

Consistent and Specific Multi-View Subspace Clustering

Shirui Luo,^{1,2} Changqing Zhang,³ Wei Zhang,¹ Xiaochun Cao^{1,2*}

¹SKLOIS, Institute of Information Engineering, Chinese Academy of Sciences

²School of Cyber Security, University of Chinese Academy of Sciences

³School of Computer Science and Technology, Tianjin University, Tianjin, China
luoshirui@iie.ac.cn, zhangchangqing@tju.edu.cn, {wzhang, caoxiaochun}@iie.ac.cn

Abstract

Multi-view clustering has attracted intensive attention due to the effectiveness of exploiting multiple views of data. However, most existing multi-view clustering methods only aim to explore the consistency or enhance the diversity of different views. In this paper, we propose a novel multi-view subspace clustering method (CSMSC), where *consistency* and *specificity* are jointly exploited for subspace representation learning. We formulate the multi-view self-representation property using a shared consistent representation and a set of specific representations, which better fits the real-world datasets. Specifically, consistency models the common properties among all views, while specificity captures the inherent difference in each view. In addition, to optimize the non-convex problem, we introduce a convex relaxation and develop an alternating optimization algorithm to recover the corresponding data representations. Experimental evaluations on four benchmark datasets demonstrate that the proposed approach achieves better performance over several state-of-the-arts.

Introduction

Subspace clustering is essential to many scientific problems, e.g., representation learning (Liu and Yan 2011), motion segmentation (Rao et al. 2010) and image processing (Ma et al. 2007). Given data from multiple categories lying in a union of subspaces, clustering a dataset into categories reduces to assigning data to their respective subspaces, where each data sample is expressed by a linear combination of other samples in the same subspace.

A number of subspace clustering methods have been developed in recent years (Parsons, Haque, and Liu 2004). For instance, sparse subspace clustering (Elhamifar and Vidal 2013) finds a sparse representation from the subspaces of the data. Besides, low-rank representation (Liu et al. 2013) explores the subspace structures by low-rank constraint to recover the data. After obtaining self-representation matrix of the data, spectral clustering (Ng, Jordan, and Weiss 2002) is applied to get the final clustering result. Additionally, innovation pursuit (Rahmani and Atia 2017) proposes another subspace discovery method, which identifies each subspace based on its novelty to other subspaces.

*Corresponding author.

Copyright © 2018, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

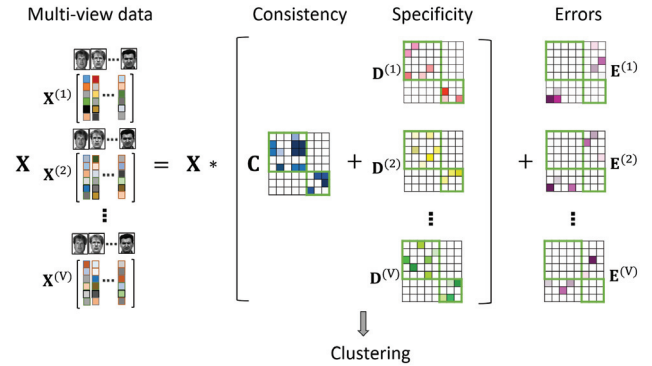


Figure 1: Illustration of our CSMSC approach. Given data samples with V views $\mathbf{X}^{(1)}, \mathbf{X}^{(2)}, \dots, \mathbf{X}^{(V)}$, our method pursues a view-consistent self-representation matrix \mathbf{C} and a set of view-specific self-representation matrices $\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \dots, \mathbf{D}^{(V)}$. The affinity matrix produced with \mathbf{C} and $\{\mathbf{D}^{(v)}\}_{v \in [V]}$ will be used as input to the spectral clustering method to generate the final clustering result.

Many real-world problems have representations with respect to multiple views (Blum and Mitchell 1998; Chaudhuri et al. 2009). For instance, an image is described by color, texture, edges and so on. A document can be simultaneously described by several different languages. Thus, methods only using single view information do not meet the real-world demand well. Based on a variety of theories, a lot of methods have been developed to extract comprehensive information from multiple views (Xu, Tao, and Xu 2013), including co-training (Blum and Mitchell 1998; Kumar, Rai, and Daumé 2011; Kumar and Daumé 2011), multiple kernel learning (Gönen and Alpaydın 2011) and subspace learning (Chaudhuri et al. 2009; Cao et al. 2015; Zhang et al. 2015; Xia et al. 2014).

However, there are several main deficiencies for most existing methods. On one hand, single view methods do not leverage as much information as multi-view methods do. On the other hand, most multi-view methods only consider the consistency of multi-view data (Kumar and Daumé 2011; Kumar, Rai, and Daumé 2011; Zhang et al. 2015), or only explore the diversity of different subspace representations

(Cao et al. 2015). Although we can simply concatenate all the features, this strategy ignores the correlation among views and may lead to a severe “curse of dimensionality”. In addition, the real world is by nature and merely considering consistency or diversity is not adequate. Thus, these methods do not explore the underlying data distribution among different views comprehensively.

Regarding the problems mentioned above, we propose a novel subspace learning method which takes consistency and diversity into consideration simultaneously. Considering that the multiple view features cover the information of the same data, thus there should exist many common information shared among different views. Specifically, for the consistent term, we assume that there is a low-rank common representation to excavate shared information among different views. Meanwhile there is some unique information in each view which could compensate for the accurate reconstruction. Therefore, with the consistent and specific terms, we can discover the correlation of multi-view data comprehensively, which better fits the real-world datasets. We refer to this method as Consistent and Specific Multi-view Subspace Clustering (CSMSC). Figure 1 gives the illustration of our method. The self-representation property is constructed using multiple types of feature, and the self-representation matrix is composed of the consistent representation shared by all views and the view-specific representation. The affinity matrix based on the consistent and specific representations is taken as input to spectral clustering (Ng, Jordan, and Weiss 2002) to generate final clustering result. Moreover, it is shown that the complexity of the proposed method scales only linearly in the number of views and the dimension of data, and cubically in the number of data samples. The main contributions of this work are:

- This paper presents a novel subspace representation learning method Consistent and Specific Multi-view Subspace Clustering (CSMSC) which simultaneously learns a view-consistent representation and a set of view-specific representations for multi-view subspace clustering.
- We introduce a non-convex optimization problem and propose a convex relaxed alternating optimization algorithm to recover the corresponding subspace representations, providing some further practical smoothing convergence.
- Extensive experiments on benchmark datasets demonstrate that our model outperforms several baseline methods and state-of-the-art methods.

Related Work

In this section, we review relevant subspace clustering methods, which pursue the subspace structure from the data and perform clustering on the learned self-representation matrix.

