

# JOINT LEARNING OF SELF-REPRESENTATION AND INDICATOR FOR MULTI-VIEW IMAGE CLUSTERING

Songsong Wu<sup>1,5</sup>, Zhiqiang Lu<sup>1</sup>, Hao Tang<sup>2</sup>, Yan Yan<sup>3</sup>, Songhao Zhu<sup>1</sup>, Xiao-Yuan Jing<sup>4</sup>, Zuoyong Li<sup>5</sup>

<sup>1</sup>Nanjing University of Posts and Telecommunications <sup>2</sup>University of Trento

<sup>3</sup>Texas State University <sup>4</sup> Wuhan University

<sup>5</sup> Industrial Robot Application of Fujian University Engineering Research Center Minjiang University

## ABSTRACT

Multi-view subspace clustering aims to divide a set of multi-source data into several groups according to their underlying subspace structure. Although the spectral clustering based methods achieve promotion in multi-view clustering, their utility is limited by the separate learning manner in which affinity matrix construction and cluster indicator estimation are isolated. In this paper, we propose to jointly learn the self-representation, continuous and discrete cluster indicators in an unified model. Our model can explore the subspace structure of each view and fusion them to facilitate clustering simultaneously. Experimental results on two benchmark datasets demonstrate that our method outperforms other existing competitive multi-view clustering methods.

**Index Terms**— Multi-view Clustering, Subspace Clustering, Self-representation Learning

## 1. INTRODUCTION

In real-world applications, images always own multiple views as they are represented by multiple ways. For instance, three typical views are pixel intensity, LBP feature, and Gabor coefficients. Unsupervised clustering of images based on their multiple views has attracted increasing attention recently [1, 2, 3]. Compared with single-view clustering (SVC) [4, 5, 6, 7, 8, 9, 10] that performs clustering relying on a single type of feature, multi-view clustering (MVC) exploits the complementary information among views to facilitate clustering.

Multi-view clustering can be divided into three categories, co-training MVC [11, 12, 13], multiple kernel learning MVC [14, 15], and subspace learning MVC [4, 16, 17, 18, 19, 20, 21]. Co-training based methods pursue the optimal clustering in each view while minimize the disagreement among them simultaneously. Multiple kernel learning based methods claims that the distinct information from each view can be integrated with a linear combination of multiple kernels to promote clustering performance. Subspace learning based methods model data views as distinct low-dimensional subspaces. In this paper, we focus on subspace MVC because of it's simplicity in computation and stability in performance.

Subspace learning based MVC clustering methods rely on affinity matrix learning followed by cluster indicator prediction. However, these two procedures are accomplished separately, resulting that they are hardly optimal for clustering. Besides, the desired discrete clustering indicator is approximated by a continuous matrix for the convenience in model solving. This incurs the loss in cluster information, as reported by [22, 23] for single view clustering.

To address the above issues, we proposed a joint learning model for MVC in this paper. It can jointly learn subspace representations, continuous labels, and discrete labels from the multi-view data. In solving the optimization problem, continuous clustering label is only used as intermediate product. The main contributions of our work is to combine subspace learning, continuous label learning, and discrete label learning to obtained multi-view affinity matrices constantly updated with the inherent interactions between three steps.

## 2. RELATED WORK

SPC [22] selects the most informative to perform standard spectral clustering algorithm. SSC [6] aims to find a sparse representation from the data, then the final clustering result can be obtained by spectral clustering. S3C [9] executes clustering on each view and selects the best performance. Min-Dis [11] aims to minimize the disagreement of two-view by generating a bipartite graph and uses spectral clustering to obtain the clustering results. ConvexReg SPC [24] learns a common representation for all views, and executes the standard spectral clustering to get clustering results. RMSC [25] is robust multi-view spectral clustering method and uses a Markov chain to cluster. Di-MSVC [16] adopts the HSIC criterion norm to obtain the diverse representations of different views with considering the complementary information, then uses spectral clustering to generate the clustering result. LT-MSVC [17] obtains the final clustering result by capturing high-dimension information of multi-view data. ECMSC [20] considers complementary information with exclusivity term and guarantees the consistency with a common indicator and uses the spectral clustering algorithm to obtain clustering results. CSMSC [21]

tries to explore a common representation and specific representation from multi-view data, and the clustering result can be obtained by spectral clustering.

