Enriched Robust Multi-View Kernel Subspace Clustering

Mengyuan Zhang
Clemson University
mengyuz@clemson.edu

Kai Liu
Clemson University
kail@clemson.edu

Abstract

Subspace clustering is to find underlying low-dimensional subspaces and cluster the data points correctly. In this paper, we propose a novel multi-view subspace clustering method. Most existing methods suffer from two critical issues. First, they usually adopt a two-stage framework and isolate the processes of affinity learning, multi-view information fusion and clustering. Second, they assume the data lies in a linear subspace which may fail in practice as most real-world datasets may have non-linearity structures. To address the above issues, in this paper we propose a novel Enriched Robust Multi-View Kernel Subspace Clustering framework where the consensus affinity matrix is learned from both multi-view data and spectral clustering. Due to the objective and constraints which is difficult to optimize, we propose an iterative optimization method which is easy to implement and can yield closed solution in each step. Extensive experiments have validated the superiority of our method over state-of-the-art clustering methods.

1. Introduction

In machine learning, high-dimensional data are ubiquitous. For example, images may consist of thousands of pixels and text data may have tons of features. High dimensionality requires demanding computational time and memory, and moreover, noise in the data can bring adversely influence on performance. Fortunately, recent research shows that high-dimensional data often lies in low-dimensional structures. For instance, the set of face images under all possible illumination conditions can be well approximated by a 9-dimensional linear subspace [2]. Recovering the low-dimensional structures of data can not only save computational cost, but also will improve the accuracy and effectiveness of learning methods. For data samples lie in low-dimensional subspaces instead of being uniformly distributed across ambient space, subspace clustering is to separate data according to their underlying subspaces and the basis for each subspace [45]. For the past decade, subspace clustering has been explored actively and applied in many applications such as image/motion/video segmentation [16, 50, 51], image representation [25, 60], *etc*.

Subspace clustering approaches have been developed and studied extensively, and among them are: iteration-based methods such as [44, 58] which alternates cluster assignment and subspace fitting; factorization-based algebraic approaches such as [34, 46] which hypothesizes that the subspaces are independent; statistical approaches such as Multi-stage Learning [17], Mixtures of Probabilistic PCA [43] which alternates between clustering and subspace estimation via Expectation Maximization; spectral clustering based approaches such as Local Subspace Affinity [50], Locally Linear Manifold Clustering [16] where data segmentation is obtained from spectral clustering. More recently, sparse subspace clustering (SSC) has been proposed [11, 37, 38] to find a sparse representation corresponding to the data points from the same subspace.

In the big-data era, many computer vision problems are fed with the dataset represented by multiple feature sets, which is so called ‘multi-view’ data. Different descriptors characterize various and independent information from different perspectives. For instance, an image can be described by color, texture, histogram of oriented gradients (HOG), local binary pattern (LBP), *etc*. These different features can provide useful information from different views to improve clustering performance [32]. Multi-view clustering is to integrate these multiple feature sets together to perform reliable clustering. Most existing multi-view subspace clustering methods integrate multi-view information in similarity or representation by merging multiple graphs or representation matrices into a shared one. For example, [18, 42] learn a shared sparse subspace representation by performing matrix factorization. Similarly, centroid-based multi-view low-rank sparse subspace clustering methods [5, 33, 55] induce low-rank and sparsity constraints on the shared affinity matrix across different views. Instead of obtaining a shared representation directly, Hilbert-Schmidt Independence Criterion (HSIC) and Markov chain are introduced to learn complementary subspace representations, followed by adding them together appropriately [7, 48].

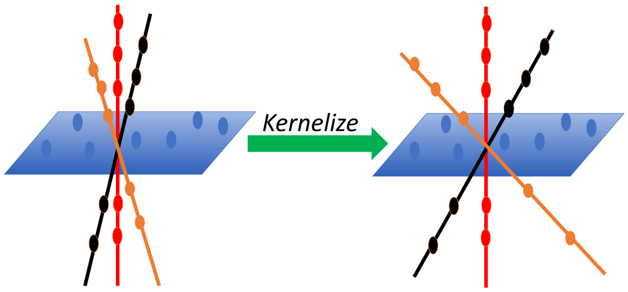


Figure 1. Using the kernel trick, data is transformed onto a high dimensional feature space so that better sparse representations can be found for subspace clustering.

Although these subspace methods mentioned above have achieved significant success, they still suffer from the following issues: 1) They assume the data is well separable in linear subspace, which may not be true and extensive researches show data can be better separated by mapping to higher dimension [15, 21]. 2) Previous approaches obtain the consensus affinity matrix by minimizing the squared Frobenius norm of its difference to each view, which may yield poor representation when a certain view is not well learned [56, 57]. 3) Most existing approaches are usually conducted in a two-step fashion [20, 24, 59], which may fail to obtain optimal clustering performance since the learning stage is separated from the subsequent clustering stage. The main contributions of this work are summarized as follows:

- To explore the nonlinear relationship in the data, we transform the data from original space to a kernel space, which improves the performance of multi-view subspace clustering when dealing with non-linearity (*e.g.* specific manifold) data distributions, as shown in Figure 1.
- We investigate a unified multi-view subspace clustering framework which jointly optimizes similarity learning and spectral clustering, where enriched consensus affinity matrix is learned from different sources.
- We provide a formulation to obtain robust consensus affinity matrix, which yields better clustering result.
- We propose an updating algorithm with closed solution in each step, which is computationally efficient.

2. Multi-View Kernel Subspace Clustering

In this section, we first provide a brief background on sparse subspace clustering. After that, we will give the motivation of our proposed method.

2.1. Sparse Subspace Clustering

Given n data points $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n\} \in \mathbb{R}^{d\times n}$, subspace clustering assumes that each data point can be approximated by a linear combination of dataset samples [11]:

$$\boldsymbol{X} = \boldsymbol{X}\boldsymbol{C} + \boldsymbol{E}, \tag{1}$$

where $\boldsymbol{C} = \{\boldsymbol{c}_1, \boldsymbol{c}_2, \dots, \boldsymbol{c}_n\} \in \mathbb{R}^{n\times n}$ is the subspace representation matrix, with each $\boldsymbol{c}_i$ representing the original data point $\boldsymbol{x}_i$ based on the subspace. $\boldsymbol{E} \in \mathbb{R}^{d\times n}$ is the error matrix.

Sparse subspace clustering formulates the objective as:

$$\min_{\boldsymbol{C}} \|\boldsymbol{X}-\boldsymbol{X}\boldsymbol{C}\|_F^2+\theta\|\boldsymbol{C}\|_1,\ \ s.t.\ diag(\boldsymbol{C}) = 0,\ \ \boldsymbol{C}^T\mathbf{1} = \mathbf{1}, \tag{2}$$

where $\|\cdot\|_F$ denotes the Frobenius norm while $\|\boldsymbol{C}\|_1 = \sum_{i,j}|\boldsymbol{C}_{ij}|$. The constraint $\boldsymbol{C}^T\mathbf{1} = \mathbf{1}$ indicates that the data point lies in a union of affine subspaces while the constraint $diag(\boldsymbol{C}) = 0$ rules out the case that a data point is represented by itself, which hints that each data point $\boldsymbol{x}_i$ can only be represented by a combination of other points $\boldsymbol{x}_j (j \neq i)$. Solving the optimization problem in Eq. (2), we will get the representation $\boldsymbol{c}_i$ for each data point $\boldsymbol{x}_i$.