Many advancements of single-view subspace clustering methods have been inspired and various methods differ in the way of extracting information from feature spaces. For the reconstruction term, Frobenius norm models the noise (Candès and Plan 2010), l_0 norm characterizes random corruptions (Candès et al. 2011) and l_2 norm deals with sample-specific corruptions and outliers (Liu, Lin, and Yu 2010).

Robust shape interaction (Wei and Lin 2011) assumes data is clean in the ideal case. For the self-representation term, nuclear norm pursues low-rank recovery (Liu et al. 2013) and l_0 norm obtains sparse representation under the widest assumption on data (Yang et al. 2016). For multi-view reconstruction error term, multi-view intact space learning (Xu, Tao, and Xu 2015) employs the Cauchy estimator as error measurement to strength robustness of outliers.

For single view methods, sparse subspace clustering (Elhamifar and Vidal 2013) finds a sparse representation from subspaces of the data, while low-rank representation (Liu et al. 2013) explores subspace structures by low-rank representation to recovery the data. Additionally, multi-subspace representation (Luo et al. 2011) discovers the number of subspaces, the dimensions of each subspace, and the samples in each subspace simultaneously. Moreover, least squares regression (Lu et al. 2012) discovers the underlying subspace segmentation of data with connectedness property.

Furthermore, based on a variety of theories, a lot of multi-view clustering methods have been designed to mine intrinsic information between views. For example, co-regularized multi-view spectral clustering (Kumar, Rai, and Daumé 2011) clusters on different views with a co-regularization constraint by the hypothesis that individual views often admit the same underlying clustering of the data. Co-Train SPC (Kumar and Daumé 2011) assumes that the underlying clustering would co-train a data sample to the same cluster regardless of the view. Low-rank tensor constrained multi-view subspace clustering (Zhang et al. 2015) captures high order correlations of multi-view data. Diversity-induced multi-view subspace clustering (Cao et al. 2015) explores the complementary information among multi-view features using a diversity term. Min-Disagreement (De Sa 2005) minimizes the disagreement of different views using bipartite graph which is based on spectral clustering algorithm. Multi-view latent representation (Zhang et al. 2017) is based on the assumption that each view is originated from one underlying latent representation. Deep non-negative matrix factorization (Zhao, Ding, and Fu 2017) builds a deep structure to seek a common feature representation with more consistent knowledge to facilitate clustering.

The Proposed Approach

In this section, we present a novel multi-view learning algorithm which effectively handles subspace clustering and enables the separation of consistency and specificity properties at self-representation.

Subspace Clustering

Given data $\mathbf{X} \in \mathbb{R}^{d \times N}$ drawn from multiple subspace, where N is the number of samples and d indicates the dimension of the data. Self-representation means that each data sample is expressed by a linear combination of other samples in the same subspace. The self-representation property can be denoted as

$$\mathbf{X} = \mathbf{X}\mathbf{Z} + \mathbf{E}, \quad (1)$$

where $\mathbf{Z} \in \mathbb{R}^{N \times N}$ is the learned self-representation matrix and $\mathbf{E} \in \mathbb{R}^{d \times N}$ is the error term. The objective function of

self-representation based subspace clustering is often of the form

$$\min \lambda \Omega(\mathbf{Z}) + \Phi(\mathbf{E}), \quad (2)$$

where $\Omega(\cdot)$ and $\Phi(\cdot)$ indicate certain regularization strategies and $\lambda > 0$ is a parameter that balances these two regularizers. After achieving the self-representation matrix \mathbf{Z} , affinity matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$ is usually constructed by

$$\mathbf{S} = \frac{|\mathbf{Z}| + |\mathbf{Z}|^T}{2}, \quad (3)$$

where $|\cdot|$ is the absolute operator. Thus, the final clustering result is generated by the spectral clustering algorithm (Ng, Jordan, and Weiss 2002) based on the affinity matrix.

Formulation

Let $\mathbf{X}^{(v)} \in \mathbb{R}^{d_v \times N}$ denote the feature matrix corresponding to the v -th view, d_v indicates the dimension of the data in v -th view and $v \in [V]$, where $[V]$ stands for $\{1, 2, \dots, V\}$. Define $\mathbf{Z}^{(v)} \in \mathbb{R}^{N \times N}$ the learned subspace representation in each view and $\mathbf{E}^{(v)} \in \mathbb{R}^{d_v \times N}$ the error term at self-representation. If we just consider the consistency term of all views, the multi-view self-representation formula can be written as

$$\mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z} + \mathbf{E}^{(v)}. \quad (4)$$

Besides the consistency term comprised in Eq. (4), we argue that $\mathbf{E}^{(v)}$ in Eq. (4) should not simply be considered as residual, it indeed contains the inherent difference in each view. Furthermore, though the common representation which remains unchanged among all views is considered in Eq. (4), the unique part in each view is not considered, and it is obviously not enough or accurate to merely leverage this small part for data modeling.

We consider

$$\mathbf{X}^{(v)} = \mathbf{X}^{(v)} (\mathbf{C} + \mathbf{D}^{(v)}) + \mathbf{E}^{(v)}, v \in [V], \quad (5)$$

where $\mathbf{C}, \mathbf{D}^{(v)} \in \mathbb{R}^{N \times N}$ are learned consistent and specific self-representation matrices with data under different views, respectively. We decompose the ordinary multi-view representation matrix $\mathbf{Z}^{(v)}$ in $\mathbf{X}^{(v)} = \mathbf{X}^{(v)} \mathbf{Z}^{(v)} + \mathbf{E}^{(v)}$ into the sum of a shared representation which encodes the part remaining unchanged on different views and a view-specific representation which corresponds to the unique part of v -th view¹.

Due to these constraints, we impose some regularization to penalize the consistent matrix and the view-specific matrices. For the consistent term, we choose nuclear norm to guarantee the low rank property in order to excavate more shared information among different views. Also, nuclear norm prevents the trivial solution. Besides, we apply l_2 norm to the specific term to ensure connectedness property, thus the representation matrices are generally dense, which alleviates the connectivity issue (Lu et al. 2012). Consequently,

¹Section **Ablation Study on Consistency and Specificity** numerically verifies that merely considering or not considering the consistent term drops many important information and has poor effect.

we can pursue the most compatible structure on all views. Thus, the regular terms are as the form of

$$\Omega(\mathbf{C}, \mathbf{D}^{(1)}, \dots, \mathbf{D}^{(V)}) = \lambda_C \|\mathbf{C}\|_* + \lambda_D \sum_{v=1}^V \|\mathbf{D}^{(v)}\|_2^2, \quad (6)$$

where $\|\cdot\|_*$ denotes nuclear norm, $\|\cdot\|_2$ is l_2 norm and $\lambda_C, \lambda_D \in (0, 1]$ are trade-off parameters.