Beyond the above methods estimate discrete cluster indicators with continuous indicators, [26, 23] learn discrete label and then use discrete cluster labels directly for clustering in single-view. s for as we have known, the issue of discrete and continuous cluster indicator has not been touched in the case of multi-view clustering. We propose to simultaneously learn subspace representation, continuous indicator and discrete indicator for multi-view clustering, rather than straightly extend [26, 23] from single view clustering to multiple view clustering.

### 3. THE PROPOSED METHOD

#### 3.1. Formulation

Given sample set  $X_v \in R^{d_v \times N}, v \in [1, 2, \dots, V]$  with  $V$  views, where  $N$  is the number of samples, and  $d_v$  is the sample dimension of view  $v$ . The self-representation of samples for a single view is written as

$$X_v = X_v Z_v + E_v, \quad \text{s.t. } Z_{v(i,i)} = 0 \quad (1)$$

where  $Z_v \in R^{n \times n}$  is self-representation matrix and  $E_v \in R^{d_v \times n}$  is error term.

Under the principle of spectral clustering, the discrete indicator is obtained by

$$\arg \min_F \text{tr}(F^T L_v F), \quad \text{s.t. } F \in I_{dx} \quad (2)$$

where  $I_{dx}$  means that  $F = [f_1, f_2, \dots, f_n] \in R^{K \times N}$  is the discrete cluster indicator matrix with each column  $f_i \in \{0, 1\}^{K \times 1}$  contains only one element 1 indicating the cluster the sample is assigned to. The  $L$  in (2) is graph Laplacian matrix with the affinity matrix  $W_v = \frac{|Z_v| + |Z_v^T|}{2}$ . In practice, the following model is solved instead as model 2 is NP-hard

$$\arg \min_P \text{tr}(P^T L_v P), \quad \text{s.t. } P^T P = I \quad (3)$$

where  $P \in R^{K \times N}$  is the relaxed continuous indicator, and the orthogonal constraint is to prevent from trivial solutions.

Take formula (1)(2)(3) into consideration, we have the following model for multi-view subspace clustering

$$\begin{aligned} \arg \min_{Z_v, E_v, F, P, Q} \sum_{v=1}^V \left\{ \|X_v - X_v Z_v - E_v\|_F^2 + \lambda_1 \|E_v\|_1 \right\} \\ \text{s.t. } Z_v(i, i) = 0, P^T P = I, Q^T Q = I, F \in I_{dx} \end{aligned} \quad (4)$$

where  $\lambda_1, \lambda_2$  and  $\lambda_3$  are positive hyper-parameters. The first and second term are associated with the self-representation in each view. The third term is related to continuous indicator

seeking so as to avoid NP-hard issue in computation. The fourth term build a linear relation between the continuous indicator the desired discrete indicator with matrix  $Q \in R^{c \times c}$  to ensure these two indicators are consistent with each other.

#### 3.2. Optimization

We solve model (4) by alternative optimization strategy, i.e. optimizing one variable by fixing the previous values of other variables. To facilitate the narrative, we omit the subscript of view tentatively.

**Update  $Z$ .** When  $E, F, P$ , and  $Q$  are fixed, we have

$$\begin{aligned} \arg \min_Z \|X - XZ - E\|_F^2 + \lambda_2 \text{Tr}(P^T L P) \\ \text{s.t. } Z_{ii} = 0 \end{aligned} \quad (5)$$