After obtaining the subspace structure, we construct the affinity matrix by setting $\boldsymbol{W} = \frac{|\boldsymbol{C}|+|\boldsymbol{C}|^T}{2}$. Therefore, we can perform spectral clustering on subspace affinity matrix:

$$\min_{\boldsymbol{F}} \mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}), \quad s.t. \quad \boldsymbol{F}^T\boldsymbol{F} = \mathbf{I}, \tag{3}$$

where $\boldsymbol{F}$ is the cluster indicator matrix, $\boldsymbol{L} := \boldsymbol{D}-\boldsymbol{W}$ where $\boldsymbol{D}$ is a diagonal matrix given by $\boldsymbol{D}(i,i) = \sum_j \boldsymbol{W}(i,j)$.

2.2. Robust Multi-View Kernel Subspace Clustering

Given the v-view dataset $\boldsymbol{X}^{(v)} \in \mathbb{R}^{d_v\times n}$, if we perform the subspace learning on each single view, we can get the subspace representation $\boldsymbol{C}^{(v)}$ for the v-th view. The fundamental challenge boils down to combine multi-view features in subspace clustering. An intuitive and naive method is to concatenate all the features together and perform clustering on the concatenated features, where the more informative view and the less informative one will be treated equally. Therefore, the solution is inevitably not optimal in many scenarios. In contrast, one can perform the clustering on each single view followed by fusing them together. In order to combine multi-view sparse subspace clustering results, we can perform the subspace learning on different views simultaneously by solving:

$$\begin{aligned}\min_{\boldsymbol{C}^{(v)},\boldsymbol{C}^*} &\sum_v \|\boldsymbol{X}^{(v)} - \boldsymbol{X}^{(v)}\boldsymbol{C}^{(v)}\|_F^2 + \theta\|\boldsymbol{C}^{(v)}\|_1 \\ &+ \lambda\|\boldsymbol{C}^{(v)} - \boldsymbol{C}^*\|_F^2,\ s.t.\ diag(\boldsymbol{C}^{(v)}) = 0,\ \ \boldsymbol{C}^{(v)T}\mathbf{1} = \mathbf{1},\end{aligned} \tag{4}$$

where $\boldsymbol{C}^*$ is the consensus affinity matrix across multiple views and spectral clustering will be performed based on it.

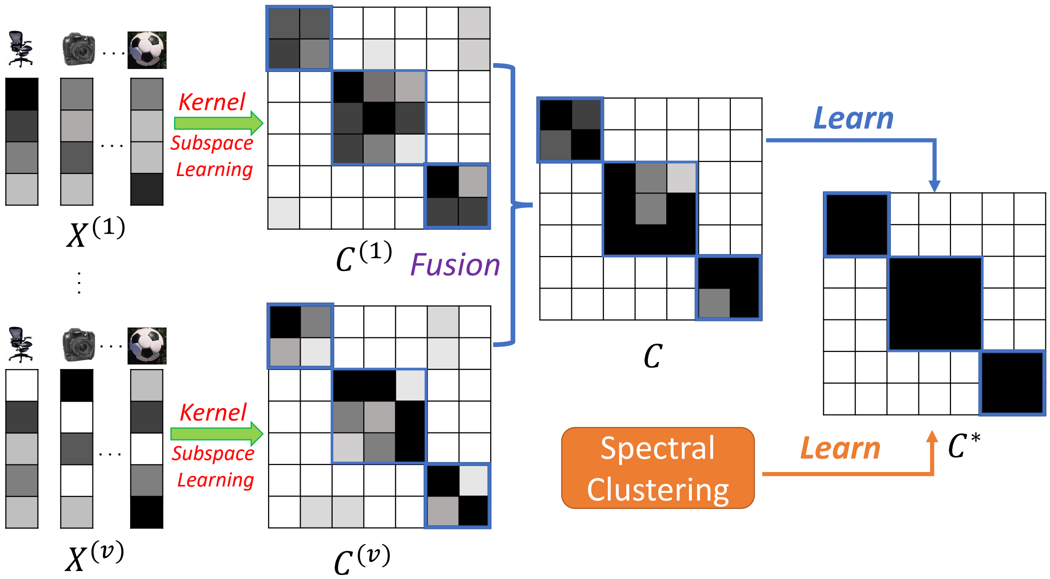


Figure 2. The framework of our proposed method.

However, experiments have demonstrated that $\boldsymbol{C}^{(v)}$ can be significantly different. One can see that if a certain view $\boldsymbol{C}^{(v)}$ is not learned well, to minimize the objective, $\boldsymbol{C}^*$ will deviate from optimal solution due to the squared Frobenius norm which is known to be sensitive to noise/outliers. Inspired by the observation above, to mitigate its adverse influence, we propose a more robust [4, 31, 52] formulation:

$$\min_{\boldsymbol{C}^{(v)},\boldsymbol{C}^*} \sum_v \|\boldsymbol{X}^{(v)} - \boldsymbol{X}^{(v)}\boldsymbol{C}^{(v)}\|_F^2 + \theta\|\boldsymbol{C}^{(v)}\|_1 + \lambda\|\boldsymbol{C}^{(v)} - \boldsymbol{C}^*\|_1, \ s.t. \ diag(\boldsymbol{C}^{(v)}) = 0, \ \ \boldsymbol{C}^{(v)T}\mathbf{1} = \mathbf{1}. \tag{5}$$

On the other hand, kernel tricks have been played in various machine learning techniques/algorithms such as PCA [35], SVM [41], K-means [9], *etc*. Those kernel version methods yield very promising results especially when the data in original space is not well separable, but can be separated by projecting into higher dimension space via $\Phi(\cdot)$ where $\Phi : \mathbb{R}^d \rightarrow \mathbb{R}^m (m > d)$ [14]. Therefore, we introduce the robust multi-view kernel subspace clustering by optimizing:

$$\min_{\boldsymbol{C}^{(v)},\boldsymbol{C}^*} \sum_v \|\Phi(\boldsymbol{X}^{(v)}) - \Phi(\boldsymbol{X}^{(v)})\boldsymbol{C}^{(v)}\|_F^2 + \theta\|\boldsymbol{C}^{(v)}\|_1 + \lambda\|\boldsymbol{C}^{(v)} - \boldsymbol{C}^*\|_1, s.t. \ diag(\boldsymbol{C}^{(v)}) = 0, \ \ \boldsymbol{C}^{(v)T}\mathbf{1} = \mathbf{1}. \tag{6}$$

2.3. Enriched Multi-View Subspace Clustering

Most existing multi-view subspace learning will do spectral clustering after obtaining $\boldsymbol{C}^*$ which ignores the potential connection between the two stages. As a contribution of this paper, we propose an enriched procedure by combining the learning with clustering stage via:

$$\begin{aligned} &\min_{\boldsymbol{C}^{(v)},\boldsymbol{C}^*,\boldsymbol{F}} \sum_v \|\Phi(\boldsymbol{X}^{(v)}) - \Phi(\boldsymbol{X}^{(v)})\boldsymbol{C}^{(v)}\|_F^2 + \theta\|\boldsymbol{C}^{(v)}\|_1 \\ &+ \lambda\|\boldsymbol{C}^{(v)} - \boldsymbol{C}^*\|_1 + \gamma\, \mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}), \\ &s.t. \quad \boldsymbol{F}^T\boldsymbol{F} = \mathbf{I}, \ \ diag(\boldsymbol{C}^{(v)}) = 0, \ \ \boldsymbol{C}^{(v)T}\mathbf{1} = \mathbf{1}. \end{aligned} \tag{7}$$

It is worth noting that here $\boldsymbol{L}$ is constructed based on the affinity matrix $\boldsymbol{W}$ from $\boldsymbol{C}^*$ instead of $\boldsymbol{C}^{(v)}$ in each view. Therefore, different from Eq. (6) which learns $\boldsymbol{C}^*$ from each view, Eq. (7) also learns $\boldsymbol{C}^*$ from spectral clustering.