Moreover, real-world datasets are often with many noisy information, thus we introduce an error term to deal with noisy data. Analogously to (Liu et al. 2013; Zhang et al. 2015), we formulate the error term as the form of

$$\Phi(\mathbf{E}^{(1)}, \dots, \mathbf{E}^{(V)}) = \sum_{v=1}^V \|\mathbf{E}^{(v)}\|_{2,1}, \quad (7)$$

where $\|\cdot\|_{2,1}$ denotes $l_{2,1}$ norm, which encourages the columns of $\mathbf{E}^{(V)}$ to be zero. Consequently, combining all the conditions discussed above together leads to the objective function of the proposed method:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{D}^{(v)}, \mathbf{E}^{(v)}} \lambda_C \|\mathbf{C}\|_* + \lambda_D \sum_{v=1}^V \|\mathbf{D}^{(v)}\|_2^2 + \sum_{v=1}^V \|\mathbf{E}^{(v)}\|_{2,1} \\ \text{s.t. } \mathbf{X}^{(v)} = \mathbf{X}^{(v)} (\mathbf{C} + \mathbf{D}^{(v)}) + \mathbf{E}^{(v)}, v \in [V]. \end{aligned} \quad (8)$$

Therefore, by solving Eq. (8) with multi-view features, a consistent and a series of specific representations can be extracted and the data is represented more naturally.

With the learned representations, we construct the affinity matrix with respect to the consistency and the specificities

$$\mathbf{S} = \frac{|\mathbf{C}| + |\mathbf{C}|^T}{2} + \frac{1}{V} \sum_{v=1}^V \frac{|\mathbf{D}^{(v)}| + |\mathbf{D}^{(v)}|^T}{2}, \quad (9)$$

and subsequently apply the spectral clustering algorithm to the affinity matrix to pursue the final clustering results.

Optimization

Our objective function Eq. (8) simultaneously learns the consistent and specific representations from multiple views. However, directly finding the optimal solution of the problem (8) is extremely difficult. Thus, we leverage a convex relaxation and develop an alternating optimization algorithm to jointly recover the corresponding data representations. Firstly, we introduce variables $\mathbf{K}, \mathbf{W}^{(v)} \in \mathbb{R}^{N \times N}$ as surrogates of \mathbf{C} in nuclear norm and $\mathbf{E}^{(v)}$ in $l_{2,1}$ norm:

$$\begin{aligned} \min_{\mathbf{C}, \mathbf{D}^{(v)}, \mathbf{E}^{(v)}, \mathbf{W}^{(v)}, \mathbf{K}} \lambda_C \|\mathbf{K}\|_* + \lambda_D \sum_{v=1}^V \|\mathbf{D}^{(v)}\|_2^2 + \sum_{v=1}^V \|\mathbf{W}^{(v)}\|_{2,1} \\ \text{s.t. } \mathbf{X}^{(v)} = \mathbf{X}^{(v)} (\mathbf{C} + \mathbf{D}^{(v)}) + \mathbf{E}^{(v)}, \\ \mathbf{C} = \mathbf{K}, \mathbf{E}^{(v)} = \mathbf{W}^{(v)}, v \in [V]. \end{aligned} \quad (10)$$

This problem can be solved by the Augmented Lagrange Multiplier (ALM) (Lin, Chen, and Ma 2010) method, which

minimizes the augmented Lagrange function of the form:

$$\begin{aligned}
& \lambda_C \|\mathbf{K}\|_* + \lambda_D \sum_{v=1}^V \|\mathbf{D}^{(v)}\|_2^2 + \sum_{v=1}^V \|\mathbf{W}^{(v)}\|_{2,1} + \\
& \sum_{v=1}^V \langle \mathbf{Y}_1^{(v)}, \mathbf{X}^{(v)} - \mathbf{X}^{(v)}(\mathbf{C} + \mathbf{D}^{(v)}) - \mathbf{E}^{(v)} \rangle + \\
& \frac{\mu}{2} \sum_{v=1}^V \|\mathbf{X}^{(v)} - \mathbf{X}^{(v)}(\mathbf{C} + \mathbf{D}^{(v)}) - \mathbf{E}^{(v)}\|_F^2 + \\
& \langle \mathbf{Y}_2, \mathbf{C} - \mathbf{K} \rangle + \frac{\mu}{2} \|\mathbf{C} - \mathbf{K}\|_F^2 + \\
& \sum_{v=1}^V \langle \mathbf{Y}_3^{(v)}, \mathbf{E}^{(v)} - \mathbf{W}^{(v)} \rangle + \frac{\mu}{2} \sum_{v=1}^V \|\mathbf{E}^{(v)} - \mathbf{W}^{(v)}\|_F^2,
\end{aligned} \tag{11}$$

where $\langle \cdot, \cdot \rangle$ is the standard Euclidean inner product of two matrices, $\|\cdot\|_F$ denotes Forbenius norm, $\{\mathbf{Y}_1^{(v)}, \mathbf{Y}_3^{(v)}\}_{v \in [V]}$, \mathbf{Y}_2 are Lagrange multipliers and $\mu > 0$ is a penalty parameter. To tackle this issue, we divide the above unconstrained problem into six sub-problems and alternatively optimize them by fixing the other variables. The procedure depicted as in Algorithm 1 solves problem (8). Note that Step 4 has closed-form solution thanks to the singular value shrinkage operator (Cai, Candès, and Shen 2010) and Step 8 can be efficiently solved by Lemma 4.1 in (Liu et al. 2013). Detailed procedure of optimization can be found in the supplementary material.

Analysis

Convergence analysis. The convergence of exact ALM method has been well studied in (Bertsekas 2014) when the objective function is smooth. The convergence of inexact ALM method with at most two pending matrices has also been proved (Lin, Chen, and Ma 2010). Unfortunately, ensuring the convergence of inexact ALM method with three or more pending matrices is still difficult. Since our method has $3V + 2$ pending matrices and the objective function Eq. (11) is not smooth, as a result, the convergence is hard to prove theoretically. Owing to the gap generated in each iteration is monotonically decreasing, as shown in Section **Performance Evaluation**, the validity of the proposed method could be guaranteed by the convexity of the Lagrange function to some extent (Eckstein and Bertsekas 1992). Thus it could be well expected that the proposed method has good convergence properties. It is worth mentioning that μ should be upper bounded by Step 12 in Algorithm 1 in order to guarantee the convergence. This is derived from traditional theory of the alternating direction method (Lin, Chen, and Ma 2010).