By defining  $Y_{ij} = \|P_i - P_j\|_2^2$  and using the definition of  $L$ , we solve (5) by seek a optimal column of  $Z$  at one time based on the following optimal model

$$\arg \min_{Z_i} \|Z_i - V_i\|_2^2 + \frac{\lambda_2}{2} |Z_i|^T Y_i, \quad \text{s.t. } Z_{ii} = 0 \quad (6)$$

where  $Z_i, Y_i$  and  $X_i$  are the  $i$ -th column of  $Z, Y$ , and  $X$  respectively, and  $V_i = \frac{K^T X_i}{X_i^T X_i}$  with  $K = X - (XZ - X_i Z_i^T) - E$ . So far, we can get closed form solution based on (6) for that if  $j = i$ , we have  $Z_{ji} = 0$ , and if  $j \neq i$ , we have

$$Z_{ji} = \begin{cases} \text{sign}(V_{ji}) \left( |V_{ji}| - \frac{\lambda_2 Y_{ji}}{4} \right) + \\ \left\{ \begin{array}{l} V_{ji} - \frac{\lambda_2 Y_{ji}}{4}, \quad \text{if } V_{ji} > + \frac{\lambda_2 Y_{ji}}{4} \\ V_{ji} + \frac{\lambda_2 Y_{ji}}{4}, \quad \text{if } V_{ji} < - \frac{\lambda_2 Y_{ji}}{4} \\ 0, \quad \text{otherwise} \end{array} \right. \end{cases} \quad (7)$$

**Update  $E$ .** When  $Z, F, P$ , and  $Q$  are fixed, we seek the optimal  $E$  using the same column-wise strategy as  $Z$ . Specifically, for a single column  $E_i$  of  $E$ , we have

$$\arg \min_{E_i} \|(X - XZ)_i - E_i\|_F^2 + \lambda_1 \|E_i\|_1 \quad (8)$$

Where  $(X - XZ)_i$  is the  $i$ -th column in the matrix. The solution is provided by

$$\begin{aligned} E_{ji} = \text{sign}(X - XZ)_{ji} \left( |(X - XZ)_{ji}| - \frac{\lambda_1}{2} \right) \\ + \left\{ \begin{array}{l} (X - XZ)_{ji} - \frac{\lambda_1}{2}, \quad \text{if } (X - XZ)_{ji} > \frac{\lambda_1}{2} \\ (X - XZ)_{ji} + \frac{\lambda_1}{2}, \quad \text{if } (X - XZ)_{ji} < - \frac{\lambda_1}{2} \\ 0, \quad \text{otherwise} \end{array} \right. \end{aligned} \quad (9)$$

**Update  $P$ .** When  $Z, F, E$ , and  $Q$  are fixed, the model becomes

$$\arg \min_P \lambda_2 \text{Tr}(P^T L P) + \lambda_3 \|F - PQ\|_F^2, \quad \text{s.t. } P^T P = I \quad (10)$$

The model can be efficiently solved by the algorithm proposed by [27].

---

**Algorithm 1** The proposed MVC method

---

**Input:**  $V$  view image sets  $\{X =_v\}_{v=1}^V$ , positive parameters  $\lambda_1, \lambda_2, \lambda_3$ .

Initialize  $Z_1, Z_2, \dots, Z_V, P, Q$  with random values, and set  $F, E = 0$

**repeat**

1. Update  $Z_v$  with (7)
2. Update  $E_v$  with (9)
3. Update  $P$  with (10).
4. Update  $Q$  with (12)
5. Update  $F$  with (14)

**until** Stopping criterion

Compute affinity matrix by  $W = \sum_{v=1}^V \frac{|Z_v| + |Z_v|^T}{2}$

Perform spectral clustering using affinity matrix  $W$ .

**Output:** Clustering result.

---

**Update Q.** When  $F, Z, E$  and  $P$  are fixed, the model becomes

$$\arg \min_Q \lambda_3 \|F - PQ\|_F^2, \quad \text{s.t. } Q^T Q = I \quad (11)$$

It is a orthogonal Procrustes problem [28], whose solution is given as

$$Q = UV^T \quad (12)$$

where  $U$  and  $V$  are the left and right singular value matrices of  $F^T P$ .

**Update F.** When  $Z, E, P$  and  $Q$  are fixed, the model becomes

$$\arg \max_F \lambda_3 \text{Tr}(F^T PQ), \quad \text{s.t. } F \in I_{dx} \quad (13)$$

whose optimal solution can be obtained as follows

$$F_{ij} = \begin{cases} 1, & \text{if } j = \arg \max_k (PQ)_{ik} \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

In summary, the proposed method can be expressed in Algorithm 1.