The proposed optimization model consists of two parts. The first part is the intra-view structure learning, which aims to learn the subspace structure in each view. The second part is the inter-view consistency learning, which measures the correlation across different views. By exploring both the view-specific property and view-consistency across multi-view data, our unified model can learn both the intra-view subspace structure and common cluster structure simultaneously. In this way, the proposed method can achieve the optimal consensus affinity matrix across multiple views that produces promising clustering results. Fig. 2 shows the main framework of the proposed method.

3. Optimization

Considering the constraints and non-differential property of the above objective for $\boldsymbol{C}^{(v)}$, we propose an updating algorithm based on Alternating Direction Method of Multipliers (ADMM) [3, 19, 28]. By introducing $\boldsymbol{A}^{(v)} = \boldsymbol{C}^{(v)} \in$

$\mathbb{R}^{n\times n}$, we reformulate the objective as:

$$\min_{\boldsymbol{A}^{(v)},\boldsymbol{C}^{(v)},\boldsymbol{C}^*,\boldsymbol{F}} \sum_v \|\Phi(\boldsymbol{X}^{(v)}) - \Phi(\boldsymbol{X}^{(v)})\boldsymbol{C}^{(v)}\|_F^2 + \theta\|\boldsymbol{A}^{(v)}\|_1$$
$$+ \lambda\|\boldsymbol{A}^{(v)} - \boldsymbol{C}^*\|_1 + \gamma\, \mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}),$$
$$s.t. \quad \boldsymbol{F}^T\boldsymbol{F} = \mathbf{I}, diag(\boldsymbol{C}^{(v)}) = 0, \boldsymbol{C}^{(v)T}\mathbf{1} = \mathbf{1}, \boldsymbol{A}^{(v)} = \boldsymbol{C}^{(v)}. \tag{8}$$

The corresponding augmented Lagrangian function is:

$$\begin{aligned}
&\mathcal{L}(\boldsymbol{A}^{(v)}, \boldsymbol{C}^{(v)}, \boldsymbol{C}^*, \boldsymbol{F}, \boldsymbol{\delta}^{(v)}, \boldsymbol{\Sigma}^{(v)}) \\
&= \sum_v \|\Phi(\boldsymbol{X}^{(v)}) - \Phi(\boldsymbol{X}^{(v)})\boldsymbol{C}^{(v)}\|_F^2 + \lambda\|\boldsymbol{A}^{(v)} - \boldsymbol{C}^*\|_1 \\
&+ \frac{\rho}{2}\|\boldsymbol{C}^{(v)T}\mathbf{1} - \mathbf{1}\|_2^2 + \langle\boldsymbol{\delta}^{(v)}, C^{(v)T}\mathbf{1} - \mathbf{1}\rangle + \theta\|\boldsymbol{A}^{(v)}\|_1 \\
&+ \frac{\rho}{2}\|\boldsymbol{C}^{(v)} - \boldsymbol{A}^{(v)} + diag(\boldsymbol{A}^{(v)})\|_F^2 + \gamma\, \mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}) \\
&+ \langle\boldsymbol{\Sigma}^{(v)}, \boldsymbol{C}^{(v)} - \boldsymbol{A}^{(v)} + diag(\boldsymbol{A}^{(v)})\rangle, \quad s.t. \quad \boldsymbol{F}^T\boldsymbol{F} = \mathbf{I},
\end{aligned} \tag{9}$$

where $\boldsymbol{\delta} \in \mathbb{R}^n, \boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ are the *Lagrangian Multipliers*.

3.1. Updating $\boldsymbol{F}$

When fixing $\boldsymbol{C}^*$, $\boldsymbol{L}$ is fixed and $\boldsymbol{F}$ can be optimized via:

$$\min_{\boldsymbol{F}} \mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}), \quad s.t. \quad \boldsymbol{F}^T\boldsymbol{F} = \mathbf{I}, \tag{10}$$

where $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$ and $\boldsymbol{W} = \frac{|\boldsymbol{C}^*| + |\boldsymbol{C}^*|^T}{2}$. Apparently the solutions are the eigenvectors corresponding to the smallest k eigenvalues of the Laplacian matrix $\boldsymbol{L}$ where k is the number of clusters [29].

3.2. Updating $\boldsymbol{A}^{(v)}$

For optimized $\boldsymbol{A}$ in each view, it is obtained via:

$$\min_{\boldsymbol{J}} \beta\|\boldsymbol{J}\|_1 + \alpha\|\boldsymbol{J} - \boldsymbol{C}^*\|_1 + \frac{1}{2}\|\boldsymbol{J} - \boldsymbol{Y}\|_F^2, \tag{11}$$

with $\boldsymbol{A} = \boldsymbol{J} - diag(\boldsymbol{J})$, where $\boldsymbol{Y} = \boldsymbol{C} + \frac{\boldsymbol{\Sigma}}{\rho}, \beta = \frac{\theta}{\rho}, \alpha = \frac{\lambda}{\rho}$. Apparently, $diag(\boldsymbol{A}) = 0$ and the above equation can be solved through element-wise optimization:

$$\min_j \beta|j| + \alpha|j - c^*| + \frac{1}{2}(j - y)^2. \tag{12}$$

Due to space limit, we provide the closed solution for $c^* \neq 0$ (otherwise, it degenerates into well known standard soft-thresholding with $j^* = sgn(y)max\{|y| - \alpha - \beta, 0\}$) by leaving the details to supplemental file:

$$j^* = \begin{cases}
y - \alpha - \beta, \ if\ c^* > 0 \wedge y \geq \alpha + \beta + c^*; \\
y + \alpha - \beta, \ if\ c^* > 0 \wedge 0 < y + \alpha - \beta < c^*; \\
y + \alpha + \beta, \ if\ c^* > 0 \wedge 0 \geq y + \alpha + \beta; \\
y - \alpha - \beta, \ if\ c^* < 0 \wedge y \geq \alpha + \beta; \\
y - \alpha + \beta, \ if\ c^* < 0 \wedge 0 > y - \alpha + \beta > c^*; \\
y + \alpha + \beta, \ if\ c^* < 0 \wedge c^* \geq y + \alpha + \beta; \\
0, else,
\end{cases} \tag{13}$$

where '$\wedge$' denotes *logical conjunction*.