Complexity analysis. Our algorithm consists of $3V + 2$ sub-problems and the complexities of the subproblems are analyzed as follows. The complexity of updating \mathbf{C} , $\mathbf{D}^{(v)}$ and \mathbf{K} derived from matrix inversion and nuclear norm are $O(N^3)$, where N is the number of data samples. The complexity of updating $\mathbf{E}^{(v)}$ and $\mathbf{W}^{(v)}$ is $O(dN)$, where d is the maximal dimension of data towards all views. Overall, the

Algorithm 1 CSMSC: Consistent and Specific Multi-view Subspace Clustering

Input: The data set \mathbf{X} with multiple types of feature $\{\mathbf{X}^{(v)}\}_{v \in [V]}$, the parameters λ_C and λ_D .

- 1: Initialize coefficient matrices $\mathbf{C} = \mathbf{K} = \mathbf{Y}_1^{(v)} = \mathbf{Y}_2 = \mathbf{Y}_3^{(v)} = \mathbf{0}$, $v \in [V]$, generate $\{\mathbf{D}^{(v)}\}_{v \in [V]}$ as in (Cao et al. 2015), set parameters $\mu = 10^{-6}$, $\mu_{max} = 10^6$, $\rho = 1.1$, maximum iterations $M = 60$ and stopping threshold $\varepsilon = 10^{-6}$.
 - 2: **for** $t \in [M]$ **do**
 - 3: Fix the others and update \mathbf{C} using
$$\mathbf{C} = \left(\sum_{v=1}^V (\mathbf{X}^{(v)})^T \mathbf{X}^{(v)} + \mathbf{I} \right)^{-1} \times \left(\sum_{v=1}^V (\mathbf{X}^{(v)})^T (\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{D}^{(v)} - \mathbf{E}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu}) + \mathbf{K} - \frac{\mathbf{Y}_2}{\mu} \right);$$
 - 4: Fix the others and update \mathbf{K} using
$$\mathbf{K} = \arg \min \frac{\lambda_C}{\mu} \|\mathbf{K}\|_* + \frac{1}{2} \|\mathbf{K} - (\mathbf{C} + \frac{\mathbf{Y}_2}{\mu})\|_F^2;$$
 - 5: **for each** $v \in [V]$ **do**
 - 6: Fix the others and update $\mathbf{D}^{(v)}$ using
$$\mathbf{D}^{(v)} = (\mu (\mathbf{X}^{(v)})^T \mathbf{X}^{(v)} + 2\lambda_D \mathbf{I})^{-1} \times \mu (\mathbf{X}^{(v)})^T (\mathbf{X}^{(v)} - \mathbf{X}^{(v)} \mathbf{C} - \mathbf{E}^{(v)} + \frac{\mathbf{Y}_1^{(v)}}{\mu});$$
 - 7: Fix the others and update $\mathbf{E}^{(v)}$ using
$$\mathbf{E}^{(v)} = \frac{1}{2} (\mathbf{X}^{(v)} - \mathbf{X}^{(v)} (\mathbf{C} + \mathbf{D}^{(v)}) + \mathbf{W}^{(v)} + \frac{\mathbf{Y}_1^{(v)} - \mathbf{Y}_3^{(v)}}{\mu});$$
 - 8: Fix the others and update $\mathbf{W}^{(v)}$ using
$$\mathbf{W}^{(v)} = \arg \min \frac{1}{\mu} \|\mathbf{W}^{(v)}\|_{2,1} + \frac{1}{2} \|\mathbf{W}^{(v)} - (\mathbf{E}^{(v)} + \frac{\mathbf{Y}_3^{(v)}}{\mu})\|_F^2;$$
 - 9: Update the multipliers
$$\mathbf{Y}_1^{(v)}_{(t+1)} = \mathbf{Y}_1^{(v)}_{(t)} + \mu (\mathbf{X}^{(v)} - \mathbf{X}^{(v)} (\mathbf{C} + \mathbf{D}^{(v)}) - \mathbf{E}^{(v)});$$

$$\mathbf{Y}_3^{(v)}_{(t+1)} = \mathbf{Y}_3^{(v)}_{(t)} + \mu (\mathbf{E}^{(v)} - \mathbf{W}^{(v)});$$
 - 10: **end for**
 - 11: Update the multiplier
$$\mathbf{Y}_2_{(t+1)} = \mathbf{Y}_2_{(t)} + \mu (\mathbf{C} - \mathbf{K});$$
 - 12: Update the parameter μ by $\mu = \min(\rho\mu, \mu_{max})$;
 - 13: **if** converges **then**
 - 14: break;
 - 15: **end if**
 - 16: Construct the affinity matrix \mathbf{S} using Eq. (9).
 - 17: Apply the spectral clustering algorithm (Ng, Jordan, and Weiss 2002) to \mathbf{S} .
 - 18: **end for**
- Output:** The cluster label of \mathbf{X} .
-

complexity is $O((V + 2)N^3 + 2VdN)$ for each iteration, where V is the number of views. Considering the number of iterations, the complexity of Algorithm 1 is

$$O(MVN(N^2 + d)),$$

where M is the number of iterations.

Table 1: Results on four datasets (mean \pm standard deviation). Higher value indicates better performance.

