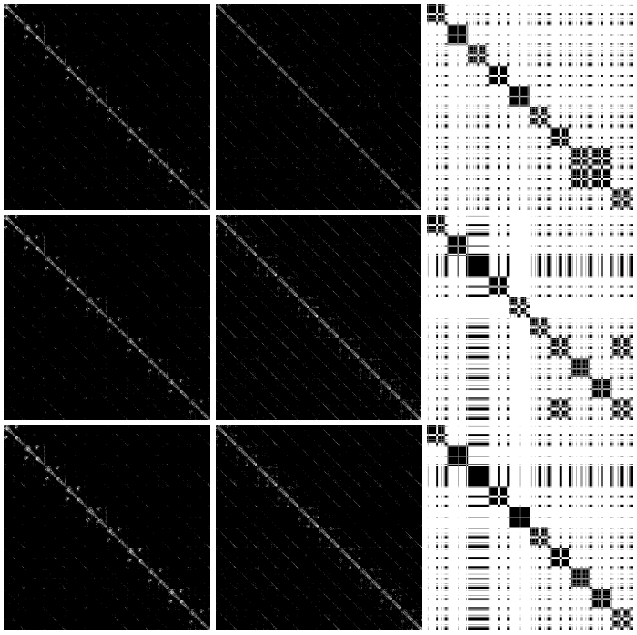


Figure 4. **From left to right**: The columns are the visualization of subspace representations $\mathbf{Z}_1$, $\mathbf{Z}_2$ and the indicator matrix $\mathbf{\Theta}$, respectively. **From top to bottom**: The rows are the results of ECMSC$_{\alpha=0}$ (ACC = 0.701), ECMSC$_{\beta=0}$ (ACC = 0.689) and ECMSC (ACC = 0.781), respectively.

5. Conclusion

In this paper, we have proposed a novel multi-view subspace clustering model namely ECMSC. Different from previous works, we simultaneously consider the complementary representation and consistent indicator into one framework. Moreover, a novel position-aware exclusivity term has been introduced to measure the diversity between different representations. In addition, an efficient alternative based algorithm has been proposed to seek the optimal solution. Extensive experimental results on several datasets have demonstrated the significant advantage of our method.

6. Acknowledgments

This work was supported by the National Key Research and Development Plan (Grant No.2016YFC0801002), the Chinese National Natural Science Foundation Projects #61473291, #61572501, #61502491, #61572536, #61672521 and AuthenMetric R&D Funds.

References

[1] X. Cai, F. Nie, and H. Huang. Heterogeneous image feature integration via multi-modal spectral clustering. *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*, pages 1977–1984, 2011.

[2] X. Cao, C. Zhang, and H. Fu. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision (CVPR)*, pages 486–594, 2015.

[3] X. Cao, C. Zhang, and C. Zhou. Constrained multi-view video face clustering. *IEEE Transactions on Image Processing (TIP)*, 24(11):4381–4393, 2015.

[4] N. Chen, J. Zhu, and E. P. Xing. Predictive subspace learning for multi-view data: a large margin approach. In *Proceedings of Advances in neural information processing systems (NIPS)*, pages 361–369, 2010.

[5] M. Collins, J. Liu, and J. Xu. Spectral clustering with a convex regularizer on millions of images. In *Proceedings of IEEE Conference on European Conference on Computer Vision (ECCV)*, pages 282–298, 2014.

[6] J. P. Costeira and T. Kanade. A multibody factorization method for independently moving objects. *International Journal of Computer Vision (IJCV)*, 29(3):159–179, 1998.

[7] S. De and R. Virginia. Spectral clustering with two views. In *Proceedings of ICML workshop on learning with multiple views*, 2005.

[8] E. Elhamifar and R. Vidal. Sparse subspace clustering. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2790–2797, 2009.

[9] H. Gao, F. Nie, and X. Li. Multi-view subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 4238–4246, 2015.

[10] A. Gretton, O. Bousquet, and A. Smola. Measuring statistical dependence with hilbert-schmidt norms. In *Proceedings of International conference on algorithmic learning theory*, pages 63–77, 2005.

[11] X. Guo. Robust subspace segmentation by simultaneously learning data representations and their affinity matrix. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 119–128, 2015.

[12] X. Guo. Exclusivity regularized machine. *arXiv preprint arXiv:1603.08318.*, 2016.

[13] J. Ho, M. H. Yang, and J. Lim. Clustering appearances of objects under varying illumination conditions. *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*, 2003.

[14] K. Kanatani. Motion segmentation by subspace separation and model selection. In *Proceedings of IEEE Conference on Computer Vision (ICCV)*, 2001.

[15] A. Kumar, P. Rai, and H. Daume. Co-regularized multi-view spectral clustering. In *Proceedings of Advances in neural information processing systems (NIPS)*, pages 1413–1421, 2011.

[16] M. Lades, J. Vorbrggen, and J. Buhmann. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[17] C. Lang, Z. Sun, W. Jia, and R. Hu. Saliency detection by multitask sparsity pursuit. *IEEE Transactions on Image Processing (TIP)*, 21(3):1327–1338, 2012.

[18] C. Li and R. Vidal. Structured sparse subspace clustering: a unified optimization framework. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*, 2015.

[19] Z. Lin, R. Liu, and Z. Su. Linearized alternating direction method with adaptive penalty for low-rank representation. In *Proceedings of Advances in neural information processing systems (NIPS)*, pages 612–620, 2011.

[20] G. Liu, Z. Lin, S. Yan, J. Sun, and Y. Yu. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 35(1):171–184, 2013.

[21] J. Liu, C. Wang, and J. Gao. Multi-view clustering via joint nonnegative matrix factorization. *Proceedings of SDM.*, 2013.

[22] S. Liu, X. Liang, and L. Liu. Matching-cnn meets knn: Quasi-parametric human parsing. *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*, 2015.

[23] S. Liu, Z. Song, and G. Liu. Street-to-shop: Cross-scenario clothing retrieval via parts alignment and auxiliary set. *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR)*, 2012.

[24] A. Ng, M. Jordan, and Y. Weiss. On spectral clustering: Analysis and an algorithm. In *Proceedings of Advances in neural information processing systems (NIPS)*, pages 849–856, 2002.

[25] F. Nie, X. Wang, and H. Huang. Clustering and projected clustering with adaptive neighbors. *Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).*, 2014.

[26] T. Ojala and M. Pietikainen. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 24(7):971–987, 2002.

[27] J. Shi and J. Malik. Normalized cuts and image segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(8):888–905, 2000.

[28] P. Tseng. Nearest q-flat to m points. *Journal of Optimization Theory and Applications*, 105(1):249–252, 2000.

[29] Y. Xie, D. Tao, and W. Zhang. Multi-view subspace clustering via relaxed l_1-norm of tensor multi-rank. *arXiv preprint arXiv:1610.07126.*, 2016.

[30] C. Zhang, H. Fu, and Q. Hu. Flexible multi-view dimensionality co-reduction. *IEEE Transactions on Image Processing (TIP)*, 26(2):648–659, 2017.

[31] C. Zhang, H. Fu, S. Liu, C. Liu, and X. Cao. Low-rank tensor constrained multiview subspace clustering. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, pages 1582–1590, 2015.

[32] Q. Zhao, D. Meng, and Z. Xu. Robust principal component analysis with complex noise. *Proceedings of the International Conference on Machine Learning (ICML).*, 2014.