3.3. Updating $\boldsymbol{C}^{(v)}$

For optimized $\boldsymbol{C}$ in each view, it can be obtained via (by skipping $diag(\boldsymbol{A})$ as it is 0 aforementioned):

$$\min_{\boldsymbol{C}} \|\Phi(\boldsymbol{X}) - \Phi(\boldsymbol{X})\boldsymbol{C}\|_F^2 + \frac{\rho}{2}\|\boldsymbol{C}^T\mathbf{1} - \mathbf{1} + \frac{\boldsymbol{\delta}}{\rho}\|_2^2 + \frac{\rho}{2}\|\boldsymbol{C} - \boldsymbol{A} + \frac{\boldsymbol{\Sigma}}{\rho}\|_F^2. \tag{14}$$

By taking the derivative and set it to be 0, we have [1]:

$$\boldsymbol{C} = (\mathcal{K} + \rho\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T)^{-1}(\mathcal{K} + \rho\mathbf{1}\mathbf{1}^T - \mathbf{1}\boldsymbol{\delta}^T + \rho\boldsymbol{A} - \boldsymbol{\Sigma}), \tag{15}$$

where $\mathcal{K} = \Phi(\boldsymbol{X})^T\Phi(\boldsymbol{X})$. One can see that with different kernel chosen, $\mathcal{K}$ is different but always are computationally efficient. For example, when polynomial kernel is applied, then $\mathcal{K}(i, j) = (\langle\boldsymbol{x}_i, \boldsymbol{x}_j\rangle + c)^d$.

3.4. Updating $\boldsymbol{C}^*$

As $\boldsymbol{C}^*$ is enriched, which is related with 2 terms, it can be optimized via:

$$\min_{\boldsymbol{C}^*} \gamma\, \mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}) + \sum_v \lambda\|\boldsymbol{A}^{(v)} - \boldsymbol{C}^*\|_1. \tag{16}$$

Before we optimize the above equation, we first introduce a useful lemma which is critical for $\boldsymbol{C}^*$:

Lemma 1. *For Laplacian matrix $\boldsymbol{L}$ and the matrix $\boldsymbol{F}$, we have [6]:*

$$\boldsymbol{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}) = \frac{1}{2}\sum_{i,j}\boldsymbol{W}(i, j)\|\boldsymbol{f}_i - \boldsymbol{f}_j\|_2^2. \tag{17}$$

We turn to optimize $\boldsymbol{C}^*$ by noticing the above equation can be written in a more compact formulation: $\mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}) = \frac{1}{2}\langle\boldsymbol{W}, \boldsymbol{Q}\rangle$, where $\boldsymbol{Q}$ is symmetric and $\boldsymbol{Q}(i, j) = \|\boldsymbol{f}_i - \boldsymbol{f}_j\|_2^2$. On the other hand, by definition $\boldsymbol{W} = \frac{|\boldsymbol{C}^*| + |\boldsymbol{C}^*|^T}{2}$, by simple algebraic operation we have:

$$\mathbf{tr}(\boldsymbol{F}^T\boldsymbol{L}\boldsymbol{F}) = \frac{1}{2}\langle|\boldsymbol{C}^*|, \boldsymbol{Q}\rangle. \tag{18}$$

Therefore, $\boldsymbol{C}^*$ can be optimized by:

$$\min \frac{\gamma}{2}\langle|\boldsymbol{C}^*|, \boldsymbol{Q}\rangle + \sum_v \lambda\|\boldsymbol{A}^{(v)} - \boldsymbol{C}^*\|_1. \tag{19}$$

Similar to $\boldsymbol{A}$, we can optimize $\boldsymbol{C}^*$ by element-wise:

$$\min \gamma q|c^*| + \sum_v 2\lambda|a^{(v)} - c^*|. \tag{20}$$

[1] We note that for Linear Kernel case, which is $\Phi(\boldsymbol{X}) = \boldsymbol{X} \in \mathbb{R}^{d\times n}$ and $\mathcal{K}(\boldsymbol{X}, \boldsymbol{X}) = \boldsymbol{X}^T\boldsymbol{X}$. When $n \gg d$, we have accelerated updating algorithm for inversion calculation. First we denote $\boldsymbol{X}^T\boldsymbol{X} + \rho\mathbf{I} + \rho\mathbf{1}\mathbf{1}^T = \boldsymbol{Z}^T\boldsymbol{Z} + \rho\mathbf{I}$, where $\boldsymbol{Z} = [\boldsymbol{X}; \sqrt{\rho}\mathbf{1}^T] \in \mathbb{R}^{(d+1)\times n}$. Then by matrix inversion lemma (aka Sherman-Morrison-Woodbury Formula), $(\boldsymbol{Z}^T\boldsymbol{Z} + \rho\mathbf{I}_n)^{-1} = \rho^{-1}\mathbf{I}_n - \rho^{-2}\boldsymbol{Z}^T(\mathbf{I}_{d+1} + \rho\boldsymbol{Z}\boldsymbol{Z}^T)^{-1}\boldsymbol{Z}$, the complexity can be reduced from $\mathcal{O}(n^3)$ to $\mathcal{O}(d^3 + dn^2)$ which is a significant improvement for $n \gg d$.

Algorithm 1 Algorithm for Enriched Robust Multi-View Kernel Subspace Clustering to solve Eq. (7).

Input: data $\boldsymbol{X}^{(v)} \in \mathbb{R}^{d_v \times n}$, number of clusters k, regularization parameters λ, γ, θ, number of iterations T.
Initialization: $\boldsymbol{C}^{(v)}, \boldsymbol{\Sigma}^{(v)}, \boldsymbol{A}^{(v)}, \boldsymbol{C}^* \in \mathbb{R}^{n \times n}, \boldsymbol{\delta}^{(v)} \in \mathbb{R}^n, \rho = 0.2, t = 1$.
while $t \leq T$ **do**
  Optimize $\boldsymbol{F}$ by solving Eq. (10);
  Optimize $\boldsymbol{A}$ in each view by solving Eq. (11);
  Optimize $\boldsymbol{C}$ in each view by solving Eq. (14);
  Optimize each $\boldsymbol{C}^*$ by solving Eq. (19);
  Update $\boldsymbol{\delta}, \boldsymbol{\Sigma}$ in each view as Eq. (22);
  Update $\rho = 1.2\rho$;
  $t = t + 1$.
end while
Output: $\boldsymbol{F}$, based on which K-means will be conducted after row normalization.

Without loss of generality, we sort $[a^{(1)}, a^{(2)}, \ldots, a^{(v)}]$ in non-decreasing order as $[a_1, a_2, \ldots, a_v]$ and none is zero (or it can be transferred into this case by simple operation). Due to space limit, we leave the derivative details to supplemental file and directly give the solution[2]:

$$c^* = \begin{cases} a_{\lceil \frac{2v\lambda - \gamma q}{4\lambda} \rceil}, \; if \; 2v\lambda > \gamma q \wedge a_{\lceil \frac{2v\lambda - \gamma q}{4\lambda} \rceil} > 0; \\ a_{\lceil \frac{2v\lambda + \gamma q}{4\lambda} \rceil}, \; if \; 2v\lambda > \gamma q \wedge a_{\lceil \frac{2v\lambda + \gamma q}{4\lambda} \rceil} < 0; \\ 0, else, \end{cases} \tag{21}$$

where $\lceil \cdot \rceil$ denotes the ceiling function.

3.5. Updating Lagrangian Multipliers in Each View

Following ADMM framework [3], we can simply update Lagrangian Multipliers by gradient ascent:

$$\begin{aligned} \boldsymbol{\delta}^{(v)} &= \boldsymbol{\delta}^{(v)} + \rho(\boldsymbol{C}^{(v)T}\mathbf{1} - \mathbf{1}), \\ \boldsymbol{\Sigma}^{(v)} &= \boldsymbol{\Sigma}^{(v)} + \rho(\boldsymbol{C}^{(v)} - \boldsymbol{A}^{(v)}). \end{aligned} \tag{22}$$

We summarize the above algorithm in Alg. 1.

4. Experiments

In this section, we will evaluate our proposed algorithm on several widely used benchmark datasets to illustrate its potential in multi-view clustering.