| Datasets | Methods | NMI | ACC | AR | F-score | Precision | Recall |
|--------------|--|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Yale | LRR _{best} (Liu et al. 2013) | 0.709 \pm 0.011 | 0.697 \pm 0.000 | 0.515 \pm 0.004 | 0.547 \pm 0.007 | 0.529 \pm 0.003 | 0.567 \pm 0.004 |
| | LRR _{con} (Liu et al. 2013) | 0.702 \pm 0.002 | 0.667 \pm 0.007 | 0.456 \pm 0.002 | 0.492 \pm 0.002 | 0.538 \pm 0.002 | 0.453 \pm 0.005 |
| | Co-Reg SPC(Kumar, Rai, and Daumé 2011) | 0.648 \pm 0.002 | 0.564 \pm 0.000 | 0.436 \pm 0.002 | 0.466 \pm 0.000 | 0.455 \pm 0.004 | 0.491 \pm 0.003 |
| | RMSC(Xia et al. 2014) | 0.684 \pm 0.033 | 0.642 \pm 0.036 | 0.485 \pm 0.046 | 0.517 \pm 0.043 | 0.500 \pm 0.043 | 0.535 \pm 0.044 |
| | LT-MSC(Zhang et al. 2015) | 0.765 \pm 0.008 | 0.741 \pm 0.002 | 0.570 \pm 0.004 | 0.598 \pm 0.006 | 0.569 \pm 0.004 | 0.629 \pm 0.005 |
| | DiMSC(Cao et al. 2015) | 0.727 \pm 0.010 | 0.709 \pm 0.003 | 0.535 \pm 0.001 | 0.564 \pm 0.002 | 0.543 \pm 0.001 | 0.586 \pm 0.003 |
| | CSMSC | 0.784 \pm 0.001 | 0.752 \pm 0.001 | 0.615 \pm 0.005 | 0.640 \pm 0.004 | 0.673 \pm 0.002 | 0.610 \pm 0.006 |
| Notting-Hill | LRR _{best} (Liu et al. 2013) | 0.579 \pm 0.005 | 0.794 \pm 0.001 | 0.558 \pm 0.012 | 0.653 \pm 0.001 | 0.672 \pm 0.004 | 0.636 \pm 0.004 |
| | LRR _{con} (Liu et al. 2013) | 0.602 \pm 0.003 | 0.685 \pm 0.002 | 0.485 \pm 0.008 | 0.601 \pm 0.008 | 0.625 \pm 0.002 | 0.579 \pm 0.003 |
| | Co-Reg SPC(Kumar, Rai, and Daumé 2011) | 0.853 \pm 0.003 | 0.715 \pm 0.000 | 0.602 \pm 0.004 | 0.615 \pm 0.000 | 0.567 \pm 0.004 | 0.666 \pm 0.004 |
| | RMSC(Xia et al. 2014) | 0.585 \pm 0.004 | 0.807 \pm 0.035 | 0.496 \pm 0.013 | 0.603 \pm 0.005 | 0.621 \pm 0.029 | 0.586 \pm 0.031 |
| | LT-MSC(Zhang et al. 2015) | 0.779 \pm 0.007 | 0.868 \pm 0.003 | 0.777 \pm 0.002 | 0.825 \pm 0.007 | 0.830 \pm 0.006 | 0.814 \pm 0.004 |
| | DiMSC(Cao et al. 2015) | 0.799 \pm 0.001 | 0.843 \pm 0.021 | 0.787 \pm 0.001 | 0.834 \pm 0.001 | 0.822 \pm 0.005 | 0.836 \pm 0.009 |
| | CSMSC | 0.908 \pm 0.001 | 0.965 \pm 0.001 | 0.923 \pm 0.000 | 0.940 \pm 0.001 | 0.931 \pm 0.001 | 0.949 \pm 0.000 |
| ORL | LRR _{best} (Liu et al. 2013) | 0.895 \pm 0.006 | 0.773 \pm 0.003 | 0.724 \pm 0.020 | 0.731 \pm 0.004 | 0.701 \pm 0.001 | 0.754 \pm 0.002 |
| | LRR _{con} (Liu et al. 2013) | 0.830 \pm 0.005 | 0.695 \pm 0.000 | 0.563 \pm 0.000 | 0.573 \pm 0.007 | 0.623 \pm 0.009 | 0.530 \pm 0.009 |
| | Co-Reg SPC(Kumar, Rai, and Daumé 2011) | 0.853 \pm 0.003 | 0.715 \pm 0.000 | 0.602 \pm 0.004 | 0.615 \pm 0.000 | 0.567 \pm 0.004 | 0.666 \pm 0.004 |
| | RMSC(Xia et al. 2014) | 0.872 \pm 0.012 | 0.723 \pm 0.025 | 0.645 \pm 0.029 | 0.654 \pm 0.028 | 0.607 \pm 0.033 | 0.709 \pm 0.027 |
| | LT-MSC(Zhang et al. 2015) | 0.909 \pm 0.009 | 0.759 \pm 0.028 | 0.709 \pm 0.024 | 0.717 \pm 0.023 | 0.654 \pm 0.026 | 0.793 \pm 0.023 |
| | DiMSC(Cao et al. 2015) | 0.940 \pm 0.003 | 0.838 \pm 0.001 | 0.802 \pm 0.000 | 0.807 \pm 0.003 | 0.764 \pm 0.012 | 0.856 \pm 0.004 |
| | CSMSC | 0.942 \pm 0.005 | 0.868 \pm 0.012 | 0.827 \pm 0.002 | 0.831 \pm 0.001 | 0.860 \pm 0.002 | 0.804 \pm 0.003 |
| BBCSport | LRR _{best} (Liu et al. 2013) | 0.690 \pm 0.019 | 0.832 \pm 0.026 | 0.667 \pm 0.008 | 0.774 \pm 0.023 | 0.726 \pm 0.009 | 0.762 \pm 0.004 |
| | LRR _{con} (Liu et al. 2013) | 0.733 \pm 0.005 | 0.873 \pm 0.001 | 0.756 \pm 0.001 | 0.813 \pm 0.005 | 0.803 \pm 0.004 | 0.823 \pm 0.009 |
| | Co-Reg SPC(Kumar, Rai, and Daumé 2011) | 0.718 \pm 0.003 | 0.564 \pm 0.000 | 0.696 \pm 0.001 | 0.766 \pm 0.002 | 0.786 \pm 0.008 | 0.748 \pm 0.012 |
| | RMSC(Xia et al. 2014) | 0.608 \pm 0.007 | 0.737 \pm 0.003 | 0.723 \pm 0.025 | 0.655 \pm 0.002 | 0.702 \pm 0.001 | 0.608 \pm 0.005 |
| | LT-MSC(Zhang et al. 2015) | 0.556 \pm 0.000 | 0.717 \pm 0.000 | 0.496 \pm 0.000 | 0.634 \pm 0.000 | 0.552 \pm 0.000 | 0.743 \pm 0.000 |
| | DiMSC(Cao et al. 2015) | 0.885 \pm 0.002 | 0.960 \pm 0.000 | 0.920 \pm 0.000 | 0.929 \pm 0.001 | 0.914 \pm 0.000 | 0.930 \pm 0.001 |
| | CSMSC | 0.881 \pm 0.000 | 0.963 \pm 0.001 | 0.909 \pm 0.001 | 0.931 \pm 0.000 | 0.928 \pm 0.001 | 0.933 \pm 0.000 |

Experimental Results

In this section, we extensively evaluate the clustering property of the proposed method on three widely used face datasets (with three views) and one well known document dataset (with two views). The performance of CSMSC is compared with six state-of-the-art subspace clustering methods and two baseline methods in term of six evaluation metrics.

Four benchmark datasets are adopted in our evaluation. **Yale** is a widely used face dataset which contains 165 gray-scale images, 15 individuals with 11 images of each category. Variations of the data include center light, with glasses, happy, left light, without glasses, normal, right light, sad, sleepy, surprised and wink. **Notting-Hill** video face dataset (Zhang et al. 2009) is derived from the movie Notting-Hill. The faces of five main casts are collected, including 4,660 faces in 76 tracks. We randomly sample 110 images of each cast. **ORL** face dataset contains 400 images of 40 distinct subjects. For each category, images were taken at different times, lights, facial expressions (open / closed eyes, smiling or not) and facial details (with glasses / without glasses). **BBCSport** (Xia et al. 2014) contains 544 documents from the BBC Sport website of sports news articles in five topical areas in 2004-2005.