## 4. EXPERIMENT

### 4.1. Experimental Settings

**Dataset Description.** We adopted two face image datasets in the experiments which are used widely in image clustering [17, 18, 25]. In our experiment, each image sample has three views: intensity, LBP [28], and Gabor [29]. Statistics of the datasets are summarized in Table 1. (i) ORL is divided into 40 different themes, each containing 10 images. All images are in a uniform black background and are All images are taken from the top of the front. For some of these themes, the images are taken at different times. In light conditions, facial expressions (such as open eyes, closed eyes, laugh, and

non-laugh), facial details (glasses) and other aspects are different. (ii) Yale is consisted of 165 gray-scale images of 15 individuals, each individual has 11 images, including: center light, glasses, happy, left light, no glasses, normal, right light, sad, sleepy, surprised and wink.

**Comparison Algorithm.** We compare our method with three single-view methods (SPC [22], SSC [6], S3C [9]) and seven multi-view methods (Min-Dis [11], ConvexReg SPC [24], RMSC [25], Di-MS C [16], LT-MS C [17], ECMSC [20], CSMSC [21]). The optimal parameters of the proposed method are set empirically based on grid searching. Specifically, for ORL dataset, the parameters are set as:  $\lambda_1=0.002, \lambda_2=0.5, \lambda_3=0.1$  and for Yale dataset, they are set as:  $\lambda_1=0.003, \lambda_2=0.7, \lambda_3=0.2$ . The parameter values of other methods are set according to the corresponding papers.

**Evaluation Metrics.** In our experiment, we use six metrics to evaluate the clustering performance, i.e., Normalized Mutual Information (NMI), Accuracy (ACC), Adjusted Rand Index (ARI), F-score, Precision and Recall. ACC and NMI have been adopted to evaluate clustering result [30, 31]. Precision, F-score, ARI and Recall have been widely used in [32] to measure the clustering quality. These six metrics evaluate the different performance of clustering. For all metrics, higher values indicate better cluster quality.

**Performance Comparison.** We repeat the random experiments for 30 rounds and report the mean and standard deviation as final results. The clustering results on ORL and Yale are provided in Table 1 and Table 2 respectively.

From Table 1, we observe that compared with Di-MS C, we achieve improvements around 0.3%, 4.8%, 2.9%, 2.8%, 4.0%, and 1.2% in terms of NMI, ACC, ARI, F, P, and Re. Compared with ECMSC, we achieve improvements around 3.2%, 2.1%, 1.4%, 2.0%, and 0.9% in terms of ACC, ARI, F, P, and Re, because the affinity matrix is constantly updated due to add learning about discrete labels. Compared with latest method (CSMSC), our experimental results improve around 1.8%, 0.4%, 0.4%, and 6.4% in terms of ACC, ARI, F, Re. The reason is that we consider the relationship between subspace learning and spectral clustering. But our results are lower than CSMSC in NMI and P, it seems that we did not consider the complementary information of multiple views.

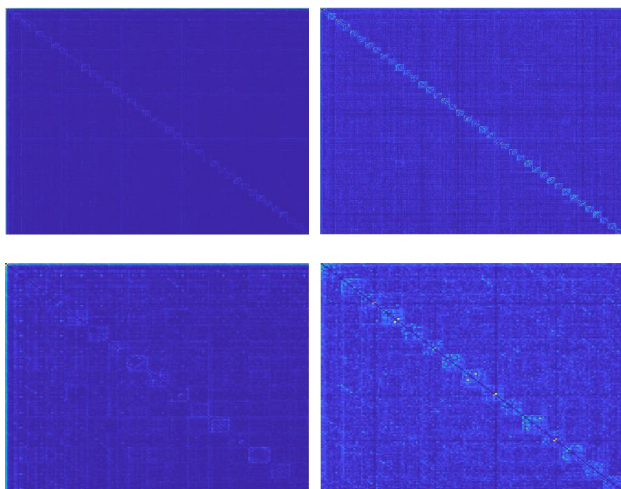
Table 2 shows that the multi-view clustering results on Yale dataset. The method LT-MS C has a significant improvement. It seems that low-rank tensor is suitable for images clustering. Di-MS C considers the diversity representations of different views. Compared with Di-MS C, we achieve improvements around 5.5%, 8.3%, 9.5%, 8.0%, 7.3%, 7.5% in terms of NMI, ACC, ARI, F, P, Re, respectively. The major reason is that we put subspace representation and spectral clustering into an optimization model. Compared with ECMSC, we achieve improvements around 0.9%, 2.1%, 3.0%, 3.7%, 3.2%, 0.8% in six metrics, respectively. The reason is that our method does not use continuous label to up-