Six benchmark datasets are used in the experiment, including MSRC-v1, UCI Handwritten digits [10], Caltech101-7 [13], Caltech101-20 [13], ORL [40] and Yale [1]. For each dataset, multiple feature sets are available to describe the images from various aspects. The detailed information is summarized in Table 3.

Throughout the experiments, we use Matlab R2019a on a laptop with 1.4 GHz QuadCore Intel Core i5 processor. The clustering quality is measured by clustering accuracy (ACC), which is the percentage of items correctly clustered with the maximum bipartite matching [49], and normalized mutual information (NMI) [23, 30]. We repeat each experiment 10 times and report the average performance with the standard deviation.

4.1. Feature Descriptions

Features adopted in this paper is shown as the following, each feature captures quite different information from images:

1. CENTRIST [47] stands for census transform histogram, is a holistic representation of images, which can be applied to capture the structural and textural properties from images.
2. HOG [8] is based on oriented gradients, so it has great power to capture edge structures in images naturally.
3. Color moment (CMT) [53] represents the color distribution in images. The mathematical basis of this descriptor is that the color distribution can be represented efficiently by some low-order moments.
4. Local binary pattern captures the texture information from an image by computing the histogram of local binary patterns [12].
5. GIST [39] is a global image feature. Gist features represent scene information from images well. It relates to the gradient information for different parts in an image, including scales and orientations.
6. Gabor, which is extracted by a Gabor filter and can do texture analysis and object detection in images.
7. Intensity (IT). Pixel intensity is the primary information stored within pixels, represents the densities of a certain pixel.

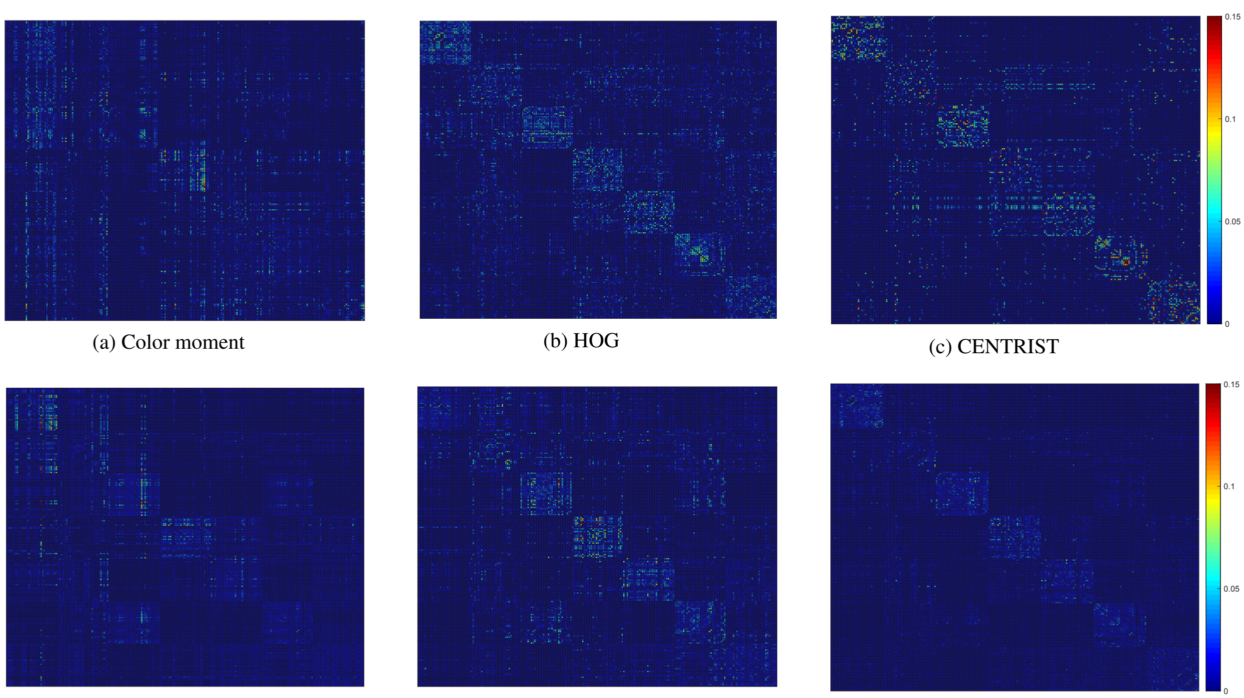
4.2. Experiment Setup

To evaluate the performance of our method, we compare our method with two subspace learning algorithms applied on single view: spectral clustering (SC) [36] and lower rank representation (LRR) [26], and five state-of-the-art multi-view methods including: pairwise co-regularized multi-view spectral clustering (P-CoReg) [22], centroid co-regularized multi-view spectral clustering (C-CoReg) [22], robust multi-view spectral clustering via low-rank and sparse decomposition (RMSC) [48], multi-view consensus

[2] It is worth noting that the optimal solution may not be unique. In practice, one can visit all $a_i (1 \leq i \leq v)$ in addition to 0, and simply set c^* as a_i or 0 which yields the lowest objective in Eq. (20).

Table 1. Subspace clustering results on various benchmark datasets

	MSRC-v1		Handwritten		Caltech101-7		Caltech101-20	
Method	ACC	NMI	ACC	NMI	ACC	NMI	ACC	NMI
HOG-SC	0.597±0.057	0.502±0.027	0.632±0.071	0.518±0.074	0.609±0.051	0.561±0.039	0.307±0.062	0.287±0.025
CEN-SC	0.618±0.038	0.556±0.017	0.711±0.027	0.641±0.041	0.657±0.032	0.587±0.031	0.501±0.018	0.536±0.019
CMT-SC	0.331±0.052	0.203±0.056	0.213±0.087	0.198±0.077	0.391±0.097	0.281±0.011	0.214±0.051	0.258±0.028
LBP-SC	0.587±0.061	0.525±0.038	0.303±0.043	0.321±0.076	0.458±0.082	0.322±0.059	0.301±0.022	0.317±0.028
GIST-SC	0.309±0.049	0.281±0.027	0.288±0.036	0.212±0.081	0.402±0.073	0.378±0.054	0.262±0.057	0.209±0.062
Gabor-SC			0.328±0.059	0.277±0.066	0.397±0.071	0.306±0.039	0.296±0.062	0.309±0.039
HOG-LRR	0.611±0.015	0.572±0.018	0.512±0.056	0.413±0.072	0.622±0.013	0.508±0.023	0.312±0.011	0.257±0.011
CEN-LRR	0.457±0.018	0.323±0.007	0.581±0.019	0.512±0.017	0.615±0.005	0.479±0.012	0.467±0.012	0.472±0.004
CMT-LRR	0.343±0.031	0.201±0.009	0.182±0.072	0.131±0.025	0.322±0.005	0.282±0.019	0.276±0.012	0.297±0.005
LBP-LRR	0.627±0.012	0.477±0.016	0.219±0.057	0.238±0.052	0.423±0.009	0.327±0.014	0.281±0.005	0.315±0.009
GIST-LRR	0.318±0.021	0.196±0.011	0.256±0.061	0.291±0.015	0.399±0.012	0.301±0.009	0.269±0.011	0.291±0.007
Gabor-LRR			0.302±0.049	0.256±0.068	0.307±0.019	0.217±0.011	0.302±0.011	0.272±0.013
P-CoReg	0.781±0.008	0.691±0.015	0.767±0.017	0.711±0.039	0.678±0.031	0.677±0.022	0.551±0.021	0.601±0.015
C-CoReg	0.767±0.006	0.678±0.031	0.758±0.015	0.704±0.041	0.658±0.052	**0.687±0.017**	0.502±0.029	0.558±0.021
RMSC	0.771±0.012	0.652±0.017	0.775±0.009	0.714±0.012	0.667±0.019	0.676±0.029	0.451±0.018	0.389±0.017
MCGC	0.682±0.012	0.601±0.009	0.768±0.011	0.721±0.021	0.649±0.035	0.533±0.071	0.425±0.033	0.392±0.019
MNMF	0.657±0.007	0.597±0.017	0.689±0.052	0.518±0.039	0.652±0.027	0.502±0.019	0.455±0.012	0.388±0.021
Our	**0.822±0.037**	**0.712±0.022**	**0.831±0.028**	**0.798±0.037**	**0.693±0.057**	0.671±0.038	**0.576±0.025**	**0.657±0.022**