Following (Zhang et al. 2015), for face datasets, we resize the images to 48×48 and extract three types of features: View1 intensity (4,096 dimensions), View2 LBP (Ojala, Pietikainen, and Maenpaa 2002) (3,304 dimensions) and View3 Gabor (Lades et al. 1993) (6,750 dimensions). The standard LBP feature is extracted from 72×80 loosely cropped images with a histogram size of 59 over 910 pixel patches. The Gabor feature is extracted with one scale $\lambda = 4$

at four orientations $\theta = \{0^\circ, 45^\circ, 90^\circ, 135^\circ\}$ with a loose face crop at the resolution of 25×30 pixels. BBCSport dataset only has two views, View1: 3,183 dimensions and View2: 3,203 dimensions, respectively. All descriptors except the intensity are scaled to have unit norm.

Compared Methods and Evaluation Metrics

To evaluate the performance of CSMSC, we compare our method with single-view approach LRR (Liu et al. 2013) and a series of state-of-the-art multi-view methods. **LRR_{best}** (Liu et al. 2013) is a single view algorithm. It seeks the lowest rank representation of data to describe samples as linear combinations with the most informative view. **LRR_{con}** (Liu et al. 2013) firstly concatenates the data from different views and then feeds into the traditional subspace clustering method LRR. **Co-Reg SPC** (Kumar, Rai, and Daumé 2011) is a pairwise multi-view spectral clustering method, which co-regularizes the clustering hypotheses for clustering consistent representations across views. **RMSC** (Xia et al. 2014) stands for robust multi-view spectral clustering method. It recovers a shared low-rank transition probability matrix and uses a Markov chain to cluster. **LT-MSC** (Zhang et al. 2015) represents low-rank tensor constrained multi-view subspace clustering. It captures high order underlying correlations in multi-view data. **DiMSC** (Cao et al. 2015) is on behalf of diversity-induced multi-view subspace clustering which explores complementary information across different views by enforcing the diversity.

We also compare with two baselines which are special cases of the proposed CSMSC method. **CMSC** represents consistent multi-view subspace clustering which uses the consistent part as self-representation matrix **Z** and the view-

Table 2: Ablation study of consistency and specificity on four datasets (mean \pm standard deviation).

| Datasets | Methods | NMI | ACC | AR | F-score | Precision | Recall |
|--------------|---------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| Yale | CMSC | 0.628 \pm 0.000 | 0.485 \pm 0.001 | 0.340 \pm 0.001 | 0.387 \pm 0.001 | 0.462 \pm 0.000 | 0.334 \pm 0.001 |
| | SMSC | 0.677 \pm 0.001 | 0.636 \pm 0.003 | 0.482 \pm 0.000 | 0.515 \pm 0.000 | 0.540 \pm 0.001 | 0.491 \pm 0.000 |
| | CSMSC | 0.784 \pm 0.001 | 0.752 \pm 0.001 | 0.615 \pm 0.005 | 0.640 \pm 0.004 | 0.673 \pm 0.002 | 0.610 \pm 0.006 |
| Notting-Hill | CMSC | 0.209 \pm 0.000 | 0.465 \pm 0.000 | 0.157 \pm 0.000 | 0.377 \pm 0.000 | 0.470 \pm 0.000 | 0.314 \pm 0.000 |
| | SMSC | 0.776 \pm 0.002 | 0.889 \pm 0.001 | 0.770 \pm 0.001 | 0.819 \pm 0.001 | 0.812 \pm 0.000 | 0.828 \pm 0.001 |
| | CSMSC | 0.908 \pm 0.001 | 0.965 \pm 0.001 | 0.923 \pm 0.000 | 0.940 \pm 0.001 | 0.931 \pm 0.001 | 0.949 \pm 0.000 |
| ORL | CMSC | 0.834 \pm 0.000 | 0.667 \pm 0.001 | 0.542 \pm 0.001 | 0.554 \pm 0.001 | 0.643 \pm 0.001 | 0.487 \pm 0.002 |
| | SMSC | 0.865 \pm 0.001 | 0.720 \pm 0.000 | 0.625 \pm 0.002 | 0.634 \pm 0.001 | 0.705 \pm 0.000 | 0.576 \pm 0.001 |
| | CSMSC | 0.942 \pm 0.005 | 0.868 \pm 0.012 | 0.827 \pm 0.002 | 0.831 \pm 0.001 | 0.860 \pm 0.002 | 0.804 \pm 0.003 |
| BBCSport | CMSC | 0.379 \pm 0.000 | 0.550 \pm 0.000 | 0.285 \pm 0.000 | 0.517 \pm 0.000 | 0.812 \pm 0.000 | 0.379 \pm 0.000 |
| | SMSC | — | — | — | — | — | — |
| | CSMSC | 0.881 \pm 0.000 | 0.963 \pm 0.001 | 0.909 \pm 0.001 | 0.931 \pm 0.000 | 0.928 \pm 0.001 | 0.933 \pm 0.000 |

Table 3: Results on Yale with different views (mean \pm standard deviation).

| Views | NMI | ACC | AR | F-score | Precision | Recall |
|-------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|-------------------------------------|
| View1 | 0.684 \pm 0.003 | 0.642 \pm 0.030 | 0.470 \pm 0.001 | 0.503 \pm 0.010 | 0.527 \pm 0.003 | 0.482 \pm 0.004 |
| View2 | 0.487 \pm 0.001 | 0.412 \pm 0.004 | 0.187 \pm 0.004 | 0.239 \pm 0.007 | 0.251 \pm 0.007 | 0.228 \pm 0.002 |
| View3 | 0.512 \pm 0.001 | 0.406 \pm 0.008 | 0.227 \pm 0.001 | 0.277 \pm 0.005 | 0.298 \pm 0.002 | 0.259 \pm 0.004 |
| View1+View2 | 0.757 \pm 0.000 | 0.727 \pm 0.001 | 0.584 \pm 0.005 | 0.610 \pm 0.004 | 0.625 \pm 0.005 | 0.596 \pm 0.005 |
| View2+View3 | 0.738 \pm 0.000 | 0.701 \pm 0.008 | 0.529 \pm 0.006 | 0.559 \pm 0.009 | 0.597 \pm 0.002 | 0.526 \pm 0.000 |
| View1+View3 | 0.719 \pm 0.006 | 0.697 \pm 0.001 | 0.532 \pm 0.008 | 0.562 \pm 0.001 | 0.579 \pm 0.007 | 0.545 \pm 0.003 |
| View1+View2+View3 | 0.775 \pm 0.001 | 0.752 \pm 0.001 | 0.618 \pm 0.000 | 0.642 \pm 0.003 | 0.661 \pm 0.002 | 0.624 \pm 0.000 |

specific term $\mathbf{D}^{(v)}$ is removed as in Eq. (4). **SMSC** stands for specific multi-view subspace clustering which runs LRR on each view as Eq. (1), and then sums these self-representation matrices $\mathbf{Z}^{(v)}$ together as matrix \mathbf{Z} . The affinity matrices of two baselines are constructed following Eq. (3).

