Image Projection Ridge Regression for Subspace Clustering

Chong Peng, Zhao Kang, Fei Xu, Yongyong Chen, and Qiang Cheng, *Senior Member, IEEE*

***Abstract*—Subspace clustering methods have been widely studied recently. When the inputs are two-dimensional (2-D) data, existing subspace clustering methods usually convert them into vectors, which severely damages inherent structures and relationships from original data. In this letter, we propose a novel subspace clustering method for 2-D data. It directly uses 2-D data as inputs such that the learning of representations benefits from inherent structures and relationships of the data. It simultaneously seeks image projection and representation coefficients such that they mutually enhance each other and lead to powerful data representations. An efficient algorithm is developed to solve the proposed objective function with provable decreasing and convergence property. Extensive experimental results verify the effectiveness of the new method.**

***Index Terms*—Subspace clustering, unsupervised learning, 2-diemnsional data, spatial information.**

I. Introduction

HIGH-DIMENSIONAL data are ubiquitous and it has been increasingly common to represent and handle such data in many real-world applications such as computer vision and image processing. Often times, such data can be represented in a low-dimensional subspace due to their latent low-dimensional structures [1], [2]. Recovering such low-dimensional subspaces usually requires clustering data points into different groups such that each group can be fitted with a subspace, which is known as subspace clustering or subspace segmentation.

During the last decade, various subspace clustering algorithms have been developed, among which spectral clustering-based methods [3], [4] have been popular. Among them, low-rank representation (LRR) [4], [5] and sparse subspace clustering (SSC) [3] have drawn considerable attention due to their efficiencies and elegant theories. Both LRR and SSC seek a linear representation of the data with respect to a dictionary of the data itself, where the representation coefficient matrix is required to be low-rank or sparse, respectively. Some recent works have theoretically analyzed parameter selection [6] and row-coherence dependence [7] for LRR. Recently studies point out that the nuclear norm used in LRR is not accurate, and some non-convex approximations of the rank can significantly improve the learning performance [1], [8]. Some studies demonstrate the importance of feature selection and data projection for subspace clustering [9], [10]. For example, [9], [11] seek a LRR with respect to a subset of features, which alleviates the importance of rank approximations; Patel *et al.* [10] seeked a sparse representation of projected data in a latent low-dimensional space such that hidden structures of the data provide useful information.

When the input samples are two-dimensional (2-D) data, such as images, most of the existing methods apply vectorization as preprocessing. However, the approaches using vectorized data do not consider the inherent structure [12] and correlations in the original data and building a model based on original high-dimensional features is not effective to filter the noise, occlusions or redundant information in the original feature space [13]. To overcome these limitations, we propose a novel subspace clustering method for 2-D data with enhanced capability of capturing inherent structures and relationships from original 2-D data. We should note the stark difference between our approach and tensor method such as [13]. In our approach, instead of using tensor structure, we seek projection directions that captures the most variational 2-D features from the data, such that the learning of representation only depends the most expressive and essential 2-D features.

We summarize the key contributions of this letter as follows.

1) Instead of being converted to vectors, 2-D data are directly used such that spatial information is preserved.
2) Covariance matrix is constructed using 2-D data, which effectively avoids the curse of dimensionality of vectorized data.
3) Projection directions are sought to simultaneously extract the most variational 2-D features and help construct the representation matrix, such that they mutually enhance each other and lead to powerful data representations.
4) Efficient optimization algorithm is developed, which theoretically guarantees the decreasing and convergence of the value sequence of the objective function.