(a) Color moment (b) HOG (c) CENTRIST

(d) LBP (e) GIST (f) Consensus

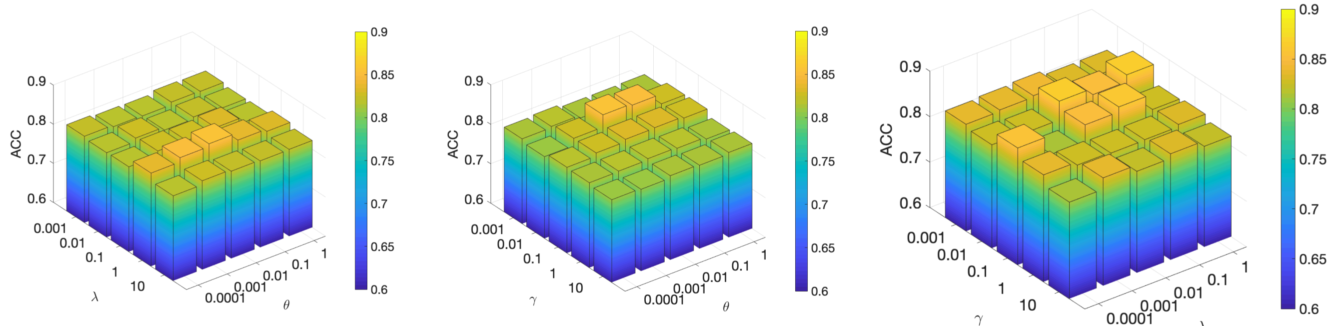
Figure 3. View-specific and consensus subspace representation matrices on MSRC-v1 dataset. (**Please zoom in to observe.**)

graph clustering (MCGC) [54], and multi-view clustering via joint nonnegative matrix factorization (MNMF) [27].

Detailed description about the methods mentioned above and the experiment process is as the following:

1. Single view with SC. We run spectral clustering on each view-specific affinity matrix independently to get clustering results based on different features.

2. Single view with LRR. We run LRR on each single feature set to get the low-rank subspace representation first, and then apply spectral clustering on each such



(a) Fixing γ to 0.01 (b) Fixing λ to 1 (c) Fixing θ to 0.01

Figure 4. Ablation study – the influence of regularization parameters on clustering accuracy.

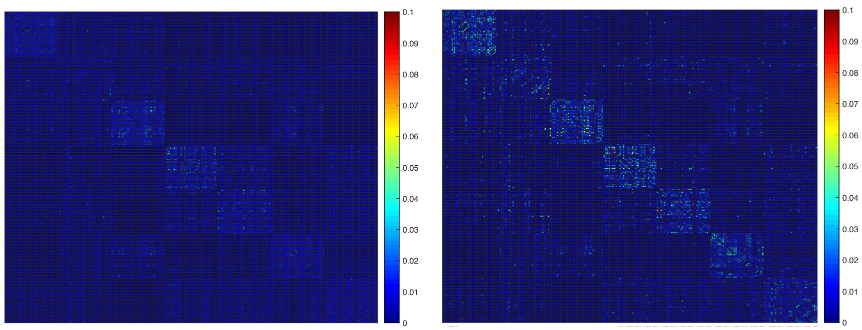


Figure 5. $\boldsymbol{C}^*$ obtained via Linear Kernel and Polynomial Kernel. (**Please zoom in to observe.**)

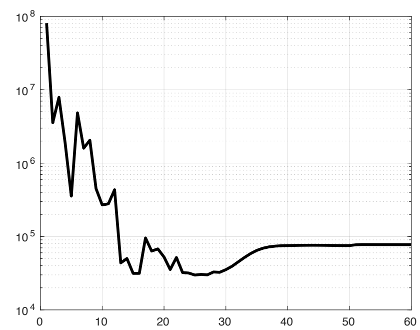


Figure 6. Objective with update.

Table 2. Subspace clustering results on benchmark datasets

	Yale		ORL	
Method	ACC	NMI	ACC	NMI
IT-SC	0.271±0.026	0.256±0.031	0.353±0.012	0.559±0.015
LBP-SC	0.617±0.036	0.637±0.017	0.713±0.009	0.798±0.011
Gabor-SC	0.653±0.012	0.643±0.011	0.647±0.011	0.813±0.006
IT-LRR	0.153±0.022	0.101±0.007	0.108±0.013	0.138±0.017
LBP-LRR	0.631±0.019	0.602±0.022	0.722±0.018	0.855±0.015
Gabor-LRR	0.625±0.023	0.637±0.018	0.721±0.021	0.823±0.012
P-CoReg	0.635±0.015	0.667±0.037	0.723±0.005	**0.868±0.012**
C-CoReg	0.655±0.009	0.637±0.011	0.731±0.007	0.852±0.008
RMSC	0.630±0.012	0.644±0.014	0.725±0.015	0.825±0.011
MCGC	0.649±0.035	0.533±0.071	0.425±0.033	0.392±0.019
MNMF	0.564±0.031	0.571±0.027	0.625±0.013	0.798±0.008
Our	**0.668±0.012**	**0.672±0.009**	**0.732±0.025**	0.863±0.019

representation we obtained.

3. P-CoReg, which makes the eigenvector matrix in standard spectral clustering method related to different views be close to each other, by employing pair-wise co-regularizers in the objective function.

4. C-CoReg, similar to P-CoReg, it regularizes the eigenvectors related to each specific view feature towards a consensus set.

5. RMSC, which is based on Markov chain method for clustering. A shared low-rank transition probability matrix is used as a crucial input to the standard Markov chain method for clustering.

6. MCGC, where a consensus graph structure is learned by minimizing disagreement between diverse views and constraining the rank of the Laplacian matrix, it's able to obtain the cluster assignment directly from the consensus graph without any post-processing steps.

7. MNMF, a joint nonnegative matrix factorization algorithm to regularize coefficient matrices learnt from different views towards a consensus, followed by K-means on the consensus matrix.