Benchmarking clustering is generally difficult. Following the convention (Cao et al. 2015; Christopher, Prabhakar, and Hinrich ; Hubert and Arabie 1985; Zhang et al. 2015), we evaluate the above approaches using six different metrics: normalized mutual information (NMI), accuracy (ACC), adjusted rand index (AR), F-score, precision and recall. Higher value indicates better clustering performance for all metrics.

Experimental Setup

In our experiments, we tune the parameter λ_C and λ_D in the range of (0,1] and report the best performing results. We run each experiment 30 times and report the average score and standard deviation. For all the compared methods, we have tuned the parameters to the best.

Performance Evaluation

The experimental results on the four datasets are presented in Table 1. As shown, multi-view methods mostly outperform single view methods, which demonstrates the necessity of extracting multiple view information for clustering. Our proposed method significantly outperforms other methods on Yale, Notting-Hill and ORL datasets, and shows very competitive performance on BBCSport dataset. On Notting-Hill, our method gains large improvements around 10.9%, 12.2%, 13.6%, 10.6%, 10.9% and 11.3% over the second-best method in terms of NMI, ACC, AR, F-score, Precision and Recall, respectively. Another attraction is that our

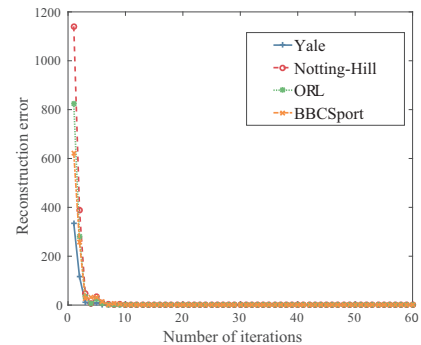


Figure 2: Number of iterations till convergence on each dataset.

method performs substantially better than the baseline of LRR_{con} , which simply concatenates all features together and then performs LRR on the new feature. This demonstrates the effectiveness of the proposed multi-view subspace clustering technique. Note that LRR_{con} performs even worse than LRR_{best} on Yale and ORL datasets, which numerically proves that simply combining all the features is not a good choice. We further present visualized affinity matrices produced by LRR_{con} and CSMSC on four datasets in supplementary material.

Moreover, we investigate the stopping criterion of our method. Figure 2 plots reconstruction error (vertical axis) versus number of iterations (horizontal axis) on all four datasets. As indicated, the error drops quickly in the beginning and then remains stable. On most datasets, CSMSC

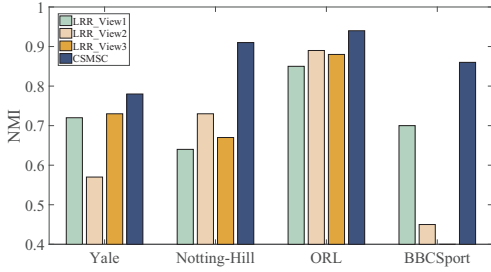


Figure 3: Performance comparison between LRR with each view and our method versus NMI.

converges within only a small number of iterations². This result further empirically confirms our convergence analysis in Section **Analysis**.

Ablation Study on Consistency and Specificity

We further analyse the improvement of the proposed CSMSC by comparing to CMSC which merely considers consistency, and SMSC which does not consider the consistent part. The experimental results on the four datasets are presented in Table 2. Note that features of BBCSport dataset are too sparse to run SVD, thus we can not perform LRR on BBCSport. It is observed that our CSMSC substantially outperforms CMSC on four datasets and thus numerically indicates that considering the view-specific terms of each view has merits. Besides, our proposed method performs better than SMSC, consequently declares that isolating the consistent representation from representations of all views can enhance performance. The two baselines do not discover the underlying subspace structure of different views exhaustively and usually incur inferior clustering performance. In conclusion, the proposed method efficiently combines consistency and specificity associated with multi-view data and thus outperforms baselines which only consider either term.

Single View versus Multiple Views

In this section, we compare our multi-view method CSMSC with a respective single-view method LRR. Figure 3 shows the detailed results by different configurations. Note that BBCSport dataset only has two bars for two different views. From Figure 3, we have following observations. First, as we mentioned before, multi-view is generally more robust and performs better than single-view, since more perspectives of information are considered. In fact, CSMSC indeed effectively combines information from different views with the consistent and specific constraints, demonstrating the advantage of our proposed method. Second, one view excels in some datasets might not win in the other datasets. For example, View2 acts as the best view on Notting-Hill and ORL, but performs the worst on Yale. View3 is the best view on Yale but shows degenerated performances on Notting-Hill and ORL. As a result, fixing a specific view for all datasets

²During our experiments, we found that 20 iterations are generally applicable to ensure decent results.

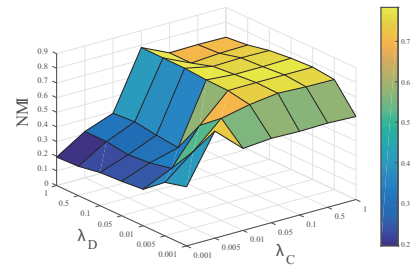


Figure 4: Sensitivity test on λ_C and λ_D versus NMI on Yale.

is unreliable, and it is difficult to choose the best view, since the best view varies from dataset to dataset. In summary, multi-view approach demonstrates clear advantage over single view approach.

To further investigate the improvement of the proposed method, we conduct LRR on each single view and CSMSC on multiple views respectively. The parameters in LRR and CSMSC are set as 0.05. According to Table 3, the clustering performance with multiple views are usually better than those of each single view, which empirically proves that clustering with multiple views is more robust than with single view. Also, the performance generally improves when the number of views is increased. Note that comparing to LRR_{best} in Table 1, View1, View2 and View3 in Table 3 are much worse, which declares that LRR is extremely sensitive to the parameter.

Parameter Sensitivity

In this section, we conduct a sensitivity test for the parameter λ_C and λ_D by varying from 0.001 to 1. Figure 4 shows the influence of different parameter values with respect to NMI on Yale dataset. It can be observed that, when λ_C or λ_D approaches to zero, the performance drops sharply, which is consistent with Section **Ablation Study on Consistency and Specificity**. Moreover, our method performs much stable when λ_C and λ_D becomes larger.

Conclusion

In this paper, we have introduced a novel subspace representation learning method of subspace clustering under the multi-view setting. The proposed method, named Consistent and Specific Multi-view Subspace Clustering, formulates the self-representation property using a consistent representation and a set of specific representations, which better fits the real-world datasets. We introduce a non-convex optimization problem and develop a convex relaxed optimization algorithm to recover the corresponding data representations, which provides some further practical smoothing convergence. The performance of our method is quantitatively evaluated on four datasets, which shows that our method outperforms baseline methods and state-of-the-art methods for multi-view subspace clustering.