II. 2-D Principal Component Analysis (2DPCA)

Given a 2-D sample $X \in \mathcal{R}^{a\times b}$ and a unitary vector $p \in \mathcal{R}^b$, the projected feature vector of X can be obtained by $y = Xp$

Manuscript received March 2, 2017; revised April 23, 2017; accepted May 1, 2017. Date of publication May 3, 2017; date of current version May 23, 2017. This work is supported in part by the National Science Foundation under Grant IIS-1218712 and in part by the Foundation Program of Yuncheng University under Grant SWSX201603 and Grant YQ-2012020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Sanghoon Lee. *(Corresponding author: Chong Peng.)*

C. Peng is with the Department of Applied Mathematics, Yuncheng University, Yuncheng 044000, China, and also with the Department of Computer Science, Southern Illinois University Carbondale, Carbondale, IL 62901 USA (e-mail: pchong@siu.edu).

Z. Kang and Q. Cheng are with the Department of Computer Science, Southern Illinois University Carbondale, Carbondale, IL 62901 USA (e-mail: zhao.kang@siu.edu; qcheng@siu.edu).

F. Xu is with the College of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China (e-mail: xf_em@163.com).

Y. Chen is with the College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China (e-mail: yongyongchen.cn@hotmail.com).

Color versions of one or more of the figures in this letter are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2017.2700852

[14], [15]. The covariance matrix of projected feature vectors is $\mathbf{S}_x = \mathbf{E}[(y - \mathbf{E}(y))(y - \mathbf{E}(y))^T]$, where $\mathbf{E}$ is the expectation operator. Then p can be obtained by

$$\max_{p^T p=1} \mathrm{Tr}(\mathbf{S}_x) \Leftrightarrow \max_{p^T p=1} \mathrm{Tr}(p^T G_t p). \tag{1}$$

where $G_t = \mathbf{E}[(X - \mathbf{E}X)^T (X - \mathbf{E}X)]$. Usually, a single optimal projection direction is not enough [14] and multiple directions $P = [p_1, p_2, \ldots, p_r] \in \mathcal{R}^{b\times r}$ are needed. Mathematically, they can be found by:

$$\max_{P^T P=I_r} \mathrm{Tr}(P^T G_t P) \tag{2}$$

where I_r is an identity matrix of size $r \times r$.

III. Image Projection Ridge Regression

Given a set of 2-D samples $\mathbf{X} = \{X_i \in \mathcal{R}^{a\times b}\}_{i=1}^n$ and a projection vector $p \in \mathcal{R}^b$, we define an operation as follows:

$$\mathbf{X} \circ p = [X_1 p, X_2 p, \ldots, X_n p] \in \mathcal{R}^{a\times n} \tag{3}$$

which transforms each image to an a-dimensional vector and generates a data matrix with each column containing a projected feature vector of image X_i. We assume that after this liner projection, the projected data points have the self-expressive property, i.e., $\mathbf{X} \circ p \approx (\mathbf{X} \circ p)Z$, where Z is the coefficient matrix with z_i, $z_{(j)}$, and z_{ij} being its ith column, jth row, and ijth element, respectively. Inspired by thresholding ridge regression (TRR) [16], Z can be constructed by

$$\min_Z \|\mathbf{X} \circ p - (\mathbf{X} \circ p)Z\|_F^2 + \lambda \|Z\|_F^2 \tag{4}$$

where $\lambda > 0$ is a balancing parameter. Recent studies have shown that different norms can be used for Z, such as the ℓ_1, nuclear, or Frobenius norms [3], [2], [17]. Incorporating the Frobenius norm on Z has two benefits: 1) As pointed out in [17], it better measures data correlations exhibited in real world data; 2) It allows more efficient optimization. Here, in contrast to TRR, we do not require $z_{ii} = 0$ because:

1) It ensures more efficient optimization;
2) $\lambda > 0$ already excludes potentially trivial solutions such as I_n;
3) X_i is within the intrasubspace of X_i.

The above discussed procedure involves two steps, which, unfortunately, might be problematic for some problems as noted in [18] and [19]. Moreover, performing multitasks in a single model has been shown successful [9]–[11], [20], [21]. These successes as well as the consideration of avoiding potential problems of using a two-step approach inspire us to simultaneously learn a projection direction and a coefficient matrix in a single, seamlessly integrated framework, such that the learned projection direction helps not only retain the major information from the original 2-D images, but also facilitate the construction of the coefficient matrix.