**Table 1:** Clustering performances on ORL dataset (meanstandard deviation).

	Method	NMI	ACC	ARI	F	P	Re
single	SPCbest	0.884(0.002)	0.726(0.025)	0.655(0.005)	0.664(0.005)	0.610(0.006)	0.728(0.005)
single	SSCbest	0.893(0.007)	0.765(0.008)	0.694(0.013)	0.682(0.012)	0.673(0.007)	0.764(0.005)
single	S3Cbest	0.902(0.012)	0.784(0.009)	0.705(0.019)	0.698(0.018)	0.688(0.012)	0.791(0.011)
Multiple	Min-Dis	0.876(0.002)	0.748(0.051)	0.654(0.004)	0.663(0.004)	0.615(0.004)	0.718(0.003)
Multiple	RMSC	0.872(0.012)	0.723(0.025)	0.644(0.029)	0.654(0.028)	0.607(0.033)	0.709(0.027)
Multiple	ConReg	0.883(0.003)	0.734(0.031)	0.668(0.032)	0.676(0.035)	0.628(0.041)	0.731(0.030)
Multiple	LTMSC	0.930(0.002)	0.795(0.007)	0.750(0.003)	0.768(0.007)	0.766(0.009)	0.837(0.004)
Multiple	DiMSC	0.940(0.003)	0.838(0.001)	0.802(0.000)	0.807(0.003)	0.764(0.012)	0.856(0.004)
Multiple	ECMSC	0.947(0.009)	0.854(0.011)	0.810(0.012)	0.821(0.015)	0.783(0.008)	0.859(0.012)
Multiple	CSMSC	0.942(0.005)	0.868(0.012)	0.827(0.002)	0.831(0.001)	0.860(0.002)	0.804(0.003)
Proposed	Ours	0.943(0.005)	0.886(0.016)	0.831(0.019)	0.835(0.019)	0.804(0.023)	0.868(0.018)

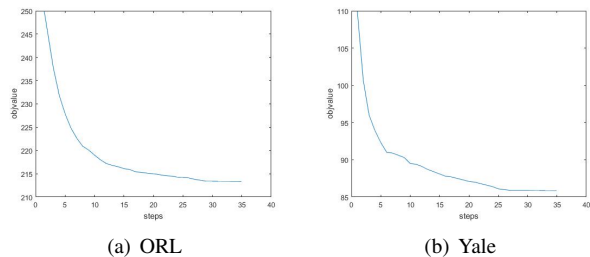
**Table 2:** Clustering performances on ORL dataset (meanstandard deviation).