Acknowledgments

The authors would like to thank all the anonymous reviewers for their valuable comments and suggestions to improve the quality of this paper. This work is supported by National Key Research and Development Plan(No.2016YFB0800603), National Natural Science Foundation of China (No.U1636214, 61422213).

References

- Bertsekas, D. P. 2014. *Constrained optimization and Lagrange multiplier methods*. Academic Press.
- Blum, A., and Mitchell, T. 1998. Combining labeled and unlabeled data with co-training. In *Proc. Conf. Learn. Theory*, 92–100.
- Cai, J.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM J. Optim.* 20(4):1956–1982.
- Candès, E. J., and Plan, Y. 2010. Matrix completion with noise. *Proc. IEEE* 98(6):925–936.
- Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *J. ACM* 58(3):11.
- Cao, X.; Zhang, C.; Fu, H.; Liu, S.; and Zhang, H. 2015. Diversity-induced multi-view subspace clustering. In *Proc. IEEE Int. Conf. Comput. Vis.*, 586–594.
- Chaudhuri, K.; Kakade, S. M.; Livescu, K.; and Sridharan, K. 2009. Multi-view clustering via canonical correlation analysis. In *Proc. Int. Conf. Mach. Learn.*, 129–136.
- Christopher, D. M.; Prabhakar, R.; and Hinrich, S. *Introduction to information retrieval*.
- De Sa, V. R. 2005. Spectral clustering with two views. In *Proc. Int. Conf. Mach. Learn. Workshop: Learning with Multiple Views*, 20–27.
- Eckstein, J., and Bertsekas, D. P. 1992. On the douglasrachford splitting method and the proximal point algorithm for maximal monotone operators. *Math. Program.* 55(1):293–318.
- Elhamifar, E., and Vidal, R. 2013. Sparse subspace clustering: Algorithm, theory, and applications. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(11):2765–2781.
- Gönen, M., and Alpaydm, E. 2011. Multiple kernel learning algorithms. *J. Mach. Learn. Res.* 12(7):2211–2268.
- Hubert, L., and Arabie, P. 1985. Comparing partitions. *J. Classif.* 2(1):193–218.
- Kumar, A., and Daumé, H. 2011. A co-training approach for multi-view spectral clustering. In *Proc. Int. Conf. Mach. Learn.*, 393–400.
- Kumar, A.; Rai, P.; and Daumé, H. 2011. Co-regularized multi-view spectral clustering. In *Proc. Adv. Neural Inf. Process. Syst.*, 1413–1421.
- Lades, M.; Vorbruggen, J. C.; Buhmann, J.; Lange, J.; Von Der Malsburg, C.; Wurtz, R. P.; and Konen, W. 1993. Distortion invariant object recognition in the dynamic link architecture. *IEEE Trans. Comput.* 42(3):300–311.
- Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.
- Liu, G., and Yan, S. 2011. Latent low-rank representation for subspace segmentation and feature extraction. In *Proc. IEEE Int. Conf. Comput. Vis.*, 1615–1622.
- Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Trans. Pattern Anal. Mach. Intell.* 35(1):171–184.
- Liu, G.; Lin, Z.; and Yu, Y. 2010. Robust subspace segmentation by low-rank representation. In *Proc. Int. Conf. Mach. Learn.*, 663–670.
- Lu, C.; Min, H.; Zhao, Z.; Zhu, L.; Huang, D.; and Yan, S. 2012. Robust and efficient subspace segmentation via least squares regression. *Proc. Eur. Conf. Comput. Vis.* 347–360.
- Luo, D.; Nie, F.; Ding, C.; and Huang, H. 2011. Multi-subspace representation and discovery. *Mach. Learn. Knowl. Discovery Databases* 405–420.
- Ma, Y.; Derksen, H.; Hong, W.; and Wright, J. 2007. Segmentation of multivariate mixed data via lossy data coding and compression. *IEEE Trans. Pattern Anal. Mach. Intell.* 29(9).
- Ng, A. Y.; Jordan, M. I.; and Weiss, Y. 2002. On spectral clustering: Analysis and an algorithm. In *Proc. Adv. Neural Inf. Process. Syst.*, 849–856.
- Ojala, T.; Pietikainen, M.; and Maenpaa, T. 2002. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(7):971–987.
- Parsons, L.; Haque, E.; and Liu, H. 2004. Subspace clustering for high dimensional data: a review. *ACM SIGKDD Explor. Newsl.* 6(1):90–105.
- Rahmani, M., and Atia, G. 2017. Innovation pursuit: A new approach to subspace clustering. *Proc. Int. Conf. Mach. Learn.*
- Rao, S.; Tron, R.; Vidal, R.; and Ma, Y. 2010. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. *IEEE Trans. Pattern Anal. Mach. Intell.* 32(10):1832–1845.
- Wei, S., and Lin, Z. 2011. Analysis and improvement of low rank representation for subspace segmentation. *arXiv preprint arXiv:1107.1561*.
- Xia, R.; Pan, Y.; Du, L.; and Yin, J. 2014. Robust multi-view spectral clustering via low-rank and sparse decomposition. In *Proc. Assoc. Adv. Artif. Intell.*, 2149–2155.
- Xu, C.; Tao, D.; and Xu, C. 2013. A survey on multi-view learning. *arXiv preprint arXiv:1304.5634*.
- Xu, C.; Tao, D.; and Xu, C. 2015. Multi-view intact space learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 37(12):2531–2544.
- Yang, Y.; Feng, J.; Jojic, N.; Yang, J.; and Huang, T. S. 2016. l_0 -Sparse Subspace Clustering. 731–747.
- Zhang, Y.; Xu, C.; Lu, H.; and Huang, Y. 2009. Character identification in feature-length films using global face-name matching. *IEEE Trans. Multimedia* 11(7):1276–1288.
- Zhang, C.; Fu, H.; Liu, S.; Liu, G.; and Cao, X. 2015. Low-rank tensor constrained multiview subspace clustering. In *Proc. IEEE Int. Conf. Comput. Vis.*, 1582–1590.
- Zhang, C.; Hu, Q.; Fu, H.; Zhu, P.; and Cao, X. 2017. Latent multi-view subspace clustering. In *Proc. IEEE Conf. Comput. Vis. Pattern Rec.*, volume 30, 4279–4287.
- Zhao, H.; Ding, Z.; and Fu, Y. 2017. Multi-view clustering via deep matrix factorization. In *Proc. Assoc. Adv. Artif. Intell.*, 2921–2927.