To integrate (4) with a measure such as (1), we propose to solve the following model

$$\min_{Z, p^T p=1} \|\mathbf{X} \circ p - (\mathbf{X} \circ p)Z\|_F^2 + \lambda \|Z\|_F^2 + \gamma p^T G p \tag{5}$$

where $G = G_t^{-1}$, and $\gamma > 0$ is a balancing parameter. Therefore, minimizing $p^T G p$ essentially retains the dominant variance of the image covariance matrix and thus helps keep the most variational information from the original 2-D images. Here, we should note that G_t can estimated by $\frac{1}{n}\sum_{i=1}^n (X_i - \frac{1}{n}\sum_{j=1}^n X_j)^T (X_i - \frac{1}{n}\sum_{j=1}^n X_j)$ in practice, which is usually invertible in real world applications. In singular case, in the implementation, we define $G = (G_t + \epsilon I_n)^{-1}$, where $\epsilon > 0$ is a small value. In the experiments, we observe that G_t is always invertible and thus positive definite.

As mentioned above, a single projection direction may be not enough to keep major information. To tackle this problem, we define the following operator:

$$\mathbf{X} \circ P = [(\mathbf{X} \circ p_1)^T, \ldots, (\mathbf{X} \circ p_r)^T]^T \in \mathcal{R}^{ar\times n} \tag{6}$$

where each image is projected to an $(a \times r)$-dimensional vector. Therefore, (5) is generalized to the following image projection ridge regression (IPRR) model:

$$\min_{Z, P^T P=I_r} \|\mathbf{X} \circ P - (\mathbf{X} \circ P)Z\|_F^2 + \lambda \|Z\|_F^2 + \gamma \mathrm{Tr}(P^T G P). \tag{7}$$

We will discuss its optimization in the next section.

IV. Optimization

The subproblem of optimizing P is

$$\min_{P^T P=I_r} \|\mathbf{X} \circ P - (\mathbf{X} \circ P)Z\|_F^2 + \gamma \mathrm{Tr}(P^T G P). \tag{8}$$

It is seen that (8) is equivalent to minimize

$$\begin{aligned}
&\sum_{s=1}^r \|\mathbf{X} \circ p_s - (\mathbf{X} \circ p_s)Z\|_F^2 + \gamma p_s^T G p_s \\
&= \sum_{s=1}^r \left\{ \sum_{i=1}^n \left\| X_i p_s - \sum_{j=1}^n z_{ji} X_j p_s \right\|_2^2 + \gamma p_s^T G p_s \right\} \\
&= \sum_{s=1}^r \left\{ p_s^T \left(\sum_{i=1}^n X_i^T X_i \right) p_s - 2 p_s^T \left(\sum_{i=1}^n \sum_{j=1}^n z_{ji} X_i^T X_j \right) p_s \right. \\
&\qquad \left. + p_s^T \left(\sum_{i=1}^n \sum_{u=1}^n \sum_{v=1}^n z_{ui} z_{vi} X_u^T X_v \right) p_s + \gamma p_s^T G p_s \right\} \\
&= \mathrm{Tr}(P^T (H_1 - 2H_2 + H_3 + \gamma G) P) > 0
\end{aligned} \tag{9}$$

where $H_1 = \sum_{i=1}^n X_i^T X_i$, $H_2 = \sum_{i=1}^n \sum_{j=1}^n z_{ji} X_i^T X_j$, and $H_3 = \sum_{i=1}^n \sum_{j=1}^n z_{(i)} z_{(j)}^T X_i^T X_j$. Then, it is easy to see that $H_1 - 2H_2 + H_3 + \gamma G$ is positive definite and seeking P is reduced to the eigen decomposition problem

$$\mathbf{eig}_r (H_1 - 2H_2 + H_3 + \gamma G) \tag{10}$$

where $\mathbf{eig}_r(\cdot)$ returns eigenvectors of a matrix associated with its smallest r eigenvalues.