	Method	NMI	ACC	ARI	F	P	Re
single	SPCbest	0.654(0.009)	0.616(0.030)	0.440(0.011)	0.475(0.011)	0.457(0.011)	0.495(0.010)
single	SSCbest	0.671(0.011)	0.627(0.000)	0.475(0.004)	0.517(0.007)	0.509(0.003)	0.547(0.004)
single	S3Cbest	0.678(0.013)	0.634(0.016)	0.471(0.005)	0.508(0.012)	0.512(0.005)	0.568(0.025)
Multiple	Min-Dis	0.645(0.005)	0.615(0.043)	0.433(0.006)	0.470(0.006)	0.446(0.005)	0.496(0.006)
Multiple	RMSC	0.684(0.033)	0.642(0.036)	0.485(0.046)	0.517(0.043)	0.500(0.043)	0.535(0.044)
Multiple	ConReg	0.673(0.023)	0.611(0.035)	0.466(0.032)	0.501(0.030)	0.476(0.032)	0.532(0.029)
Multiple	LTMSC	0.765(0.008)	0.741(0.002)	0.570(0.004)	0.598(0.006)	0.569(0.004)	0.629(0.005)
Multiple	DiMSC	0.727(0.010)	0.709(0.003)	0.535(0.001)	0.564(0.002)	0.543(0.001)	0.586(0.003)
Multiple	ECMSC	0.773(0.010)	0.771(0.014)	0.590(0.014)	0.617(0.012)	0.584(0.013)	0.653(0.013)
Multiple	CSMSC	0.784(0.001)	0.752(0.001)	0.615(0.005)	0.640(0.004)	0.673(0.002)	0.610(0.006)
Proposed	Ours	0.782(0.005)	0.792(0.026)	0.620(0.008)	0.644(0.007)	0.616(0.009)	0.661(0.006)

**Fig. 1:** Affinity matrix visualization on ORL (top) and Yale (bottom). From left to right: The affinity matrix without joint learning (corresponding to  $\lambda_2 = 0, \lambda_3 = 0$ ) and the affinity matrix of our method.

date the subspace representation, instead simultaneously uses discrete label and continuous label to update the subspace representation. Compared with latest method (CSMSC), our experimental results improve around 4.0%, 0.5%, 0.4%, 5.0% in terms of ACC, ARI, F, and Re. CSMSC does not put subspace learning and spectral clustering method in an optimization model.

**Visualization and Convergence Analysis.** To validate the effectiveness of our method, we show the visualization of affinity matrices on Yale and ORL dataset. As observed in Fig. 1, we can clearly see that the affinity matrix of our method presents multiple block structures. Each block structure represents a subspace. From the two figures on the right,

**Fig. 2:** Convergence curve of our method on ORL and Yale.

can see simultaneously learn subspace representations of each view, continuous labels, and discrete labels will be beneficial for multi-view subspace clustering. As shown in Fig. 2, “objvalue” represents the value of the objective function; “steps” is the number of iterations. We show the convergence of our method on ORL and Yale datasets. We can see that our algorithm converge quickly on both datasets.

## 5. CONCLUSIONS

We propose a novel multi-view image clustering approach by jointly learning the self-representation of images and cluster indicators. The two processes promote each other to achieve complementary information confusion for multi-view clustering. The experimental results of multi-view face images clustering demonstrate that our method outperforms other existing competitive multi-view clustering algorithms.

## 6. ACKNOWLEDGEMENT

This work was supported in part by Industrial Robot Application of Fujian University Engineering Research Center, Minjiang University. (Grant No. MJUKF-IRA201806).