4.3. Experiment Results

Fig. 3 shows the subspace representation matrices $\boldsymbol{C}$ obtained by different feature descriptors and the final consensus $\boldsymbol{C}^*$ of the MSRC-v1 dataset. A good $\boldsymbol{C}$ should have a clear block diagonal structure, since the data in the MSRC-v1 dataset is grouped by object classes. In Fig. 3, view-specific $\boldsymbol{C}$ vary a lot from each other since they are capturing different characteristic from images. And for some of

Table 3. Datasets information and available feature sets

Dataset	# images	# classes	HOG	CENTRIST	Color Moment	LBP	GIST	Intensity	Gabor
MSRC-v1	210	7	✓	✓	✓	✓	✓		
Handwritten	2000	10	✓	✓	✓	✓	✓		✓
Caltech101-7	1474	7	✓	✓	✓	✓	✓		✓
Caltech101-20	2386	20	✓	✓	✓	✓	✓		✓
ORL	400	40				✓		✓	✓
Yale	165	15				✓		✓	✓

them, there is no obvious block structure, noise exists over the whole matrix. It is apparent that only relying on one single view-specific $\boldsymbol{C}$ has a high chance to achieve poor result. But for the consensus $\boldsymbol{C}^*$, the block diagonal structure is well-established and there is almost no noise off the diagonal blocks, which means each sample is well represented by the remaining data from the same object class, thus a great clustering result can be obtained based on it. We also utilize polynomial kernels to further improve the subspace learning performance. Fig. 5 shows the consensus $\boldsymbol{C}^*$ obtained with linear kernel and polynomial kernel for the MSRC-v1 dataset, **please zoom in to observe the details and differences**. The block diagonal structure gets more recognizable with an appropriate polynomial kernel. Thus on complex datasets where nonlinear relationships exist, polynomial kernels can have superior performance compared to simple linear kernel. And from Fig. 6 we can see that the objective of the ADMM solver is converging as the iteration increases.

Experiment results on six datasets are shown in Table 1 and 2, highest ACC and NMI for each dataset is highlighted. From the results it's not hard to conclude that certain view-specific $\boldsymbol{C}$ cannot have a satisfying clustering performance, this may be caused by the fact that images from different clusters have great similarity in the characteristic captured by that view, for example, the color moment feature doesn't work well on the Handwritten dataset. But with multi-view clustering methods, independent feature sets are combined together to construct a view-consistent $\boldsymbol{C}^*$, the clustering performance is greatly improved. What's more, in our proposed method, the consensus $\boldsymbol{C}^*$ will not deviate from the optimal solution when a view-specific $\boldsymbol{C}$ is not well learned, so our proposed method can achieve best clustering performance in most cases over the comparison methods.

In addition, we investigate the performance of our method with varying parameter settings. There are three important parameters in our method: λ, γ, and θ. We explore the effect of two parameters by fixing another. We present the results on MSRC-v1 dataset as Fig. 4 demonstrates. From the figure we see that with the setting of λ in the range of (0.001, 10), γ in the range of (0.001, 10) and θ in the range of (0.001, 1), promising performance can be

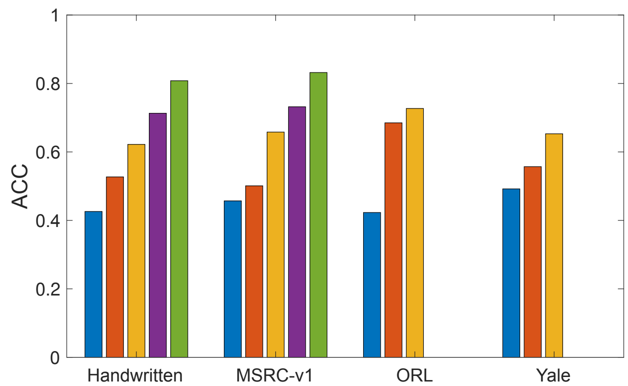


Figure 7. Accuracy comparison with increasing views, 1-5 views for Handwritten and MSRC-v1, 1-3 views for ORL and Yale.

achieved. To show the advantage of combining multi-view feature sets in subspace clustering, we run our proposed method with increasing number of views on four datasets. The result is averaged on all the possible combinations of views, and for each combination we run the experiment 5 times. Comparison is shown in Fig. 7. Apparently subspace clustering performance is improved significantly as number of views increases for all the datasets, since the data is described in a more comprehensive and extensive way.

5. Conclusion

In this paper, we propose an *Enriched Robust Multi-View Kernel Subspace Clustering* model. Different from most existing multi-view clustering methods, our method obtains an enriched consensus affinity matrix from both the learning and clustering stages. Besides, the proposed method extends linear space to kernel space to capture the nonlinear structure hidden in the multi-view data. To optimize the objective with various constraints, we propose ADMM to obtain the optimal solution where in each step a closed solution is provided. Extensive experimental results on six benchmark datasets have demonstrated the superiority of our method over several *SOTA* clustering methods.

References

[1] Yale face dataset. 5

[2] Ronen Basri and David W Jacobs. Lambertian reflectance and linear subspaces. *IEEE transactions on pattern analysis and machine intelligence*, 25(2):218–233, 2003. 1

[3] Stephen Boyd, Neal Parikh, and Eric Chu. *Distributed optimization and statistical learning via the alternating direction method of multipliers*. Now Publishers Inc, 2011. 3, 5

[4] Lodewijk Brand, Xue Yang, Kai Liu, Saad Elbeleidy, Hua Wang, Hao Zhang, and Feiping Nie. Learning robust multi-label sample specific distances for identifying hiv-1 drug resistance. *Journal of Computational Biology*, 27(4):655–672, 2020. 3

[5] Maria Brbić and Ivica Kopriva. Multi-view low-rank sparse subspace clustering. *Pattern Recognition*, 73:247–258, 2018. 1

[6] Deng Cai, Xiaofei He, Jiawei Han, and Thomas S Huang. Graph regularized nonnegative matrix factorization for data representation. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1548–1560, 2010. 4

[7] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang. Diversity-induced multi-view subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–594, 2015. 1

[8] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. Ieee, 2005. 5

[9] Inderjit S Dhillon, Yuqiang Guan, and Brian Kulis. Kernel k-means: spectral clustering and normalized cuts. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 551–556, 2004. 3

[10] Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. 5

[11] Ehsan Elhamifar and René Vidal. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE transactions on pattern analysis and machine intelligence*, 35(11):2765–2781, 2013. 1, 2

[12] Li Fei-Fei, Rob Fergus, and Pietro Perona. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. In *2004 conference on computer vision and pattern recognition workshop*, pages 178–178. IEEE, 2004. 5

[13] Li Fei-Fei, Rob Fergus, and Pietro Perona. One-shot learning of object categories. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):594–611, 2006. 5

[14] Jerome H Friedman. *The elements of statistical learning: Data mining, inference, and prediction*. springer open, 2017. 3

[15] Hongchang Gao, Feiping Nie, Xuelong Li, and Heng Huang. Multi-view subspace clustering. In *Proceedings of the IEEE international conference on computer vision*, pages 4238–4246, 2015. 2

[16] Alvina Goh and René Vidal. Segmenting motions of different types by unsupervised manifold clustering. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. IEEE, 2007. 1

[17] Amit Gruber and Yair Weiss. Multibody factorization with uncertainty and missing data using the em algorithm. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 1, pages I–I. IEEE, 2004. 1

[18] Yuhong Guo. Convex subspace representation learning from multi-view data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 27, 2013. 1

[19] Deren Han and Xiaoming Yuan. A note on the alternating direction method of multipliers. *Journal of Optimization Theory and Applications*, 155(1):227–238, 2012. 3

[20] Reinhard Heckel and Helmut Bölcskei. Subspace clustering via thresholding and spectral clustering. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 3263–3267. IEEE, 2013. 2

[21] Zhao Kang, Wangtao Zhou, Zhitong Zhao, Junming Shao, Meng Han, and Zenglin Xu. Large-scale multi-view subspace clustering in linear time. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 4412–4419, 2020. 2