When P is fixed, optimizing Z gives

$$Z = \Big((\mathbf{X} \circ P)^T (\mathbf{X} \circ P) + \lambda I_n \Big)^{-1} \Big((\mathbf{X} \circ P)^T (\mathbf{X} \circ P) \Big). \tag{11}$$

Theorem 1: Denote the objective function in model (7) as $F(P, Z)$. Then, under the updating rules of (10) and (11), $\{F(P^t, Z^t)\}_{t=1}^\infty$ is nonincreasing and converges, where t represents the iteration number.

Proof: It is easy to verify that $F(P^{t+1}, Z^{t+1}) \leq F(P^{t+1}, Z^t) \leq F(P^t, Z^t)$, because at each iteration (10) and

(11) return the optimal solution to the corresponding subproblems. Also, it is obvious $F(P, Z) \geq 0$ by its definition. Therefore, $\{F(P^t, Z^t)\}_{t=1}^{\infty}$ converges. ■

V. Subspace Clustering via IPRR

A postprocessing step is commonly used for many spectral clustering-based subspace clustering methods [1], [4]. Based on the solution to (7), we construct an affinity matrix $\mathbf{A}$ as follows:

1) Define the weighted column space of Z as $\bar{Z} = U\Sigma^{1/2}$, where $Z = U\Sigma V^T$ is the skinny SVD of Z.
2) Normalize each row of $\bar{Z}$ and obtain $\bar{U}$.
3) Construct the affinity matrix $\mathbf{A}$ as $[\mathbf{A}]_{ij} = \left(|[\bar{U}\bar{U}^T]_{ij}|\right)^{\phi}$, where $\phi \geq 1$ controls the sharpness of the affinity matrix between two data points.[1]

Subsequently, we perform Normalized Cut [23] on $\mathbf{A}$ in a way similar to [24].

VI. Experiments

In this section, we present the experimental results, where we focus on two specific applications: Face clustering and hand written digit clustering. To demonstrate the effectiveness of the proposed method, we compare it with several state-of-the-art and some recently developed methods, including local subspace affinity (LSA) [25], spectral curvature clustering (SCC) [26], LRR [4], SSC [3], kernel SSC (KSSC) [27], low-rank subspace clustering (LRSC) [28], latent LRR (LatLRR) [29], robust LatLRR (RLatLRR) [30], block-diagonal LRR (BDLRR) [31], block-diagonal SSC (BDSSC) [31], structured SSC (S^3C) [32], and latent space SSC (LS3C) [10]. We adopt clustering accuracy or clustering error rate to measure the performance.

For fair comparison, we specify the number of clusters K for all methods. Without specification, the results for baseline methods are obtained from [32], [3], [27], where parameters have been finely tuned; otherwise, we tune the parameters with the best performance obtained. For IPRR, the parameters are tuned with the best performance reported. For purpose of reproductivity, we provide code of IPRR online.[2]

A. Face Clustering

Face clustering is to find groups from a given set of face images, such that each group corresponds to an individual. We use the Extended Yale B (EYaleB) data [33] to evaluate IPRR on this task. Face images of 38 persons, each of whom has 64 frontal face images taken under varying lighting conditions, are contained in this data. These images are cropped to 192 × 168 pixels. To save the high cost in computation, as commonly done [3], [8], [34], we down-sample these images to 48 × 42 pixels. To better investigate the performance of IPRR, we conduct experiments using different subsets with different K values. To reduce the potentially combinatorially large number of subsets, we divide these 38 objects into four groups that contain objects 1–10, 11–20, 21–30, 31–38, respectively. Then, all possible combinations of $K \in \{2, 3, 5, 8, 10\}$ objects within each group are combined together as collections of subsets for different K values. We conduct experiments on all these subsets and report the results in Table I in mean and median clustering error rates within each collection.