## 7. REFERENCES

- [1] Changqing Zhang, Huazhu Fu, Qinghua Hu, Pengfei Zhu, and Xiaochun Cao, "Flexible multi-view dimensionality co-reduction," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 648–659, 2017.
- [2] Ying Cui, Xiaoli Z Fern, and Jennifer G Dy, "Non-redundant multi-view clustering via orthogonalization," in *ICDM*, 2007.
- [3] Hua Wang, Feiping Nie, and Heng Huang, "Multi-view clustering and feature learning via structured sparsity," in *ICML*, 2013.
- [4] Kamalika Chaudhuri, Sham M Kakade, Karen Livescu, and Karthik Sridharan, "Multi-view clustering via canonical correlation analysis," in *ICML*, 2009.
- [5] Hao Tang, Hong Liu, Wei Xiao, and Nicu Sebe, "Fast and robust dynamic hand gesture recognition via key frames extraction and feature fusion," *Neurocomputing*, vol. 331, pp. 424–433, 2019.
- [6] Ehsan Elhamifar and René Vidal, "Sparse subspace clustering," in *CVPR*, 2009.
- [7] Can-Yi Lu, Hai Min, Zhong-Qiu Zhao, Lin Zhu, De-Shuang Huang, and Shuicheng Yan, "Robust and efficient subspace segmentation via least squares regression," in *ECCV*, 2012.
- [8] 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*, vol. 35, no. 1, pp. 171–184, 2013.
- [9] Chun-Guang Li and Rene Vidal, "Structured sparse subspace clustering: A unified optimization framework," in *CVPR*, 2015.
- [10] Hao Tang and Hong Liu, "A novel feature matching strategy for large scale image retrieval," in *IJCAI*, 2016.
- [11] Virginia R De Sa, "Spectral clustering with two views," in *ICML workshop on learning with multiple views*, 2005.
- [12] Abhishek Kumar and Hal Daumé, "A co-training approach for multi-view spectral clustering," in *ICML*, 2011.
- [13] Abhishek Kumar, Piyush Rai, and Hal Daume, "Co-regularized multi-view spectral clustering," in *NIPS*, 2011.
- [14] Mehmet Gönen and Ethem Alpaydın, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, no. Jul, pp. 2211–2268, 2011.
- [15] Grigorios Tzortzis and Aristidis Likas, "Kernel-based weighted multi-view clustering," in *ICDM*, 2012.
- [16] Xiaochun Cao, Changqing Zhang, Huazhu Fu, Si Liu, and Hua Zhang, "Diversity-induced multi-view subspace clustering," in *CVPR*, 2015.
- [17] Changqing Zhang, Huazhu Fu, Si Liu, Guangcan Liu, and Xiaochun Cao, "Low-rank tensor constrained multiview subspace clustering," in *ICCV*, 2015.
- [18] Hongchang Gao, Feiping Nie, Xuelong Li, and Heng Huang, "Multi-view subspace clustering," in *ICCV*, 2015.
- [19] Yuan Xie, Dacheng Tao, Wensheng Zhang, and Lei Zhang, "Multi-view subspace clustering via relaxed  $l_1$ -norm of tensor multi-rank," *arXiv preprint arXiv:1610.07126*, 2016.
- [20] Xiaobo Wang, Xiaojie Guo, Zhen Lei, Changqing Zhang, and Stan Z Li, "Exclusivity-consistency regularized multi-view subspace clustering," in *CVPR*, 2017.
- [21] Shirui Luo, Changqing Zhang, Wei Zhang, and Xiaochun Cao, "Consistent and specific multi-view subspace clustering," in *AAAI*, 2018.
- [22] Andrew Y Ng, Michael I Jordan, and Yair Weiss, "On spectral clustering: Analysis and an algorithm," in *NIPS*, 2002.
- [23] Zhao Kang, Chong Peng, Qiang Cheng, and Zenglin Xu, "Unified spectral clustering with optimal graph," in *AAAI*, 2018.
- [24] Maxwell D Collins, Ji Liu, Jia Xu, Lopamudra Mukherjee, and Vikas Singh, "Spectral clustering with a convex regularizer on millions of images," in *ECCV*, 2014.
- [25] Rongkai Xia, Yan Pan, Lei Du, and Jian Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *AAAI*, 2014.
- [26] Yang Yang, Fumin Shen, Zi Huang, and Heng Tao Shen, "A unified framework for discrete spectral clustering," in *IJCAI*, 2016.
- [27] Zaiwen Wen and Wotao Yin, "A feasible method for optimization with orthogonality constraints," *Mathematical Programming*, vol. 142, no. 1-2, pp. 397–434, 2013.
- [28] Timo Ojala, Matti Pietikainen, and Topi Maenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [29] Martin Lades, Jan C Vorbruggen, Joachim Buhmann, Jörg Lange, Christoph Von Der Malsburg, Rolf P Wurtz, and Wolfgang Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Transactions on Computers*, no. 3, pp. 300–311, 1993.
- [30] Xiaochun Cao, Changqing Zhang, Chengju Zhou, Huazhu Fu, and Hassan Foroosh, "Constrained multi-view video face clustering," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4381–4393, 2015.
- [31] Xiaojie Guo, "Robust subspace segmentation by simultaneously learning data representations and their affinity matrix," in *IJCAI*, 2015.
- [32] Qian Zhao, Deyu Meng, Zongben Xu, Wangmeng Zuo, and Lei Zhang, "Robust principal component analysis with complex noise," in *ICML*, 2014.