[22] Abhishek Kumar, Piyush Rai, and Hal Daume. Co-regularized multi-view spectral clustering. *Advances in neural information processing systems*, 24:1413–1421, 2011. 5

[23] Andrea Lancichinetti, Santo Fortunato, and János Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New journal of physics*, 11(3):033015, 2009. 5

[24] Chun-Guang Li and Rene Vidal. Structured sparse subspace clustering: A unified optimization framework. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 277–286, 2015. 2

[25] Chun-Guang Li, Chong You, and René Vidal. Structured sparse subspace clustering: A joint affinity learning and subspace clustering framework. *IEEE Transactions on Image Processing*, 26(6):2988–3001, 2017. 1

[26] Guangcan Liu, Zhouchen Lin, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma. Robust recovery of subspace structures by low-rank representation. *IEEE transactions on pattern analysis and machine intelligence*, 35(1):171–184, 2012. 5

[27] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han. Multi-view clustering via joint nonnegative matrix factorization. In *Proceedings of the 2013 SIAM international conference on data mining*, pages 252–260. SIAM, 2013. 6

[28] Kai Liu, Lodewijk Brand, Hua Wang, and Feiping Nie. Learning robust distance metric with side information via ratio minimization of orthogonally constrained l21-norm distances. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*, 2019. 3

[29] Kai Liu, Qiuwei Li, Hua Wang, and Gongguo Tang. Spherical principal component analysis. In *Proceedings of the 2019 SIAM International Conference on Data Mining*, pages 387–395. SIAM, 2019. 4

[30] Kai Liu and Hua Wang. High-order co-clustering via strictly orthogonal and symmetric l1-norm nonnegative matrix tri-factorization. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, 2018. 5

[31] Kai Liu, Hua Wang, Feiping Nie, and Hao Zhang. Learning multi-instance enriched image representations via non-greedy ratio maximization of the l1-norm distances. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7727–7735, 2018. 3

[32] Kai Liu, Hua Wang, Shannon Risacher, Andrew Saykin, and Li Shen. Multiple incomplete views clustering via non-negative matrix factorization with its application in alzheimer's disease analysis. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 1402–1405. IEEE, 2018. 1

[33] Shirui Luo, Changqing Zhang, Wei Zhang, and Xiaochun Cao. Consistent and specific multi-view subspace clustering. In *Thirty-second AAAI conference on artificial intelligence*, 2018. 1

[34] Yi Ma, Allen Y Yang, Harm Derksen, and Robert Fossum. Estimation of subspace arrangements with applications in modeling and segmenting mixed data. *SIAM review*, 50(3):413–458, 2008. 1

[35] Sebastian Mika, Bernhard Schölkopf, Alexander J Smola, Klaus-Robert Müller, Matthias Scholz, and Gunnar Rätsch. Kernel pca and de-noising in feature spaces. In *NIPS*, volume 11, pages 536–542, 1998. 3

[36] Andrew Y Ng, Michael I Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. In *Advances in neural information processing systems*, pages 849–856, 2002. 5

[37] Vishal M Patel and René Vidal. Kernel sparse subspace clustering. In *2014 ieee international conference on image processing (icip)*, pages 2849–2853. IEEE, 2014. 1

[38] Xi Peng, Lei Zhang, and Zhang Yi. Scalable sparse subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 430–437, 2013. 1

[39] Mary C Potter and Ellen I Levy. Recognition memory for a rapid sequence of pictures. *Journal of experimental psychology*, 81(1):10, 1969. 5

[40] FS Samaria and AC Harter. In: Proceedings of the 2nd ieee workshop on applications of computer vision. december 1994, sarasota (florida). parameterisation of a stochastic model for human face identification. 5

[41] KP Soman, R Loganathan, and V Ajay. *Machine learning with SVM and other kernel methods*. PHI Learning Pvt. Ltd., 2009. 3

[42] Chang Tang, Xinzhong Zhu, Xinwang Liu, Miaomiao Li, Pichao Wang, Changqing Zhang, and Lizhe Wang. Learning a joint affinity graph for multiview subspace clustering. *IEEE Transactions on Multimedia*, 21(7):1724–1736, 2018. 1

[43] Michael E Tipping and Christopher M Bishop. Mixtures of probabilistic principal component analyzers. *Neural computation*, 11(2):443–482, 1999. 1

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

[45] René Vidal. Subspace clustering. *IEEE Signal Processing Magazine*, 28(2):52–68, 2011. 1

[46] Rene Vidal, Yi Ma, and Shankar Sastry. Generalized principal component analysis (gpca). *IEEE transactions on pattern analysis and machine intelligence*, 27(12):1945–1959, 2005. 1

[47] Jianxin Wu and Jim M Rehg. Centrist: A visual descriptor for scene categorization. *IEEE transactions on pattern analysis and machine intelligence*, 33(8):1489–1501, 2010. 5

[48] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *Proceedings of the AAAI conference on artificial intelligence*, volume 28, 2014. 1, 5

[49] Wei Xu, Xin Liu, and Yihong Gong. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, pages 267–273. ACM, 2003. 5

[50] Jingyu Yan and Marc Pollefeys. A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate and non-degenerate. In *European conference on computer vision*, pages 94–106. Springer, 2006. 1

[51] Allen Y Yang, John Wright, Yi Ma, and S Shankar Sastry. Unsupervised segmentation of natural images via lossy data compression. *Computer Vision and Image Understanding*, 110(2):212–225, 2008. 1

[52] Haoxuan Yang, Kai Liu, Hua Wang, and Feiping Nie. Learning strictly orthogonal p-order nonnegative laplacian embedding via smoothed iterative reweighted method. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence*, pages 4040–4046, 2019. 3

[53] Hui Yu, Mingjing Li, Hong-Jiang Zhang, and Jufu Feng. Color texture moments for content-based image retrieval. In *Proceedings. International Conference on Image Processing*, volume 3, pages 929–932. IEEE, 2002. 5

[54] Kun Zhan, Feiping Nie, Jing Wang, and Yi Yang. Multiview consensus graph clustering. *IEEE Transactions on Image Processing*, 28(3):1261–1270, 2018. 6

[55] Changqing Zhang, Qinghua Hu, Huazhu Fu, Pengfei Zhu, and Xiaochun Cao. Latent multi-view subspace clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4279–4287, 2017. 1

[56] Guang-Yu Zhang, Yu-Ren Zhou, Xiao-Yu He, Chang-Dong Wang, and Dong Huang. One-step kernel multi-view subspace clustering. *Knowledge-Based Systems*, 189:105126, 2020. 2

[57] Pei Zhang, Xinwang Liu, Jian Xiong, Sihang Zhou, Wentao Zhao, En Zhu, and Zhiping Cai. Consensus one-step multi-view subspace clustering. *IEEE Transactions on Knowledge and Data Engineering*, 2020. 2

[58] Teng Zhang, Arthur Szlam, and Gilad Lerman. Median k-flats for hybrid linear modeling with many outliers. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pages 234–241. IEEE, 2009. 1

[59] Qinghai Zheng, Jihua Zhu, Zhongyu Li, Shanmin Pang, Jun Wang, and Yaochen Li. Feature concatenation multi-view subspace clustering. *Neurocomputing*, 379:89–102, 2020. 2

[60] Wencheng Zhu, Jiwen Lu, and Jie Zhou. Nonlinear subspace clustering for image clustering. *Pattern Recognition Letters*, 107:131–136, 2018. 1