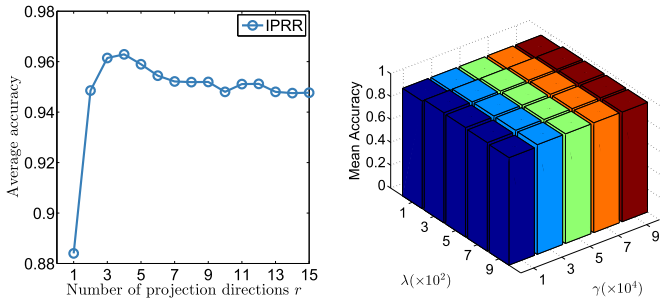

Fig. 1. Sensitivity of IPRR to parameters.

From Table I, it is observed that IPRR outperforms most of the baseline methods, and especially the improvement is significant with large K values. For example, the improvement of IPRR over LRR is around 20% with $K = 10$. S^3C and TRR show more competitive performance than the other baseline methods, where the best results are obtained in some cases. However, the clustering error rate tends to grow faster than IPRR as K increases. In this test, IPRR shows the robustness in clustering performance with respect to the number of clusters. We set $r = 3$ for IPRR and fix $\lambda = 5$, $\gamma = 100$. For TRR, we project data to 10 K dimensions and fix regularization and thresholding parameters to be 100 and 9, respectively. For RLatLRR, we fix the parameter to 0.014 as suggested in the original paper.

B. Handwritten Digit Clustering

In this section, we test IPRR on Binary Alphadigits dataset[3] with the application to handwritten digits clustering. This dataset contains binary digits and capital letters of 0–9 and A–Z, with each of which regarded as a class. There are 39 images for each class with size of 20 × 16 pixels. In a similar experimental setting as in face clustering, the classes are divided into four groups containing 0–9, A–J, K–T, and U–Z, respectively. Similar to EYaleB data, five collections of subsets are obtained with $K \in \{2, 3, 5, 8, 10\}$. We conduct experiments within each collection and report the mean and median error rates in Table II.

It is seen that IPRR has the best performance, where the clustering performance has been significantly improved. It is noted that S^3C, which achieves some of the best results on EYaleB data, shows inferior performance to IPRR, and the improvement from IPRR ranges around 3–10% and 4–10% in mean and median accuracy, respectively. In this experiment, we fix $r = 3$, $\lambda = 500$, and $\gamma = 5e4$ for IPRR, respectively. For LS3C, we project data to $2K$ dimensions and fix 0.2 and 0.1 for two regularization parameters. For S^3C, we project data to $2K$ dimensions and fix $\alpha = 0.25$ and $\lambda = 1000$. For TRR, we project data to 8 K dimensions and fix regularization and thresholding parameters to be 1000 and 13, respectively.

C. Parameter Sensitivity

We use Alphadigit data to test how IPRR depends on the parameters where, without loss of generality, we use $K = 2$. We first show the effects of r on clustering performance in left side of Fig. 1 with λ and γ fixed as in Section VI-B. It is observed that IPRR achieves the best performance with only a few projection directions and competitive performance with a wide range of values for r. Meanwhile, the performance tends to

[1]Following [2] and for fair comparison, we set $\phi = 4$ in this work.
[2]https://www.researchgate.net/publication/316609151_IPRR_SPL
[3]http://www.cs.nyu.edu/~roweis/data.html

TABLE I
CLUSTERING PERFORMANCE ON EYALEB DATA

no. of Subjects Error Rate (%)	Two Subjects		Three Subjects		Five Subjects		Eight Subjects		Ten Subjects	
	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
LSA	32.80	47.66	52.29	50.00	58.02	56.87	59.19	58.59	60.42	57.50
SCC	16.62	07.82	38.16	39.06	58.90	59.38	66.11	64.65	73.02	75.78
LRR	09.52	05.47	19.52	14.58	34.16	35.00	41.19	43.75	38.85	41.09
LRR-H	02.54	00.78	04.21	02.60	06.90	05.63	14.34	10.06	22.92	23.59
LRSC	05.32	04.69	08.47	07.81	12.24	11.25	23.72	28.03	30.36	28.75
SSC	01.86	**00.00**	03.10	01.04	04.31	02.50	05.85	04.49	10.94	05.63
LatLRR	02.54	00.78	04.21	02.60	06.90	05.63	14.34	10.06	22.92	23.59
BDLRR	03.91	–	10.02	–	12.97	–	27.70	–	30.84	–
BDSSC	03.90	–	17.70	–	27.50	–	33.20	–	39.53	–
RLatLRR	11.87	09.37	10.40	09.37	85.35	88.75	19.39	19.63	22.19	23.12
S^3C	**01.43**	**00.00**	03.09	**00.52**	04.08	**02.19**	04.84	04.10	06.09	05.16
TRR	02.13	00.78	02.93	01.56	03.83	02.50	04.31	03.52	04.90	04.22
IPRR	01.88	01.56	**02.57**	02.08	**03.11**	02.81	**03.34**	**03.12**	**03.39**	**02.93**

'–" means that the result is not reported in the corresponding letter. The best performance is boldfaced.

TABLE II
CLUSTERING PERFORMANCE ON ALPHADIGITS DATA

no. of Subjects Error Rate (%)	Two Subjects		Three Subjects		Five Subjects		Eight Subjects		Ten Subjects	
	Average	Median	Average	Median	Average	Median	Average	Median	Average	Median
LSA	10.70	03.85	22.69	22.22	33.81	33.85	40.76	40.06	42.65	41.28
SSC	05.70	02.56	13.58	08.54	23.26	25.12	30.00	30.01	32.14	32.82
LRR	07.76	03.84	14.21	11.11	23.34	23.59	30.50	30.44	33.67	32.56
LRSC	15.81	08.97	25.65	25.64	37.77	37.95	47.98	48.08	50.77	51.03
KSSC (P)	05.42	02.56	12.85	07.69	22.64	23.08	31.06	32.05	33.85	34.36
KSSC (G)	05.93	02.56	13.64	08.55	23.84	26.15	31.19	31.09	32.48	33.33
LS3C	06.23	03.85	12.67	09.83	25.24	26.15	34.06	33.97	37.01	36.49
S^3C	06.66	05.13	13.66	11.97	27.11	28.72	35.83	34.94	37.18	38.21
TRR	04.40	02.56	09.29	06.41	18.98	16.41	27.94	**27.56**	31.62	30.51
IPRR	**03.85**	**01.28**	**08.49**	**05.56**	**17.17**	**14.87**	**27.19**	**27.56**	**28.89**	**27.95**

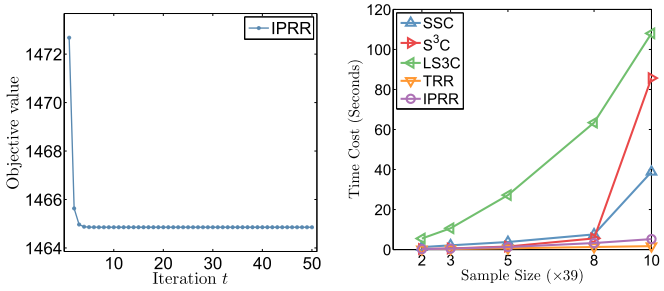

Fig. 2. Left: Convergence curve of IPRR. Right: Comparison of computation time.

degrade for large r values. This is because large r may involve noisy features in building representations. Then, we show how IPRR performs with different combinations of λ and γ in right side of Fig. 1, where we fix $r = 3$. It is seen that IPRR is quite insensitive to λ and γ. Similar patterns are also seen for median performance. These observations suggest potential of IPRR in real world applications.

D. Convergence and Computation Time

To empirically testify the convergence stated in Theorem 1, in left side of Fig. 2, we show the curve of IPRR on a subset of Alphadigit data. It is seen that IPRR converges within ten iterations, revealing fast convergence and efficiency of IPRR. Also, to further show the efficiency of IPRR, we compare it with several competitive methods on Alphadigit data and show the average computation time, which is reported in right side of Fig. 2. Realizing the less effectiveness in clustering performance and for clearer representation, we omit some methods in this figure. It is seen that IPRR is competitive to TRR in speed, and is much faster than methods such as SSC, S^3C, and LS3C. This observation verify the efficiency of IPRR and suggests its applicability in real world applications.

VII. CONCLUSION

In this letter, we propose a novel subspace clustering method for 2-D data, which directly use 2-D data as input such that the inherent structures and relationships of the original data are better captured. Orthogonal projection directions are sought, which helps retain major information of the data and construct representation coefficients of projected data points. The projection and representation are learned jointly, hence they enhance each other and lead to powerful data representations. Efficient optimization algorithm is developed with provably nonincreasing and convergent property. Extensive experimental results confirm the effectiveness of IPRR. As special a special case, IPRR can also be applied for vector data.

REFERENCES

[1] C. Peng, Z. Kang, H. Li, and Q. Cheng, "Subspace clustering using log-determinant rank approximation," in *Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2015, pp. 925–934.

[2] G. Liu, Z. Lin, and Y. Yu, "Robust subspace segmentation by low-rank representation," in *Proc. 27th Int. Conf. Mach. Learning*, 2010, pp. 663–670.

[3] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2765–2781, Nov. 2013.

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

[5] P. Favaro, R. Vidal, and A. Ravichandran, "A closed form solution to robust subspace estimation and clustering," in *Proc. 2011 IEEE Conf. Comput. Vis. Pattern Recog.*, 2011, pp. 1801–1807.

[6] G. Liu, H. Xu, J. Tang, Q. Liu, and S. Yan, "A deterministic analysis for LRR," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 38, no. 3, pp. 417–430, Mar. 2016.

[7] G. Liu, Q. Liu, and P. Li, "Blessing of dimensionality: Recovering mixture data via dictionary pursuit," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 47–60, Jan. 2017.

[8] Z. Kang, C. Peng, and Q. Cheng, "Robust subspace clustering via smoothed rank approximation," *IEEE Signal Process. Lett.*, vol. 22, no. 11, pp. 2088–2092, Nov. 2015.

[9] C. Peng, Z. Kang, M. Yang, and Q. Cheng, "Feature selection embedded subspace clustering," *IEEE Signal Process. Lett.*, vol. 23, no. 7, pp. 1018–1022, Jul. 2016.

[10] V. M. Patel, H. Van Nguyen, and R. Vidal, "Latent space sparse subspace clustering," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 225–232.

[11] C. Peng, Z. Kang, and Q. Cheng, "Integrating feature and graph learning with low-rank representation," *Neurocomputing*, vol. 249, no. 2, pp. 106–116, Aug. 2017.

[12] X. Li, Y. Pang, and Y. Yuan, "L1-norm-based 2DPCA," *IEEE Trans. Syst., Man, Cybern, Part B, Cybern.*, vol. 40, no. 4, pp. 1170–1175, Aug. 2010.

[13] Y. Fu, J. Gao, D. Tien, Z. Lin, and X. Hong, "Tensor LRR and sparse coding-based subspace clustering," *IEEE Trans. Neural Netw. Learning Syst.*, vol. 27, no. 10, pp. 2120–2133, Oct. 2016.

[14] J. Yang and J.-y. Yang, "From image vector to matrix: a straightforward image projection techniqueimpca vs. pca," *Pattern Recognit.*, vol. 35, no. 9, pp. 1997–1999, 2002.

[15] J. Yang, D. Zhang, A. F. Frangi, and J.-y. Yang, "Two-dimensional pca: a new approach to appearance-based face representation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, Jan. 2004.

[16] X. Peng, Z. Yi, and H. Tang, "Robust subspace clustering via thresholding ridge regression," in *Proc. 29th AAAI Conf. Artif. Intell.*, 2015, pp. 3827–3833.

[17] C.-Y. Lu, H. Min, Z.-Q. Zhao, L. Zhu, D.-S. Huang, and S. Yan, "Robust and efficient subspace segmentation via least squares regression," *Eur. Conf. Comput. Vis.*, 2012, pp. 347–360.

[18] W.-C. Chang, "On using principal components before separating a mixture of two multivariate normal distributions," *Appl. Statist.*, vol. 32, pp. 267–275, 1983.

[19] P. Arabie, "Cluster analysis in marketing research," in *Advanced Methods of Marketing Research*, R. P. Bagozzi, Ed. Cambridge, MA, USA: Blackwell Business, 1994.

[20] K. Allab, L. Labiod, and M. Nadif, "Simultaneous semi-NMF and PCA for clustering," in *Proc. 2015 IEEE Int. Conf. Data Mining*, 2015, pp. 679–684.

[21] C. Peng, Z. Kang, Y. Hu, J. Cheng, and Q. Cheng, "Nonnegative matrix factorization with integrated graph and feature learning," *ACM Trans. Intell. Syst. Technol.*, vol. 8, no. 3, pp. 42:1–42:29, Jan. 2017. [Online]. Available: http://doi.acm.org/10.1145/2987378

[22] V. D. M. Nhat and S. Lee, "Kernel-based 2DPCA for face recognition," in *Proc. 2007 IEEE Int. Symp. Signal Process. Inform. Technol.*, 2007, pp. 35–39.

[23] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.

[24] P. K. Agarwal and N. H. Mustafa, "k-means projective clustering," in *Proc. 23rd ACM SIGMOD-SIGACT-SIGART Symp. Principles Database Syst.*, 2004, pp. 155–165.

[25] J. Yan and M. Pollefeys, "A general framework for motion segmentation: Independent, articulated, rigid, non-rigid, degenerate, and non-degenerate," in *Computer Vision–ECCV 2006*. New York, NY, USA: Springer, 2006, pp. 94–106.

[26] G. Chen and G. Lerman, "Spectral curvature clustering (SCC)," *Int. J. Comput. Vis.*, vol. 81, no. 3, pp. 317–330, 2009.

[27] V. M. Patel and R. Vidal, "Kernel sparse subspace clustering," in *Proc. 2014 IEEE Int. Conf. Image Process.*, 2014, pp. 2849–2853.

[28] R. Vidal and P. Favaro, "Low rank subspace clustering (LRSC)," *Pattern Recog. Lett.*, vol. 43, pp. 47–61, 2014.

[29] G. Liu and S. Yan, "Latent low-rank representation for subspace segmentation and feature extraction," in *Proc. 2011 IEEE Int. Conf. Comput. Vis.*, 2011, pp. 1615–1622.

[30] H. Zhang, Z. Lin, C. Zhang, and J. Gao, "Robust latent low rank representation for subspace clustering," *Neurocomputing*, vol. 145, pp. 369–373, 2014.

[31] J. Feng, Z. Lin, H. Xu, and S. Yan, "Robust subspace segmentation with block-diagonal prior," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 3818–3825.

[32] C.-G. Li and R. Vidal, "Structured sparse subspace clustering: A unified optimization framework," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 277–286.

[33] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[34] Z. Kang, C. Peng, and Q. Cheng, "Robust subspace clustering via tighter rank approximation," in *Proc. 24th ACM Int. Conf. Inform. Knowl. Manage*, 2015, pp. 393–401.